

University of Wisconsin Milwaukee

UWM Digital Commons

Cell and Molecular Biology 5e: What We Know
and How We Found Out - All Versions

Biological Sciences

4-2022

Basic Cell and Molecular Biology 5e: What We Know and How We Find Out

Gerald Bergtrom

University of Wisconsin - Milwaukee, bergtrom@uwm.edu

Follow this and additional works at: https://dc.uwm.edu/biosci_facbooks_bergtrom



Part of the [Biology Commons](#), and the [Cell Biology Commons](#)

Recommended Citation

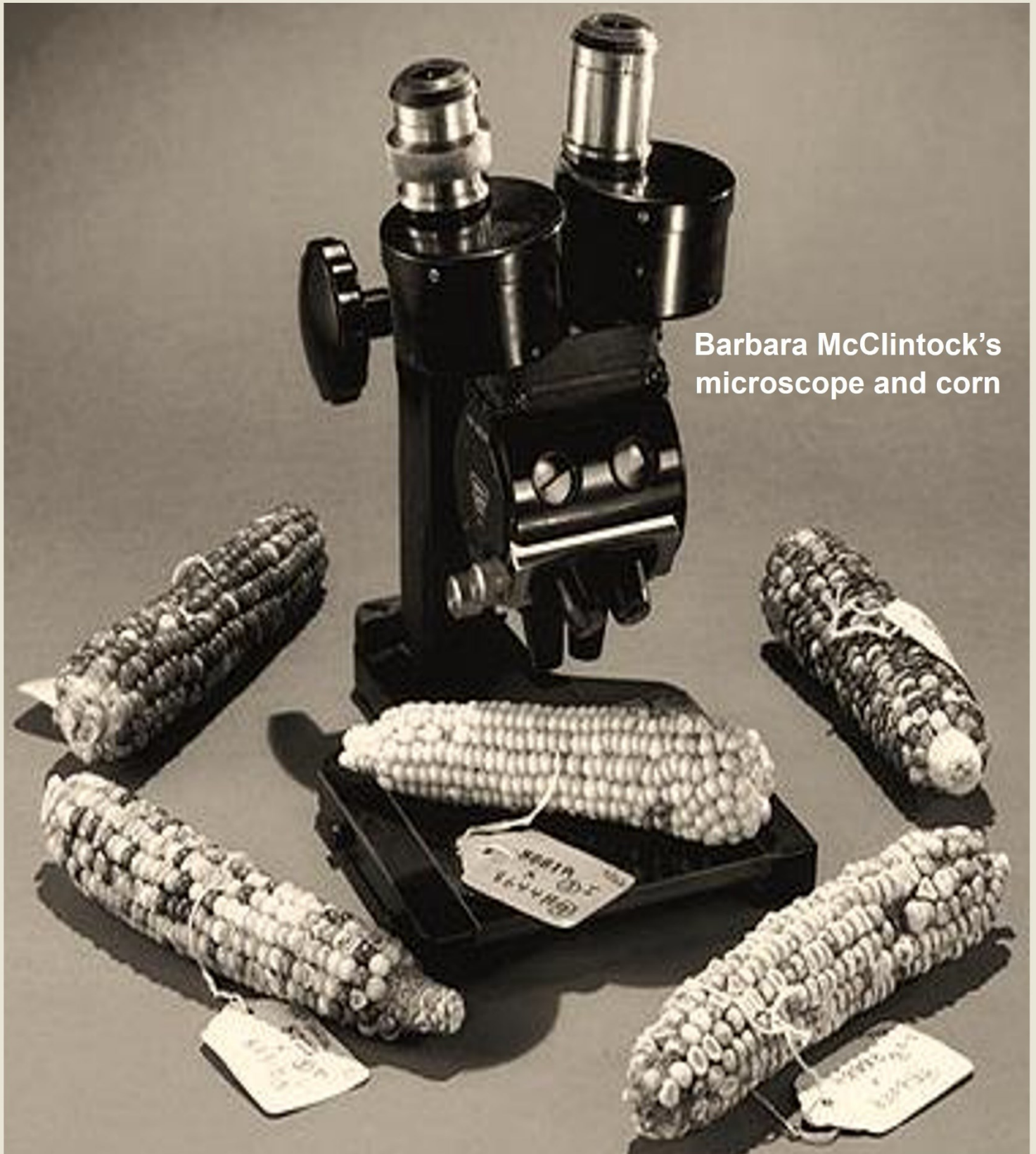
Bergtrom, Gerald, "Basic Cell and Molecular Biology 5e: What We Know and How We Find Out" (2022). *Cell and Molecular Biology 5e: What We Know and How We Found Out - All Versions*. 14.
https://dc.uwm.edu/biosci_facbooks_bergtrom/14

This Book is brought to you for free and open access by UWM Digital Commons. It has been accepted for inclusion in Cell and Molecular Biology 5e: What We Know and How We Found Out - All Versions by an authorized administrator of UWM Digital Commons. For more information, please contact scholarlycommunicationteam-group@uwm.edu.

Cell and Molecular Biology

What We Know & How We Found Out

Basic CMB5e iText (Digital Edition)



Barbara McClintock's
microscope and corn

by
Gerald Bergtrom

Cell and Molecular Biology

What We Know & How We Found Out

Basic CMB5e iText (Digital Edition)

An OER hardcopy (print) version of the *Annotated CMB5e iText* (CMB5p) is available; search *Bergtrom* at <https://www.lulu.com/shop/>.

Open Access, high resolution copies of the images in this book are available. Click <https://dc.uwm.edu/> and search Bergtrom.

Cell and Molecular Biology What We Know & How We Found Out Basic CMB5e iText (Digital Edition)

An interactive textbook published under a Creative Commons (CC-BY) license

**By
Gerald Bergtrom**
Revised October, 2022

ISBN# 978-0-9961502-7-9

New in CMB5:

- ✓ *Reformatted to include*
 - *New typeface*
 - *QR codes to enable access to external websites from printed pages*
 - *Larger figures, now accompanied by Alternative Text to increase accessibility*
- ✓ *Many content updates*

Published by UWM Digital Commons at <https://dc.uwm.edu/>

An OER hardcopy (print) version of the *Annotated CMB5e iText* (CMB5p) is available; search *Bergtrom* at <https://www.lulu.com/shop/>.

Creative Commons Licensure and Permissions

Written, Compiled and Curated Under a *Creative Commons with Attribution* (CC-BY) License

The following is a human-readable summary of (and not a substitute for) the **CC-BY** 4.0 ([Creative Commons — Attribution 4.0 International — CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

You are free to:

- **Share** — copy and redistribute the material in any medium or format
- **Adapt** — remix, transform, and build upon the material for any purpose, even commercially. The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

- **Attribution** — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- **No additional restrictions** — You may not apply legal terms or [technological measures](#) that legally restrict others from doing anything the license permits.

Notices:

- You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable [exception or limitation](#).
- No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as [publicity, privacy, or moral rights](#) may limit how you use the material.

CMB5e Published 2022



Cover: Public Domain image from [Microscope and Corn- Barbara McClintock](#)

Dedicated to:

Sydell, Aaron, Aviva, Edan, Oren, Tamar and our extended family, whose patience and encouragement made this work possible, my students from whose curiosity I received as much as I gave, and the memory of my mentor Herbert Oberlander, who gave me the time, opportunity, and tools to "do science."

Preface to CMB5

Details, goals, and hopes for teaching and learning

A grasp of the logic and practice of science is essential to understanding the world around us. So, all editions/versions of **CMB** focus on experimental support for what we know about cell and molecular biology, and on showing students the relationship of cell structure and function. Rather than trying to be comprehensive reference, the book selectively details investigative questions, methods and experiments that lead to our understanding of cell biology. This focus is nowhere more obvious than in the chapter **learning objectives** and in links to the author's short YouTube voice-over PowerPoint (VOP) videos. Numbered links to each of these VOPs are embedded near relevant text and includes edited, optional closed captions. These are easily launched by clicking a *play-video* symbol or descriptive title on a computer or tablet app or by using *QR codes*, as in the example below:



[102 Golgi Vesicles & the Endomembrane System](#)

Other **external online resource* links are all numbered. The numbers refer to the QR codes for the links at the end of each chapter for use by students using a print edition or printed pages from the digital text.

All digital (the *Basic*, *Annotated*, and *Instructors CMB5e*) as well as the hard-copy version (the *Annotated CMB5p*) include these interactive features. In addition, the *Annotated CMB5e* and *CMB5p* have **Challenge boxes** that typically include questions about significant new science that is not necessarily definitive and still subject to confirmation. Finally, the *Instructors' CMB5e* adds interactive short **25 Words or Less** short writing assignments. The *Instructors' CMB5e* is available on request. All interactive elements are intended to expand on concepts discussed in the text. My hope is that that you will engage and experience some of them. In writing and updating **CMB**, I tried to make it *user-friendly*, current, and accurate. I invite you to use the interactive features of the iText to think about:

- how good and great experiments were inspired and designed.
- how alternative experimental results were predicted.
- how data was interpreted.
- how investigators (and we!) arrive at the most interesting “next questions.”

Along the way, I hope active learning about cell and molecular biology will be an exciting way to reinforce your critical thinking and writing skills across the board.

**External online resources* links are active at the time of publication but may not remain so in the future.

A mission Statement

Ever since I wrote my first published research paper my goal was to make my science clear enough to be read by anyone interested in biology. Finding jargon-free ways to explain research is not always easy. Sometimes it involves metaphor or analogy and sometimes it requires linguistic precision. I've heard that writers of fiction and truth (novelists, storytellers, playwrights, poets...) spend a lot of time revising to make themselves understood. So, I agree with Eve. L. Ewing, a poet, Professor of Sociology of Education and *Marvel* comic writer, who said in an interview that *academic writing is a creative act* (Toor, R. 2020, *Scholars Talk Writing*, The Chronical of Higher Education). I have used interactive components of the early editions of **CMB** to create assessable homework assignments in my *flipped, blended* Cell Biology course. **CMB5** is the latest edition, the product of many revisions, updates, additions, and corrections. In aggregate, all editions and versions of **CMB** have been downloaded more than 60,000 times and have been adopted or made recommended reading in college and even high school and medical school courses in the U.S. and around the world. I would like to think that this universal appeal is due to some measure of imagination, clarity, and creative effort.

Some Notes on Using CMB5e and CMB5p

If you are reading the hardcopy *Annotated CMB5p* iText, you can also download the digital **CMB5e** version at <https://dc.uwm.edu/> into your browser's URL line. If your instructor has uploaded a digital version of CMB5 to your course site, expect further instructions on how you to access and use the book. **CMB5p** is an open access hardcopy version of the *Annotated CMB5e*, for students that would like to study from a printed book. It is virtually identical in content and format to its digital cousin. All numbered text and QR code links should be active in the latest digital (**CMB5e**) versions. Hardcopy readers, please use QR codes to access web links. Links to the author's POVs should be stable (please notify the author or your instructor at of any exceptions). On the other hand, remember that links to *external online resources* that are active at the time of publication may not remain so in the future.

Special to Instructors from the Author

The complete digital *Instructors CMB5e* iText is available at <https://dc.uwm.edu/>. To get the complete *Instructors CMB5e* (with additional interactive features), you will need to fill out a short questionnaire identifying yourself as an instructor, your institutional and departmental affiliation, and course information (title, number). When you submit the form, you will get pdf as well as MS-Word files for all digital versions of the iText. When you have download the digital *iText(s)* of your choice, you should find it an easy matter to add, subtract, modify or enhance any parts of it to suit your purposes (in accordance with the **Creative Commons CC-BY license** under which it is published ([Creative Commons — Attribution 4.0 International — CC BY 4.0](https://creativecommons.org/licenses/by/4.0/))). You are free to provide the original or a customized version to your students as a small pdf file (recommended) or the larger MS-Word file. If students access a digital iText through your CMS (Course Management System (e.g., Blackboard, D2L, Canvas), you can link assignments to Discussion Fora, a DropBox, or Quiz directly in the iText. Of

course, remember to provide instructions to students on how they should use the book in your class! Here are a few notes on the versioning of **CMB5**.

1. If you modify the digital *Annotated* or *Instructors'* versions, the index (which is not *dynamic*) may no longer function as intended.
2. The *Basic* version of the digital iText has no index but can (like all versions) be searched online or after downloading.
3. An interactive OER, low-cost print version of the *Annotated CMB5 iText (CMB5p)* is now available for anyone that prefers a hardcopy textbook. **CMB5p** is still open access, with the a **Creative Commons CC-BY-NC-SA license** ([Creative Commons — Attribution-NonCommercial-ShareAlike 4.0 International — CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)).
4. And once again, please remember that links to **external online resources** that are active at the time of publication may not remain so in the future.

Finally, to assist instructors, *Open Access*, high resolution copies of the images in this book are provided with the *Instructors' CMB5e* or at <https://dc.uwm.edu/> (search Bergtrom).

I hope that you (and even your students!) will enjoy customizing interactive elements and digging into some of the more current research included in the latest edition of the **CMB**. Above all, I hope that your students will achieve a better understanding of how scientists use skills of inductive and inferential logic to ask questions and formulate hypotheses..., and how they apply concept and method to test those hypotheses.

Acknowledgements

I thank my erstwhile UW-M LTC (now CETL) colleagues Matthew Russell, Megan Haak, Melissa Davey Castillo, Jessica Hutchings, and Dylan Barth for ideas on how to make OER content more interactive and engaging. Thanks to Ann Hanlon and all my Golda Meir Library colleagues for enabling publication of all digital **CMB** versions on UW-M Digital Commons. I am most grateful to Ms. M. Terry Bott for reviewing and vetting the images in the book, confirming their public domain or Creative Commons license status. I also owe a debt of gratitude to our departmental lab manager Jordan Gonnering for much hardware and software assistance during the preparation of CMB5, including the print edition. I also thank my copy editor at Elite Authors for her thorough and knowledgeable copy-editing and the team at Index Busters for their patient effort that eased the path to completion of the latest edition of my CMB textbook, especially the print edition. Last among these, but not least, thank you to all the artists, educators and scientists who generously gave Creative Commons license to others to use their work, without which this book would have been impossible. They are individually listed in Appendix I herein. Finally, I must acknowledge the University of Wisconsin-Milwaukee for the opportunity to research and learn both science and pedagogy for more than 35 years. My UW-M experience has left its mark on the content, concept and purpose all CMB editions and versions.

An OER hardcopy (print) version of the *Annotated CMB5e iText* (CMB5p) is available; search *Bergtrom* at <https://www.lulu.com/shop/>.

About the Author

Dr. Bergtrom is Professor (Emeritus) of Biological Sciences at the UW-Milwaukee and a former *Teaching and Learning* consultant in the UW-M Center for Excellence in Teaching and Learning. Scientific interests include cell, molecular and evolutionary biology. Pedagogic interests include the use of technology to serve active and engaged teaching and learning. He has taught face-to-face, fully online, *blended* and *flipped* classes. He also developed and co-instructed *Teaching with Technology*, an interdisciplinary graduate course for graduate students that might someday be teaching. In 40+ years of teaching and research, he has tested and adopted pedagogically proven teaching technologies in his courses. His scientific publications are supplemented by publications on active learning in all teaching modes¹⁻³. The first edition of his *Cell and Molecular Biology–What We Know & How We Found Out* appeared in 2015⁴. Further editions came out in 2016⁵, 2018⁶, 2019⁷ and 2022^{8,9}.

1. Bergtrom, G. (2006) *Clicker Sets as Learning Objects*. Int. J. Knowl. & Learn. Obj.

2:105-110. (<http://www.ijello.org/Volume2/v2p105-110Bergtrom.pdf>)

2. Bergtrom, G. (2009) *On Offering a Blended Cell Biology Course*. J. Res. Center Ed.

Tech. 5(1) (<http://www.rcetj.org/?type=art&id=91609&>)

3. Bergtrom, G. (2011) *Content vs Learning: An Old Dichotomy in Science Courses*. J.

Asynchr. Learning Networks 15:33-44 (http://jaln_v15n1_bergtrom.pdf)

4. Bergtrom, G. (2015) *Cell and Molecular Biology: What We Know & How We Found Out (CMB)* (If necessary, please contact the author to access this "first" edition)
5. Bergtrom, G. (2016) *Cell and Molecular Biology: What We Know & How We Found Out (CMB2e)*.
6. Bergtrom, G. (2018) *Cell and Molecular Biology: What We Know & How We Found Out (CMB3e)*
7. Bergtrom, G. (2020) *Cell and Molecular Biology: What We Know & How We Found Out (CMB4e)*
8. Bergtrom, G. (2022) *Cell and Molecular Biology: What We Know & How We Found Out (Basic, Annotated and Instructors OER digital CMB5e versions, available at*

https://dc.uwm.edu/biosci_facbooks_bergtrom/)

9. Bergtrom, G. (2022) *Cell and Molecular Biology: What We Know & How We Found Out*;



(the Annotated hardcover *CMB5p*; Search Bergtrom at <https://Lulushop.com/>)

Table of Chapters

Use these links to navigate CMB5e (an online version of CMB5p)

Preface	v
Chapter 1: Cell Tour, Life's Properties and Evolution, Studying Cells	1
Chapter 2: Basic Chemistry, Organic Chemistry and Biochemistry	38
Chapter 3: Details of Protein Structure	58
Chapter 4: Bioenergetics	80
Chapter 5: Enzyme Catalysis and Kinetics	91
Chapter 6: Glycolysis, the Krebs Cycle and the Atkins Diet	106
Chapter 7: Electron Transport, Oxidative Phosphorylation and Photosynthesis	130
Chapter 8: DNA Structure, Chromosomes, and Chromatin	149
Chapter 9: Details of DNA Replication & DNA Repair	174
Chapter 10: Transcription and RNA Processing	200
Chapter 11: The Genetic Code and Translation	225
Chapter 12: Regulation of Transcription and Epigenetic Inheritance	244
Chapter 13: Posttranscriptional Regulation of Gene Expression	275
Chapter 14: Repetitive DNA, A Eukaryotic Genomic Phenomenon	299
Chapter 15: DNA Technologies	331
Chapter 16: Membrane Structure	369
Chapter 17: Membrane Function	391
Chapter 18: The Cytoskeleton and Cell Motility	434
Chapter 19: Cell Division and the Cell Cycle	469
Chapter 20: The Origins of Life	498
Epilogue	531
Appendix I: List of Figures & Sources	537
Appendix II: Context-Embedded YouTube Videos	563

Table of Contents

Use these links to navigate CMB5e (an online version of CMB5p)

<u>Preface</u>	v
<u>Chapter 1: Cell Tour, Life's Properties and Evolution, Studying Cells</u>	1
1.1 <u>Introduction</u>	1
1.2 <u>Scientific Method – The Practice of Science</u>	3
1.2.1 The Method as it is Really Practiced	
1.2.2 Logic and the Origins of the Scientific Method	
1.3 <u>Domains of Life</u>	6
1.3.1 Viruses: Dead or Alive; Big and Small; A History of Surprises	
1.3.2 The Prokaryotes (Eubacteria = <i>Bacteria</i> and <i>Cyanobacteria</i>)	
Bacterial Reproduction	
Cell Motility and the Possibility of a Prokaryotic Cytoskeleton	
Some Bacteria Have Internal Membranes	
Bacterial Ribosomes Do the Same Thing as Eukaryotic Ribosomes... and Look Like Them!	
1.3.3 The Archaeobacteria (Archaea)	
1.3.4 The Eukaryotes	
1.4 <u>Tour of the Eukaryotic Cell</u>	16
1.4.1 The Nucleus	
Structure of the Interphase Nucleus	
Every Cell (i.e., Nucleus) of an Organism Contains the Same Genes	
1.4.2 Ribosomes	
1.4.3 Internal Membranes and the Internal Membrane System	
1.4.4 Mitochondria and Plastids	
1.4.5 Cytoskeletal Structures	
1.5 <u>How we Know Functions of Organelles and Cell Structures: Cell Fractionation</u>	27
1.6 <u>The Origins, Evolution, Speciation, Diversity and Unity of Life</u>	29
1.6.1 Random Acts of Genetic Variation, the Basis of Natural Selection	
1.6.2 The Genome: An Organism's Complete Genetic Instructions	
1.6.3 Genomic 'Fossils' Can Confirm Evolutionary Relationships	
1.7 <u>Microscopy Reveals Life's Diversity of Structure and Form</u>	33
1.7.1 Light Microscopy	
1.7.2 Electron Microscopy	
<u>Chapter 2: Basic Chemistry, Organic Chemistry and Biochemistry</u>	38
2.2 <u>Introduction</u>	38
2.2 <u>Atoms and Basic Chemistry</u>	40
2.2.1 Overview of Elements and Atoms	
2.2.2 Electron Configuration – Shells and Subshells	
2.3 <u>Chemical Bonds</u>	43
2.3.1 Covalent bonds	
2.3.2 Ionic Bonds	
2.3.3 Hydrogen Bonds	
2.4 <u>Water Chemistry</u>	46
2.5 <u>Basic Biochemistry: Monomers, Polymers, Macromolecular Synthesis, Degradation</u>	48
2.5.1 Isomerism in Organic Molecules and the Diversity of Shape	
2.5.2 Monomers to Polymers and Back: Dehydration Synthesis and Hydrolysis	
2.5.3 A Tale of Chirality Gone Awry	

<u>Chapter 3: Details of Protein Structure</u>	58
3.1 <u>Introduction</u>	58
3.2 <u>Levels (Orders) of Protein Structure</u>	60
3.2.1 Protein Primary Structure; Amino Acids and the C-N-C-N... Polypeptide Backbone Formation of peptide bonds leads to polypeptide primary structure Determining Protein Primary Structure - Polypeptide Sequencing	
3.2.2 Protein Secondary Structure	
3.2.3 Protein Tertiary Structure	
3.3 <u>Changes in Protein Shape Can Cause Disease</u>	66
3.3.1 Sickle Cell Anemia	
3.3.2 The Role of Misshapen and Mis-folded Proteins in Alzheimer's Disease The <i>amyloid beta</i> ($A\beta$) peptide Relatives of Alzheimer's Disease	
3.4 <u>Protein Quaternary Structure, Prosthetic Groups Chemical Modifications</u>	72
3.5 <u>Domains, Motifs, and Folds in Protein Structure</u>	73
3.6 <u>Proteins, Genes and Evolution: How Many Proteins are We?</u>	74
3.7 <u>Directed Evolution: Getting Cells to Make New Proteins for our use and pleasure</u>	75
3.8 <u>View Animated 3D Protein Images in the NCBI Database</u>	76
<u>Chapter 4: Bioenergetics</u>	80
4.1 <u>Introduction</u>	80
4.2 <u>Kinds of Energy</u>	81
4.3 <u>Deriving Simple Energy Relationships</u>	82
4.3.1 Energy in the Universe: the Universe is a <i>Closed System</i>	
4.3.2 Energy is Exchanged Between Systems in the Universe	
4.3.3 How is Enthalpy Change (ΔH) Determined?	
4.3.4 How is Standard Free Energy Change (ΔG_0) Determined?	
4.3.5 Working an Example Using These Equations for Closed Systems	
4.4 <u>Summary: The Properties of Closed Systems</u>	88
4.5 <u>Open Systems and Actual Free Energy Change</u>	89
<u>Chapter 5: Enzyme Catalysis and Kinetics</u>	91
5.1 <u>Introduction</u>	91
5.2 <u>Enzymes and the Mechanisms of Enzyme Catalysis</u>	93
5.2.1 Structural Considerations of Catalysis	
5.2.2 Energetic Considerations of Catalysis	
5.3 <u>Enzyme Regulation</u>	97
5.4 <u>Enzyme Kinetics</u>	99
5.4.1 Why Study Enzyme Kinetics	
5.4.2 How We Determine Enzyme Kinetics and Interpret Kinetic Data	
<u>Chapter 6: Glycolysis, the Krebs Cycle and the Atkins Diet</u>	106
6.1 <u>Introduction</u>	106
6.2 <u>Glycolysis, a Key Pathway in Energy Flow through Life</u>	108
6.3 <u>Some Details of Glycolysis</u>	112
6.3.1 Glycolysis, Stage 1	
6.3.2 Glycolysis, Stage 2	
6.4 <u>A Chemical and Energy Balance Sheet for Glycolysis</u>	120
6.5 <u>Gluconeogenesis</u>	121
6.6 <u>The Atkins Diet and Gluconeogenesis</u>	125
6.7 <u>The Krebs/TCA/Citric Acid Cycle</u>	126
<u>Chapter 7: Electron Transport, Oxidative Phosphorylation and Photosynthesis</u>	130
7.1 <u>Introduction</u>	130
7.2 <u>Electron Transport Chains</u>	131

7.3	<u>Electron Transport in Respiration</u>	131
7.4	<u>Oxidative Phosphorylation in Respiration</u>	133
7.5	<u>Photosynthesis</u>	135
7.5.1	The Light-Dependent Reactions	
7.5.2	The Light-Independent ("Dark" Reactions")	
	C3 Photosynthesis - Light-Independent Reactions	
	CAM Photosynthetic Pathway - Light-Independent Reactions	
	C4 Photosynthetic Pathway - Light-Independent Reactions	
7.6	<u>On the Origins and Evolution of Respiration and Photosynthesis</u>	146
<u>Chapter 8: DNA Structure, Chromosomes and Chromatin</u>		149
8.1	<u>Introduction</u>	149
8.2	<u>The Stuff of Genes</u>	150
8.2.1	Griffith's Experiment	
8.2.2	The Avery-MacLeod-McCarty Experiment	
8.2.3	The Hershey-Chase Experiment	
8.3	<u>DNA Structure</u>	156
8.3.1	X-Ray Crystallography and the Beginnings of Molecular Biology	
8.3.2	Wilkins, Franklin, Watson & Crick – DNA Structure Revealed	
8.3.3	Meselson and Stahl's Experiment – Replication is Semiconservative	
8.4	<u>Chromosomes</u>	161
8.5	<u>Genes and Chromatin in Eukaryotes</u>	163
8.6	<u>Structure and Organization of Bacterial DNA... and Bacterial Sex</u>	168
8.7	<u>Phage Can Integrate Their DNA Into the Bacterial Chromosome</u>	171
<u>Chapter 9: Details of DNA Replication & DNA Repair</u>		174
9.1	<u>Introduction</u>	174
9.2	<u>DNA Replication</u>	175
9.2.1	Visualizing Replication and Replication Forks	
9.2.2	Visualizing Bidirectional Replication	
9.3	<u>DNA Polymerases Catalyze Replication</u>	178
9.4	<u>The Process of Replication</u>	179
9.4.1	Initiation	
9.4.2	Elongation	
9.4.3	Termination	
9.4.4	Is Replication <i>Processive</i> ?	
9.4.5	One more Problem with Replication	
9.5	<u>DNA Repair</u>	188
9.5.1	Germline vs Somatic Mutations; A Balance Between Mutation and Evolution	
9.5.2	What Causes DNA Damage?	
9.6	<u>Molecular Consequences of Uncorrected DNA Damage</u>	190
9.6.1	Depurination	
9.6.2	Pyrimidine Dimerization	
9.6.3	Deamination	
9.7	<u>DNA Repair Mechanisms</u>	192
9.7.1	Base Excision Repair	
9.7.2	Nucleotide Excision Repair	
9.7.3	Mismatch Repair	
9.7.4	Transcription Coupled Repair (in Eukaryotes)	
9.7.5	Non-homologous End-Joining	
9.7.6	Homologous Recombination	
	Repair of a Single-Stranded Break	
	Repair of a Double-Stranded Break	

<u>Chapter 10: Transcription and RNA Processing</u>	200
10.1 <u>Introduction</u>	200
10.2 <u>Overview of Transcription</u>	201
10.2.1 The Major Types of Cellular RNA	
10.2.2 Key Steps of Transcription	
10.2.3 RNAs are Extensively Processed after Transcription in Eukaryotes	
10.3 <u>Details of Transcription</u>	207
10.3.1 Details of Transcription in Prokaryotes	
10.3.2 Details of Transcription in Eukaryotes	
Eukaryotic mRNA Transcription	
Eukaryotic tRNA and 5SRNA Transcription	
Transcription of the Other Eukaryotic rRNAs	
10.4 <u>Details of mRNA Processing in Eukaryotic nuclei</u>	213
10.4.1 Spliceosomal Introns	
10.4.2 Specific Nuclear Body Proteins Facilitate SnRNP Assembly and Function	
10.4.3 Group I and Group II <i>Self-Splicing</i> Introns	
10.4.4 So, Why Splicing?	
10.4.5 5-Prime Capping	
10.4.6 3-Prime Polyadenylation	
10.5 <u>Ribosomal RNA Processing in Eukaryotic Nuclei</u>	219
10.6 <u>tRNA Processing in Eukaryotic Nuclei</u>	220
10.7 <u>Export of mRNA and Ribosomes from the Nucleus</u>	222
 <u>Chapter 11: The Genetic Code and Translation</u>	 225
11.1 <u>Introduction</u>	225
11.2 <u>Overview of the Genetic Code</u>	227
11.2.1 The (Nearly) Universal, Degenerate Genetic Code	
11.2.2 Comments on the Nature and Evolution of Genetic Information	
11.3 <u>Gene and Protein Colinearity and Triplet Codons</u>	229
11.3.1 Colinearity of Gene and Protein	
11.3.1 How is a Linear Genetic Code 'Read' to Account for All of An Organisms' Genes?	
11.4 <u>Breaking the Genetic Code</u>	231
11.5 <u>Translation</u>	235
11.5.1 Even Before Initiation - Making Aminoacyl-tRNAs	
11.5.2 Translation Initiation	
11.5.3 Translation Elongation	
Translation Elongation-1	
Translation Elongation-2	
Translation Elongation-3	
11.5.4 Translation Termination	
11.6 <u>How Can the Genetic Code be Degenerate and Accurate at the Same Time?</u>	243
 <u>Chapter 12: Regulation of Transcription and Epigenetic Inheritance</u>	 244
12.1 <u>Introduction</u>	244
12.2 <u>Gene Regulation in Prokaryotes: the Lactose (<i>lac</i>) Operon</u>	245
12.2.1 Working Out Regulation of the Lactose (<i>lac</i>) Operon in <i>E. coli</i>	
12.2.2 Negative Regulation of the lac Operon by Lactose	
12.2.3 Positive Regulation of the Lac Operon; Induction by Catabolite Activation	
12.2.4 Lac Operon Regulation by Inducer Exclusion	
12.2.5 Structure of the lac Repressor Protein and Additional Operator Sequences	
12.3 <u>Gene Regulation in Prokaryotes: the Tryptophan (<i>trp</i>) Operon</u>	254

12.4	<u>The Problem with Unregulated (<i>Housekeeping</i>) Genes in All Cells</u>	255
12.5	<u>Gene Regulation in Eukaryotes</u>	256
12.5.1	Complexities of Eukaryotic Gene Regulation	
12.5.2	Regulation of Gene Expression by Hormones that Enter Cells and Those That Don't	
	How Steroid Hormones Regulate Transcription	
	How Protein Hormones Regulate Transcription	
12.6	<u>Regulating Eukaryotic Gene Expression Means Contending with Chromatin</u>	263
12.7	<u>Regulating all genes on a Chromosome at Once</u>	266
12.8	<u>Mechanoreceptors: Capturing Non-Chemical Signals</u>	267
12.9	<u>Epigenetics</u>	268
12.9.1	Epigenetic Inheritance in Somatic Cells	
12.9.2	Epigenetic Inheritance in the Germ-Line	
	 <u>Chapter 13: Posttranscriptional Regulation of Gene Expression</u>	 275
13.1	<u>Introduction</u>	275
13.2	<u>Posttranscriptional Control of Gene Expression</u>	276
13.2.1	Riboswitches	
13.2.3	CRISPR/Cas: a Prokaryotic Adaptive Immune System	
	The CRISPR/Cas 'Immune' Response	
	Using CRISPR/Cas to Edit/Engineer Genes	
	CRISPR..., the Power and the Controversy	
13.2.3	The Small RNAs: miRNA and siRNA <u>in Eukaryotes</u>	
	Small Interfering RNA (siRNA)	
	Micro RNAs (miRNA)	
	Piwi-Interacting RNAs	
13.2.4	Long Non-Coding RNAs	
13.2.5	Circular RNAs (CircRNA)	
13.3	<u>"Junk DNA" in Perspective</u>	288
13.4	<u>The RNA Methylome</u>	288
13.5	<u>Eukaryotic Translation Regulation</u>	288
13.5.1	Specific Translation Control by mRNA Binding Proteins	
13.5.2	Coordinating Heme and Globin Synthesis	
13.5.3	Translational Regulation of Yeast GCN4	
13.6	<u>Protein Turnover in Eukaryotic Cells: Regulating Protein Half-Life</u>	295
	 <u>Chapter 14: Repetitive DNA, A Eukaryotic Genomic Phenomenon</u>	 299
14.1	<u>Introduction</u>	299
14.2	<u>The Complexity of Genomic DNA</u>	300
14.1.1	The Renaturation Kinetic Protocol	
14.2.2	Renaturation Kinetic Data	
14.2.3	Genomic Complexity	
14.2.4	Functional Differences between Cot Classes of DNA	
14.3	<u>The 'Jumping Genes' of Maize</u>	305
14.3.1	Discovering the Genes of Mosaicism; the Unstable Ds Gene	
14.3.2	The Discovery of Mobile Genes: the Ac/Ds System	
14.4	<u>Since McClintock: Transposons in Bacteria, Plants and Animals</u>	310
14.4.1	Bacterial Insertion Sequences (IS Elements)	
14.4.2	<i>Composite</i> Bacterial Transposons: Tn Elements	
14.4.3	<i>Complex</i> Transposons that Can Act Like Bacteriophage	
14.5	<u>Overview of Eukaryotic Transposable Elements</u>	314
14.6	<u>The Structure of Eukaryotic DNA (Class II) Transposons</u>	316
14.6.1	Cut-and-Paste Transposition	
14.6.2	Replicative Transposition	

14.7	<u>The Structure of Eukaryotic RNA (Class I) Transposons</u>	318
14.7.1	LTR retrotransposons: The Yeast Ty element	
14.7.2	Non-LTR Retrotransposons: LINEs	
14.7.3	Non-LTR Retrotransposons: LINEs	
14.8	<u>Mechanisms of Retrotransposition</u>	320
14.8.1	Extrachromosomally Primed Retrotransposition (e.g., of a LINE)	
14.8.2	Target-Site Primed SINE Retrotransposition (e.g., of a SINE)	
14.9	<u>On the Evolution of Transposons, Genes and Genomes</u>	322
14.9.1	A Common Ancestry DNA and RNA (i.e., All) Transposons	
14.9.2	Retroviruses and LTR Retrotransposons Share a Common Ancestry	
14.9.3	Transposons Can Be Acquired by <i>Horizontal Gene Transfer</i>	
14.10	<u>Evolutionary Roles of Transposition in Genetic Diversity</u>	327
14.10.1	Transposons and <i>Exon Shuffling</i>	
14.10.2	Transposon Genes and Immune System Genes Have History	
14.11	<u>Coping with the Dangers of Rampant Transposition</u>	329
	<u>Chapter 15: DNA Technologies</u>	331
15.1	<u>Introduction</u>	331
15.2	<u>Make and Screen a cDNA Library</u>	333
15.2.1	cDNA Construction	
15.2.2	Cloning cDNAs into plasmid vectors	
15.2.3	Preparing recombinant plasmid vectors containing cDNA inserts	
15.2.4	Recombining plasmids and cDNA inserts and transforming host cells	
15.3	<u>DNA sequencing</u>	342
15.3.1	Manual DNA Sequencing	
15.3.2	Automated First Generation DNA Sequencing	
15.3.3	Shotgun Sequencing	
15.3.4	Next Generation DNA Sequencing	
15.4	<u>Genomic Libraries</u>	349
15.4.1	Preparing Specific Length Genomic DNA for Cloning; the Southern Blot	
15.4.2	Recombining Size-Restricted Genomic DNA with Phage DNA	
15.4.3	Creating Infectious Viral Particles with Recombinant phage DNA	
15.4.4	Screening a Genomic Library; Titering Recombinant Phage Clones	
15.4.5	Screening a Genomic Library; Probing the Genomic Library	
15.4.6	Isolating the Gene	
15.5	<u>The Polymerase Chain Reaction (PCR)</u>	356
15.5.1	PCR – the Basic Process	
15.5.2	The Many Uses of PCR	
15.6	<u>Genomic Approaches: The DNA Microarray</u>	360
15.7	<u>Ome-Sweet-Ome</u>	363
15.8	<u>From Genetic Engineering to Genetic Modification</u>	364
	<u>Chapter 16: Membrane Structure</u>	369
16.1	<u>Introduction</u>	369
16.2	<u>Plasma Membrane Structure</u>	370
16.2.1	The Phospholipid Bilayer	
16.2.2	Models of Membrane Structure	
16.2.3	Evidence for Membrane Structure	
16.2.4	Chemical Factors Affecting Membrane Fluidity	
16.2.5	Making and Experimenting with Artificial Membranes	
16.2.6	Separate Regions of a Plasma Membrane with Unique Fluidity and Permeability Properties	
16.3	<u>Membrane Proteins</u>	381
16.4	<u>A Diversity of Membrane Protein Functions</u>	384
16.5	<u>Glycoproteins and Glycolipids</u>	386

16.6	<u>Glycoproteins and Human Health</u>	387
Chapter 17: <u>Membrane Function</u>		391
17.1	<u>Introduction</u>	391
17.2	<u>Membrane Transport</u>	393
17.2.1	Passive Diffusion of Solutes	
17.2.2	Facilitated Diffusion of Solutes and Ions	
	Carrier Proteins	
	Ion Channels	
17.3	<u>Osmosis</u>	397
17.3.1	Osmosis in Plant and Animal Cells	
17.3.2	Osmosis in Plant Life	
17.3.3	Osmosis in Animal Life	
17.3.4	Summing Up	
17.4	<u>Active Transport</u>	401
17.5	<u>Ligand and Voltage Gated Channels in Neurotransmission</u>	403
17.5.1	Measuring Ion Flow and Membrane Potential with a Patch-Clamp Device	
17.5.2	Ion Channels in Neurotransmission	
17.6	<u>Endocytosis and Exocytosis</u>	407
17.6.1	Endocytosis	
17.6.2	Exocytosis and the Formation of Protein Storage Organelles	
17.7	<u>Directing the Traffic of Proteins in Cells</u>	411
17.7.1	Proteins Packaged in RER are Made as Larger Precursor Proteins	
17.7.2	Testing the Signal Hypothesis for Packaging Secreted Protein in RER	
17.8	<u>Synthesis of Integral Membrane Proteins</u>	415
17.9	<u>Moving and Sorting Proteins to Their Final Destinations</u>	415
17.9.1	Traffic on the Endomembrane Highway	
17.9.2	Nuclear Protein Traffic	
17.9.3	Mitochondrial Protein Traffic	
17.10	<u>How Cells are Held Together and How They Communicate</u>	420
17.10.1	Cell Junctions	
17.10.2	Microvesicles and Exosomes	
17.10.3	Cancer and Cell Junctions	
17.11	<u>Signal Transduction</u>	425
17.11.1	G-Protein Mediated Signal Transduction by PKA (Protein Kinase A)	
17.11.2	Signal Transduction using PKC (Protein Kinase C)	
17.11.3	Receptor Tyrosine Kinase-Mediated Signal Transduction	
17.12	<u>Signal Transduction in Evolution</u>	431
Chapter 18: <u>The Cytoskeleton and Cell Motility</u>		434
18.1	<u>Introduction</u>	434
18.2	<u>Overview of Cytoskeletal Filaments and Tubules</u>	435
18.3	<u>The Molecular Structure and Organization of Cytoskeletal Components</u>	436
18.4	<u>Microtubules: Dynamic Structures Composed of Tubulin Monomers</u>	438
18.4.1	The Two Kinds of Microtubules in Spindle Fibers	
	Kinetochore Microtubules	
	Polar Microtubules	
18.4.2	Microtubules in Cilia and Flagella	
18.4.3	Microtubule Motor Proteins Move Cargo from Place to Place in Cells	
18.4.4	Demonstrating Sliding Microtubules	
18.4.5	The Motor Protein Dynein Enables Axonemes to Bend	
18.5	<u>Microfilaments – Structure and Role in Muscle Contraction</u>	447
18.5.1	The Thin (Micro-) Filaments and Thick Filaments of Skeletal Muscle	
18.5.2	The Sliding Filament Model of Skeletal Muscle Contraction	

18.5.3	The Contraction Paradox: Contraction and Relaxation Require ATP	
18.6	<u>Actin-Myosin Interactions In Vitro: Dissections and Reconstitutions</u>	450
18.7	<u>Allosteric Change and the Microcontraction Cycle</u>	455
18.8	<u>The Microcontraction Cycle Resolves the Contraction Paradox</u>	456
18.9	<u>Ca⁺⁺ Ions Regulate Skeletal Muscle Contraction</u>	457
18.9.1	Muscle Contraction Generates Force	
18.9.2	The Elastic Sarcomere: Do Myosin Rods Just Float in the Sarcomere?	
18.10	<u>Actin Microfilaments in NonMuscle Cells</u>	463
18.11	<u>Both Actins and Myosins are Encoded by Large Gene Families</u>	465
18.12	<u>Intermediate Filaments</u>	465
 <u>Chapter 19: Cell Division and the Cell Cycle</u>		469
19.1	<u>Introduction</u>	469
19.2	<u>Cell Division in a Prokaryote</u>	470
19.3	<u>Cell Division in Eukaryotes</u>	471
19.3.1	Defining the Phases of the Eukaryotic Cell Cycle	
19.3.2	When Cells Stop Dividing...	
19.4	<u>Regulation of the Cell Cycle</u>	477
19.4.1	Discovery and Characterization of <i>Maturation Promoting Factor (MPF)</i>	
19.4.2	Other Cyclins, CDKs and Cell Cycle Checkpoints	
	The G ₁ Checkpoint	
	The G ₂ Checkpoint	
	The M Checkpoint	
	The G ₀ State	
19.5	<u>When Cells Die</u>	483
19.6	<u>Disruption of the Cell Cycle Checkpoints Can Cause Cancer</u>	485
19.7	<u>p53 Protein Mediates Normal Cell Cycle Control</u>	486
19.7.1	p53 is a DNA-Binding Protein	
19.7.2	How p53 Works to Salvage Cells	
19.7.3	How p53 Works When Cells Can't be Saved	
19.8	<u>The Centrality of p53 Action in Cell Cycle Regulation</u>	491
19.8.1	'Oncogenic Viruses'	
19.8.2	p53 and Signal Transduction	
19.9	<u>Cancer Cell Growth and Behavior; Cancer Treatment Strategies</u>	492
19.9.1	Cancer Cell Origins, Growth and Behavior	
19.9.2	Cancer Treatment Strategies	
 <u>Chapter 20: The Origins of Life</u>		498
20.1	<u>Introduction</u>	498
20.2	<u>Thinking about Life's Origins: A Short Summary of a Long History</u>	503
20.3	<u>Formation of Organic Molecules in an Earthly <i>Reducing Atmosphere</i></u>	504
20.3.1	Origins of Organic Molecules and a Primordial Soup	
20.3.2	The Tidal Pool Scenario for an Origin of Polymers and Replicating Chemistries	
20.4	<u>Origins of Organic Molecules in a <i>NON-Reducing Atmosphere</i></u>	508
20.4.1	Panspermia – an Extraterrestrial Origin of Earthly Life	
20.4.2	Extraterrestrial Origins of Organic molecules	
20.5	<u>Organic Molecular Origins of Life Closer to Home</u>	510
20.5.1	Origins in a High-Heat Hydrothermal Vent (<i>Black Smoker</i>)	
20.5.2	Origins in an Alkaline Deep-Sea Vent (<i>White Smoker</i>)	
20.6	<u>Heterotrophs-First vs Autotrophs-First: Some Evolutionary Considerations</u>	514
20.7	<u>Life Origins, A Summing Up</u>	514
20.8	<u>Origins of Life Chemistries in an RNA World</u>	515
20.9	<u>Experimental Evidence for an RNA World</u>	517
20.10	<u>Molecules Talk: Selecting Molecular Communication and Complexity</u>	518

20.10.1	Intermolecular Communication: Establishment of Essential Interconnected Chemistries	
20.10.2	Origins of Coordination	
20.10.3	An RNA World: Origins of Information Storage and Retrieval	
30.10.4	From Self-Replicating RNAs to Ribozymes to Enzymes; From RNA to DNA Ribozymes Branch Out: Replication, Transcription and Translation Transfer of Information Storage from RNA to DNA	
20.11	<u>The Evolution of Biochemical Pathways</u>	527
20.12	<u>A Grand Summary and Some Conclusions</u>	528
	<u>Epilogue</u>	531
	<u>Appendix I: List of Figures and Sources</u>	537
	<u>Appendix II: Context-Embedded YouTube Videos</u>	563

Chapter 1

Cell Tour, Life's Properties and Evolution, Studying Cells

Life's domains, scientific method, cell structures, Study methods (microscopy, cell fractionation, functional analyses); Common ancestry, Genetic variation, Evolution, Species diversity

Reminder: For inactive *links*, google key words/terms for alternative resources.



CELLS: *Left*, Robert Hooke's drawing of cork slices seen through a microscope from his 1665 *Micrographia*; *Right*, a monk's cell.

1.1 Introduction

You will read in this book about experiments that revealed secrets of cell and molecular biology, many of which earned their researchers Nobel and other prizes. But let's begin here with a *Tale of Roberts*, two among many giants of science in the renaissance and age of enlightenment whose seminal studies came too early to win such coveted prizes.

One of these, **Robert Boyle**, was born in 1627 to wealthy, aristocrat parents. In his teens, after the customary *Grand Tour* of renaissance Europe (France, Greece, Italy...) and the death of his father, he returned to England in 1644, heir to great wealth. In the mid 1650s he moved from his estates where he had already set about studying physics and chemistry, to Oxford. There he built a laboratory with his own money to do experiments on the behavior of gasses under pressure. With some help, he discovered *Boyle's Law*, confirming that the gasses obey mathematical rules. He is also credited with showing that light could travel through a vacuum, that something in air enables combustion, that sound travels through air in waves, that heat and particulate motion were related, and that the practice of alchemy was bogus! In fact, Boyle pretty much converted alchemy to chemistry by performing *chemical analysis*, a term he coined.

As a chemist, he also rejected the old Greek concept of the elements: earth, air, fire, and water. Instead, he defined elements as we still do today: the element is the smallest component of a substance that cannot be further chemically subdivided. He did this a century before Antoine Lavoisier listed and defined the first elements! Based on his physical studies

and chemical analyses, Boyle even believed that the indivisible units of elements were atoms, and that the behavior of elements could be explained by the motion of atoms. Finally, Boyle codified in print the scientific method that made him a successful experimental scientist.

The second of our renaissance Roberts was **Robert Hooke**, born in 1635. In contrast to Boyle parents, Hooke's parents were of modest means. They managed nonetheless to nurture their son's interest in things mechanical. While he never took the *Grand Tour*, he learned well and began studies of chemistry and astronomy at Christ Church College, Oxford in 1653. To earn a living, he took a position as Robert Boyle's assistant. It was with Hooke's assistance that Boyle did the experiments leading to the formulation of *Boyle's Law*.

While at Oxford, he made other friends and useful connections. One friend was the architect Christopher Wren. In 1662, Boyle, a founding member of the Royal Society of London, supported Hooke to become the society's *curator of experiments*. However, to support himself, Hooke also hired on as professor of geometry at Gresham College (London). After "the great fire" of London in 1666, Hooke, as city surveyor and builder, participated with Christopher Wren in the design and reconstruction of the city. Ever interested in things mechanical, he also studied the elastic property of springs, leading him to *Hooke's Law*, namely that the force required to compress a spring was proportional to the length that the spring was compressed. Later, these studies led Hooke to imagine how a coil spring might substitute for a pendulum to regulate a clock. While he never invented such a clock, he was appointed to a Royal Commission to find the first reliable method to determine longitude at sea. He must have been gratified to know that the solution to accurate determination of longitude at sea turned out to involve a coil-spring clock! Along the way in his 'practical' studies, he also looked at little things, publishing his observations in *Micrographia* in 1665. Therein he described microscopic structures of animal parts and even snowflakes. He also described fossils as having once been alive and compared microscopic structures he saw in thin slices of cork to monk's cells (rooms, chambers) in a monastery. Hooke is best remembered for his law of elasticity and of course, for coining the word *cell*, which we now know as the smallest unit of living things.

Now fast-forward almost two hundred years to observations of plant and animal cells early in the nineteenth century. These observations revealed structural features common to all cells including a nucleus and a boundary (membrane or wall) and the common organization of cells in groups to form multicellular structures in lower life forms as well as in plants and animals. By the 1830s an enriched understanding of cell structure and the role of cells in the structure of animals and plants led botanist **Matthias Schleiden** and zoologist **Theodor Schwann** to propose the first two precepts of a unified ***Cell Theory***: (1) *Cells are the basic unit of living things*; (2) *Cells can have an independent existence*. Later in the century, when Louis Pasteur finally disproved *spontaneous generation* and German histologists observed *mitosis* and *meiosis* (the underlying events of eukaryotic cell division), **Rudolf Virchow** added a third precept to round out ***Cell Theory***: (3) *Cells come from pre-existing cells*. That is, they reproduce. We begin this chapter with a reminder of the ***scientific method***, that way of thinking about our world that emerged formally in the seventeenth century. Then we'll take a tour of the cell, reminding ourselves of basic structures and organelles. After the 'tour' we'll consider the ***origin of life*** from a common ancestral cell and the subsequent ***evolution*** of cellular complexity and the incredible diversity of life forms.

Finally, we consider some of the *methods* we use to study cells. Since cells are small, several techniques of microscopy, cell fractionation (in essence a biochemical dissection of the cell) and functional/biochemical analysis are described to illustrate how we come to understand cell function.

Learning Objectives

When you have mastered the information in this chapter, you should be able to do the following:

1. Compare and contrast *hypotheses* and *theories* and place them (and other elements of the scientific enterprise) into their place in the cycle of the *scientific method*.
2. Compare and contrast structures common to, and that distinguish *prokaryotes*, *eukaryotes*, and *archaea*, and groups within these *domains of life*.
3. Articulate the function of different cellular substructures.
4. Explain how *prokaryotes* and *eukaryotes* accomplish the same functions, i.e., have the same *properties of life*, even though prokaryotes lack most structures found in eukaryotes.
5. Outline a procedure to study a specific cell *organelle* or another substructure.
6. Describe or speculate on how the different structures (particularly in eukaryotic cells) relate/interact with each other to accomplish specific functions.
7. Describe some structural and functional features that distinguish prokaryotes (eubacteria), eukaryotes, and archaea.
8. Place cellular organelles and other substructures in their evolutionary context, i.e., describe their origins and the selective pressures that could have led to their *evolution*.
9. Distinguish between the roles of random *mutations* and *natural selection* in evolution.
10. Relate archaea to other life forms and speculate on their origins in evolution.
11. Suggest why evolution leads to more complex ways of sustaining life.
12. Explain how *fungi* are more like animals than plants.

1.2 Scientific Method – The Formal Practice of Science

Let's focus here on the essentials of the scientific method originally inspired by Robert Boyle, and then on how science is practiced today. *Scientific method* is one or another standardized protocol for observing, asking questions about, and investigating natural phenomena. Its simplest expressions are look/listen, infer, and test your inference. According to the Oxford English Dictionary, all scientific practice relies on the *systematic observation, measurement, and experiment, and the formulation, testing and modification of hypotheses*. Here is the scientific method as you might read it a typical science textbook:

- ***Read*** the science of others and ***observe*** natural phenomena on your own.
- ***Infer*** and state a ***hypothesis*** (explanation) based on logic and reason.
- Hypotheses are declarative sentences that sound like fact but aren't! Good hypotheses are testable ***predictions***, easily turned into *if/then statements* or *yes-or-no questions*.
- ***Design experiments*** to test the hypothesis: results must be ***measurable evidence*** for or against the hypothesis.
- ***Perform that experiment*** and then observe, measure, collect data, and test for statistical validity (where applicable). Then, ***repeat the experiment***.

- Consider how your data supports or does not support your hypothesis and then ***integrate your experimental results*** with earlier hypotheses and prior knowledge.
- Finally, ***publish*** (i.e., make public) your experiments, results, and conclusions. In this way, shared data and experimental methods can be evaluated (and repeated) by other scientists.

We'll return to the scientific method and how it is practiced shortly.

So, what are scientific *hypotheses*, *theories* and *laws* and how do they fit into the scientific method? A scientific *hypothesis*, as suggested above, is an inference, and *educated guess* about what might be going on based on evidence and logic. A hypothesis is a declarative sentence, for example "The Sun revolves around the Earth." This hypothesis was stated by **Aristotle** (among others)! Remember, a good hypothesis can be easily turned into a yes-or-no question, in this case "Does the sun revolve around the Earth?" By its nature, such yes-or-no questions can be answered (i.e., a good hypothesis can be tested) by gathering more evidence by observation and experiment. When Aristotle's hypothesis was finally tested by the observations and measurements of **Nicolaus Copernicus**, **Galileo Galilei** and others, it proved to be false! But you knew that, didn't you?

Contrary to what many people think, a ***scientific theory is not a guess***, neither an *educated* nor an *uneducated* one. Rather, a theory is a statement well supported by experimental evidence and widely accepted by the scientific community. Nevertheless, theories are not facts. Scientists know that theories are subject to further test and modification and may even be overturned. Even scientific ***laws*** can be questioned. Astrophysicists actively test otherwise universally accepted physical ***laws***, occasionally threatening to modify them. In biology, Mendel's *Law of Independent Assortment* shouldn't even be called a law. Indeed, it was not factual as he stated it, or for that matter when he stated it. Check the Mendelian Genetics section of an introductory textbook to see how chromosomal crossing over violates this law, and a history of science book to see what happens when observations or experimental results are inexplicable or as we might say today, 'too far out.'

Even Darwin's ***Theory of Evolution***, one of the most enduring and tested biological theories, has been modified over time. But in this case, the modifications have only strengthened our understanding that biological diversity is the result of natural selection. For commentary on the evolutionary underpinnings of biology, check out Dobzhansky T (1973, *Nothing in biology makes sense except in the light of evolution*. Am. Biol. Teach. 35:125-129), and Gould, S. J. (2002, *The Structure of Evolutionary Theory*. Boston, Harvard University Press). Or, check out some of Darwin's own work at ^{1,1}[Origin of Species](#).

Do you think that Darwin's *Theory of Evolution* by natural selection should be promoted to a law? To sum up, a Wikipedia entry states that *the goal of a scientific inquiry is to obtain knowledge in the form of testable explanations (hypotheses) that can predict the results of future experiments. This allows scientists to gain an understanding of reality, and later use that understanding to intervene in its causal mechanisms (such as to cure disease)*. The better a hypothesis is at making predictions, the more useful it is. In the last analysis, think of hypotheses as *educated guesses* and think of theories and/or laws not as proofs of anything, but as one or more experimentally supported hypothesis that everyone agrees should serve as *guideposts* to help us evaluate new observations and hypotheses.

But do not making the mistake of placing hypotheses at the low end of a hierarchy of ideas. They are in fact are the bread and butter of the scientific enterprise. Good ones are testable and should predict either/or results of well-designed experiments. Those results (observations, experimental data) should support or nullify the hypotheses being tested. In either case, scientific data generates conclusions that inevitably lead to new hypotheses whose predictive value will also be tested. If you get the impression that scientific discovery is a cyclic process, that's the point! Exploring scientific questions reveals more questions than answers!

A word about well-designed experiments. **Erwin Schrödinger** (winner of a **Nobel Prize** in physics in 1933) once proposed a *thought experiment*. He wanted his audience to understand the requirements of scientific investigation but gained a fame (and notoriety) far beyond the world of theoretical physics. Perhaps you have heard of his cat! Considered a founding father of quantum physics, he recognized that adherence to scientific method is not strict and that we can (and should) occasionally violate adherence to the dictates of scientific method. In the now popular story of *Schrödinger's Cat*, Schrödinger stated that if you sealed a cat in a box with a toxic substance, how could you know if the cat was alive or dead unless you open the box. Wearing his philosopher's hat (yes, he had one!), he postulated that until you open the box, the cat is both "dead and alive." That is, until the box was opened, the cat was in a sense, neither dead nor alive, but both! Often presented as little more than an amusing puzzle, Schrödinger was in fact illustrating that there were two alternate hypotheses: (1) *the cat exposed to toxin survived*, **or** (2) *the cat exposed to toxin died*. Note that either hypothesis is a declarative sentence, and that either one could be tested. Just open the box!

In a twist however, Schrödinger added that by opening the box, the investigator would become a factor in the experiment. For example, let's say (for the sake of argument) that you find a dead cat in the box. Is it possible that instead of dying from a poison, the cat was scared to death by your act of opening the box? Or that the toxin made the cat more likely to die of fright but was not lethal by itself? How then to determine whether it was the toxin or your action that killed the cat? This made the puzzle even more beguiling, and to the many laypersons, his greatest scientific contribution! But to a scientist, the solution to the puzzle just means that a scientist must take all possible outcomes of the experiment into account, including the actions of the experimenter, ensuring sound experimental design with all necessary controls. The bottom line, and often the reason that scientific manuscripts suffer negative peer review, is the absence or inadequacy of control experiments. See more about *Schrödinger's cat* at ^{1,2}[A Cat Video](#).

1.2.1 The Method as It Is Really Practiced!

If you become a scientist, you may find that adherence to the 'rules' of scientific method are honored as much in the breach as in their rigorous observance. An understanding of those rules, or more appropriately principles of scientific method guide prudent investigators to balance personal bias against the leaps of intuition that successful science requires. Deviations from protocol are allowed!

I think that we would all acknowledge that the actual practice of science by would be considered a success by almost any measure. *Science is a way of knowing* the world around us

through constant test, confirmation, and rejection that ultimately reveals new knowledge, integrating that knowledge into our worldview.

An element often missing but integral to any scientific method is that *doing science is collaborative*. Less than a century ago, many scientists worked alone. Again, Gregor Mendel is an example, and his work was not appreciated until decades after he published it. Today, most publications have two or more coauthors. And the inherent collaborative nature of science doesn't end with the investigators in a study. When a paper (or a research grant for that matter) is submitted for consideration, other scientists are recruited to evaluate the quality of hypotheses, experimental design, and soundness of conclusions reported in a manuscript. This *peer review* of fellow scientists is part and parcel of good scientific investigation.

1.2.2 Logic and the Origins of the Scientific Method

The scientist, defined as a both observer and investigator of natural phenomena, is only a few centuries old. Long before that, philosophers developed formal rules of *deductive* and *inferential logic* to try and understand nature, humanity's relationship to nature, and the relationship of humans to each other. We owe to those *philosophers* the logical basis of the scientific enterprise. They came up with the rules and systems of *deductive* and *inductive logic* now integral to the practice of science. Scientific method grew from those beginnings, along with increasing empirical observation and experimentation. We recognize these origins when we award the Ph.D. (*Doctor of Philosophy*), our highest academic degree! We are now going to learn about the life of cells, their structure and function, and their classification or grouping based on those structures and functions. Everything we know about life comes from applying the principles of scientific method to our intuition. For a bemused take on how scientists think, check out *The Pleasure of Finding Things Out: The Best Short Works of Richard Feynman* (1999, New York, Harper Collins).

1.3 Domains of Life

We believe with good reason that all life on Earth evolved from a common ancestral cell that existed soon after the origins of life on our planet. At one time, all life was divided into two groups: the true bacteria and everything else! Now we group life into one of three ***domains***:

- ***Prokaryotes*** are among the first descendants of that common ancestral cell. They lack nuclei (*pro* meaning *before* and *karyon* meaning *kernel*, or *nucleus*). They include *bacteria* and *cyanobacteria* (blue-green algae).
- ***Eukaryotes*** include all higher life forms, characterized by cells with true nuclei (*Eu*, true; *karyon*, *nucleus*).
- ***Archaeobacteria***, (meaning "old" bacteria) include many ***extremophile*** bacteria ('lovers' of life at extreme temperatures, high salinity, and the like). Originally classified as ancient prokaryotes, *Archaeobacteria* were shown by 1990 to be separate from prokaryotes and eukaryotes, in fact a third domain of life.

The archaea are found in such inhospitable environments as boiling hot springs or arctic ice, though some also live in conditions that are more temperate. Based on comparison of the DNA sequences of genes for ribosomal RNAs in eukaryotes, normal bacteria (*eubacteria*) and extremophiles, Carl Woese proposed the three-domain phylogeny illustrated in Figure 1.1.

Carl Woese's Three-Domain Phylogeny

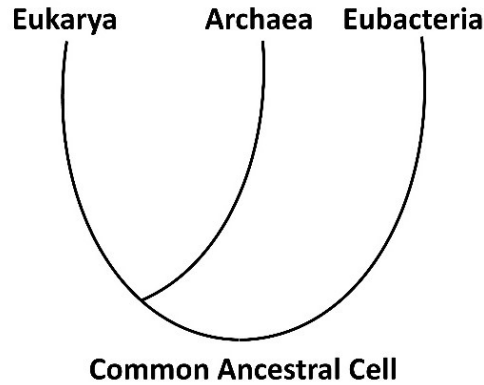


Fig. 1.1: A three-domain phylogeny showing a closer relationship between archaea and eukaryotes.

Based on sequence similarities and differences, Woese concluded that the archaeobacteria (Archaea) are not just a domain separate from the rest of the bacteria and from eukaryotes but are more closely related to eukarya than eubacteria! In fact, the Archaea share genes, proteins, and metabolic pathways found in eukaryotes but *not* in bacteria, supporting their close evolutionary relationship to eukaryotes. This unique sharing is further testimony to their domain status. Understanding that all living organisms belong to one of three domains has dramatically changed our understanding of evolution. Archaea may be prokaryotes, but their immediate ancestors are more closely related to us than to *E. coli*! For a review, see (Woese, C. 2004, *A new biology for a new century*. Microbiol. Mol. Biol. Rev. 68:173-186) The three domains of life (**Archaea**, **Eubacteria** and **Eukarya**) quickly replaced the older division of living things into Five Kingdoms, the *Monera* (*prokaryotes*), *Protista*, *Fungi*, *Plants*, and *Animals* (*all eukaryotes!*).

At this point you may be asking, "What about viruses?" Where are they on, or do they even belong in the tree of life? You may already know that viruses require live cellular hosts to reproduce, but that they are not themselves alive. In fact, much about the place of viruses in evolution is an open question that we will consider in a later chapter. For now, let's look at how we come to know about viruses and some of their peculiarities.

1.3.1. Viruses: Dead or Alive; Big and Small - A History of Surprises

Viruses that infect bacteria are called *bacteriophage* (phage meaning eaters, hence *bacteria eaters*). Eukaryotic viruses include DNA and RNA viruses, with DNA and RNA genomes. Smallpox, hepatitis B, herpes, chicken pox/shingles, and adenovirus are caused by DNA viruses. Common colds, influenza, SARS, and COVID-19 are caused by *positive strand RNA viruses* that upon infecting a cell, replicate their RNA genome to make RNA *negative strand RNAs* encoding all necessary information to make new viruses. HIV AIDS, Ebola, Zika, yellow fever, and some cancers are caused by *retroviruses*, RNA viruses whose genome is reverse-transcribed into a cDNA intermediate that replicates and is transcribed to generate new

viruses. Viruses were not identified as agents of disease until late in the nineteenth century, and we have learned much in the ensuing century. In 1892, **Dmitri Ivanofsky**, a Russian botanist, was studying plant diseases. One that damaged tobacco (and was therefore of agricultural significance) was the *mosaic disease* (Figure 1.2, below).

Tobacco Mosaic Virus – Infected Leaf



Fig. 1.2: Tobacco mosaic virus symptoms (white patches) on a tobacco leaf.

Ivanofsky showed that extracts of infected tobacco leaves were themselves infectious. The assumption was that the extracts would contain infectious bacteria. But his extracts remained infectious even after passing them through a *Chamberland-Pasteur* filter with a pore size so small that bacteria would not pass into the filtrate. Thus, the infectious agent(s) couldn't be bacterial. Since the infectious material was not cellular and depended on a host for reproduction with no independent life of its own, they were soon given the name *virus*, a term that originally just meant *toxin*, or *poison*. This marked the start of **virology**, the study of viruses. The virus that Ivanofsky studied is now called *Tobacco Mosaic Virus*, or TMV.

Invisible by light microscopy, viruses are sub-microscopic non-cellular bits of life-chemistry that only become reproductive (come alive) when they parasitize a host cell. Since many viruses cause disease in humans, we have learned much about how they are similar and how they differ. In other chapters, we'll learn how viruses have even become tools for the study of cell and molecular biology. Let's start with a recent surprise from the study of viruses.

As eventually seen in the electron microscope, viruses (called virions or viral particles) are typically 150 nm or less in diameter. And that is how we have thought of viruses for over a century! But in 2002, a *particle* inside an amoeba, originally believed to be a bacterium, was shown by electron microscopy to be a *giant virus*! Since then, several more *giant*, or *Megavirales* were discovered.

Megavirales fall into two groups, *pandoraviruses* and *mimiviruses*. At 1000nm (1 μm) *Megavirus chilensis* (a *pandoravirus*) may be the largest. Compare a few giant viruses to a bacterium (*E. coli*) and the AIDS virus in Figure 1.3 below.

An *E. Coli* Bacterium and the *HIV* AIDS Virus Compared to Some *Giant Viruses* at Scale ()

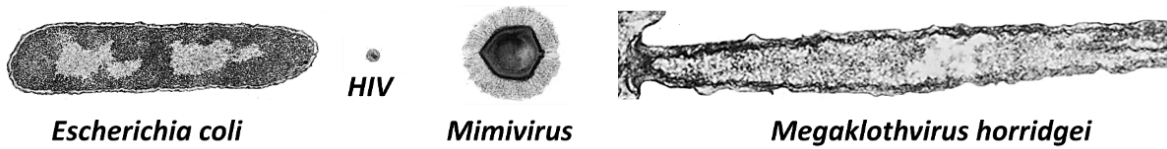


Fig. 1.3: Transmission electron micrographs of 2 giant viruses, the AIDS (HIV) virus and an *E. coli* bacterium. *M. horridgei* is twice the size of *E. coli*. All giant viruses dwarf HIV, a typical eukaryotic virus.

Consider that a typical virus contains a small genome, encoding an average of ten genes. In contrast, the *Megaviurus chilensis* genome contains 2.5×10^6 base pairs (bp) encoding up to 1,100 proteins. Still, it requires host cell proteins to infect and replicate. More surprising is that 75% of the sequenced 1.2×10^6 bp *mimivirus* genome code putative proteins *with no counterparts* in other viruses or cells! Equally surprising, some mimiviruses genes encode proteins homologous to those used for translation in prokaryotes and eukaryotes. If all viruses, including the *Megavirales*, only use host cell enzymes and ribosomal machinery to synthesize proteins, what are these genes doing in a mimivirus genome? Think of the surprises here as questions. The big ones concern where and when *Megavirales* (giant viruses) evolved:

- What are those genes with no cellular counterparts all about?
- What were the selective advantages to a virus of large size and a large genome?
- Were *Megavirales* once large cells that invaded other cells, eventually becoming viral parasites and losing most but not all their genes? Or were they once small viruses that incorporated host cell genes, increasing their genome size and coding capacity?

Viruses are typically identified because they are harmful. The discovery of HIV earned Luc Montagnier, Françoise Barré-Sinoussi, and Harald Z. Hausen the 2008 Nobel Prize in Physiology or Medicine. More recently, as this is written, we are (we hope) near the end of the COVID-19 viral pandemic caused by the SARS-CoV-2 retrovirus. In fact, few viruses resident in humans are known to be beneficial.

In contrast, some bacteria are harmful to humans and animals but many are beneficial (even necessary) as symbionts in our many *microbiomes*.

Let's now turn our attention to cells, entities that we define as living, with *all* the *properties of life...*, starting with *eubacteria*.

1.3.2 The Prokaryotes (Eubacteria = *Bacteria* and *Cyanobacteria*)

Prokaryotic cells lack nuclei and other eukaryotic organelles, such as mitochondria, chloroplasts, endoplasmic reticulum, and assorted eukaryotic vesicles and internal membranes. Transmission and scanning electron micrographs and an illustration of rod-shaped bacteria are shown below (Figure 1.4).

Structure of a Bacterium

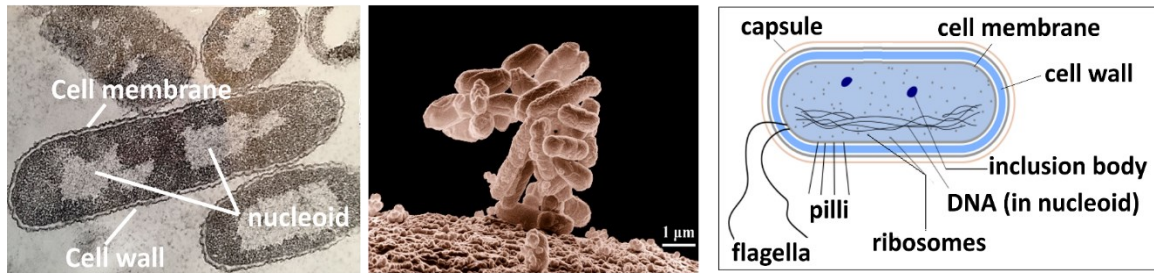


Fig. 1.4: Transmission electron micrograph (left) and drawing of the gram-negative *E. coli* bacterium (right) labeling its structural components, and a scanning electron micrograph of a cell cluster (middle).

Bacteria do contain *bacterial microcompartments* (BMCs), but these are made up entirely of protein and are **not** surrounded by a phospholipid membrane. These function for example in CO₂ fixation to sequester metabolites toxic to the cells. Check out ^{1,5}[Bacterial Organelles](#) for more information. Bacteria are typically unicellular, although a few (like some cyanobacteria) live colonial lives at least some of the time.

1.3.2.a Bacterial Reproduction

Without the compartments afforded by the internal membrane systems of eukaryotic cells, all intracellular chemistries, (reproduction and gene expression (DNA replication, transcription, translation, and all the metabolic biochemistry of life) happen in the cytoplasm. Bacterial DNA is a circular double helix that duplicates as the cell grows. While not enclosed in a nucleus, bacterial DNA is concentrated in a region of the cell called the **nucleoid**. When not crowded at high density, bacteria replicate their DNA throughout the life of the cell, dividing by **binary fission**. The result is the equal partition of duplicated bacterial *chromosomes* into new cells. The bacterial chromosome is basically naked DNA, unassociated with proteins.

1.3.2.b Cell Motility and the Possibility of a Cytoskeleton

Movement of cells is a response to environment. Some respond to chemicals (chemotaxis), some to light (*phototaxis*) or even gravity (*geotaxis*). Bacteria move to or from nutrients, noxious chemicals, light, dark, gravitational force, etc., by one of several mechanisms. Some use a flagellum made up largely of the bacterial protein **flagellin**. The main proteins of eukaryotic cell flagella and cilia are the **tubulins**. Together with *actin* and other proteins, tubulins are also part of the eukaryotic cell cytoskeleton of rods and tubes. Prokaryotes were long thought to lack similar cytoskeletal components. But two bacterial homologues of eukaryotic actin and tubulin genes were recently discovered. **MreB** is the actin homologue. Like actin, *MreB* monomers polymerize to form filaments that lie under the cell membrane of bacteria (e.g., *E. coli*), helping to maintain their rod-like shape. In fact, *E. coli* with a mutant *MreB* gene is spherical..., and normally spherical bacteria lack an *MreB* gene! *MreB* was also thought to form an actin-like **cortical ring** that in dividing eukaryotic cells constricts to pinch off two new cells. But this function seems to be served by the **FtsZ** protein that encodes a eukaryotic *tubulin homologue*. *FtsZ* polymers form filaments that are seen in a **Z ring** at the center of a bacterial cell during *binary fission*. *FtsZ* mutants divide, but abnormally. Thus, the

role of *FtsZ* in separating bacterial cells during *binary fission* is not yet clear. Figure 1.5 shows a micrograph and an illustration of *FtsZ* in the Z rings of dividing *E. coli* cells.

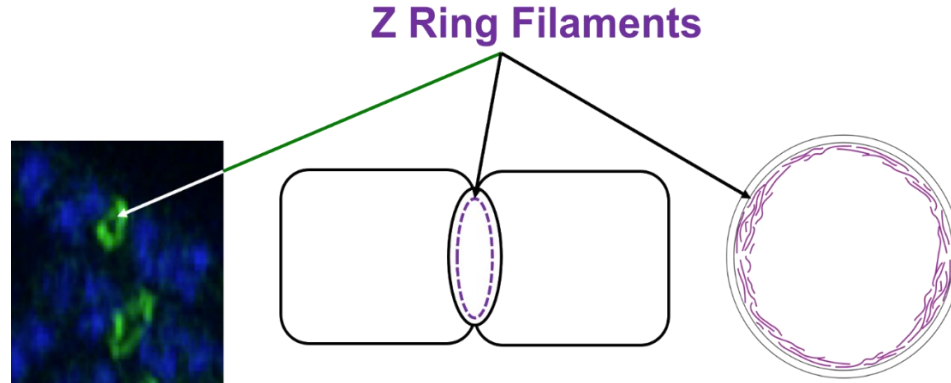


Fig. 1.5: Immunofluorescence localization of *FtsZ* proteins in Z rings in cross-sections in the middle of dividing *E. coli* cells (left). Drawings represent the location of *FtsZ* protein in Z rings between dividing cells (middle) and in cross section (right).

It seems that together with flagellin, the *MreB* and *FtsZ* proteins may be part of a primitive prokaryotic *cytoskeleton* involved in cell structure and motility, from which our own evolved!

1.3.2.c Some Bacteria Have Internal Membranes

While bacteria lack organelles (the membrane-bound structures of eukaryotic cells), internal membranes in some bacteria form as inward extensions, or *invaginations* of plasma membrane. Some of these capture energy from sunlight (photosynthesis) or from inorganic molecules (*chemolithotrophy*). Photosynthetic vesicles called *Carboxysomes* (Figure 1.6) are membrane bound structures in which CO_2 is fixed (reduced) in cyanobacteria. Photosynthetic bacteria have less elaborate internal membrane systems.

Cyanobacterial Carboxysomes

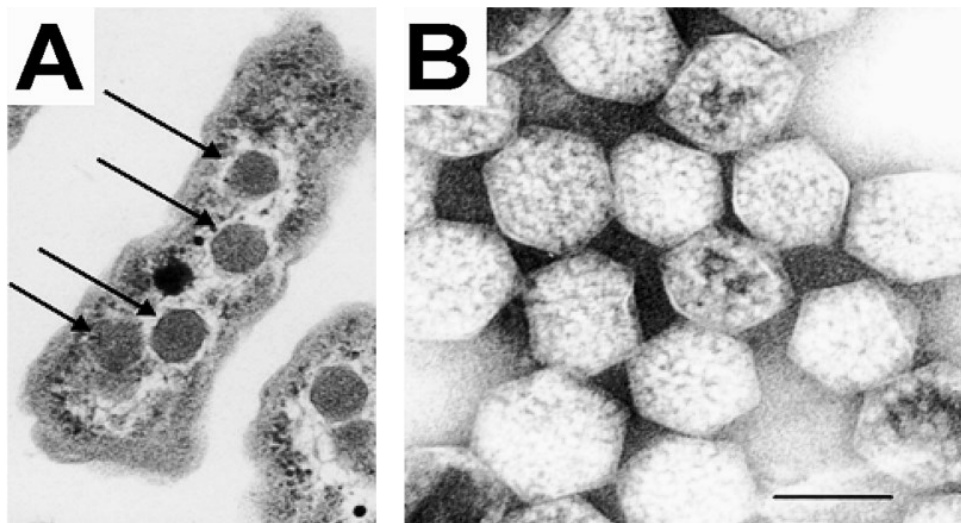


Fig. 1.6: Transmission electron micrograph of carboxysomes in a cyanobacterium (arrows, left) and isolated from a cyanobacterium (right).

1.3.2.d Bacterial Ribosomes Do the Same Thing as Eukaryotic Ribosomes... and Look Like Them!

Ribosomes are protein-synthesizing machines. Those of prokaryotes are smaller than those of eukaryotes but can translate eukaryotic messenger RNA (mRNA) in vitro. This is because the sequences and structures of ribosomal RNAs are shared by all species, indicating long conserved evolutionary relationships. Recall that it was ribosomal sequence similarities that revealed our closer relationship to archaea than bacteria.

The prokarya (*eubacteria*) are a diverse group, occupying almost every wet, dry, or hot and cold nook-and-cranny of our planet. Yet, all prokaryotic cells share structural and functional metabolic properties with each other and with archaea and eukaryotes! As we've seen with ribosomes, this sharing supports the common ancestry of all life.

Finally, we share not only common ancestry, but living arrangements with bacteria. There are microbiomes in our gut, on our lips, in belly buttons, and in fact all over our skin (see ^{1.6}[Our Skin Microbiome](#) for more about *that!*). Gut microbiome bacteria alone number ~10X more than our own cells! And microbiomes are invisible but not quiet (^{1.7}[The Human Microbiome](#)). Interest in our microbiomes even earned them their own ^{1.8}[The NIH Human Microbiome Project](#).

1.3.3 The Archaeobacteria (Archaea)

Allessandro Volta, a physicist who gave his name to the 'volt' (electrical potential energy), discovered methane producing bacteria (*methanogens*) way back in 1776! He found them living in the extreme environment at the bottom of Lago Maggiore, a lake shared by Italy and Switzerland. These unusual bacteria are *chemoautotrophs* that get energy from H₂ and CO₂ and generate methane gas in the process.

It was not until the 1960s that Thomas Brock (at the University of Wisconsin-Madison) discovered *thermophilic* bacteria living at temperatures approaching 100°C in Yellowstone National Park in Wyoming.

Organisms living in any extreme environment were soon nicknamed *extremophiles*. One of the thermophilic bacteria, now called *Thermus aquaticus*, became the source of *Taq* polymerase, the heat-stable DNA polymerase that made the *polymerase chain reaction* (PCR) possible scalable (i.e., rapid!). PCR is now a household name in labs around the world!

Extremophile and "normal" bacteria are similar in size and shape(s) and lack nuclei. This initially suggested that most extremophiles were prokaryotes. But as Carl Woese demonstrated, it is the archaea and eukarya that share a more recent common ancestry! While some bacteria and eukaryotes can live in extreme environments, the archaea include the most diverse extremophiles. Here are some extremophilic archaeobacteria:

- *Acidophiles* grow at acidic (low) pH.

- *Alkaliphiles* grow at high pH.
- *Halophiles* require high [salt] (e.g., *Halobacterium salinarium* (Figure 1.7)).
- Heat-loving *Thermophiles* and *hyperthermophiles* live at high temperatures. *Thermus aquaticus* (Figure 1.8) is a thermophile that lives at 70°C and is noted for its role in developing the polymerase chain reaction. *Pyrolobus fumarii* is a hyperthermophile, lives at 113°C!

Thermophilic Bacteria

Fig. 1.7: *Pyrolobus fumarii* **Fig. 1.8: *Thermus Aquaticus***

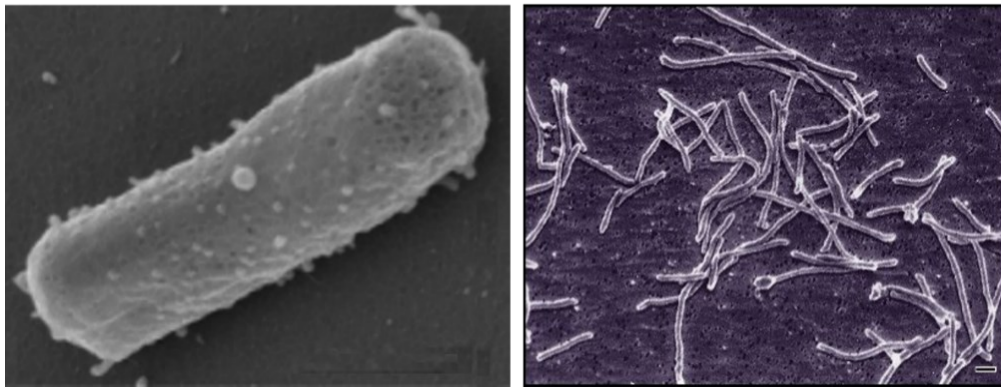


Fig. 1.7 (left): Scanning electron micrograph of *Halobacterium salinarium*, a 'salt-loving' bacterium (**Fig. 1.7**). Right: Scanning electron micrograph of 'heat-loving' *Thermus aquaticus* bacteria (**Fig. 1.8**).

- *Methanogens* produce methane.
- *Barophiles* grow best at high hydrostatic pressure.
- *Psychrophiles* grow best at temperature 15 °C or lower.
- *Xerophiles* grow at very low *water activity* (i.e., drought or near drought conditions).
- *Toxicolerants* grow in the presence of high levels of damaging chemicals, for example, pools of benzene, nuclear waste.

Archaea were originally seen as oddities of life, thriving in unfriendly environments. But they include organisms living in less extreme environments, including soils, marshes, and even in the human colon. They are also abundant in the oceans where they are a major part of plankton, participating in the carbon and nitrogen cycles. In the guts of cows, humans, and other mammals, methanogens facilitate digestion, generating methane gas in the process. In fact, cows have even been cited as a major cause of global warming because of their prodigious methane emissions!

On the plus side, methanogenic Archaea are being exploited to create biogas and to treat sewage. Other extremophiles are the source of enzymes that function at high temperatures or in organic solvents. As already noted, some of these have become part of the biotechnology toolbox.

1.3.4 The Eukaryotes

The volume of a typical eukaryotic cell is some 1000 times that of a typical bacterial cell. Imagine a bacterium as a 100 square foot room (the size of a small bedroom, or a large walk-in closet!) with one door. Now imagine a room 1000 times as big. That is, imagine a 100,000 square foot 'room'. You might expect many smaller rooms inside this room for such a large space to be functional. The eukaryotic cell is a lot like that large space, with lots of interior *rooms* (i.e., organelles) with their own entryways and exits. In fact, eukaryotic life would not even be possible without a division of labor of eukaryotic cells among different *organelles* (the equivalence to the small rooms in our metaphor).

The smaller prokaryotic "room" has a much larger plasma membrane *surface area-to-volume ratio* than a typical eukaryotic cell. This enables required environmental chemicals to enter and quickly diffuse throughout the cytoplasm of e.g., an *E. coli* cell. The communication between chemicals and structures in a small cell is therefore rapid. In contrast, the communication over a larger expanse of cytoplasm inside a eukaryotic cell requires the coordinated (not to mention regulated!) activities of subcellular components and compartments. Such communication can be relatively slow in a large space. In fact, eukaryotic cells have lower rates of metabolism, growth, and reproduction than prokaryotic cells. Thus, the existence of large cells required the evolution of divided labors supported by *compartmentalization*.

Fungi are more closely related to animal than plant cells. They are a curious beast for several reasons! For one thing, the organization of fungi and fungal cells is somewhat less defined than animal cells. *Septa* are structures between cells that separate fungal *hyphae*. They allow passage of cytoplasm and even organelles between cells. On the other hand, some primitive fungi have few or no septa, in effect creating *coenocytes*, which are single giant cell with multiple nuclei. Fungal cells are surrounded by a wall, whose principal component is *chitin*. Chitin is the same material that makes up the exoskeleton of *arthropods* (which includes insects and lobsters!).

We end this look at the domains of life by noting that, while eukaryotes are a tiny minority of all living species, "their collective worldwide biomass is estimated to be equal to that of prokaryotes" (Wikipedia). And we already noted that the bacteria living commensally with us humans represent 10 times as many cells as our own human cells! Clearly, each of us (and probably most animals and even plants) owes our existence to its microbiome as much we do to our own human cells. For now, keeping in mind that plants and animal cells share many internal structures and organelles that perform the same or similar functions, let's look at them and briefly describe their functions.

Typical animal and plant cells with organelles and other structures are illustrated below, in Figure 1.9 and in Figure 1.10).

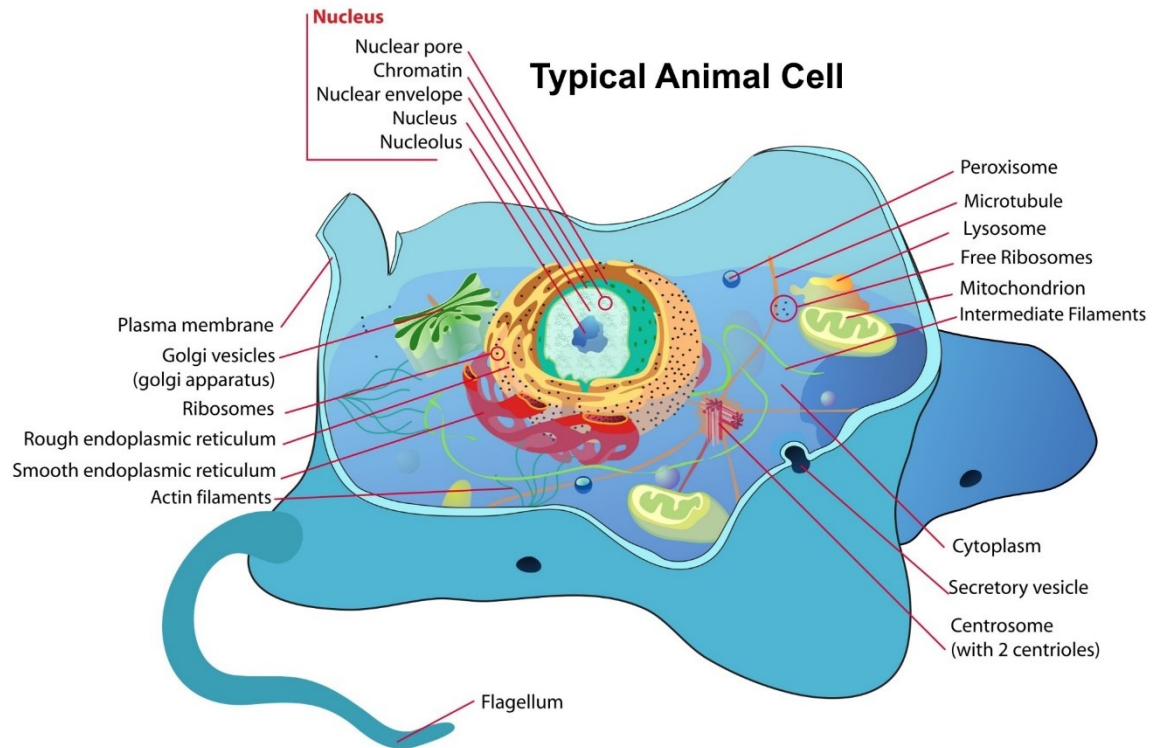


Fig. 1.9: Labeled drawing of the structural components of a typical animal cell.

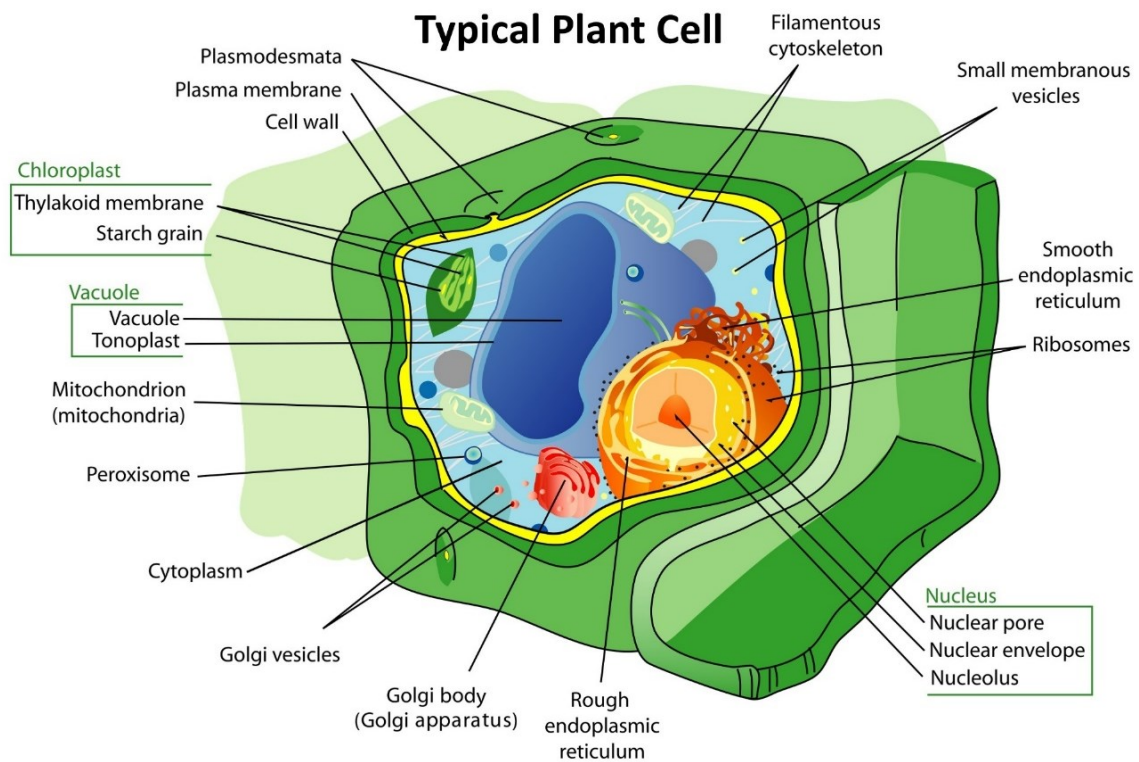


Fig. 1.10: Labeled drawing of the structural components of a typical plant cell.

1.4 Tour of the Eukaryotic Cell

Here we take a closer look at the division of labors among the organelles and structures within eukaryotic cells. We'll look at cells and their compartments in a microscope and see how the organelles and other structures were isolated from cells and identified not only by microscopy, but by biochemical and molecular analysis of their isolates.

1.4.1 The Nucleus

The nucleus is the largest organelle in the cell, separating the genetic blueprint (DNA) from the cell cytoplasm. Although the eukaryotic nucleus breaks down during mitosis and meiosis as chromosomes form and cells divide, it spends most of its time in its familiar form during **interphase**, the time between cell divisions. The structural organization of an interphase nucleus is shown in Figure 1.11 below.

Structure and Organization of the Eukaryotic Nucleus

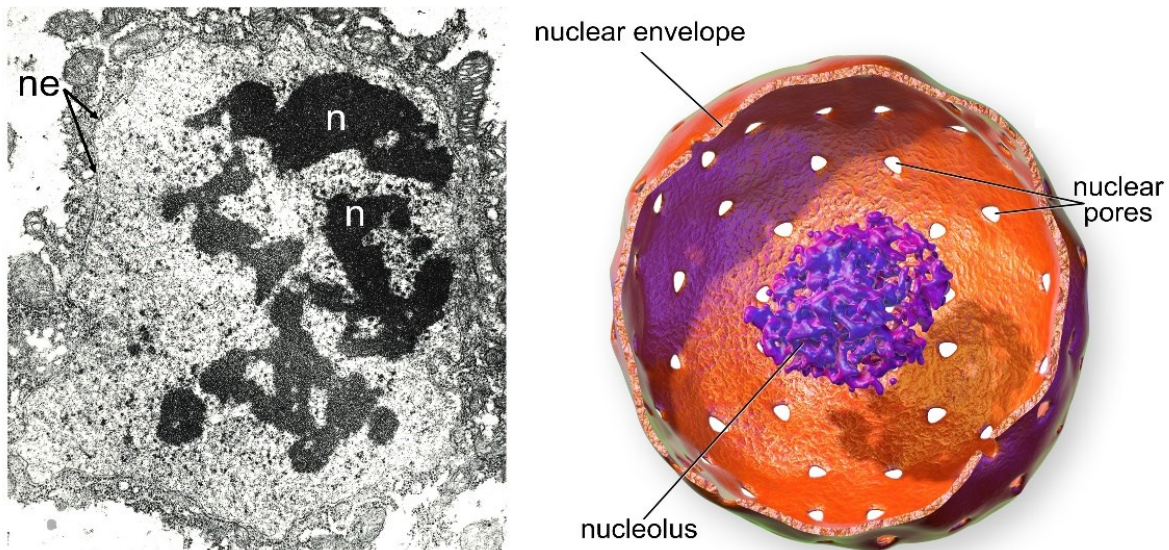


Fig. 1.11: LEFT: Transmission electron micrograph of an insect cell nucleus showing the nuclear envelope (ne) and nucleolus (n); RIGHT: Drawing of a nucleus with chromatin (purple) and nuclear pores.

The cross-section of the interphase nucleus in the electron micrograph shows a prominent **nucleolus** (labeled **n**). The nucleus is enclosed in a **nuclear envelope** and surrounded by a darkly granular **RER** (rough endoplasmic reticulum). You can make out ribosomes (small granules) bound to the RER and to the outer nuclear membrane. The space enclosed by the RER (the **lumen**) is in fact continuous with the space separating the inner and outer membranes of the nuclear envelope, as illustrated in the drawing (above right). **Nuclear pores** in the nuclear envelope (look at the drawing) let large molecules and even particles move in and out of the nucleus across both membranes. The eukaryotic nucleus is where genes and RNA transcription are regulated and thus one place where cellular protein levels are controlled. *RNAs*, once transcribed from genes and processed, are exported to the cytoplasm through the nuclear pores. Even completely assembled ribosomal subunits are exported from the nucleus. Other RNAs remain in the nucleus, often participating in the regulation of gene

activity. We learn some details of nuclear pore traffic, DNA replication, and the dynamics of cell division in later chapters.



104-2 The nucleus



Beyond its nucleolus and nuclear envelope, the nucleus is more organized than it appears in conventional transmission electron micrographs. The nucleolus is just the largest of several inclusions that seem to segregate nuclear functions. Over 100 years ago **Santiago Ramón y Cajal** reported other structures in the nuclei of neurons, including what came to be known as **Cajal bodies (CBs)**. His elegant hand-drawn illustrations of nuclear bodies (made before the advent of photomicrography) can be seen at ^{1,10}[Cajal's Nuclear Bodies](#) and ^{1,11}[Cajal's Beautiful Brain Cells](#). **Cajal** and **Camillo Golgi** shared the Nobel Prize in Physiology or Medicine 1906 for their studies of nerve cell structure. In the electron microscope, **Cajal bodies (CBs)** look like coils of tangled thread, and were thus called **coiled bodies** (conveniently, also CBs). Other nuclear bodies since identified include **Gems**, **PML bodies**, nuclear speckles (or **splicing speckles**), **histone locus bodies (HLBs)**, and more! The results of immunofluorescence localization studies show that different nuclear bodies are associated with specific proteins (Figure 1.12, below).

Immunolocalization of Markers for Three Different Nuclear Bodies

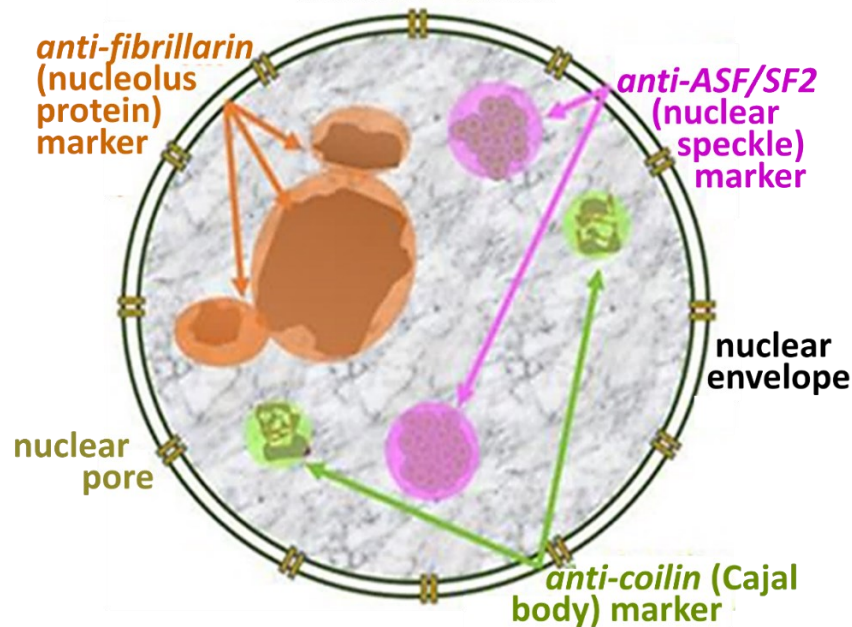


Fig. 1.12: Simulated immunolocalization of different proteins (fibrillarin, coilin and ASF/SF2) to nuclear bodies (nucleolus, Cajal Bodies and nuclear speckles, respectively), using fluorescent antibodies.

Nucleoli contain **fibrillarin** proteins, stained red by treating cells with red-fluorescence-tagged antibodies to **fibrillarin**. Pink-fluorescence-tagged **anticoilin** antibodies light up the coilin proteins of CBs. Green-fluorescing **ASF/SF2** antibodies localize to nuclear speckles. As part of or included in a nuclear matrix, nuclear bodies organize and regulate distinct aspects of nuclear activity and molecular function. The different nuclear bodies perform specific

functions and interact with each other and with proteins DNA and RNA to do so. We will revisit nuclear bodies in their working context later.

1.4.1.a Every Cell (i.e., Every Nucleus) of an Organism Contains the Same Genes

We read earlier that bacteria are busy doubling and partitioning their naked DNA chromosomes at the same time as they grow and divide by binary fission. In eukaryotic cells, a **cell cycle** divides life into discrete consecutive events. During most of the cell cycle, cells are in **interphase** and DNA is wrapped up in proteins in **chromatin** inside a nucleus. It is not merely the DNA, but chromatin that must be duplicated when cells reproduce. Duplication of DNA also involves disturbing and rearranging the chromatin proteins resting on the DNA. This occurs before cell division. As the time of cell division nears, chromatin associates with even more proteins, condensing to form **chromosomes**, while the nuclear envelope dissolves, marking the start of **mitosis** (**meiosis** in germline cells) and **cytokinesis**. You may recall that each somatic cell of a eukaryotic organism has paired **homologous chromosomes** and thus two copies of every gene the organism owns. But sperm and eggs emerge from meiosis with one of each pair of chromosomes and only one copy of each gene. Whether by mitosis or meiosis, duplicated chromosomes (chromatids) lined up at metaphase attach to spindle fibers (as seen in Figure 1.13) to be separated and drawn into new daughter cells formed during cytokinesis.

The Mitotic Spindle

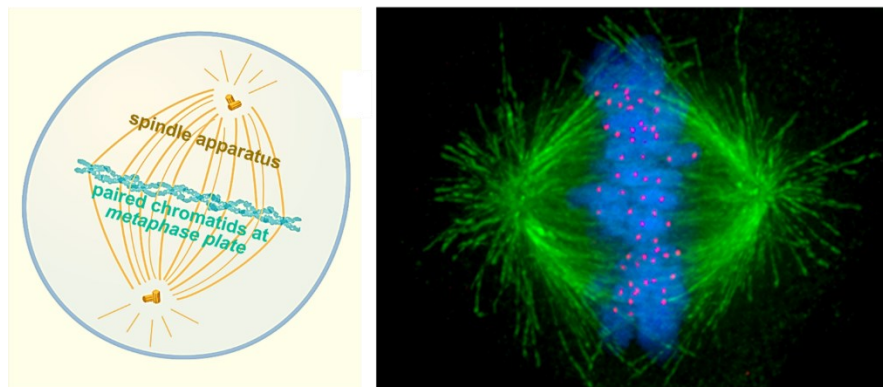


Fig. 1.13: Drawing (left) and fluorescence micrograph (right) of a cell in metaphase of mitosis: aligned chromosomes (chromatids) at the center of the cell (blue in the micrograph) are just about to be pulled apart by microtubules of the spindle apparatus (green) extending from the poles to the center of the cell.

As chromosomes separate and daughter cells form, nuclei reappear and chromosomes de-condense. These events mark the major visible difference between cell division in bacteria and eukaryotes. Cytokinesis begins near the end of mitosis. *Sexual reproduction*, a key characteristic of eukaryotes, involves **meiosis** rather than mitosis. The mechanism of **meiosis**, the division of *germ cells* leading to production of sperm and eggs, is like mitosis except that the ultimate daughter cells have just one each of the parental chromosomes, eventually to become the gametes (eggs or sperm). Google *meiosis* and/or *mitosis* to remind yourself about the differences between the two processes, meiosis and mitosis. A key take-home message here is that every cell in a multicellular organism, whether egg, sperm or somatic, contains the same genome (genes) in its nucleus. This was already understood from the time that mitosis and meiosis were first described in the late nineteenth century.

That every cell of an organism really does contain copies of all of its genes was finally demonstrated by John Gurdon and Shinya Yamanaka in 1962. They transplanted nuclei from the intestinal cells the frog *Xenopus laevis* into *enucleated eggs* (eggs from which their own nuclei had been removed). These 'eggs' grew and developed into normal tadpoles, proving that no genes are lost during development, but are just expressed differentially. For these cloning experiments, Gurdon and Yamanaka shared the 2012 Nobel Prize for Physiology or Medicine. We'll revisit animal cloning later. For now, it's enough to know that Molly the cloned frog was followed by Dolly, the first cloned sheep (1966) and then other animals, all cloned from enucleated eggs transplanted with differentiated cell nuclei. See ^{1,12}[Cuarteterra](#) to read about the cloning a champion polo mare whose clones are also champions!

1.4.2 Ribosomes

On the tiny end of the size spectrum, ribosomes are protein-making machines found in all cells. They consist of large and small subunits, each made up of proteins and ribosomal RNAs (rRNAs). Ribosomes bind to messenger RNAs (mRNAs), moving along the mRNA to translate 3-base code words (codons) into polypeptides. Multiple ribosomes can move along the same mRNA, forming *polyribosomes* (or *polysomes*) that simultaneously translate the same polypeptide encoded by the mRNA as shown in Figure 1.14.

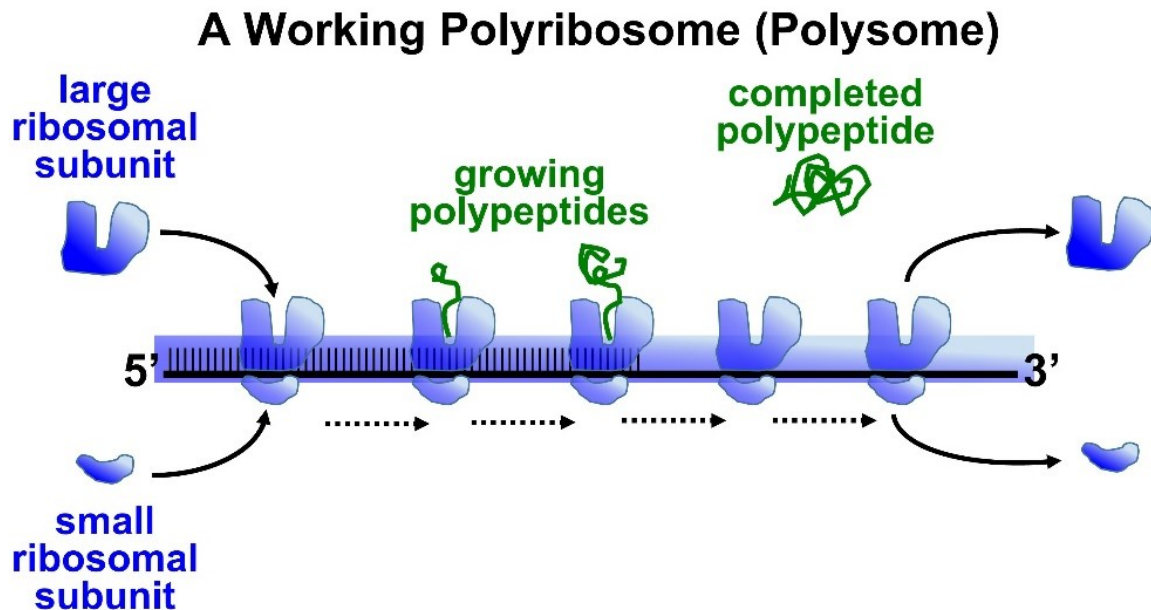


Fig. 1.14: Polysomes form when ribosomes (blue) assembled at the left on an mRNA molecule move. As they move from left to right more ribosomes can assemble at the left, each translating the message; a polypeptide (green) grows, emerging from the ribosomes. To accommodate size the allosteric changes during polypeptide synthesis, there are at least 35 nucleotides separating ribosomes on polysomes.

In the illustration, ribosomes assemble on the left (5') end of the messenger RNA to form the polysome. When they reach the other (3') end of the mRNA, the ribosomes disassemble from the RNA and release the finished polypeptide. The granular appearance of cytoplasm in electron micrographs is largely due to the ubiquitous distribution of ribosomal subunits and polysomes in cells. In electron micrographs of leaf cells from a dry, desiccation-tolerant desert plant, *Selaginella lepidophylla* (Figure 1.15), you can make out randomly distributed ribosomes

and ribosomal subunits (arrows, below left). In cells from a fully hydrated plant, you can see *polysomes* as more organized strings of ribosomes (arrows, below right).

The difference between inactive and active ribosomes...

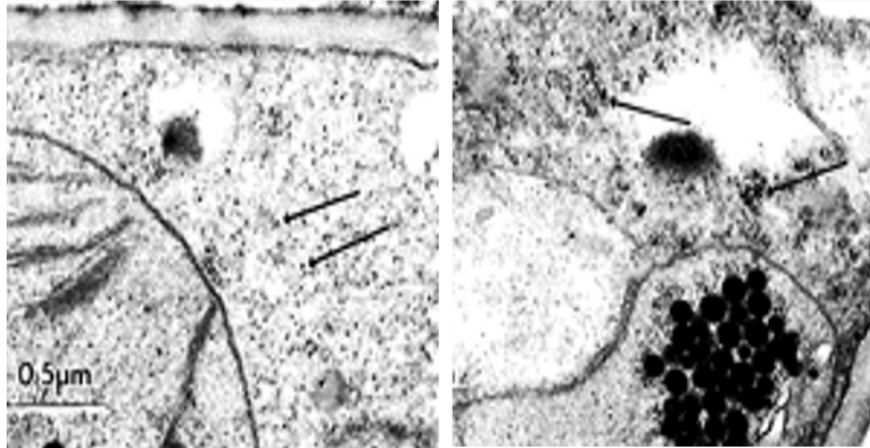
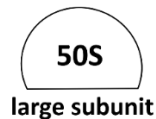


Fig. 1.15: Transmission electron micrographs of *Selaginella lepidophylla* plant cells. Small, free ribosomal subunits in desiccated cells (left) seem to have formed larger structures (ribosomes) aligned on polysomes in the hydrated plant cells (right). Zoom in to see these details.

Isolated ribosomes and subunits can be separated by sucrose-density-gradient centrifugation based on differences in mass. Figure 1.16 compares ribosomal subunit 'size', protein, and ribosomal RNA (rRNA) composition in eukaryotes and prokaryotes.

Ribosome Composition

PROKARYOTES

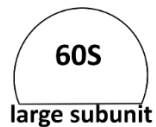


23S rRNA + 5S rRNA + 31 ribosomal proteins



16S rRNA + 31 ribosomal proteins

EUKARYOTES



28S rRNA + 5S rRNA + 5.8S rRNA + 50 ribosomal proteins



18S rRNA + 33 ribosomal proteins

Fig. 1.16: Comparison of prokaryotic (upper) and eukaryotic (lower) ribosomal subunit, rRNA and protein composition. The subunits and their component rRNAs were separated by mass (i.e., size) by sucrose density gradient *ultracentrifugation*. Proteins were isolated from separated subunits for analysis.

S (*Svedberg*) units are calculated from the position of particles and molecules in the gradient after separation. Theodor Svedberg earned the 1926 Nobel Prize in Chemistry for among other

things, applying *analytical ultracentrifugation* to the separation and determination of particulate and molecular masses.



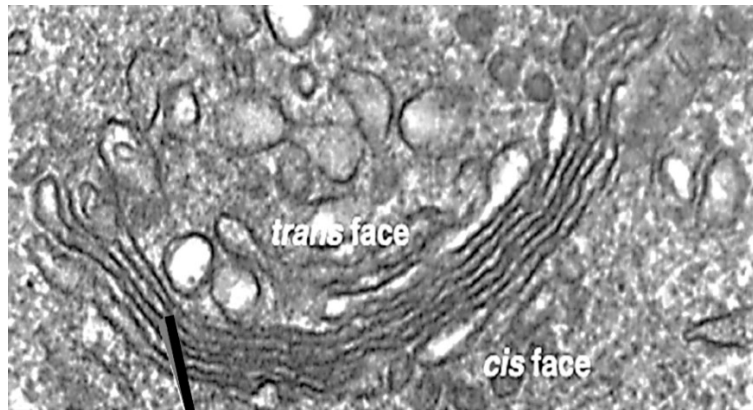
[101 Ribosomes & Polysomes](#)



1.4.3 Internal membranes and the Endomembrane System

Microscopists of the nineteenth century saw many subcellular structures using the art of histology, staining cells to increase the visual contrast between cell parts. One of these microscopists was the early neurobiologist, *Camillo Golgi*. He developed a silver (black) stain that first detected a network of vesicles which we now call *Golgi bodies* (or *Golgi vesicles*) in nerve cells. For his studies of the membranes now named after him, *Camillo Golgi* shared the 1906 Nobel prize for Medicine or Physiology with *Santiago Ramón y Cajal*.

Golgi vesicles along with other *vesicles* and *vacuoles* in cells, including, comprise the *endomembrane system*. Proteins made by ribosomes of the *rough endoplasmic reticulum* (RER) either enter the interior space (*lumen*) or become part of the RER membrane itself. The syntheses of *RER*, *smooth endoplasmic reticulum* (*SER*), *Golgi bodies*, *microbodies*, *lysosomes*, and other vesicular membranes (and their protein content) all start in the RER. *transport vesicles* that bud off from RER fuse with *Golgi Vesicles* at their *cis face* (Figure 1.17).



Golgi apparatus (vesicles)

Fig. 1.17: Transmission electron micrograph of an insect cell Golgi body showing cis and trans faces (labeled in white). Vesicles from the RER at the cis face (below, right) will fuse with the Golgi vesicles. Vesicles emerging at the trans face (above, left) are progenitors of a variety of different organelles.

Some proteins made in the endomembrane system are secreted by *exocytosis*. Others end up in organelles such as *lysosomes* that contain hydrolytic enzymes. These enzymes are activated when the lysosomes fuse with other organelles destined for degradation. For example, *food vacuoles* form when a plasma membrane *invaginates*, engulfing food particles. They then fuse with lysosomes to digest the engulfed nutrients. Still other proteins synthesized by ribosomes on the RER are incorporated into the RER membranes, destined to become part of lysosomes, peroxisomes, and even the plasma membrane itself. In moving through the endomembrane system, *packaged proteins* undergo stepwise modifications (*maturation*) before becoming biologically active (Figure 1.18, below).

Direction of Movement of Packaged Proteins

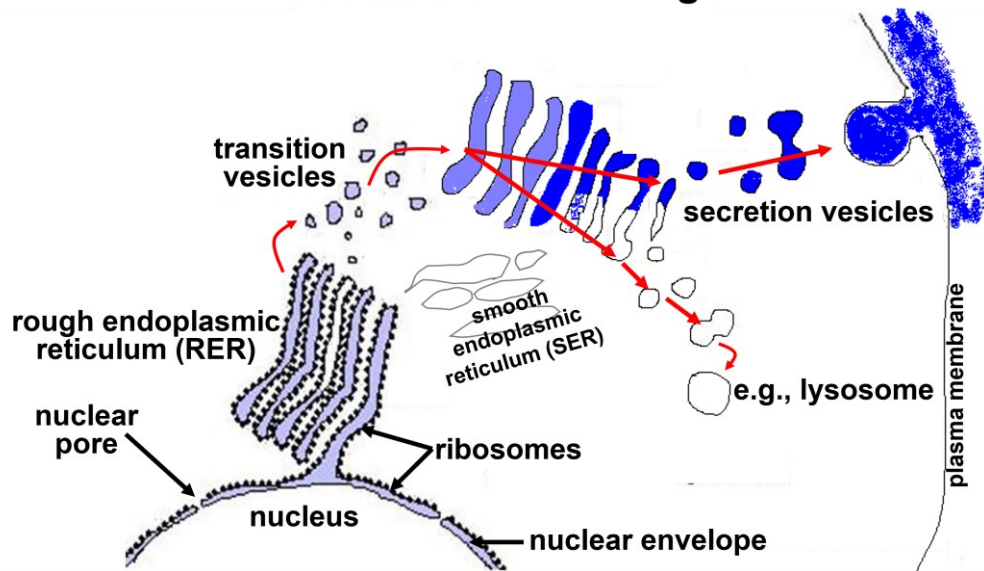


Fig. 1.18: Illustration of 'packaged' protein traffic through a cell from the RER (lower left) through Golgi vesicles (middle) to organelles (e.g., lysosomes) and the plasma membrane (upper right) for exocytosis (i.e., secretion). RER and Golgi vesicles are major sites for the modification (i.e., maturation) of packaged proteins.



[100-2 The RER-Rough Endoplasmic Reticulum](#)



[102 Golgi Vesicles & the Endomembrane System](#)



Autophagosomes are small vesicles that surround and eventually encapsulate tired organelles (for example, worn out mitochondria), eventually merging with lysosomes whose enzymes degrade their contents. In 2016, Yoshinori Ohsumi earned the Nobel Prize in Physiology and Medicine for nearly 30 years of research unraveling the cell and molecular biology of autophagy. **Microbodies** are a class of vesicles smaller than lysosomes but formed by a similar process. Among them are peroxisomes that break down toxic peroxides formed as a by-product of cellular biochemistry. Some vesicles emerging from the RER lose their ribosomes to become part of the SER, which has several different functions (e.g., alcohol detoxification in liver cells).



[103-2 Smooth Endoplasmic Reticulum](#)



Other organelles include the **contractile vacuoles** of freshwater protozoa that expel excess water that enters cells by osmosis. Some protozoa have **extrusomes**, vacuoles that release chemicals or structures that deter predators or enable prey capture. A large aqueous central vacuole dominates the volume of many higher plant cells. When filled with water, they will push all other structures against the plasma membrane. In a properly watered plant, this

water-filled vacuole exerts osmotic pressure that among other things keeps plant leaves from wilting and keeps stems upright.

1.4.4 Mitochondria and Plastids

Nearly all eukaryotic cells contain *mitochondria*, shown in Figure 1.19.

Electron Micrograph and Drawing of Mitochondria

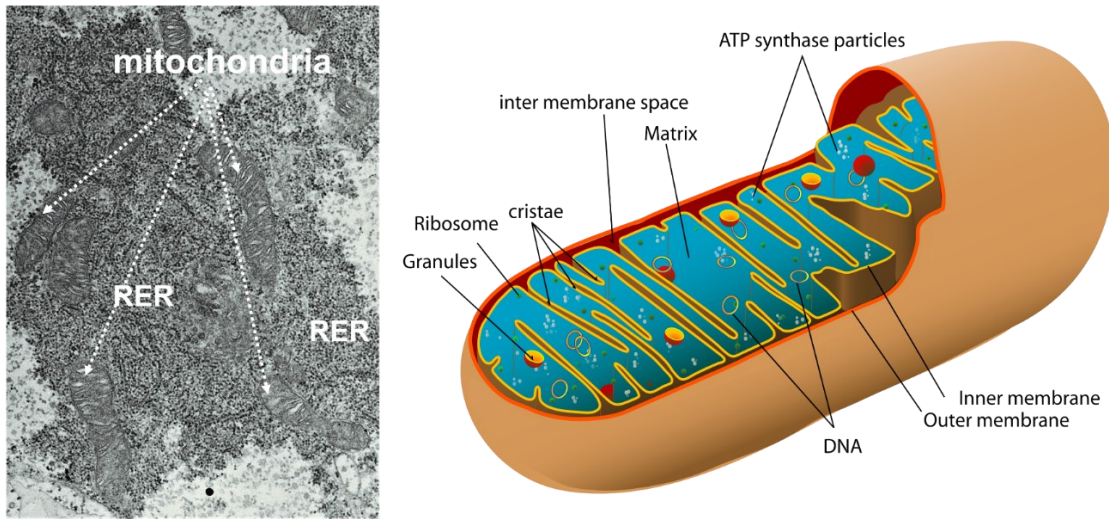


Fig. 1.19: Transmission electron micrograph of mitochondria embedded in RER (left) and a labeled drawing of a single mitochondrion (right).

The **matrix** of the mitochondrion is enclosed by a **cristal membrane** surrounded by an **outer membrane**. Each contains and replicates its own DNA, which contains genes encoding some of the mitochondrial proteins. The surface area of the inner mitochondrial membrane is increased by being folded into **cristae**, which are sites of **cellular respiration** (aerobic nutrient oxidation). Later, we'll consider the role of mitochondria in respiration in more detail.

Earlier, we speculated that some eukaryotic organelles could have originated within bacteria. But mitochondria probably evolved from an aerobic bacterium that was engulfed by another cell that escaped destruction to become an **endosymbiont** in the host cell. Lynn Margulis first proposed this in her **Endosymbiotic Theory**, in which a primitive eukaryotic cell acquired a bacterial endosymbiont (Margulis, L. [Sagan, L], 1967, *On the origin of mitosing cells*. *Journal of Theoretical Biology* **14**: 225–274). She proposed that chloroplasts also started as *endosymbionts*. Both mitochondria and the plastids of plants contain their own DNA, transcribe it into RNA and use their own translational machinery (i.e., ribosomes) to synthesize proteins, further supporting their bacterial and cyanobacterial origins. Living at first in symbiosis with the rest of the cell, these endosymbionts would eventually evolve into the organelles with which we are familiar.

Several protozoa lacking mitochondria and other organelles were discovered and suggested to be "first ingestors" of an ancestral endosymbiont, but since these cells contain other organelles (e.g., *hydrogenosomes*, *mitosomes*) it is thought more likely that these species *once had, but then lost mitochondria*.

Therefore, the descendants of ancient eukaryotic cells missing mitochondria probably no longer exist, if they ever existed at all! More evidence for the *Endosymbiotic Theory* is discussed elsewhere.

Chloroplasts, *photosynthetic protozoa*, and *cyanobacteria* contain chlorophyll and use similar photosynthetic mechanisms to make glucose. Transmission electron micrographs of chloroplasts are shown in the Figure 1.20. The one on the right shows a few starch granules.

Chloroplast Structures

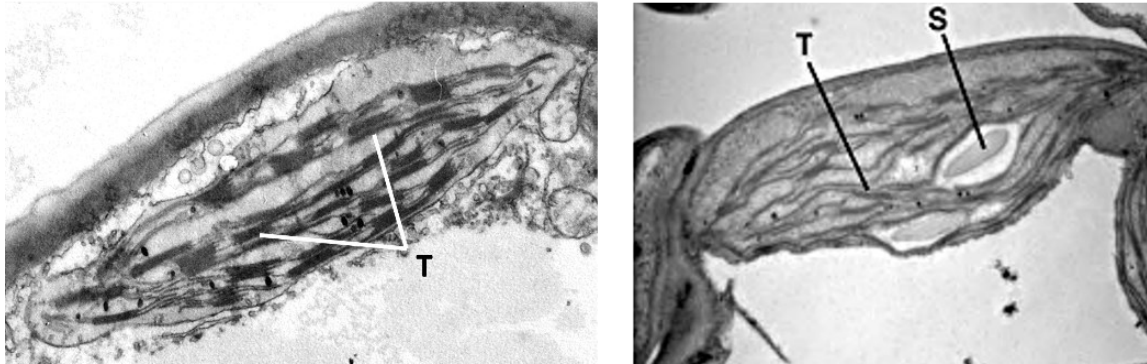


Fig. 1.20: Transmission electron micrographs of a typical photosynthetic chloroplast (LEFT), and one that has photosynthesized long enough to accumulated starch granules (RIGHT). S, starch granule; T, thylakoids.

A *leucoplast* is also a plastid, a chloroplast that has become filled with starch granules. In the electron micrograph of a leucoplast in Figure 1.21, you can see that, because of the accumulation of starch, the grana have become dispersed and indistinct.

A Leucoplast

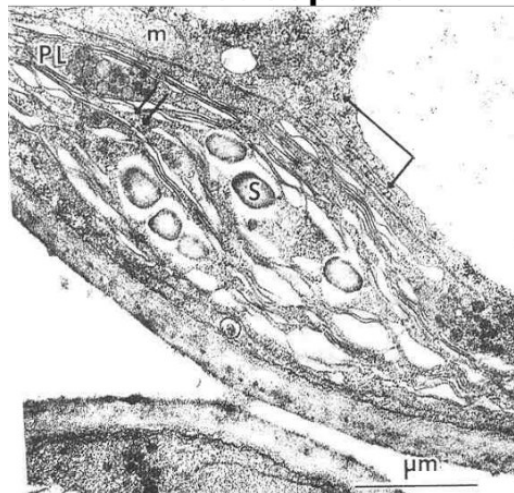


Fig. 1.21: Electron micrograph of a leucoplast, a chloroplast that has become filled with starch granules (S).



1.4.5 Cytoskeletal structures

We have come to understand that the cytoplasm of a eukaryotic cell is highly structured, permeated by rods and tubules. The three main structural components of this *cytoskeleton* are *microfilaments*, *intermediate filaments*, and *microtubules*. The structure and *polarity* of these structures are shown below in Figure 1.22.

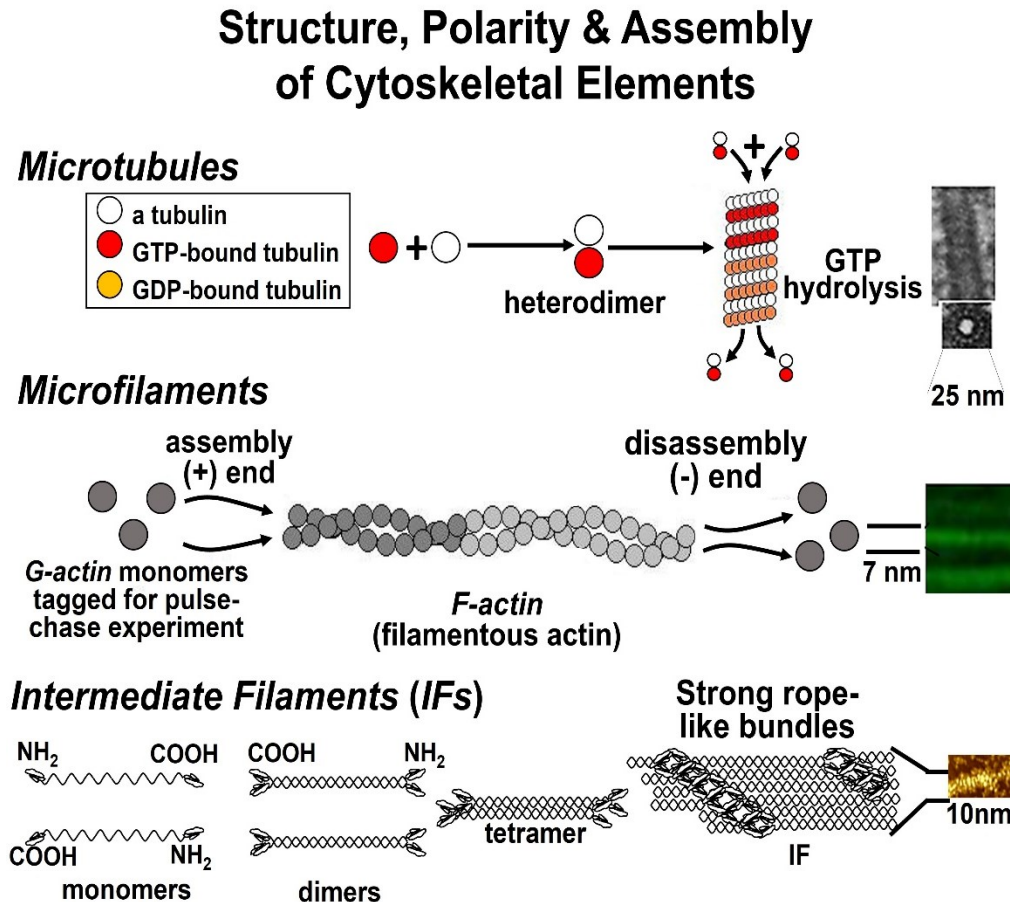


Fig. 1.22: Assembly, structure and polarity of microtubules (upper), microfilaments (middle), and intermediate filaments (lower) with electron micrographs of isolated microtubules (upper right) and immunofluorescence micrographs of using fluorescent antibodies to microfilament and intermediate filament proteins (middle and bottom right, respectively).

Microtubules are composed of α - and β -tubulin protein monomers. Monomeric *actin* proteins make up microfilaments. Intracellular intermediate filament proteins are related to the extracellular *keratin* of hair, fingernails, claws, and bird feathers. These cytoskeletal rods and tubules not only determine *cell shape*, but also play a role in *cell motility*. This includes the movement of cells from place to place and the movement of structures within cells.

We have already noted that a prokaryotic cytoskeleton is composed in part of proteins homologous to the actins and tubulins. As in a eukaryotic cytoskeleton, these bacterial proteins may play a role in maintaining or changing cell shape. On the other hand, *flagellin* (a protein not found in eukaryotic cells) powers the movement of bacterial flagella. A bacterial flagellum is a rigid hook-like structure attached to a molecular motor in the cell membrane that spins to

propel the bacterium through a liquid medium. In contrast, eukaryotic *microtubules* slide past one another causing a more flexible flagellum to undulate in wave-like motions and a cilium to beat rather than undulate. Cilia are involved not only in motility, but also in feeding and sensation. Microtubules in eukaryotic flagella and cilia arise from a *basal body* (similar to *kinetosomes* or *centrioles*) such as the one in Figure 1.23.

Longitudinal Section of a Flagellum

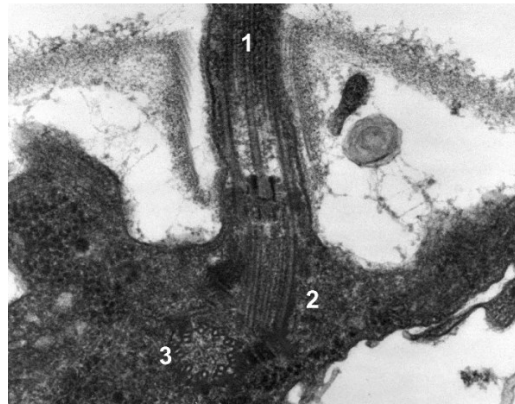


Fig. 1.23: Transmission electron micrograph of a longitudinal section of a flagellum (#1) emerging from a basal body (2). Number 3 is another basal body, this time in cross section.

Aligned in a flagellum or cilium, microtubules form an **axoneme** surrounded by plasma membrane. In electron micrographs of cross sections, a ciliary or flagellar axoneme is typically organized as a ring of nine paired microtubules (called *doublets*) around two *singlet* microtubules. Figure 1.24 shows the 9+2 microtubule arrangement of an isolated axoneme.

Axoneme of a Eukaryotic Cilium or Flagellum

9 + 2 arrangement
of microtubules

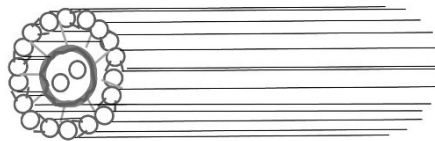


Fig. 1.24: Drawing of an axoneme isolated from a eukaryotic cilium or flagellum, by removing their plasma membrane, showing their characteristic 9+2 arrangement of microtubules in cross section (at the left) and longitudinal section projecting to the right.

Centrioles are themselves comprised of a ring of microtubules. In animal cells they participate in spindle fiber formation during mitosis and meiosis and are the point from which microtubules radiate thorough the cell to help form and maintain its shape. These structures do not involve axonemes. The spindle apparatus in plant cells typically lack centrioles but form from an amorphous *MTOC*, or *MicroTubule Organizing Center*. The MTOC serves the same purpose in mitosis and meiosis as centrioles serve in animal cells.



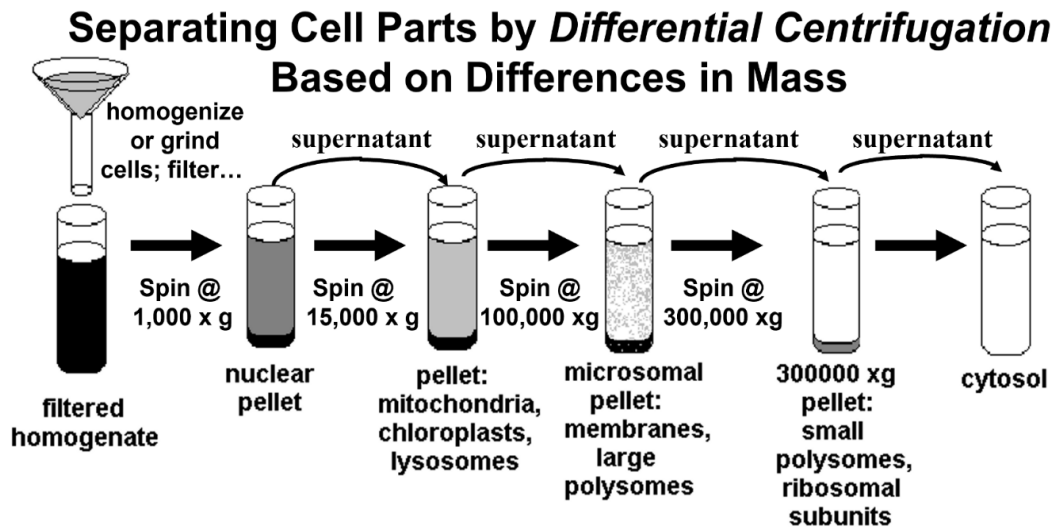
[106-2 Filaments & Tubules of the Cytoskeleton](#)



Elsewhere, we describe how microfilaments and microtubules interact with motor proteins (e.g., *dynein*, *kinesin*, and *myosin*) to generate force that results in the sliding of filaments and tubules to allow cellular movement. You'll see that motor proteins also transport molecular cargo from one place to another in a cell.

1.5 How we Know Functions of Organelles and Cell Structures: Cell Fractionation

We could see and describe cell parts in the light or electron microscope, but we could not definitively know their function until it became possible to release them from cells and separate them from one another. This became possible with the advent of differential centrifugation. Under centrifugal force generated in a spinning centrifuge rotor, subcellular structures separate by differences in mass. Structures that are more massive reach the bottom of the centrifuge tube before less massive ones. A cell fractionation scheme is illustrated in Fig 1.25. Biochemical analysis of the isolated cell fractions can reveal what different organelles and cellular substructures do.



...A cell fractionation/dissection

Fig. 1.25: Cells are broken open to release their contents and then filtered to remove unbroken cells (far left). Centrifugation at sequentially higher speed (G-force) sediments progressively smaller cellular parts (organelles, ribosomes, etc.) in *centrifugal pellets* (the 4 tubes in the middle), leaving behind a final supernatant, the soluble cell fraction or cytosol (tube at the far right). The smallest cell parts (membranes, ribosomes) require ultracentrifugation at the highest G-forces.



[107-2 Dissecting the Cell-a Cell Fractionation Scheme](#)

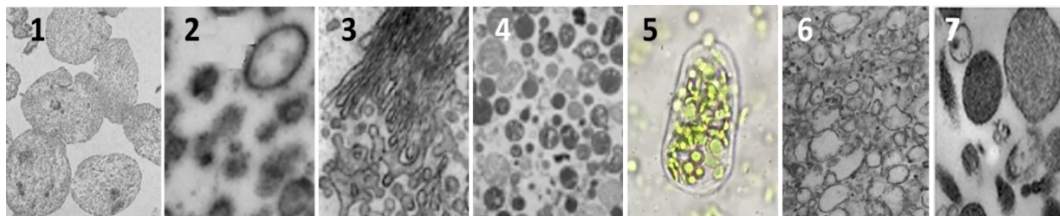


Cell fractionation separates cells into their constituent parts. The first step is to break open the cells and release their contents. This can be done by physical means such as grinding in a mortar and pestle, tissue grinder or similar device; exposure to ultrasound or high pressure; or exposure to enzymes or other chemicals that can selectively degrade the plasma membrane.

The next step is to isolate the subcellular organelles and particles from the cytoplasm (i.e., cytosol) by differential centrifugation. The centrifugation of broken cells at progressively higher centrifugal force separates (fractionates) particulate cell components based on their

mass. At the end of this process, a researcher will have isolated ribosomes, mitochondria, chloroplasts, nuclei, and other subcellular structures. After re-suspension, each pellet can be prepared for microscopy. Micrographs of some isolated subcellular fractions are shown in Figure 1.26.















Electron and Light Micrographs of Isolated Subcellular Fractions



Unlike most other cell substructures, chloroplasts (#5) are large (>5mm) and easily seen by light microscopy.

Fig. 1.26: Transmission electron micrographs of organelles isolated by eukaryotic cell fractionation: 1, nuclei; 2, RER; 3, Golgi vesicles; 4, mitochondria; 6, membrane vesicles; 7, lysosomes. The chloroplast (5) is a light micrograph.

These structures can be tentatively identified by microscopy based on their dimensions and appearance. Molecular analyses and biochemical tests on the cell fractions then help to confirm these identities.

 108-2 Isolated Nuclei		 109-2 Isolated RER	
 110-2 Isolated Golgi Vesicles		 111-2 Lysosomes & Peroxisomes	
 112-2 Isolated Mitochondria		 113-2 Isolated Chloroplasts	
 114-2 Isolated Membranes Form Vesicles			

Can you tell what organelles have been purified in each of these fractions based on the electron micrographs alone? Consider the structures on the left as an example. These were found in a low-speed centrifugal pellet, implying that they are large structures. They look a bit like nuclei, (which are in fact, the largest structures in a eukaryotic cell)—and indeed that's what they are!

Physical separation and the biochemical and molecular analysis of subcellular structures have revealed their basic functions and continue to reveal previously un-noticed structures and functions in cells. What biochemical tests might you do to confirm the identities of the

structures shown? At this point you may realize that all cell and molecular biology is devoted to understanding how prokaryotic and eukaryotic cells and organisms use their common structural and biochemical inheritance to meet very different survival strategies. As you keep studying, watch for experiments in which cell parts are separated and reconstituted. *Reconstitution* is a recurring experimental theme in the functional analysis of cell parts. Also look for another, even bigger theme: how evolution accounts for the common biochemistry and genetics of life— *and* its structural diversity!

1.6 The Origins, Evolution, Speciation, Diversity and Unity of Life

The question of how life began has been with us since the beginnings of recorded history. It is now accepted that there was a time, however brief or long, when the Earth was a lifeless (prebiotic) planet. Life's *origins* on Earth date to about four billion years ago under conditions that favored the formation of the first cell, the first entity with all of the properties of life. But couldn't those same conditions have spawned multiple cells independently, each with all the properties of life? If so, from which of these did life, as we know it today, descend? Whether there were one or more different "first cells", evolution (a property of life) could only begin with 'that or those' cells.



[115 Properties of Life](#)



The *progenote* has been defined as the first cell from which all life then descended. This implies that the origin of a cell was a unique, one-time only event. The fact that there is no evidence of multiple, independent origins of cellular life might be evidence (albeit negative evidence) that life originated only once, to produce a progenote as defined, and that multiple first cells (or potential progenotes) never existed. Alternatively, we can propose that the cell we call our ancestral progenitor originally had company, but that this progenote was evolutionarily successful at the expense of other early life forms, which thus became extinct.

Whatever our progenote may have looked like, one of its descendants later evolved the solutions to living that we see in force in all cells and organisms alive today, including a common (*universal*) genetic code to store life's information, as well as a common mechanism for retrieving the encoded information, what Francis Crick called the *Central Dogma* of biology. That ancestral cell is called our *Last Universal Common Ancestor*, or *LUCA*. We will consider ideas about life's origins, progenotes, and universal ancestors in the last chapter of this book. For now, feel free to check out the links below for more information.



[116 The Universal Genetic Code](#)



[117 Origins of Life](#)



[118 Life Origins vs Evolution](#)



For the moment, our focus is on evolution, the property of life that is the basis of speciation and life's diversity. Charles Darwin's theory of evolution was an explanation of the *structural* diversity of species. A naturalist, Darwin lived at a time of ferment where scientific discovery was challenging religion. But by 1839, Charles Darwin had published his *Narrative of the Surveying Voyages of His Majesty's Ships Adventure and Beagle*. This was the first of many reports of his careful observations of nature, with the seeds of what was to become his theory of *natural selection*. He published his more fully formed theory of evolution by natural selection in 1859 in *The Origin of Species*. There, he finally acknowledged his evidence-based belief that that new species arise when beneficial traits are selected from random genetic differences in individuals in a population. At the same time, less fit individuals would be culled from the population.

Finally, if natural selection acts on individuals, the emergence of new species (evolution) results from the persistence and spread of selected, heritable changes through successive generations in a population. In this way, evolution results in *an increase in biological diversity and complexity* at all levels of biological organization, from species to individual organisms and all the way down to biomolecules.

Darwin recognized that his theory would generate discord between science and biblical accounts of purposeful creation. He addressed the issue with great tact in introducing *The Origin of Species*: "*Although much remains obscure, and will long remain obscure, I can entertain no doubt, after the most deliberate study and dispassionate judgement of which I am capable, that the view which most naturalists entertain, and which I formerly entertained—namely, that each species has been independently created—is erroneous.*" Yet today, according to creationists, our exquisite eyes could only have formed by the intelligent design of a creator.

For the evolutionary perspective, see the article in National Geographic by E. Yong (Feb. 2016, with photography by D. Littschwager). Over time science favored Darwin. With the rediscovery of Mendel's genetic experiments at the turn of the twentieth century, it became increasingly clear that the genes of an organism are the basis of an organism's inherited physical and chemical traits, those traits that are passed down through the generations. It also became clear that Mendel had found the genetic basis for Darwin's theory and that the evolution even of *miraculous* eyes can be explained. Science and religion found ways to co-exist but, the controversy persists.

Repeated speciation occurs with the continual divergence of life forms from an ancestral cell through natural selection and evolution. Our shared cellular structures, nucleic acid, protein, and metabolic chemistries (the 'unity' of life) are testimony to our common ancestry with all life, dating back to our LUCA! Living things even share some early *behaviors*, governed at least in part by genes.

Take as an example the fact that our *biological clock* is an evolutionary adaptation to our planet's 24-hour daily cycles of light and dark. Day and night have been around since the origins of life, and all organisms studied so far seem to have a biological clock! The discovery of the genetic and molecular underpinnings of *circadian rhythms* (those daily cycles) earned Jeffrey C. Hall, Michael Rosbash and Michael W. Young the 2017 Nobel Prize in Medicine or Physiology (check out ^{1,15}[Circadian Rhythms Win Nobel Prize](#) to learn more)!

The molecular relationships common to all living things largely confirm what we have learned from the species represented in the fossil record. Morphological, biochemical, and genetic traits that are shared across species are defined as *homologous* and can be used to reconstruct evolutionary histories. The *biodiversity* that scientists (in particular, environmentalists) try to protect is the result of millions of years of adaptation (natural selection), speciation, and extinction. Biodiversity needs protection from the unwanted acceleration of evolution arising from human activity, including blatant extinctions (think passenger pigeon), and near extinctions (think American bison by the late 1800s). Think also of the consequences of the introduction of invasive aquatic and terrestrial species and the looming effects of climate change.

Now let's consider the biochemical and genetic *unity* among living things. We've already considered what happens when cells get larger when we tried to explain how larger cells divide their labors among smaller intracellular structures and organelles. When eukaryotic cells evolved further into multicellular organisms, it became necessary for the different cells to communicate with each other and to respond collaboratively to environmental cues. Some cells evolved mechanisms to "talk" directly to adjacent cells while others evolved the ability to transmit electrical (neural) signals over long distances to other cells and tissues. Still other cells produced hormones to communicate with cells, also at some distance, but to which they had no physical attachment.

As species diversified to live in very different habitats, they also evolved very different nutritional requirements, along with more extensive and elaborate biochemical pathways to digest their nutrients and capture their chemical energy. Nevertheless, through billions of years of evolution and astonishing diversification, the underlying genetics and biochemistry of living things on this planet is remarkably unchanged. Early in the twentieth century, Albert Kluyver first recognized that cells and organisms vary in form appearance in spite of an essential biochemical unity of all organisms (see ^{1.16}[Albert Kluyver](#)). This unity amidst the diversity is a life paradox that we examine in this course.

1.6.1 Random Acts of Genetic Variation, the Basis of Natural Selection

DNA contains the genetic instructions for the structure and function of cells and organisms. When and where a cell's or organism's genetic instructions are used (i.e., to make RNA and proteins) are highly regulated. Genetic variation results from random mutation. Genetic diversity arising from random mutations is in turn, the basis of natural selection during evolution.



[119 The Random Basis of Evolution](#)



1.6.2 The Genome: An Organism's Complete Genetic Instructions

Recall that every cell of an organism carries the same genome as every other cell. The genome of an organism is the entirety of its genetic material (DNA, or for some viruses, RNA), including genes and other kinds of DNA sequences. The genome of a common experimental

strain of *E. coli* was sequenced by 1997 (Blattner FR et al. 1997, *The complete genome sequence of Escherichia coli K-12*. Science 277:1452-1474). Sequencing of the human genome was completed (more or less!) by 2001, well ahead of schedule (Venter JC 2001, The sequence of the human genome. Science 291:1304-1351). Recall also that it was an analysis of rRNA gene sequences led to the dramatic re-classification of life from five kingdoms into three domains.

So, comparisons of specific gene or other DNA sequences can tell us a great deal about evolution. We now know that evolution depends on the structure of genomes as much as on individual gene sequences, Genome sequencing confirms genetic variation between species, but also much variation among individuals of the same species. It is this the latter variation that is the raw material of evolution. Genomic studies tell us that genomes have been shaped and remodeled in evolution. We'll consider genome remodeling in more detail elsewhere.

1.6.3 Genomic 'Fossils' Can Confirm Evolutionary Relationships.

We have been looking to gene and protein sequencing to find evolutionary relationships and even, familial relationships. You can read about an early demonstration of such relationships based on amino acid sequence comparisons across evolutionary time in Zuckerkandl E and Pauling L. (1965) *Molecules as documents of evolutionary theory*. J. Theor. Biol. 8:357-366. In addition, it has been possible for some time now, to extract and sequence DNA from fossil bones and teeth. allowing comparisons of extant and extinct species, including the fossil remains of humans, other hominids, and many animals, revealing our relationship to animals (from bugs to frogs to mice to chimps...) and hominids like us. In fact, we (*Homo sapiens*) share 95% of our DNA with Neanderthals. We know this from the work of Svante Paabo, whose technologies for extracting and sequencing ancient DNA, enabled him to sequence the Neanderthal genome..., and for which he earned the 2022 Nobel prize for Physiology and Medicine. Unfortunately, DNA from organisms older than 10,000 years is typically damaged or absent, so that relationship building beyond that time is almost impossible.

Using what we know from gene sequences of species alive today, investigators have recently '*reconstructed*' a genetic phylogeny suggesting the sequences of genes of some of our long-gone progenitors, including bacteria (to learn more, check out: ^{1,17}[Deciphering Genomic Fossils](#)). The comparison of these '*reconstructed*' ancestral DNA sequences suggests when photosynthetic organisms diversified and when our oxygenic planet became a reality. Closer to home, many remains of ancestral humans have been discovered in the Americas. These promise to unlock the mysteries of human settlement of the continents, though not without controversy. Native American tribal cultures treat their ancestors as sacred and argue against sampling such remains for DNA Analysis. In one example, a well-preserved mummified body was discovered in the Nevada desert in the 1940s. Tests of clothing fragments and hair revealed that this *Spirit Cave mummy* was over 10,000 years old. DNA sequence analysis was proposed to confirm the origins of the mummy. But then the *Fallon Paiute-Shoshone* tribe, which lives near the burial site, asserted a cultural relationship to the body and requested the right of its return in compliance with the *Native American Graves Protection and Repatriation Act*. Anthropologists then counter-asserted a need for further study of the body to learn more about its origins and about native American origins in general. The dispute ended only after 20 years, when the time the tribe consented DNA tests were allowed. When the DNA sequence

analysis results finally established that the remains were indeed that of an ancestor to the tribe, the *Spirit Cave mummy* was returned to the *Fallon Paiute-Shoshone* to be reburied with full tribal rites in 2018. To read more, see ^{1.18}[Resolving American Indian Ancestry](#) or ^{1.19}[Ice Age Mummy DNA Analysis Unlocks Tribal Secrets](#).



[120-2 Genomic Fossils-Molecular Evolution](#)



1.7 Microscopy Reveals Life's Diversity of Structure and Form

Broadly speaking, there are two main categories of microscopy. In *Light Microscopy*, the slide is viewed through optical glass lenses that see visible light reflected from or passing through the specimens on the slide. In *Electron Microscopy*, the viewer is looking at an image on a screen created by electrons passing through or reflected from the specimen, usually mounted on a copper grid. For a sampling of light and electron micrographs, check out this ^{1.20}[Micrograph Gallery](#). Here we compare and contrast different microscopic techniques.

1.7.1 Light Microscopy

Historically one or another version of light microscopy has revealed much of what we know of the structural diversity of cells, especially eukaryotic cells. Check out the ^{1.21}[Mitosis Drawings](#) for a reminder of how eukaryotic cells divide, and then check out ^{1.22}[The Optical Microscope](#) for descriptions of different variations of light microscopy (e.g., *bright-field*, *dark field*, *phase-contrast*, and *fluorescence*). Limits of *magnification* and *resolution* of 1200X and 2 μ m, (respectively) are common to all forms of light microscopy. Some variations of light microscopy are briefly described here:

- *Bright-Field microscopy* is the most common kind of light microscopy. The specimen is illuminated from below; contrast between regions of the specimen comes from the difference between light absorbed by the sample and light passing through it. Live specimens lack contrast in conventional bright-field microscopy because differences in refractive index between components of the specimen (e.g., organelles and cytoplasm in cells) diffuse the resolution of the magnified image. Therefore *Bright-Field microscopy* is best suited to fixed and stained specimens.
- In *Dark-field* illumination, light passing through the center of the specimen is blocked and the light passing through the periphery of the beam is *diffracted* (*scattered*) by the sample. The result is enhanced contrast for certain kinds of specimens, including live, unfixed, and unstained ones.
- In *Polarized light microscopy*, light is passed through a polarizing filter before passing through the specimen. The resulting *incident light* (the light beamed at the object on the slide) to pass through the specimen where it may be bent (diffracted), thereby increasing the contrast of the specimen. The microscopist achieves the highest contrast by rotating the plane of polarized light passing through the sample. A valuable feature of polarized light microscopy is that samples can be unfixed, unstained or even live.

- *Phase-Contrast* or *Interference microscopy* enhances contrast between parts of a specimen with higher refractive indices (e.g., cell organelles) and lower refractive indices (e.g., cytoplasm). *Phase-Contrast* microscopy optics shift the phase of the light entering the specimen from below by a half a wavelength to capture small differences in refractive index to increase contrast. Phase-Contrast microscopy is a most cost-effective tool for examining live, unfixed, and unstained specimens.
- In a *fluorescence microscope*, a specimen is, for example, treated with a molecule tagged with (covalently attached to) a *fluorophore* that fluoresces (emits visible light) when exposed to short wavelength, high-energy (usually UV) light. The tagged molecule are often fluorescent antibodies that will bind to specific molecules in a cell. In *fluorescence microscopy*, the visible fluorescent light localizes the target molecule/structure in the cell.
- *Confocal microscopy*, a variant of fluorescence microscopy, enables imaging through thick samples and sections. The result is often 3D-like, with much greater depth of focus than other light microscope methods. Look at ^{1,23}[Microscope Image Gallery-Confocal Included](#) to see a variety of confocal micrographs and related images; look mainly at the specimens.
- *Lattice Light-Sheet Microscopy* is a 100-year old variant of light microscopy that allows us to follow subcellular structures and macromolecules moving about in living cells. Read more about the renewed interest in this technique at ^{1,23}[Lattice Light Sheet Microscopy](#).

1.7.2 Electron Microscopy

Transmission Electron Microscopy (TEM) microscopy generates an image by passing electrons through a specimen treated with an electron-opaque stain (e.g., osmium), or reflecting electrons from a specimen, and capturing the electron image on a screen. TEM can achieve much higher magnification (up to 10⁶X) and resolution (2.0 nm) than any form of optical microscopy. The higher voltage of **High Voltage Electron microscopy** allows passage of electrons through thicker sections than regular (low voltage) TEM. The result is micrographs with much greater resolution, depth, and contrast. **Scanning Electron Microscopy** (SEM), with magnifying power up to 10⁵X and a 3.0-20.0 nm resolution. SEM lets us to examine the surfaces of tissues, small organisms like insects, and even of cells and organelles. Objects of SEM must be conductive, so that biological samples are usually spray-coated with a thin layer of metal (e.g., palladium, platinum) (check the link to ^{1,25}[Scanning Electron Microscopy](#) for more on SEM and look at the gallery of SEM images at the end of the entry). **Helium Ion Microscopy** is a form of SEM that substitutes helium ions for the vacuum in which SEM samples are normally viewed, eliminating the need for metal spray coating. Thus, **HIM** enables investigators to examine e.g., cells and viruses in a more natural state (you can google this on your own to see some examples of this). Electron microscopy, together with biochemical and molecular biological studies have revealed how interacting cellular and molecular components work with each other, and continue to do so, shedding light on all manner of biological processes and interactions.



[121-2 Electron Microscopy](#)



Some iText & VOP Key words and Terms

actin	eukaryotes	Natural Selection
Archaea	eukaryotic flagella	nuclear envelope
bacterial cell walls	evolution	nuclear pores
bacterial flagella	exocytosis	nucleoid
binary fission	extinction	nucleolus
cell fractionation	hypothesis	nucleus
cell theory	inference	optical microscopy
chloroplasts	intermediate filaments	plant cell walls
chromatin	keratin	progenote
chromosomes	kingdoms	prokaryotes
cilia	LUCA	Properties of life
confocal microscopy	lysosomes	rough endoplasmic reticulum (RER)
cytoplasm	meiosis	scanning electron microscopy
cytoskeleton	microbiome	Scientific Method
cytosol	microbodies	secretion vesicles
deductive logic	microfilaments	smooth endoplasmic reticulum (SER)
differential centrifugation	microtubules	speciation
diversity	mitochondria	theory
domains of life	mitosis	transmission electron microscopy
dynein	motor proteins	tubulins
endomembrane system	mutation	

CHAPTER 1 WEB LINKS



1.1



1.2



1.3



1.4



1.5



1.6





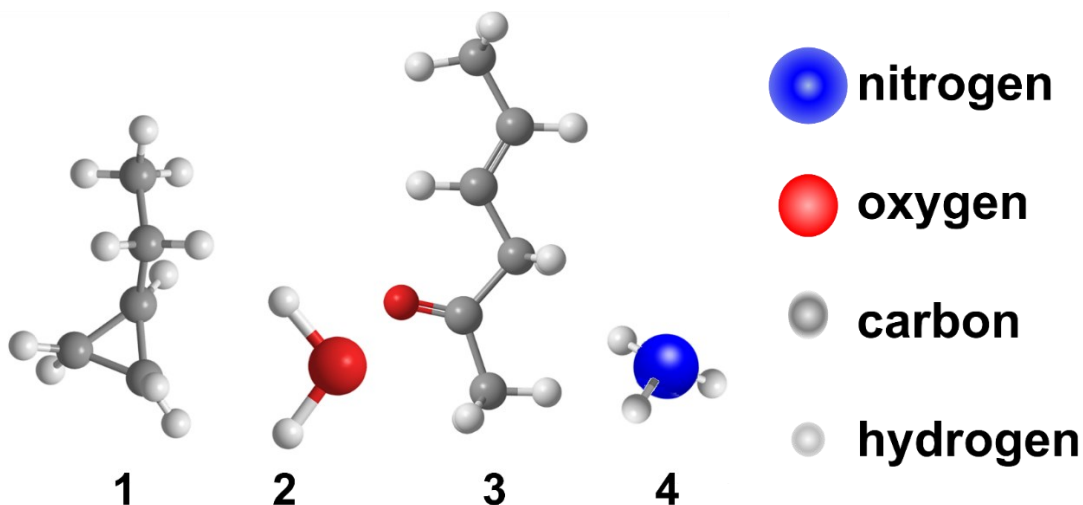
Chapter 2

Basic Chemistry, Organic Chemistry, and Biochemistry

Atoms, elements (the nucleus, atomic models); bonds (covalent, ionic, hydrogen); water properties, pH; organic molecules and biochemistry: chemical groups, monomers, polymers, condensation and hydrolysis, polysaccharides, polypeptides, proteins, DNA, RNA, lipids

Reminder: For inactive links, google key words/terms for alternative resources.

Organic or not organic?



Do you know? Did you guess correctly?

2.1 Introduction

In this chapter we start with a review of *basic chemistry* from **atomic structure** to **molecular bonds** to the structure and properties of **water**, followed by a review of key principles of *organic chemistry*—the chemistry of carbon-based molecules. You may find it useful to have your old general chemistry textbook handy or to check out the excellent introduction to general chemistry by Linus Pauling (1988, *General Chemistry* New York, Springer-Verlag). We'll see how **polar covalent bonds** define the **structure** and explain virtually all the **properties of water**. These range from the energy required to melt a gram of ice to the energy required to vaporize a gram of water and from its surface tension to its ability to hold heat, not to mention its ability to dissolve a wide variety of **solutes**, from salts to proteins and other macromolecules. We will distinguish water's **hydrophilic** interactions with solutes from its **hydrophobic** interactions with fatty molecules. Then we will review some basic biochemistry. Well-known biological molecules include monomers (such as sugars, amino acids, and nucleotides) and polymers (including polysaccharides, proteins, and nucleic acids).

Biochemical reactions that link *glucose* monomers into polymers on the one hand, and that break the polymers down on the other are essential reactions for life on Earth (and probably everywhere else!). On Earth, photosynthetic organisms link glucose monomers into starch, a polysaccharide. Amylose, a simple starch, a large **homopolymer** of repeating

glucose monomers. Likewise, polypeptides are **heteropolymers** of twenty different *amino acids*. DNA and RNA nucleic acids are also heteropolymers, made using only four different *nucleotide* monomers.

Digestive enzymes in your gut catalyze the **hydrolysis** (*i.e.*, breakdown) of the plant or animal macromolecular polymers we ate, back down to monomers. Hydrolysis adds water molecules across the bonds linking the monomers in a polymer, breaking those linkages. Our own cells then take up the monomers. Once there, **condensation (dehydration synthesis) reactions** remove water molecules from participating monomers to grow new polymers that are more useful to us. Strictly speaking, **triglycerides** (fats) and **phospholipids** are neither polymers nor macromolecules. Both are broken down by hydrolysis and synthesized in condensation reactions. Energy-rich triglyceride hydrolysis products include fatty acids that, (like sugars) can be oxidized for energy. Phospholipids (chemical relatives of triglycerides) are the basis of cellular membrane structure.

Weak interactions between macromolecules, for example **hydrogen bonds** (H-bonds), **electrostatic interactions** and **van der Waals forces**, are responsible for holding many cellular molecules and structures together. While individually these bonds are weak, in their multitudes, they are collectively strong. Familiar examples are the two complementary DNA strands held in a tight and stable double helix by millions of H-bonds, and the thousands of H-bonds that hold a polypeptide into its folded three-dimensional shape. We will see this theme—strength in numbers—repeated in other molecular and cellular structures. Monomers also serve other purposes related to energy metabolism, cell signaling, and the like. Depending on your chemistry background, you may find googling these subjects interesting and useful. The short voice-over PowerPoint recordings (VOPs) in this chapter might also help guide your understanding of the basic chemistry and biochemistry presented here.

Learning Objectives

When you have mastered the information in this chapter, you should be able to do the following:

1. Compare and contrast the definitions of *atom*, *element*, and *molecule*.
2. List differences between atoms, elements, and molecules and between *energy-* and *position-*based atomic models.
3. Describe the behavior of subatomic particles when they absorb and release energy.
4. State the difference between atomic *shells* and *orbitals*.
5. State how *kinetic* and *potential* energy apply to atoms and molecules.
6. Explain the behavior of atoms or molecules that fluoresce when excited by high-energy radiation, as well as the behavior of those that do not.
7. Distinguish *polar* and *nonpolar* covalent bonds and their physical-chemical properties.
8. Predict the behavior of electrons in compounds held together by *ionic interactions*.
9. Explain how the *properties of water* account for the solubility of macromolecules and salts and describe the role of H-bonds in supporting those properties.
10. Consider, in terms of water's properties, why some salts are not soluble in water.
11. Describe how molecular linkages form during polymer synthesis, and place hydrolytic and dehydration synthetic reactions in a *metabolic context*.
12. Distinguish between chemical bonds and linkages in polymers.
13. Categorize different chemical bonds based on their strengths.

2.2 Atoms and Basic Chemistry

The difference between *elements* and *atoms* is often confused in casual conversation. Both terms describe *matter*, substances with *mass*. Different elements are distinct kinds of matter distinguished by different physical and chemical properties. In turn, the atom is the fundamental unit of matter—that is, of an element. The *positively charged protons* and *neutral neutrons* in an atomic nucleus account for most of the mass of an atom. Each negatively charged *electron* that orbits a nucleus is about 1/2000 of the mass of a proton or neutron. Thus, they do not add much to the mass of an atom.

Electrons stay in atomic orbits because of electromagnetic forces, (i.e., their attraction to the positively charged nuclei). Its nuclear size (mass) and the space occupied by the cloud of electrons around its nucleus define structure of an atom, and that structure dictates the different properties of the elements. Let's take a closer look at the elements and their atoms and how their atomic properties account for the formation of molecules and molecular structure.

2.2.1 Overview of Elements and Atoms

We model atoms to illustrate the average physical location of electrons (the *orbital model*) on one hand and their potential energy levels (the *Bohr* or *shell model*) on the other (Figure 2.1).

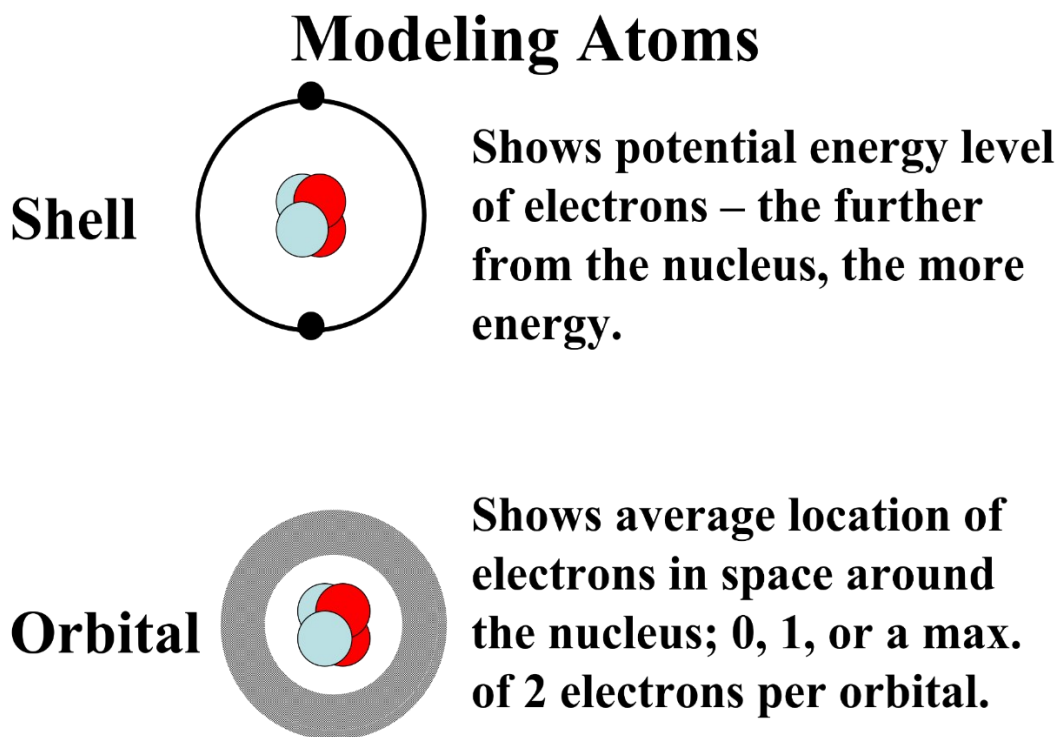


Fig. 2.1: The *Bohr* (shell) model of an atom (above) emphasizes electron *kinetic* energy; the *orbital* model (below) emphasizes the *space* occupied by electrons moving around the atomic nucleus.

Up to two electrons move in a space defined as an **orbital**. In addition to occupying different areas around the nucleus, electrons exist at different **energy levels** and move with different **kinetic energy**. As we will learn, electrons can absorb or lose energy, jumping or falling from one energy level to another.

Recall that atoms are chemically most stable when they are electrically uncharged, with an equal number of protons and electrons. **Isotopes** of the same element are atoms with the same number of protons and electrons but a different number of neutrons and thus, different masses. Isotopes of an element are chemically stable but may not be physically stable. For example, the most abundant isotope of hydrogen contains one proton, one electron, and *no neutrons*. The nucleus of the **deuterium** isotope of hydrogen contains one neutron, and that of **tritium** contains two neutrons. Both isotopes can be found in water molecules. Deuterium is stable. In contrast, the tritium atom is radioactive, subject to nuclear decay over time. Whether they are physically stable or not, all isotopes of an element share the same electromagnetic and chemical properties and behave the same way in chemical reactions. The electromagnetic forces that keep electrons orbiting their nuclei allow the formation of chemical bonds in molecules.

The partial periodic table below (Figure 2.2) shows the elements that are essential for all life (in greater or lesser amounts), as well as some that may also be essential in humans.

Elements Found in Organisms Based On the Distribution of Electrons in Shells

Distribution of Electrons in Shells																		0
1																		
1 H 1																		He
2																		
Li	Be	Valence: # of unpaired electrons in the outer shell; determines the chemical properties of the element										3	4	3	2	1		
11 Na 23	12 Mg 24											5 B 11	6 C 12	7 N 14	8 O 16	9 F 19	He	
19 K 39	20 Ca 40	Sc	Ti	23 V 51	24 Cr 52	25 Mn 55	26 Fe 56	27 Co 59	28 Ni 59	29 Cu 64	30 Zn 65	Al	14 Si 28	15 P 31	16 S 32	17 Cl 35	Ar	
Rb	Sr	Y	Zr	Nb	42 Mo 96	Tc	Ru	Rh	Pd	Ag	Cd	Ga	Ge	As	34 Se 79	Br	Kr	
Rb	Sr	Y	Zr	Nb	42 Mo 96	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	53 I 127	Xe	

blue elements: 99.0% of atoms in the human body
purple elements: 0.9% of atoms in the human body
green elements: required in trace amounts
brown elements: may be required by human cells

Fig. 2.2: Partial *Periodic Table of Elements* emphasizing elements found in living things.

The table shows the unique **atomic number** (number of protons) and **atomic mass** (usually measured in **daltons**, or *Da*) that characterize the different elements. For example, the atomic number of carbon (C) is 6, which is the number of protons in its nucleus. Its mass is 12 Da (six protons plus six neutrons, at 1 Da each). Remember that the mass of all the electrons in a C atom is negligible! Find the C atom and look at some of the other atoms of elements in the partial periodic table (Figure 2.2). Superscripted atomic numbers and subscripted atomic mass numbers uniquely define each element.



[122-2 Atoms & Elements](#)

2.2.2 Electron Configuration—Shells and Subshells

The *Bohr* model of the atom reveals how electrons can absorb and release energy. The shells indicate the energy levels of electrons. Electrons can absorb different kinds of energy, including electrical energy, radiation, and light (which is just a form of radiation—weaker than some and stronger than others). Ultraviolet (UV) light beamed at atoms can excite electrons. If an electron in an atom absorbs a full **quantum** (a **photon**) of UV radiant energy, it will be boosted from the **ground state** (the shell it normally occupies) into a higher shell, an **excited state**. Excited electrons move at greater speed around the nucleus and with more **kinetic energy** than they did at *ground state*. Excited electrons also have more **potential energy** than ground-state electrons. This is because they are unstable, releasing some of the energy gained during excitation as they return to *ground* (i.e., their starting energy level, or shell, as seen below in Figure 2.3).

The *shell* model emphasizes electron energy...

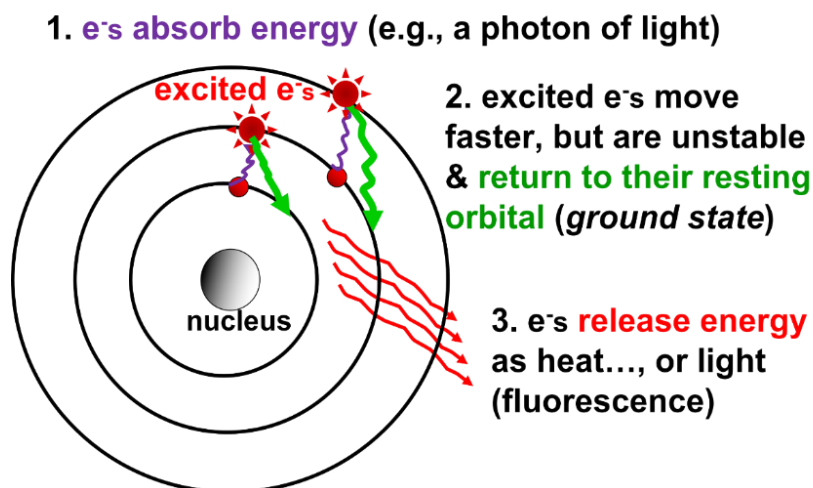


Fig. 2.3: When an electron absorbs e.g., light energy, it moves faster. Absorbance of one *photon* causes an electron to *jump* to a shell farther from the atomic nucleus. While having higher kinetic energy, excited electrons are unstable. Thus, they have *potential energy*, which is released as the electron returns to *ground state*, releasing heat or light (*fluorescence*).

Electrons falling back to ground typically release excitation energy as heat. Atoms whose excited electrons release their energy as light *fluoresce*; we say they are *fluorescent*. A fluorescent light is an example of this phenomenon. Electrical energy excites electrons out of their atomic orbitals in the molecules that coat the interior surface of the bulb. As all those excited electrons return to ground state, they *fluoresce*, releasing light. These atoms can be repeatedly excited by electricity.

As we shall see, biologists and chemists have turned fluorescence into a tool of biochemistry, molecular biology, and microscopy. The ground state is also called the *resting state*, but electrons at ground are by no means resting! They just move with less kinetic energy than excited electrons.



[123 Electron Energy and Fluorescence](#)



2.3 Chemical bonds

Atoms form bonds to make molecules, and there are three main classes of chemical bonds. There are two subsets of *covalent bonds*, both of which are *strong* bonds. They involve *unequal* or *equal* sharing of a pair of electrons. Unequal sharing of electrons results in *polar covalent bonds*. Equal sharing forms *nonpolar covalent bonds*. *Ionic bonds* are created by electrostatic interactions between elements after they gain or lose electrons, and these bonds are weaker than covalent bonds. *Hydrogen bonds (H-bonds)* are in a class by themselves. Their electrostatic interactions account for the physical and chemical properties of water. They are also involved in interactions between and within other molecules. Note that while atoms can share, gain, or lose electrons in chemical reactions, they will neither gain nor lose protons or neutrons. Let's look more closely at chemical bonds and how even the "weak" bonds are essential to life.

2.3.1 Covalent Bonds

Electrons are shared in covalent bonds. Hydrogen gas (H_2) is a molecule, not an atom! The two H atoms in the H_2 molecule share their electrons equally. Likewise, the carbon atom in methane (CH_4) shares electrons equally with four hydrogen atoms. A single pair of electrons forms the covalent bond between two H atoms in the hydrogen molecule (H_2). In methane, the carbon (C) atom has four electrons in its outer shell that it can share. Each H atom has one electron to share. If a C atom shares each of its four electrons with the electron in each of four H atoms, there will be eight (four paired) electrons moving in filled orbitals around the nucleus of the C atom some of the time and one pair moving around each of the H atomic nuclei some of the time. Thus, the outer shells of the C atom and the H atoms are filled at least some of the time. This stabilizes the molecule. Remember that atoms are most stable when their outer shells are filled and when each electron orbital is filled (i.e., with a pair of electrons). The equal sharing of electrons in *nonpolar covalent bonds* in H_2 and CH_4 is shown below in Figure 2.4.

Non-polar covalent bonds form when atoms share valence electrons equally:

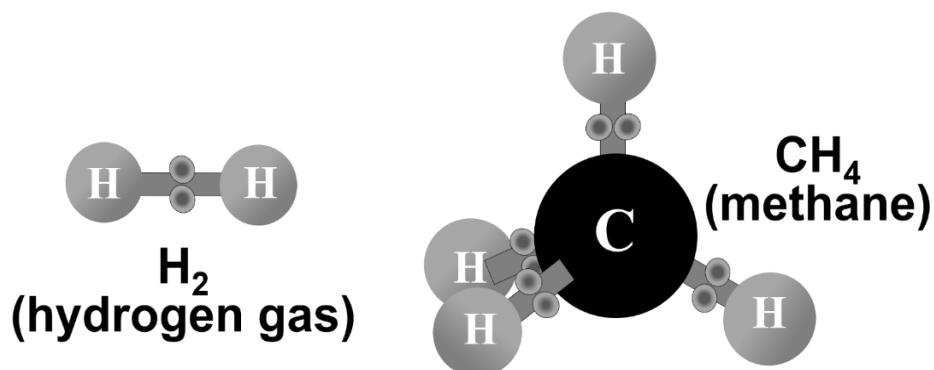
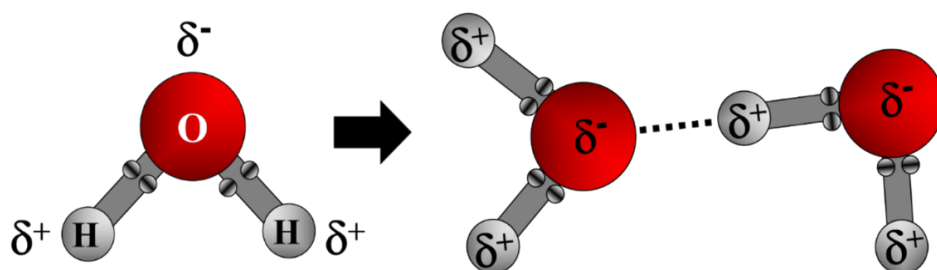


Fig. 2.4: Nonpolar covalent bonds in hydrogen gas (left) and methane (right).

Polar covalent bonds form when electrons in a molecule are shared unequally. This occurs if the atomic nuclei in a molecule are very different in size, as is the case with water (Figure 2.5).

All of the properties of water come from its **polar covalent structure...**



The paired electrons between hydrogen and oxygen are drawn to the larger oxygen atom, leaving all the atoms with a **partial charge** (δ^+ or δ^-). Thus, water molecules attract each other and other polar covalent molecules.

Fig. 2.5: Electrons on the H atoms of water molecules are drawn close to the large, positively charged nucleus of the O atom (molecule at left). As a result, H atoms “lose” electrons and acquire a partial positive charge (δ^+) while oxygen atoms “gain” those electrons and have a partial negative charge (δ^-). The polar covalent water molecules then attract and bind to other water molecules (molecules at the right).

The larger nucleus of the oxygen atom in H_2O attracts electrons more strongly than does either of the two H atoms. As a result, the shared electrons spend more of their time orbiting the O atom, such that the O atom carries a *partial negative charge* while each of the H atoms carries a *partial positive charge*. The Greek letter delta (δ) indicates partial charges in polar covalent bonds. In Figures 2.4 and 2.5, compare the position of the paired electrons in water with those illustrated for hydrogen gas or methane.

Water's polar covalent bonds allow it to attract and to interact with other polar covalent molecules, including other water molecules. The polar covalent nature of water also goes a long way to explaining its physical and chemical properties and the reason water is essential to life on this planet!



[124 Covalent Bonds](#)



Both polar and nonpolar covalent bonds play a major role in the structure of macromolecules, as in the protein hormone *insulin*, modeled in Figure 2.6.

The Insulin Molecule

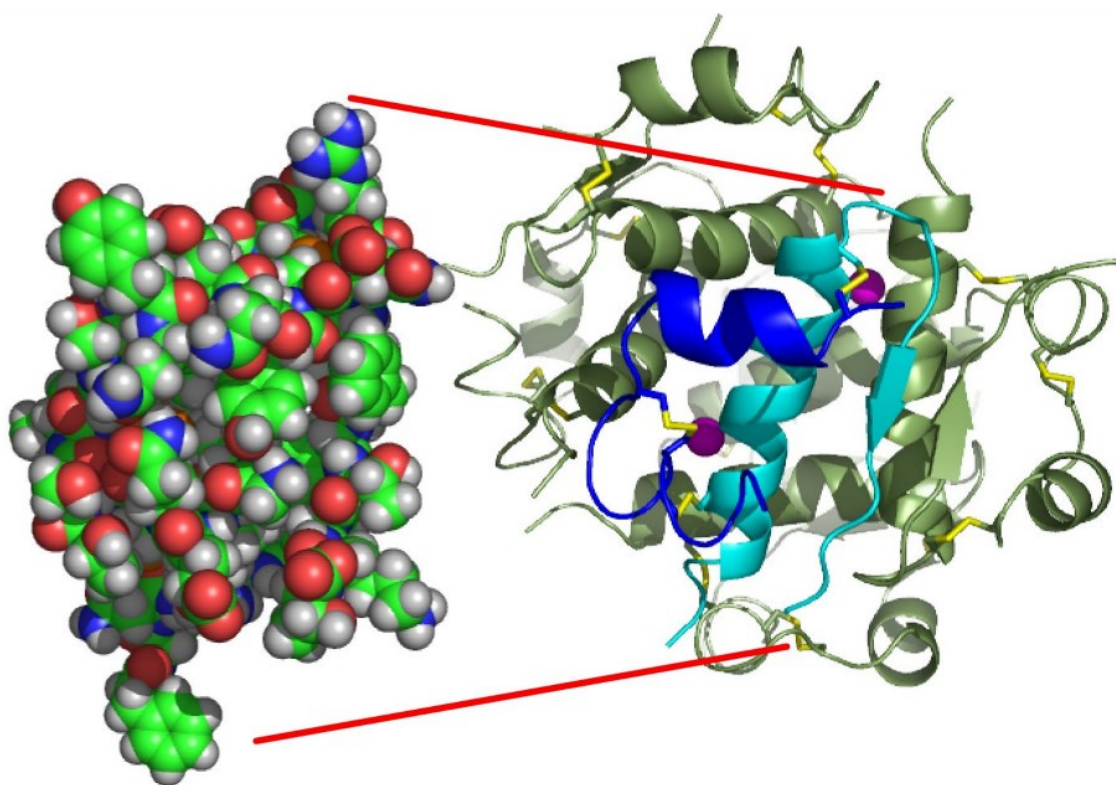


Fig. 2.6: Computer-generated space-filling (left) and “ribbon” models (right) of insulin.

The X-ray image of a space-filling model of the *hexameric* form of stored insulin (Figure 2.6, left) emphasizes its tertiary structure in detail. The ribbon diagram (Figure 2.6, right) highlights regions of internal secondary structure. When secreted from the *Islets of Langerhans* cells of the *pancreas*, active insulin is a dimer of A and B polypeptides (blue and cyan in the ribbon diagram, respectively). The subunit structure and the interactions holding the subunits together result from many electrostatic interactions (including H-bonds) and other weak interactions. The disulfide bonds or bridges (seen as yellow Vs in the ribbon diagram) stabilize the associated A and B monomers. We will look at protein structure in more detail in an upcoming chapter.

2.3.2 Ionic Bonds

Atoms that gain or lose electrons to achieve a filled outer shell acquire a negative or a positive charge (respectively) to form **ions**. Despite their electrical charge, ions are stable because their outer electron shells are filled. Common table salt (Figure 2.7) is a good example.

Ionic bonds form when an atom loses e⁻s and becomes a *cation* and another gains them, becoming an *anion*.

Table salt is a crystalline ionic compound composed of ionized sodium and chlorine atoms

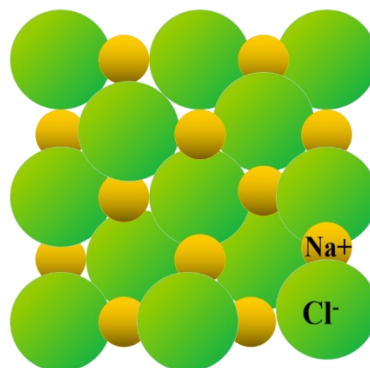


Fig. 2.7: Ionic bonds in table-salt (NaCl) crystals result from the attraction of oppositely charged small Na⁺ and large Cl⁻ ions that bind them together in a regular, crystalline array.

Na (sodium) atoms can donate a single electron to Cl (chlorine) atoms, generating partially charged Na⁺ (sodium) and Cl⁻ (chloride) ions. The oppositely charged ions then come together to form an **ionic bond**, an *electrostatic interaction* of opposite charges that holds the Na⁺ and Cl⁻ ions together in crystal salt. Look up the Bohr models of these two elements, and see how ionization of each leaves filled outer shells (energy levels) in the ions.

2.3.3 Hydrogen Bonds

The hydrogen bond is a subcategory of electrostatic interactions formed by the attraction of opposite charges. As noted above, water molecules attract one another (**cohere**) because of strong electrostatic interactions that form the H-bonds. Water's polar covalent structure enables it to attract positively and negatively charged groups of molecules, making it a good solvent. Solutes (soluble molecules) or polar (charged) molecular surfaces that are attracted to water are **hydrophilic**. Lipids, like fats and oils, are not polar molecules and therefore do not dissolve in water; they are **hydrophobic** (from *hydro*: "water;" *phobic*: "fearing"). Next, we'll take a closer look at the chemistry and properties of water.

2.4 Water Chemistry

Soluble salts like NaCl dissolve because the Cl⁻ and Na⁺ ions attract the partial positive and negative charges (respectively) of water molecules more strongly than other water molecules do. As a result, the salt ionizes; the ions separate from each other. The ionization of NaCl dissolving in water is shown below (Figure 2.8).

Its polar covalent structure makes water a *good solvent* for ionic compounds:

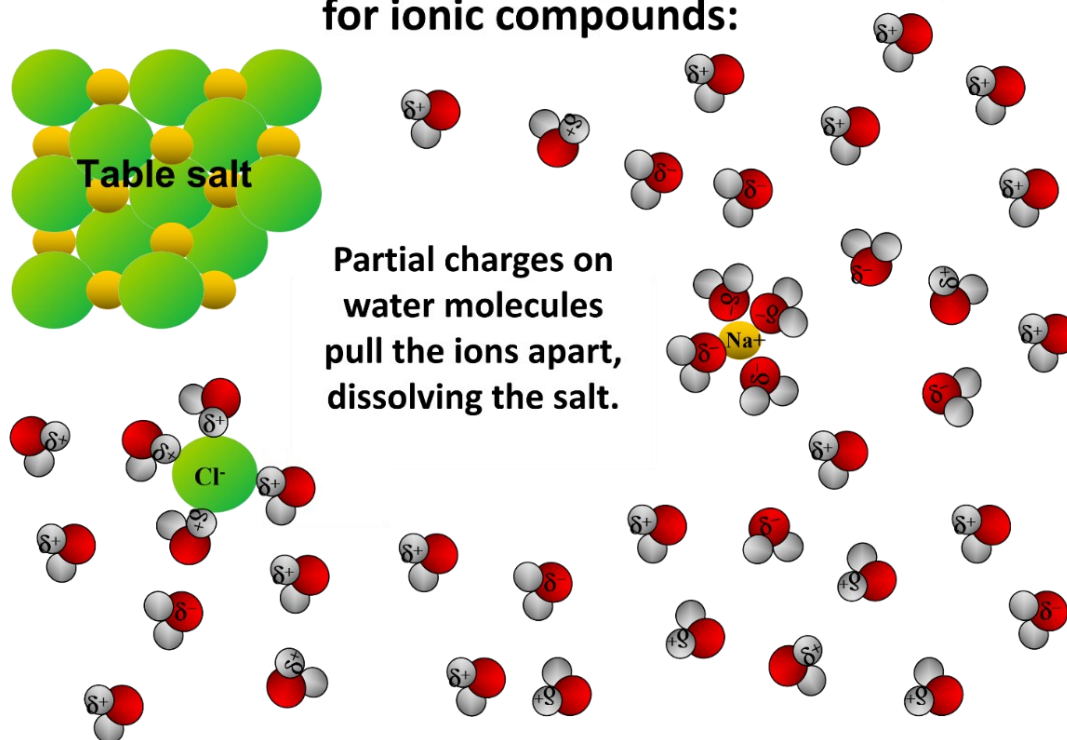


Fig. 2.8: Water's solvent properties result from its polar covalent structure, allowing electrostatic interactions between water molecules at the right and NaCl at the upper left, disrupting ionic bonds, drawing the Na⁺ and Cl⁻ ions into solution.

Water is also a good solvent for macromolecules (e.g., proteins and nucleic acids) with exposed polar chemical groups on their surface that attract water molecules (Figure 2.9).

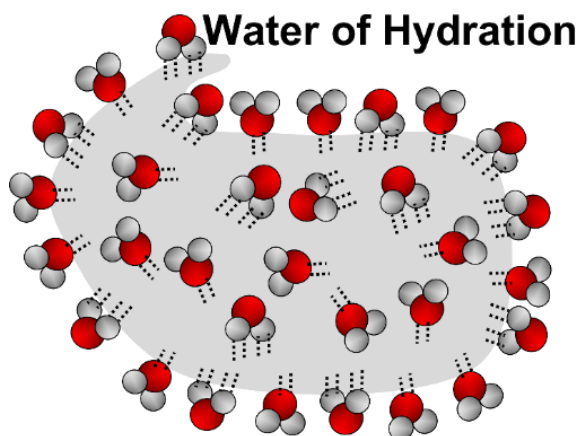


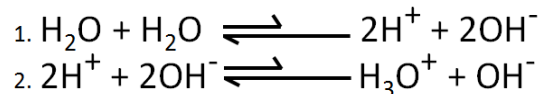
Fig. 2.9: In hydrophilic interactions, charged groups on a macromolecule (e.g., a protein) attract the partial charges on water molecules, hydrating the molecule.



In addition to its being a good solvent, we recognize the following properties of water (all of which result from its polar nature and H-bonding abilities):

- **Cohesion:** Water molecules stick together via H-bonds.
- **High surface tension:** Water's high cohesion means that it can be hard to break the surface; think of the water strider, an insect that literally walks on water.
- **Adhesion:** This results from water's electrostatic interactions with ions and with the partial charges on polar covalent molecules or functional groups. Adhesion explains water's solvent properties and (at least in part) capillary action, in which water molecules *crawl* along hydrophilic surfaces, often against the force of gravity.
- **High specific heat:** The cohesion of water molecules is so strong that it takes a lot of energy to separate them and to make them move faster, i.e., it takes a lot of energy to heat water. Specifically, it takes 1 Kcal, (1 *Calorie*, with a capital C) to heat a gram of water 1°C. High specific heat explains why water "holds heat," staying hot longer than the pot it's in!
- **High heat of vaporization:** It takes even more energy/gm to turn water into vapor (steam)!

One last property of water: it ionizes weakly to form H^+ and OH^- ions—or more correctly, H_3O^+ and OH^- ions. You can think of this as happening in the following two reactions:



Acid molecules added to water will dissociate and release protons. This drives reaction 2 to form more H_3O^+ ions in the solution, in turn driving reaction 1 forward. A pH meter measures the relative acidity or concentration of protons in a solution. Acidic solutions have a pH below 7.0 (*neutrality*). Bases ionizing in water release hydroxyl (OH^-) ions. The increase in OH^- ions removes protons from the solution, driving both reactions in reverse and raising the pH of the solution. To summarize acid-base chemistry: when dissolved in water, acids release H^+ , while bases accept H^+ . Since the pH of a solution is the negative logarithm of the hydrogen ion concentration, the following are true:

- A solution at pH 7.0 is neutral.
- A solution below a pH of 7.0 is acidic.
- A solution above a pH of 7.0 is basic.

Check a basic chemistry textbook to be reminded of the relationship between pH and the $[\text{H}^+]$ in a solution!

2.5 Basic Biochemistry: Monomers, Polymers; Macromolecular Synthesis, Degradation

Like evolution, the origin of life involved a prebiotic *chemical selection*—a kind of "natural selection" of environmental chemicals that favored increasing biochemical possibilities and diversity.

In simple terms, atoms that could interact with a maximal number of other atoms to form the largest number of stable molecules would have been most likely to accumulate in the environment. The tetravalent C atom met these criteria for chemical selection, proving ideal for building an organic chemistry set.

At the same time, water turned out to be the perfect place to launch prebiotic *chemical selection* experiments. Water persists as life's universal solvent, which explains why evidence of water in places beyond our Earth (e.g., other planets in our solar system, the moons of other planets, and other planets in other solar systems) gets us all excited!

2.5.1 Isomerism in Organic Molecules and the Diversity of Shape

The **carbon skeleton** is a perfect platform for organic molecule diversity. The different possible arrangements of atoms and functional chemical groups around C atoms result in **isomerism**. **Isomers** of an organic molecule have the same chemical formula but different shapes, and thus, potentially different chemical properties and biochemical functions. The larger the C-skeleton of an organic molecule, the greater the diversity of molecular shapes available for chemical selection. Look at the examples of **structural isomers** and **geometric isomers** in Figure 2.10.

The C-Skeleton, Isomers & Diversity of Shape

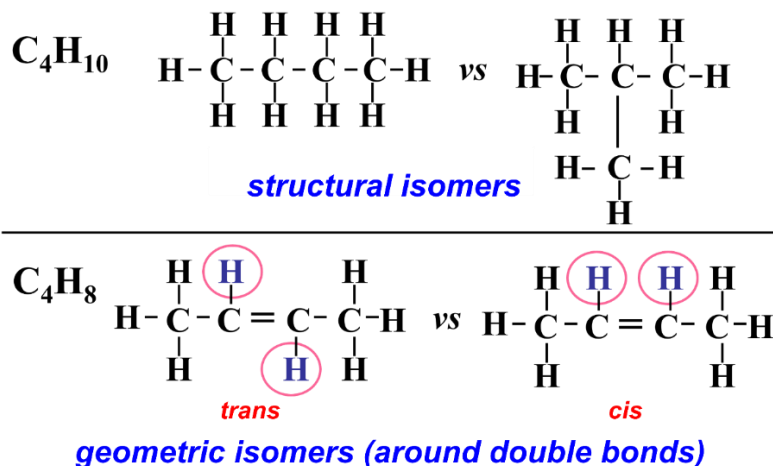


Fig. 2.10: Structural and geometric *isomers* of hydrocarbons create molecules with the same chemical formula but different shapes.

It is easy to see that the structural isomers of C_4H_{10} (Figure 2.10, top panel, left and right) have different shapes. You cannot convert one structural isomer to the other without breaking covalent bonds. In the geometric isomers of C_4H_8 in the lower panel, the H atoms on the double-bonded C atoms can be on the same (*cis*) or opposite (*trans*) side of the *planar* double bond. Geometric isomers, too, cannot be interconverted without breaking chemical bonds. **Optical isomers** are a third kind of isomer. They exist around **optically active** (**asymmetric** or **chiral**) carbons. A chiral C is one that is covalently linked to four different atoms and/or molecular groups.

The principle of chirality is illustrated in Figure 2.11.

Chiral (Asymmetric) Carbons & Optical Isomers



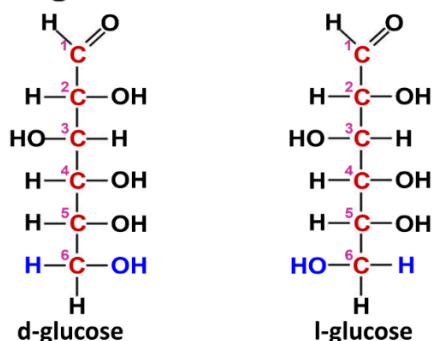
These molecules are different

Fig. 2.11: The two different generic molecules (an *enantiomeric pair*) are mirror images of one another. They form from alternate arrangements of the same molecular groups around a *chiral* carbon.

A pair of molecules that can exist as different *optical isomers* is called an *enantiomer*. Each optical isomer of an enantiomeric pair is structurally identical but differs in *shape* from its mate. Just like structural and geometric isomers, they can't be interconverted without breaking and remaking covalent bonds. A molecule is defined as *optically active* because it bends (rotates) a plane of light in opposite directions in a *polarimeter*: light passing through a solution of one optical isomer is rotated in one direction, while light passing through the other isomer is rotated in the opposite direction. These directions are referred to as *l* for *levo* (meaning *left*) and *d* for *dextro* (meaning *right*). If a molecule has more than one chiral C (e.g., glucose has four chiral carbons), it comprises 4 enantiomers! The behavior of such a *multi-chiral* molecule in a polarimeter will be based on the sum of the optical activities of all the chiral carbons.

The common enantiomer of glucose in our diet, (*d*)glucose, is the enantiomeric pair shown in Figure 2.12 below. These optical isomers are also called *D glucose* and *L glucose*, a convention based on the configuration of the four different atoms or groups around the *last optically active carbon* in a molecule C5 in this case.

Straight-Chain Forms of Glucose



Positions of H and OH on the #5C make these sugars optical isomers (enantiomers)

Fig. 2.12: Two straight-chain forms of glucose - (*d*)glucose and (*l*)glucose - are enantiomers (optical isomers), differing in the arrangement of the H atom and the OH group around C5.

For glucose, *d* and *l* in fact correspond to *D* and *L*, respectively. As we will see for some molecules, the uppercase designation of a chiral molecule does not always indicate how it bends light in a polarimeter, while the lowercase *d* and *l* always do! The chiral carbons in glucose are shown in red above.

Remember that the shape and chemical properties of a molecule dictate its function. Isomerism in organic (carbon-based) molecules would have increased the diversity of molecular shapes available for chemical selection. Early selection of isomers (specific optical isomers in particular) during chemical evolution contributed greatly to the chemical functions and reactions we recognize in cells now, even before there was life on Earth. All life uses the same isomers of glucose in energy reactions and of amino acids in protein-building, a fact that confirms the prebiotic selection of those isomers!

2.5.2 Monomers to Polymers and Back: Dehydration Synthesis and Hydrolysis

All living things build and break down polymers (macromolecules) by **dehydration synthesis** (**condensation** reactions) and **hydrolysis**, respectively. *Dehydration synthesis* and *hydrolysis* reactions are essentially the reverse of each other (Figure 2.13).

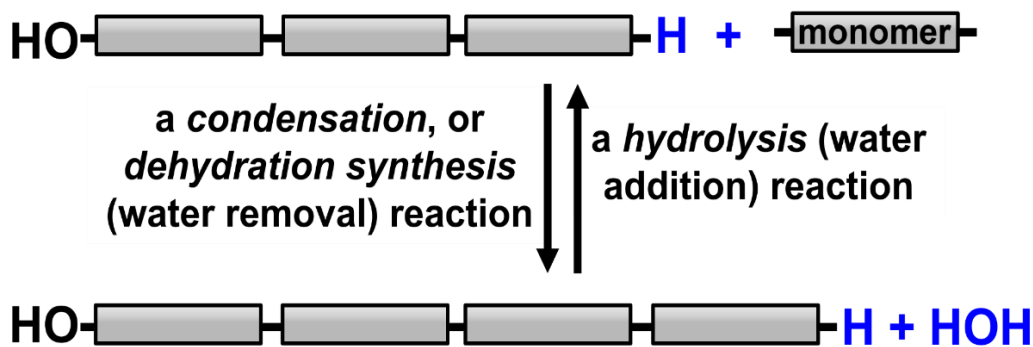


Fig. 2.13: A generic monomer is linked to a growing polymer by water removal (*dehydration synthesis*). Water addition across the linkage between monomers (*hydrolysis*) breaks the polymer back down to monomers.

Condensation reactions build macromolecules by removing a water molecule from interacting monomers. The “bond” that forms in a condensation reaction is *not a single bond*; rather, it is a **linkage** involving several bonds! Polymer synthesis happens when an OH from one monomer and an H group from another are removed and combine to form a water molecule.



[126 Organic Molecules, Monomers, & Polymers](#)



Cells perform repeated dehydration synthesis reactions to build diverse polymers, including polysaccharides and polynucleotides (the RNA and DNA nucleic acids). Repeated condensation reactions between two amino acids (Figure 2.14, below) form the **peptide linkages** that build polypeptides during translation.

Two Amino Acids

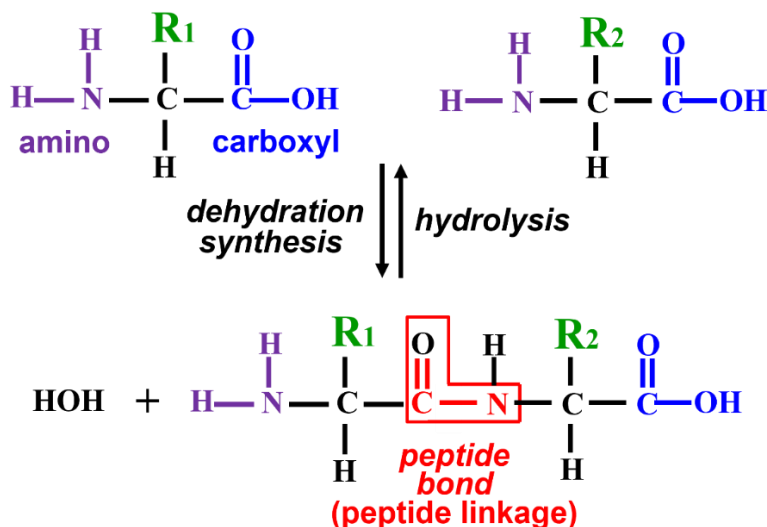


Fig. 2.14: Dehydration synthesis forms peptide linkage (circled) between amino acid monomers. Hydrolysis, the reverse reaction is also shown. Hydrolysis, the reverse reaction is also shown.

Next let's consider the polymerization of glucose. Natural selection settled on using the *(d)glucose* optical isomer for energy metabolism - all living things do it! And none of them, with one known exception (a plant bacterial pathogen) either contain or use the *(l)glucose* isomer. In solution, straight-chain *(d)glucose* becomes cyclic as shown in Figure 2.15.

In solution, glucose forms a ring, making the #1 carbon **chiral**:

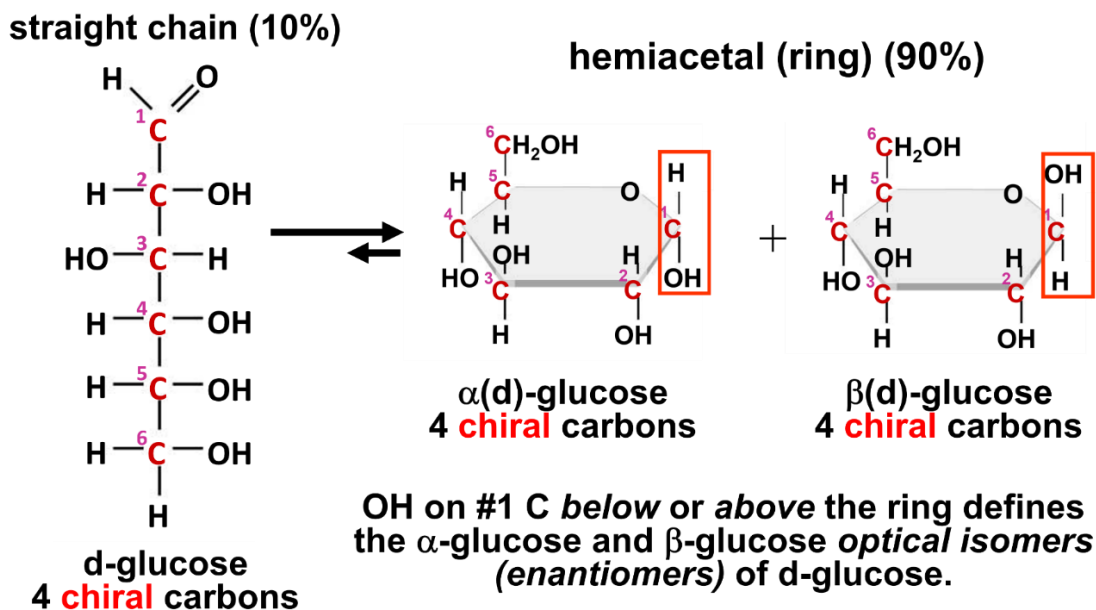


Fig. 2.15: When straight-chain glucose forms a cyclic molecule in solution in water, C1 becomes optically active (chiral), creating a racemic mixture of $\alpha(\text{d})\text{glucose}$ and $\beta(\text{d})\text{glucose}$ enantiomers.

Having selected (*d*)glucose for most cellular energy metabolism, life then exploited the cyclic (*d*)glucose to make the different polysaccharide polymers we now find in plants and animals. Recall that in glucose, C2, C3, C4 and C5 are optically active carbons. As (*d*)glucose becomes cyclic, the molecule gains a fifth chiral carbon, namely, the #1 carbon (C1), forming α (*d*)glucose and β (*d*)glucose in equal amounts.

Straight-chain and cyclic glucose exist in equilibrium in solution, with 90% of glucose monomers in the cyclic form. α (*d*)glucose and β (*d*)glucose are also in equilibrium, each at 45% of the total dissolved glucose. Furthermore, the reaction forming cyclic glucose is reversible, with α (*d*)glucose easily converting to β (*d*)glucose as cells consume the latter (and vice versa). This is an especially important property since the two enantiomers (optical isomers) are used differently, especially in plants.

It is the cyclic forms of glucose that are polymerized. The condensation reactions shown in Figure 2.16 link α (*d*)glucose monomers to form storage polysaccharides, and β (*d*)glucose to make structural polysaccharides..

α and β isomers of (*d*)glucose condense to form two classes of polymers with different physical and chemical properties

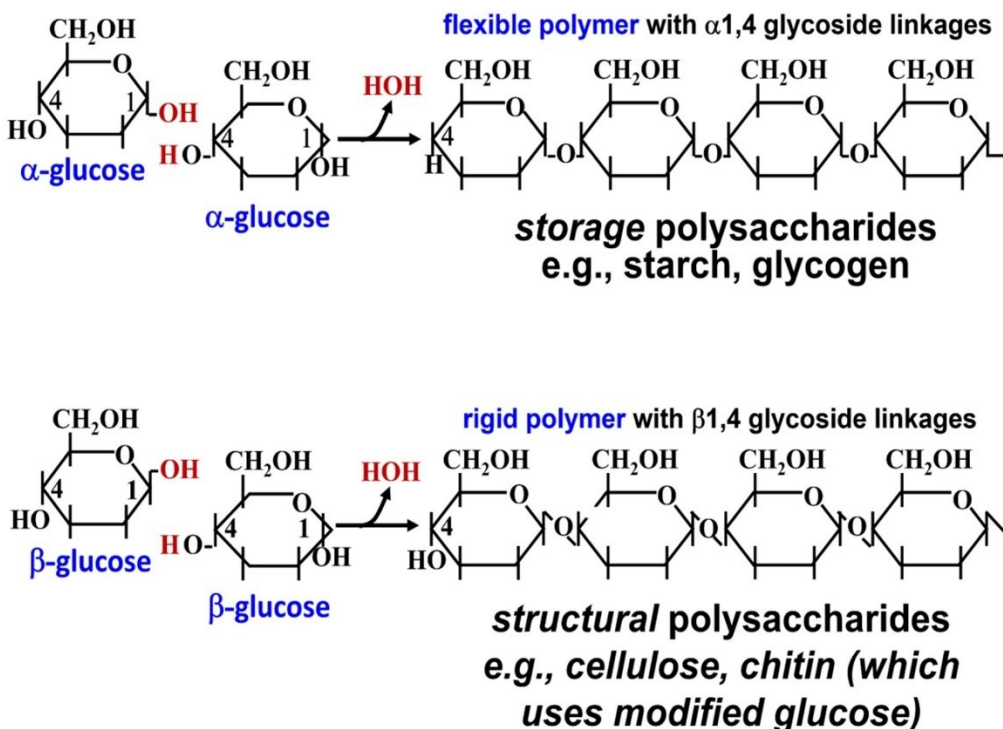


Fig. 2.16: Glucose monomers polymerize by dehydration synthesis (essentially a reversal of hydrolysis) to form polysaccharides. α (*d*)glucose monomers polymerize to form energy-storage molecules (e.g., starch and glycogen); β (*d*)glucose monomers polymerize to form structural polysaccharides (e.g., starch and cellulose).

The –OH ions (hydroxyl groups) on the C1 of ***α(d)glucose*** are *below* the glucose rings. The condensation reaction removes a water molecule, linking the sugars by an ***α1,4 glycoside linkage*** in the dimer, connecting them by their C1 and C4 carbons. Other *linkages* are possible. For example, diverse *α-glycoside* linkages characterize ***branched storage polysaccharides***—like *glycogen* in animals and *starches* in plants. On the other hand, when ***β(d)glucose*** isomers polymerize, they form rigid ***structural polysaccharides***, such as those of cellulose in plant cell walls. A modified *β-glucose* called *N-acetyl glucosamine* (not shown) polymerizes to form *chitin*, the principal component of fungal cell walls and of the tough exoskeleton of arthropods (e.g., insects and crustaceans). Later, when we look at *translation*, we'll consider why only ***L*** amino acids are used to make polypeptides! And when we look at *replication* and *transcription*, we'll see the condensation reactions that form phosphate ester (phosphodiester) linkages when DNA and RNA are synthesized from nucleotide monomers. Summarizing the role of condensation reactions in the formation of biological polymers:

- Linkages in biopolymers are broken down and re-formed daily in our lives! Digestion (the breakdown) begins after a meal. The hydrolysis of glycoside linkages starts in your mouth. Further polysaccharide digestion and the breakdown of peptide (among other) linkages continues in your stomach and small intestines. Then our cells use condensation reactions to finish the job of turning carrot- and cow-derived monomers into you and me!
- Prebiotic ***chemical evolution*** has selected only one each of the optical isomers of glucose, amino acids, and other enantiomeric monomers with which to build polymers. This is so even though some of the different isomers are available and even used by cells for different purposes. The flexible ***α(d)glucose*** polymers were selected as storage polysaccharides to be used for energy. The storage polysaccharides include plant *starches* and animal *glycogen*. Likewise, the rigid inflexibility of ***β(d)glucose*** polymers was selected precisely because it reinforced cell structure and stability. Since all organisms store carbohydrate energy in ***α(d)glucose*** polymers, and since ***β(d)glucose*** polymers are almost universally used to strengthen cell structure, these selections must have occurred early in the history of life.



[127-2 Carbohydrates: Sugars and Polysaccharides](#)



[128-2 Lipids, Triglycerides, and Phospholipids](#)



[129-2 Proteins: Amino Acids and Polypeptides](#)



[130-2 130-2 Nucleic Acids: Nucleotides, DNA, and RNA](#)



2.5.3 A Tale of Chirality Gone Awry

To conclude this chapter and to emphasize the significance of chirality to life, here is what can happen if the wrong isomer ends up in the wrong place at the wrong time....

Consider the story of *thalidomide*, a tragic example of what happens when we are unaware of enantiomeric possibilities. Introduced in 1957, thalidomide was sold as an over-the-counter anti-nausea drug for patients undergoing cancer therapies and as an effective morning-sickness remedy for pregnant women. However, by the early 1960s, the births of about ten thousand infants with severely deformed limbs were connected to the drug. These deformities characterized roughly half of these infants that survived. Once the connection was made, the response was, of course, to pull thalidomide off the market.

Thalidomide is a *teratogen*. Teratogens are substances or conditions (drugs, chemicals, radiation, illness during pregnancy...) that cause deformities during embryogenesis and fetal development. The chemical bases of thalidomide's effects are based on its *enantiomeric (chiral)* structure, in which an amine-containing ring can exist in front of or behind the rest of the molecule. The optical isomers of thalidomide are shown below in Figure 2.17.

Optical Isomers of Thalidomide

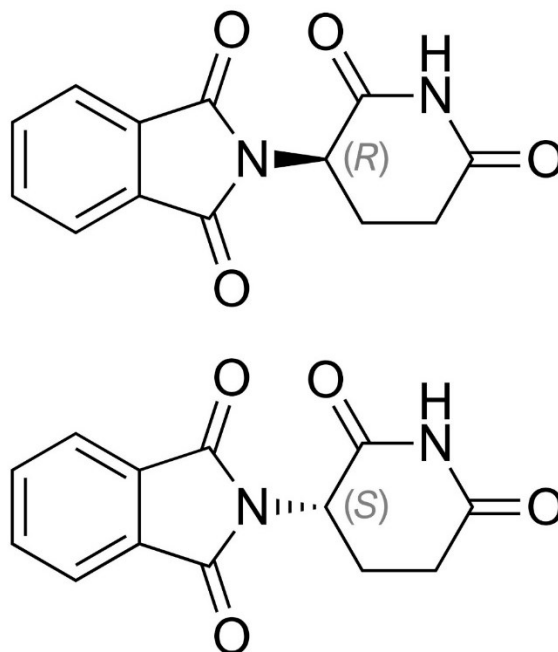


Fig. 2.17: Two enantiomers (optical isomers) of thalidomide (**R** and **S**) form in water.

The two optical isomers are referred to as "S" and "R." Of these, the S isomer is the teratogen. While synthesis of pure R is possible, when used in treatment, R and S easily interconvert, creating a **racemic mixture** (recall that α - and β -(d)glucose are such a solution. In the mother, S isomers are transferred to the embryo or fetus, with terrible consequences.

Remarkably, there were relatively few cases of thalidomide-induced birth deformities in the United States, largely because of the efforts of ^{2,1}[Frances Oldham Kelsey](#), the person in charge of the FDA's review of the drug. The German pharmaceutical company *Chemie Grünenthal* (developer of Thalidomide) and an American pharmaceutical company had applied for FDA approval for US distribution of the drug. Dr. Oldham Kelsey refused approval on multiple occasions, arguing that the safety of thalidomide had not been demonstrated. This was even before it was shown to cause birth deformities! In 1962 President John F. Kennedy presented her with the *President's Award for Distinguished Federal Civilian Service* for not allowing thalidomide to be approved for sale in the United States without sufficient safety testing—potentially saving thousands of lives.

Of course, we already knew that cells synthesized polymers from specific optical isomers of their precursor monomers. So, the sad thalidomide story resulted from the untested effects of an unexpected optical isomer. Many countries quickly tightened their preapproval drug-testing regulations because of this tragedy.

In a more hopeful twist of the tale, thalidomide has turned out to be effective in treating cancer, leprosy, rheumatoid arthritis, and other autoimmune diseases. Such therapeutic benefits may be due to its *anti-inflammatory effects*. The effects of thalidomide on tumor growth seem to be due to its inhibition of *angiogenesis* (development of blood vessels) in the tumors. Ironically, blockage of angiogenesis may also have contributed to the failure of proper limb growth during pregnancy.

To conclude, when all is normal, the shapes of molecules have been uniquely selected for the specificity of reactions essential to life.



[131-2 Shape and the Specificity of Molecular interactions](#)

Some iText & VOP Key Words and Terms

acids and bases	geometric isomers	polar covalent bonds
adhesion	glycogen	polymers
α -glucose	glycoside linkage	polynucleotides
amino acids	heat of vaporization	polypeptides
angiogenesis	hydrogen bonds	polysaccharides
atom	hydrolysis	potential energy
atomic mass	hydrophilic	properties of water
β -glucose	hydrophobic	protons
Bohr model	ionic bonds	quantum
carbohydrates	ionization	racemic mixture
cellulose	isomers	RNA
chirality	isotopes	salts
chiral carbon	kinetic energy	scanning tunneling

chitin	lipids	photon
cohesion	macromolecules	sharing electrons
condensation reaction	microscope	solutes
dehydration synthesis	molecule	specific heat
digestion	monomers	starches
DNA	neutrons	structural isomers
electron shell	nucleotides	surface tension
electrons	optical isomers	teratogen
electrostatic interaction	orbitals	thalidomide
element	partial charge	triglycerides
enantiomers	peptide linkage	valence
ester linkage	pH	van der Waals forces
excitation	phosphate ester linkage	water ions
fats	phosphodiester linkage	water of hydration
fluorescence	phospholipids	

CHAPTER 2 WEB LINKS



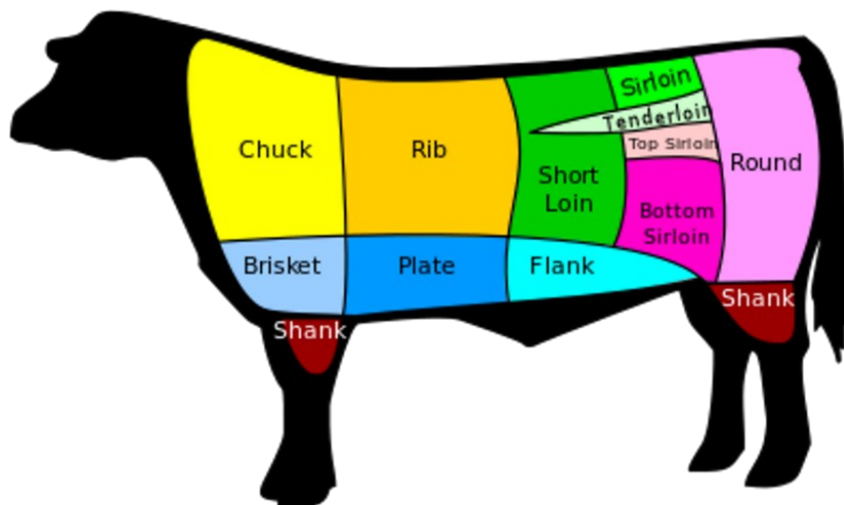
Chapter 3

Details of Protein Structure

Protein structure and configuration: primary, secondary, tertiary, quaternary; protein folding, domains and motifs, studying proteins

Reminder: For inactive *links*, google key words/terms for alternative resources.

However good it might look,



this is *not* protein primary structure!

3.1 Introduction

Protein molecules, which are each comprised of one or more *polypeptides*, are the workhorse of the cell, responsible for just about all aspects of life:

- They are the ***catalysts*** that make biochemical reactions possible.
- They are ***membrane components*** that selectively let substances in and out of cells.
- They allow ***cell-cell communication*** and cell ***response to environmental change***.
- They form the internal structure of cells (***cytoskeleton***) and nuclei (***nucleoskeleton***).
- They enable the ***motility*** of cells and e.g., particles and organelles within cells.
- They are responsible for other cell functions too numerous to summarize here!

We owe much of what we know about biomolecular structure to the development of X-ray crystallography. In fact, an early determination of the structure of insulin (as well as penicillin and vitamin B₁₂) using X-ray crystallography earned Dorothy Hodgkin the 1964

Nobel Prize in Chemistry. In this chapter, we look at the different levels (*orders*) of protein structure and at what it takes to be a functional protein. The **primary structure** (*1^o structure*) of a polypeptide is its amino acid sequence.

Interactions between amino acids near each other in the sequence will cause the polypeptide to fold into **secondary** (*2^o*) **structures**, including the α -helix and the β -pleated sheet (or just β -sheet) conformations. **Tertiary** (*3^o*) **structures** form when noncovalent interactions occur between amino acid sidechains at some distance from one another in the primary sequence, causing a polypeptide to further fold into a more complex 3D shape. Other proteins (called *chaperones*!) facilitate the accurate folding of a polypeptide into its correct, bioactive, 3D conformation. **Quaternary** (*4^o*) **structure** refers to proteins made up of two or more polypeptides. Refer to the four levels or orders of structure in Figure 3.1 as we explore how each level affects the shape and biological/biochemical function of the protein.

Levels of Protein Structure

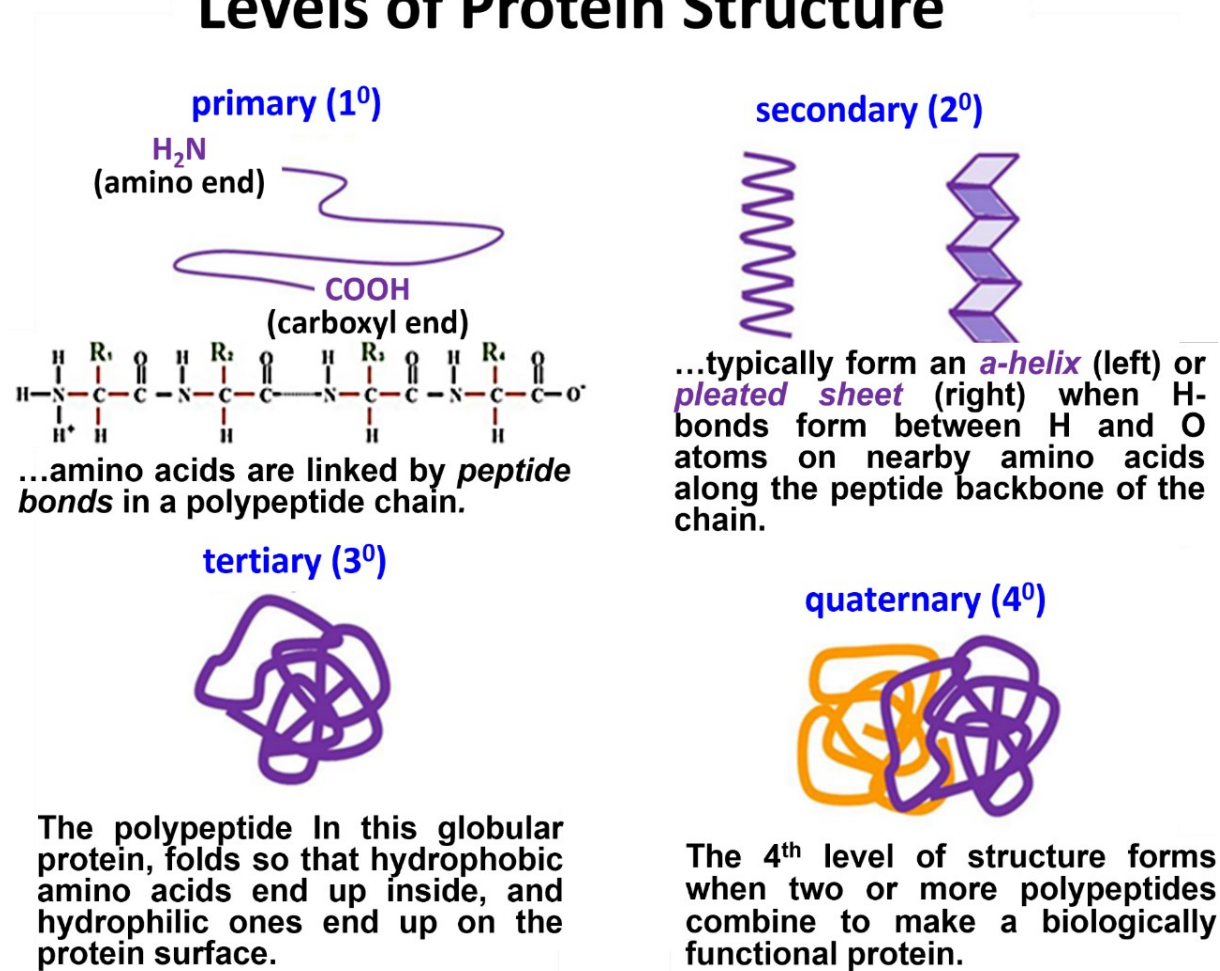


Fig. 3.1: Three orders (levels) of protein structure describe polypeptides: *primary* (upper left), defined by peptide bonds between amino acids), *secondary* (upper right), based on H-bonds between amino and carboxyl groups of nearby amino acids in a polypeptide chain, and *tertiary* (lower left), based mostly on non-covalent interactions between amino acid side-chains in a polypeptide). A fourth order of protein structure, or *quaternary structure* (lower right) applies to proteins composed of two or more polypeptides.

In addition to protein structures dictated by noncovalent interactions, we'll see covalent bonds between specific amino acids (e.g., cysteines) that end up near each other after folding. These function to stabilize tertiary and quaternary structures. Many proteins must also bind metal ions (e.g., Mg^{++} , Mn^{++}) or small organic molecules (e.g., heme) to become biologically active. Finally, we look beyond these orders of structure at protein **domains** and **motifs**, which evolved to perform specific protein functions.

As we try to understand molecular (especially macromolecular) function, a clear recurring theme emerges: a protein's function depends on its shape, its **conformation**. In turn, protein conformation is based on the location and physical and chemical properties of critical **functional groups**, usually amino acid side chains. Watch for this theme as we look at enzyme catalysis, the movement of molecules in and out of cells, the response of cells to changing environments, the ability of cells to travel, the ability of organelles to move around inside of cells, the process of DNA replication, the regulation of gene transcription and protein synthesis... and just about everything a cell does! We will conclude this chapter with a look at some techniques for studying protein structure.

Learning Objectives

When you have mastered the information in this chapter, you should be able to do the following:

1. Define and distinguish between the orders of protein structure.
2. Differentiate between β -sheet, α helix, and *random coil* structures based on the atomic interactions involved on each.
3. Trace the path to the formation of a polypeptide, define its primary structure and explain how its final structure might be determined by its amino acid sequence.
4. Describe how globular proteins arise from the hydrophobic and hydrophilic interactions that drive protein folding and how changes in protein shape can cause disease.
5. Formulate a hypothesis (or look one up) to explain why the amino acid glycine is a disruptor of alpha-helical polypeptide structure.
6. Compare and contrast motif and domain structure of proteins and polypeptides, as well as their contributions to protein function.
7. Describe different techniques for studying proteins and explain what physical/chemical differences between proteins that make each technique possible.

3.2 Levels (Orders) of Protein Structure

The three levels of polypeptide structure are primary, secondary, and tertiary structure. Quaternary structure arises from an association of two or more polypeptides that create higher-order protein structures. Superimposed on these levels, or orders, are still other features of protein structure. These are created by the specific amino acid configurations in a mature, biologically active protein. Let's begin with a look at primary structure.

3.2.1 Protein Primary Structure: L-Amino Acids and the -C-C-N- Polypeptide Backbone

The **primary structure** of a protein simply refers to the amino acid sequence of its polypeptide chain(s). Cells use only twenty amino acids to make polypeptides and proteins, although they do use a few additional amino acids for other purposes. The **peptide linkages** between amino

acids in polypeptides form in *condensation reactions* in cells during protein synthesis (i.e., **translation**). The linkages involve multiple covalent bonds. They break and rearrange between the *carboxyl* and *amino* groups of amino acids during linkage formation.

The twenty amino acids found in proteins are shown below in Figure 3.2.

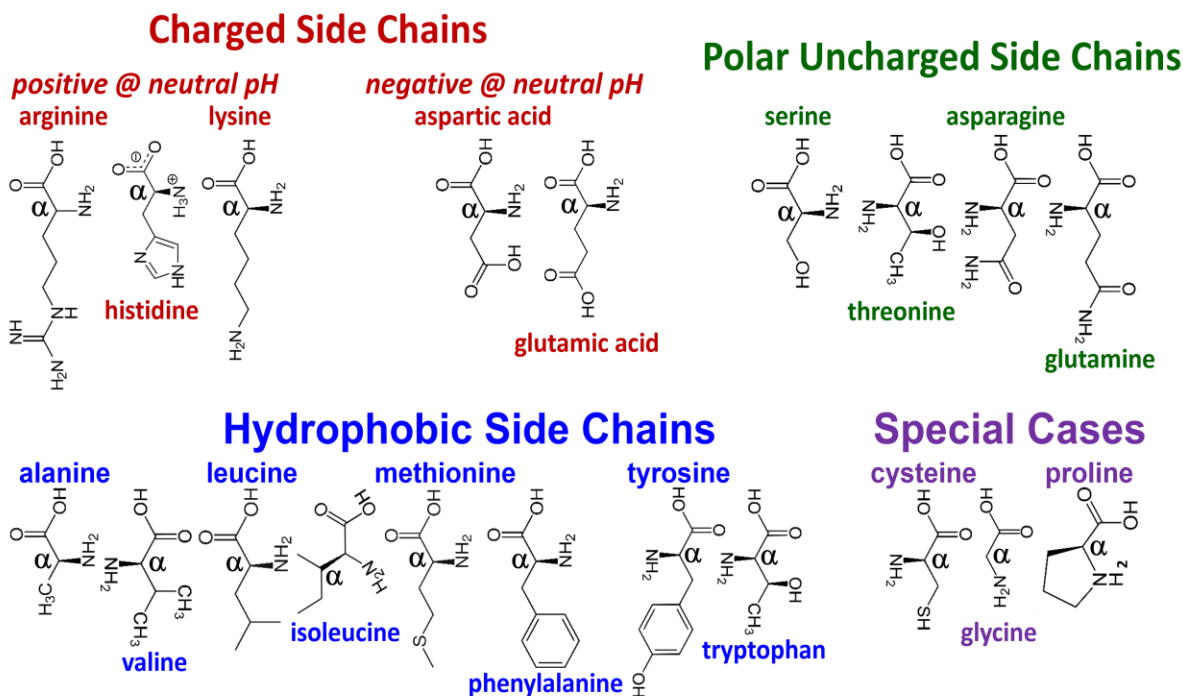


Fig. 3.2: Chemical characteristics of the twenty amino acids found in the proteins of cells. All side chains (R groups) are shown at the bottom of each structure, bound to an α carbon.

In all twenty amino acids (except glycine), the α -carbon is bound to four different groups, making them *chiral* or *optically active*.

3.2.1.a Peptide Bond Formation and Polypeptide Primary Structure

Recall that chiral carbons allow for mirror image **D** and **L** (or **d** and **l**) optical isomers. Remember, *only* the lower-case **d** and **l** define the optical properties of isomers. Just to make life interesting, **L**-amino acids are dextrorotary in a polarimeter, making them (**d**)-**amino acids**! While both (**d**)- and (**l**)-amino acid enantiomers exist in cells, only (**d**)- (i.e., **L**-) amino acids (along with glycine) are used by cells to build polypeptides and proteins by the process called **translation**.

We say that polypeptides have polarity because they have “free” **carboxyl ends** and “free” **amino ends**. Within a polypeptide, amino acid side chains end up alternating on opposite sides of a **C-C-N-C-C-N-...** **polypeptide backbone** because of covalent-bond angles along the backbone. You can prove this to yourself by making a short polypeptide with the kind of molecular modeling kit you may have used in a chemistry class! The C-C-N-C-C-N- backbone is the underlying basis of higher orders (or levels) of protein structure. A partial polypeptide is shown below in Figure 3.3.

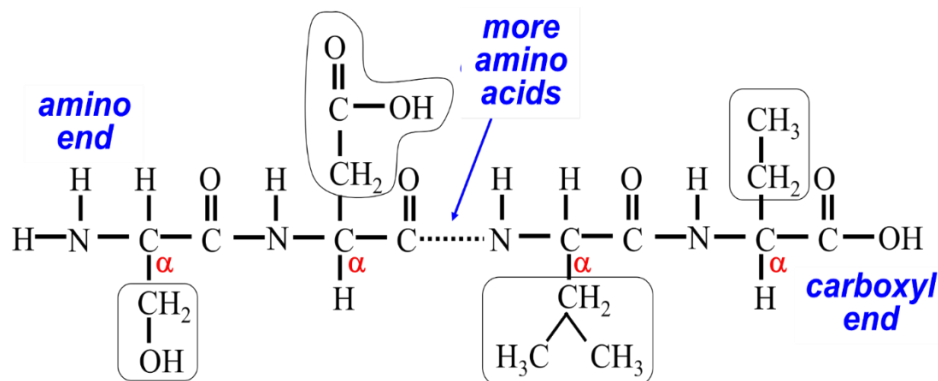


Fig. 3.3: The amino and carboxyl ends of a polypeptide define its polarity, with positively charged amino and negatively charged carboxyl ends at physiological pH. The positions of the circled amino acid R groups alternate along the polypeptide backbone. The two R groups in the carboxyl region (right) are hydrophobic; the two in the amine region (left) are polar.



[132-2 Amino Acid Sequence & Protein Primary Structure](#)



3.2.1.b Determining Protein Primary Structure—Polypeptide Sequencing

Primary structure is dictated directly by the gene encoding the protein. After transcription of a gene, a ribosome *translates* the resulting mRNA into a polypeptide. Frederick Sanger demonstrated the first practical protein sequencing method when he reported the amino-acid sequence of the two polypeptides of *bovine* (cow) *insulin*. His technique involves stepwise hydrolysis (called an *Edman Degradation*) of polypeptide fragments; each hydrolysis leaves behind a single amino acid that can be identified, and a polypeptide fragment shortened by one amino acid. Sanger received a Nobel Prize in 1958 for this feat. By convention, the display and counting of amino acids always starts at the amino end (the N-terminus, at the left); the count ends with a free NH₂-group at the right, as suggested in Figure 3.3.

For some time now, the sequencing of DNA has replaced most direct protein sequencing. The method of DNA sequencing, colloquially referred to as the *Sanger dideoxy* method (yes, this is the same Sanger who first sequenced proteins), quickly became widespread and was eventually automated, enabling rapid gene (and even whole-genome) sequencing. Now, instead of sequencing polypeptides directly, we can, by knowing the genetic code, infer amino-acid sequences from gene sequences isolated by cloning or revealed after complete genome sequencing projects. DNA sequencing is such a powerful window into gene and cell function, not to mention evolution, that Sanger won a second Nobel Prize for the DNA sequencing work in 1980!

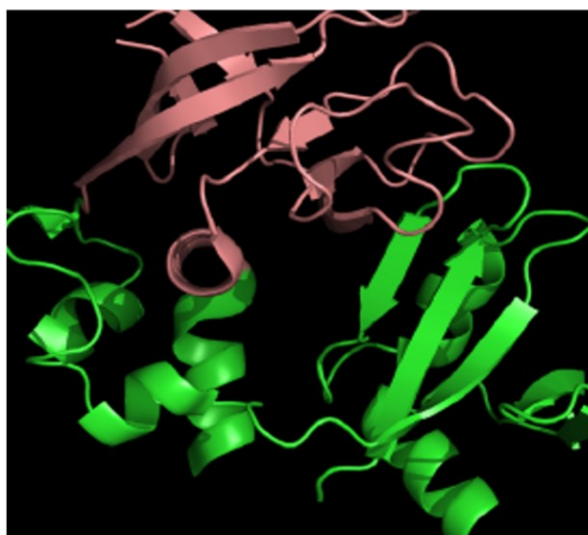
The different physical and chemical properties of each amino acid result from the side chains on its α -carbons. The unique physical and chemical properties of polypeptides and proteins are determined by their unique combination of amino acid side chains and their interactions within and between polypeptides. In this way, primary structure reflects the genetic underpinnings of polypeptide and protein function. The higher-order structures that account

for the functional *motifs* and *domains* of a mature protein derive from its primary structure. Christian Anfinsen won a half share of the 1972 Nobel Prize in Chemistry for demonstrating that this was the case for the ribonuclease (RNase) enzyme; Stanford Moore and William H. Stein earned their quarter shares of the prize for relating the structure of the active site of the enzyme to its tertiary structure and its catalytic function. See ^{3.1}[Anfinsen et al. 1972 Chemistry Nobel Prize](#) for more.

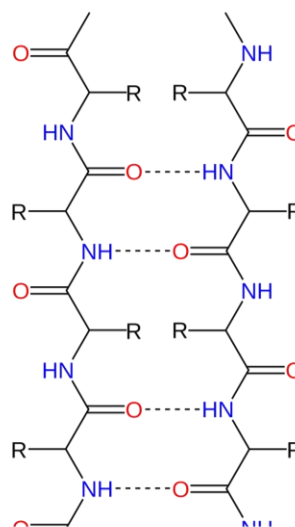
3.2.2 Protein Secondary Structure

Secondary structure refers to highly regular local structures within a polypeptide (e.g., α -*helix*) and either within or between polypeptides (β -*pleated sheets*). Linus Pauling and his coworkers suggested these two types of secondary structure in 1951. A little Linus Pauling history would be relevant here! By 1932 Pauling had developed his *electronegativity scale* of the elements, which could predict the strength of atomic bonds in molecules. He contributed much to our understanding of atomic orbitals and later of the structure of biological molecules. He earned the 1954 Nobel Prize in Chemistry for this work. Pauling and his colleagues went on to predict the α helical and β -pleated sheet secondary structure of proteins (Figure 3.4, below).

Protein Secondary Structure



α -helical (coiled ribbons), pleated sheet (aligned arrowhead ribbons), and random coil regions in a single (upper pink or lower green) protein



β -pleated sheets formed by H-bonds between polypeptide chains

Fig. 3.4: Left: a computer-generated 3D image (left) shows α -helical and pleated sheet “ribbon” regions separated by less-organized *random-coils* of amino acids in each of the two polypeptides (upper pink and lower green) of a two-subunit protein. The β -pleated sheet regions are shown as aligned arrow-ribbons. Right: Pleated sheet structures result when hydrogen bonds form between adjacent regions shown at the right.

In 1962, Max F. Perutz and John C. Kendrew earned the Nobel Prize in Chemistry for their X-ray crystallographic studies of the 3D structure of hemoglobin, confirming some of Pauling's predictions. Although Pauling did not earn a second Nobel for his studies predicting hemoglobin protein structure, he did win the 1962 Nobel Peace prize for convincing almost ten thousand scientists to sign a petition to the United Nations urging a vote to ban atmospheric nuclear bomb tests. Clearly 1962 was a good year for Nobel prizes for 'hemoglobinologists'! More about Pauling's extraordinary life (e.g., at ^{3,2}[L. Pauling-a Short Biography](#)) is worth a read!

Secondary structure conformations occur due to the spontaneous formation of hydrogen bonds between amino groups and oxygens along the polypeptide backbone, as shown in the two left panels in Figure 3.4. Note that amino acid side chains play no significant role in secondary structure. The α helix or β sheets are the most stable arrangement of H-bonds in the chain(s), and both are typically found in the same protein (on the right in Figure 3.4). These regions of ordered secondary structure in a polypeptide can be separated by varying lengths of less-structured peptides called *random coils*.

As illustrated in Fig. 3.4, all three of these protein secondary structure features can be found in a single or polypeptide (left) or in each of the polypeptides of a multi-subunit protein (right), each of which has already folded into its tertiary structure (see below). Pleated sheets are shown as *ribbons* with arrowheads representing *N-to-C* or *C-to-N* polarity of the sheets. As you can see, a pair of ribbons, (peptide regions forming a pleated sheet, may do so either in the parallel or antiparallel directions. Whether aligned ribbons are parallel or antiparallel depends on other influences on how the polypeptide(s) will fold to form tertiary structure.

Some polypeptides never go beyond their secondary structure, remaining fibrous and insoluble. Keratin is perhaps the best-known example of a *fibrous protein*, making up hair, fingernails, bird feathers, reptilian (but not fish!) scales, and even one of the filaments of the cytoskeleton. However, most polypeptides and proteins do fold and assume tertiary structure, becoming soluble *globular proteins*.



[133-2 Protein Secondary Structure](#)



3.2.3 Protein Tertiary Structure

Polypeptides acquire their *tertiary structure* when *hydrophobic* and nonpolar interactions between amino acid side chains spontaneously draw them together to exclude water. This folding is aided by the formation of *salt bridges* and H-bonds between polar side chains in the interior of the globular polypeptide, incorporating α helices and β sheets into globular shapes. Polar (*hydrophilic*) side chains with no other polar side-chain partners typically stay on the outer surface of the "globule," interacting with water to dissolve the protein (recall *water of hydration*). Though based on noncovalent bonds, tertiary structures are nonetheless strong simply because of the large numbers of otherwise-weak interactions that form them. The forces that cooperate to form and stabilize 3D polypeptide and protein structures are illustrated below in Figure 3.5.

Interactions Governing Polypeptide Tertiary Structure

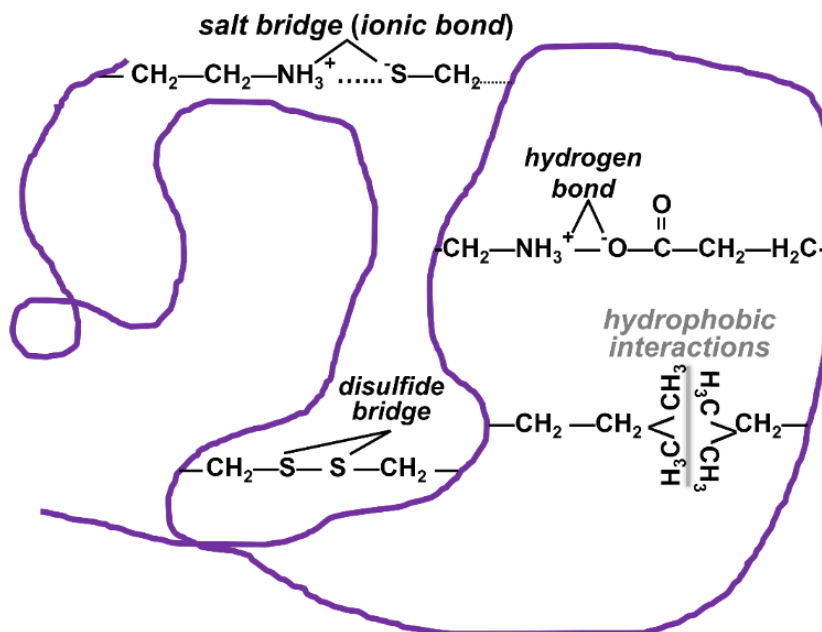


Fig. 3.5: Tertiary structure is created by noncovalent *hydrophobic* (non-polar) amino acid interactions as well as *H-bonding* and ionic interactions in the interior of a polypeptide, leaving charged (hydrophilic) amino acid side chains to interact with water on the exterior of a typical “globular” protein. Covalent disulfide bonds between cysteine amino acids help stabilize tertiary structures.

Covalent disulfide bonds between cysteine amino acids in the polypeptide (shown in Figure 3.5) can further stabilize tertiary structure. Disulfide bonds (bridges) form in a polypeptide when cysteines far apart in its primary structure of the molecule end up near each other.

Formation of a Disulfide Bridge

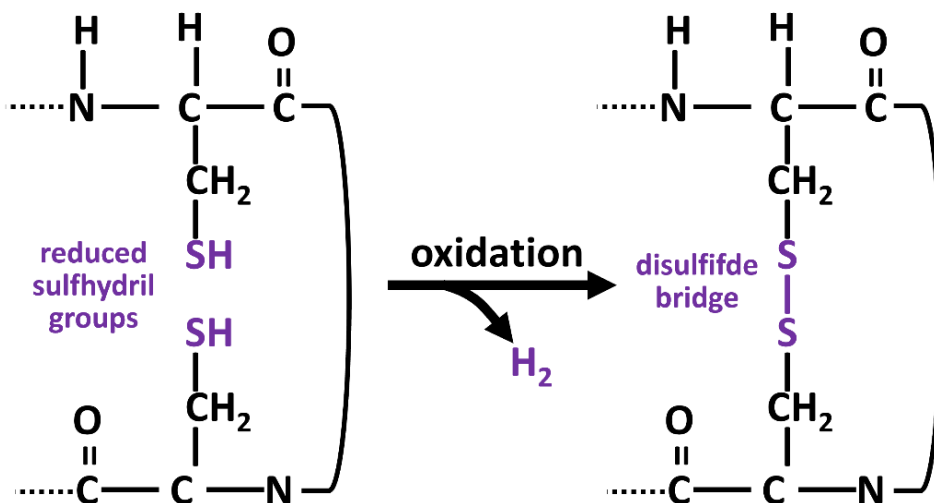


Fig. 3.6: Disulfide bonds (or bridges) form when cysteine amino acids in one or more polypeptides end up near each after it (or they) fold into higher-order structures. Oxidation reactions between the close SH (sulfhydryl) groups (left) result in linking their two sulfurs in a disulfide bond, or bridge (right).

When the **SH** (*sulphydryl*) groups in the cysteine side chains are oxidized, they form the disulfide (-S-S-) bonds shown in Figure 3.6.



[134-2 Protein Tertiary Structure](#)



[135-2 Disulfide Bridges Stabilize 3° Structure](#)



Disulfide bridges in a protein are crucial to the active protein. Imagine how changes in temperature or ionic strength would disrupt noncovalent bonds required for the correct 3D shape of the active protein. Unaffected by such changes, disulfide bridges limit the disruption and enable proteins to refold correctly when conditions return to normal (think *homeostasis*!).

As we'll see, protein activity can be regulated by chemical modification (e.g., phosphorylation).

3.3 Changes in Protein Shape Can Cause Disease

While the conformation of a protein determines its biological function, an allosteric (shape) change can moderate or disrupt its function. Under normal circumstances, cells use changes in protein shape to regulate metabolism. Such *allosteric regulation* is well documented in familiar biochemical pathways such as glycolysis and is discussed in more detail elsewhere. Less well understood is how (or why) conformational change in some protein's cells can have devastating effects.

3.3.1 Sickle Cell Anemia

Mutations of globin genes can cause hemoglobin disorders characterized by inefficient oxygen delivery by blood. In the 1940s, the British biochemist J. B. S. Haldane studied southern African regions and made a correlation between high incidences of hemoglobin disorders and low incidences of malaria, suggesting that heterozygous individuals (i.e., those that had only one copy of a mutant hemoglobin gene) were somehow protected from malaria.

Sickle cell anemia is a well-known example of a hemoglobin disorder and is caused by a single base change in the gene for human β -hemoglobin (one of the polypeptides in hemoglobin). Since red blood cells are rich in hemoglobin, the abnormal shape of the β -hemoglobin can cause the cells themselves to become sickle shaped. Sickle cells disrupt capillary flow and oxygen delivery, causing the symptoms of anemia.

While sickle cell anemia originated in Africa, it probably spread to the United States because of the slave trade. It may even have abetted the slave trade. Europeans exploiting their African colonies' natural resources were dying of malaria, but African natives seemed unaffected. Europeans, having brought stowaway malarial mosquitos to the new world in the first place, figured that enslaved Africans would survive the illness in the Americas.

We know now that individuals heterozygous for the mutant β -hemoglobin suffer *sickle-cell trait* and are generally unaffected, because at least some of their hemoglobin is normal. Homozygous individuals that make only the sickle-cell β -hemoglobin variant suffer more frequent and more severe episodes of the disease. Stressors that can trigger sickling include infection or dehydration. Compare normal red blood cells to a sickle cell below in Figure 3.7.

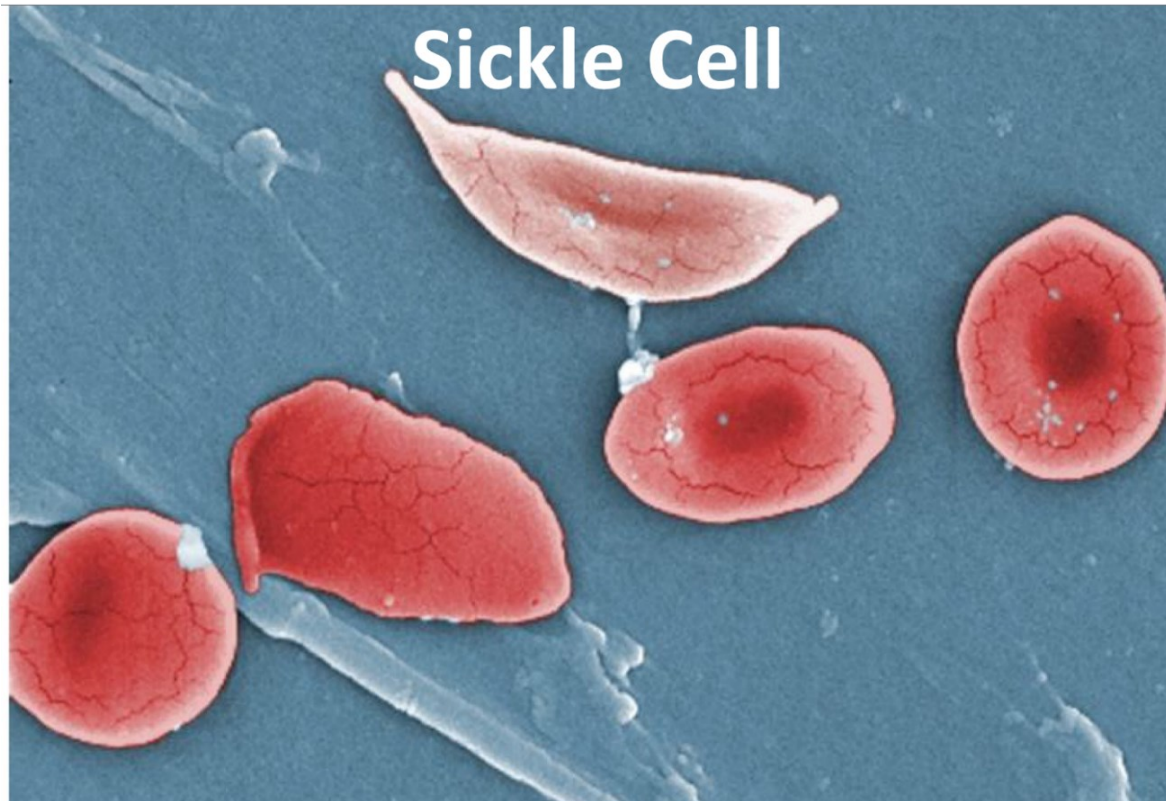


Fig. 3.7: Light micrograph of a sickled erythrocyte (the red blood cell at the top).

The sickle-cell gene mutation affects perhaps more than one hundred million people worldwide, including 8–10% of African Americans. For more demographic information, see ^{3.3}[Sickle Cell Trait Demographics](#) and ^{3.4}[Sickle Cell Data](#). In Africa, heterozygotes with sickle-cell trait are protected from malaria, confirming Haldane's hypothesis. But patients homozygous for the β -hemoglobin mutation derive little benefit from its antimalarial effects.

In the meantime, despite a 33% reduction in cases of malaria in recent years, this disease, caused by a mosquito-borne parasite, still threatens half of the people on the planet, causing over a half-million deaths per year. There are treatments (other than mosquito nets and killing mosquitos), but currently there is still no preventive vaccine.

3.3.2 The Role of Misshapen and Misfolded Proteins in Alzheimer's Disease

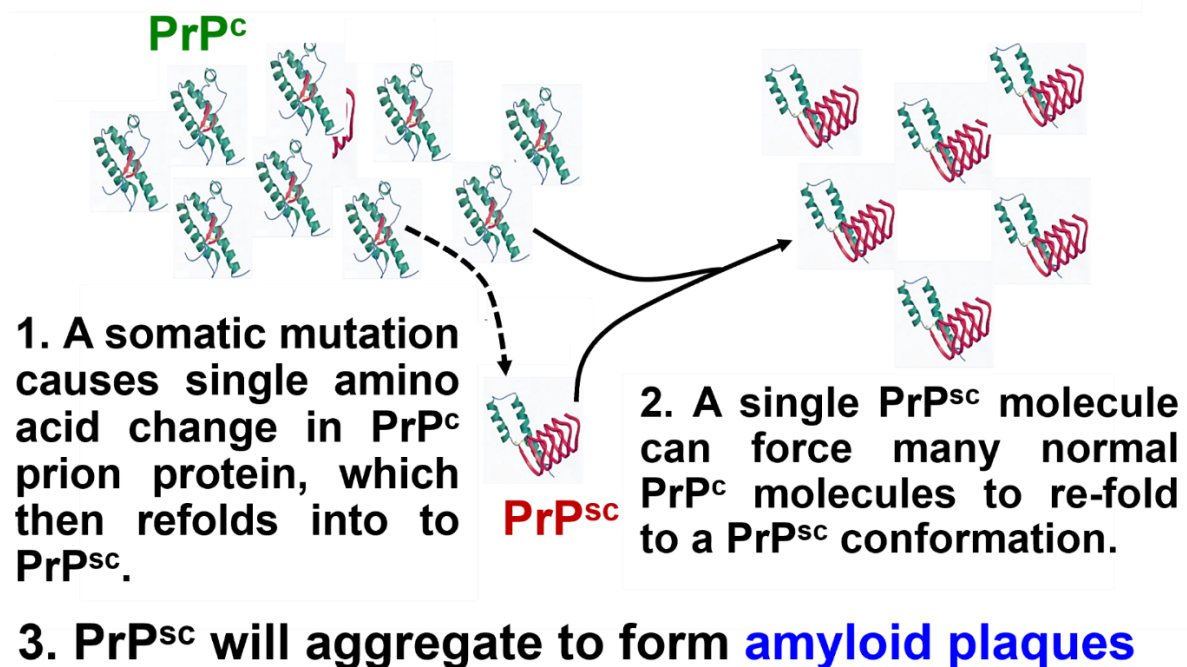
Prion proteins, when first discovered, seemed to behave as infectious agents that could reproduce without DNA or other nucleic acid. As you can imagine, this highly unorthodox and

novel hereditary mechanism generated its share of controversy. Read about research on the cellular prion (PrP^c) protein at ^{3,5}[Wikipedia Take on Prions](#).

Of course, prions turned out *not* to be reproductive agents of infection after all. Recent studies of prions have revealed several normal prion protein functions, including roles in memory formation in mice and sporulation in yeast (check out ^{3,6}[Prion Proteins and the Formation of Memory](#)).

A mutant version of the prion protein (PrP^{Sc}) was discovered that could mis-fold and to take on an abnormal shape. The deformed PrP^{Sc} could then go on and induce abnormal folding in other, normal PrP^c molecules. These events, illustrated in Figure 3.8, result in the formation of so-called *amyloid plaques*.

Transmissible Conformational Change of *Prion* Protein



...Somatic mutations are *not* heritable.

Fig. 3.8: Mutations in the gene for the normal prion PrP^c protein (left) may produce some abnormally folded prion PrP^{Sc} proteins (at the right). The misfolded PrP^{Sc} molecules interact with other (even normal) prions, causing them to mis-fold, precipitating PrP^{Sc} proteins into aggregated amyloid plaques.

Abnormally folded prions (PrP^{Sc}) have been associated with Alzheimer's disease, which affects about 5.5 million Americans. PrP^{Sc} is also associated with *mad cow disease* and *Creutzfeldt-Jakob disease* (mad cow disease in humans), as well as *scrapie* in sheep, among others. We are beginning to understand that the role of prion proteins in *Alzheimer's disease* is less causal than indirect.

3.3.2.a The *Amyloid Beta* ($A\beta$) peptide

Postmortem brains from patients Alzheimer's disease patients show the characteristic *extracellular amyloid plaques*. These plaques were shown to be composed largely of *beta-amyloid* ($A\beta$) peptides. In affected cells, an amyloid precursor protein (APP) is enzymatically digested, generating thirty-nine to forty-three extracellular amino acid ($A\beta$) amyloid peptide fragments. Normally, excess ($A\beta$) peptides are themselves digested. However, an unregulated ($A\beta$) peptide accumulation leads to the formation of the beta-amyloid plaques seen in Alzheimer's disease (Figure 3.9).

Extracellular Formation of β -Amyloid Plaques of Alzheimer's Disease

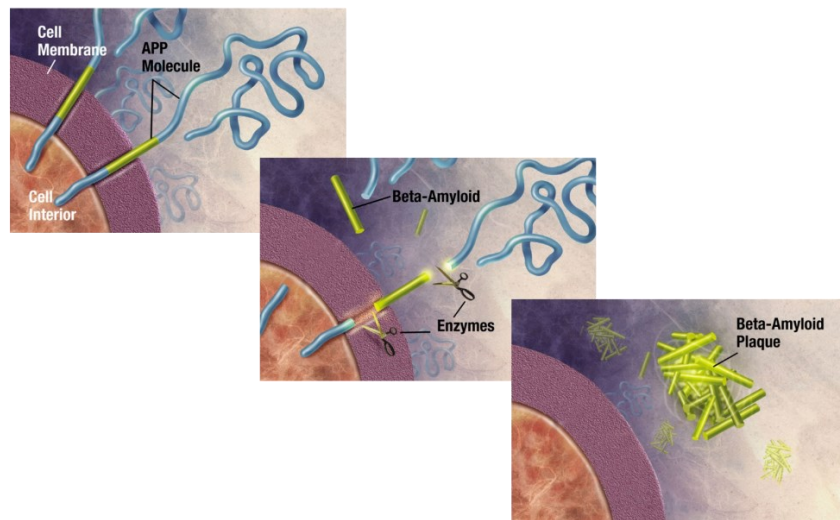


Fig. 3.9: Cell membrane *amyloid precursor protein* (APP, left) is enzymatically cleaved to release β -amyloid peptide fragments (middle) that aggregate, forming the characteristic *amyloid plaques* of Alzheimer's disease

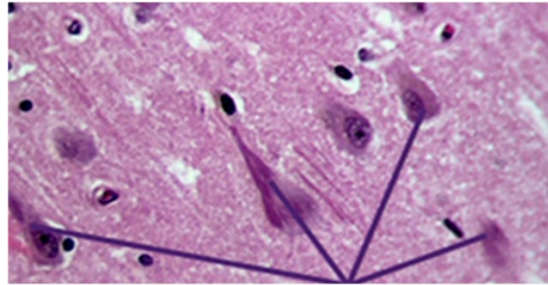
The scissors in the illustration represents two enzymes that digest the APP. Prion proteins are not a proximal cause of Alzheimer's disease but may have a role in initiating events that lead to it. Normal prion protein (PrP^c) is itself a membrane receptor and is thought to bind ($A\beta$) peptides, effectively preventing their aggregation into plaques. An experimental reduction of PrP^c was shown to increase the extracellular ($A\beta$) peptides. Presumably, prion protein aggregation induced by the mutant PrP protein (PrP^{Sc}) prevents prion proteins from binding to ($A\beta$) peptides, leading to ($A\beta$) peptide accumulation and ultimately to amyloid plaque formation and neurodegeneration.

3.3.2.b The Tau Protein

The *tau* protein is also associated with Alzheimer's disease. **MAP-T** (*Microtubule-Associated Protein-Tau*) is as its name indicates, one of several microtubule-associated proteins (MAPS). In normal neurons, **MAP-T** is phosphorylated, and the phosphorylated MAP-T binds to, and stabilizes microtubules. But misshapen *tau* that accumulates in *neurofibrillary tangles* in hippocampus brain neurons may be a more immediate cause of the neuronal dysfunction associated with Alzheimer's disease than ($A\beta$) peptides.

When neuronal *tau* becomes *hyperphosphorylated*, its conformation changes, making it unable to associate with microtubules. No longer stabilized, the microtubules disassemble and the deformed tau proteins form **neurofibrillary** tangles. Immunostaining of *hippocampal* neurons with antibodies against tau protein localizes the **neurofibrillary** tau protein tangles, as seen in the light micrograph below (Figure 3.10).

Immunostaining of Brain Cells from an Alzheimer's Patient



Tau protein tangles

Fig. 3.10: Non-fluorescent immunostaining micrograph of tau proteins in brain cells of an Alzheimer's brain; pointers indicate tau tangles.

The formation of neurofibrillary tau protein tangles in a diseased neuron is compared to normal neurons in Figure 3.11. The tangled clumps of tau proteins in this illustration are what appear as deep purple in Figure 3.10. Recent studies suggest further that at least two structurally related tau proteins isoforms make up the tau tangles (^{3,7}[Multiple tau proteins in a Tangle](#)).

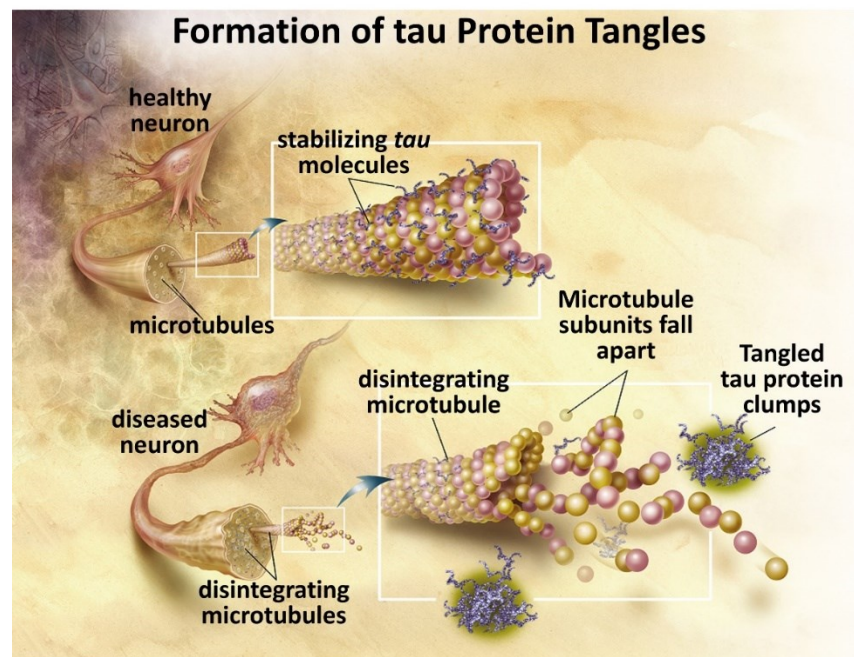


Fig. 3.11: The formation of tau tangles (lower left) occurs in diseased neurons that cannot maintain normal microtubule structure (upper middle). In the absence of stabilizing tau molecules, microtubules degenerate and tau proteins clump (lower right).

There is no cure yet for Alzheimer's disease, although treatments with *cholinesterase inhibitors* seem to slow its advancement. For example, the drug *Aricept* inhibits acetylcholine breakdown by *acetylcholinesterase*, thereby enhancing cholinergic neurotransmission, which may in turn prolong the brain's neural function.

Sadly too, there is yet no treatment to restore lost memories or the significant cognitive decline associated with Alzheimer's disease. Perhaps more promising in this respect is a recent development. Proteins or peptides associated with Alzheimer's disease have been detected in the blood and serum. Amyloid beta- ($A\beta$) and/or *tau*-protein fragments that escape into the blood stream can be detected six to eight (or more) years before Alzheimer's symptoms appear. A *neurofilament light chain* (*NfL*) that is seen in a familial form of Alzheimer's disease (among other neuropathies) is detectable sixteen years before the symptoms!

The ability to detect these marker proteins and Alzheimer's-associated peptides so far in advance of symptoms raises hopes for early monitoring of at-risk individuals and for new therapies for Alzheimer's disease. For brief reviews, see ^{3,8}[Early Detection of Circulating Ab Peptides](#) and ^{3,9}[Early Detection of Circulating Tau Peptides](#). For a recent report on tracking the *NfL* protein, see ^{3,10}[Early Detection of Circulating Neurofilament Light-Chain Protein](#).

3.3.2.c Some Relatives of Alzheimer's Disease

Some of the same protein abnormalities seen in Alzheimer's disease are also seen in other neurodegenerative diseases as well as traumatic brain damage. An abnormal accumulation of tau protein is diagnostic of *Chronic Traumatic Encephalopathy* (CTE).

First described in the early twentieth century, disoriented boxers staggering about after a fight were called "punch drunk," suffering from *dementia pugilistica*. We now know they suffered from CTE, as do other athletes exposed to repetitive mild-to-severe brain trauma, such as football players. Immunostaining of whole brains and brain tissue from autopsied CTE patients using antibodies to tau protein, show accumulations of abnormal tau proteins and tau neurofibrillary tangles very much like those found in Alzheimer's patients.

Many National Football League and other football players have been diagnosed postmortem with CTE, and many still living show signs of degenerative cognition and behavior consistent with CTE. (See a ^{3,11}[2020 List of NFL players with CTE](#) to see how many!)

Parkinson's disease is yet another example of a neurodegenerative disease that results when a single protein changes shape in brain cells. Though they are not characterized as plaques, aggregates can form in brain cells when the protein *alpha-synuclein* undergoes anomalous conformational change. The change results in *MSA* (*Multiple System Atrophy*) or Parkinson's disease. To read details of this recent research and how it may lead to treatments, see ^{3,12}[Targeting Alpha Synuclein for Parkinson's Therapy](#).

Much of the high-resolution electron microscopy that can reveal details of the structure and conformational changes in proteins came from the work of Jacques Dubochet, Joachim Frank, and Richard Henderson, who shared the 2017 Nobel Prize for developing and refining

cryoelectron microscopy for biomolecular imaging (For details, see ^{3.13}[Nobel Prize-Cryoelectron Microscopy](#)).

3.4 Protein Quaternary Structure, Prosthetic Groups, and Chemical Modifications

Quaternary structure describes proteins composed of two or more polypeptides. Like tertiary structure, such multimeric proteins are formed by noncovalent interactions and may be stabilized disulfide bonds. *Dimers* contain two, a *trimer* three, a *tetramer* four polypeptides, and so on. Multimers made up of identical subunits are referred to with a prefix of *homo-* (e.g., a homotetramer). Multimers made up of different subunits are called *heteromers*. The vertebrate hemoglobin molecule is a *heterotetramer*, with two α and two β globins (Figure 3.12).

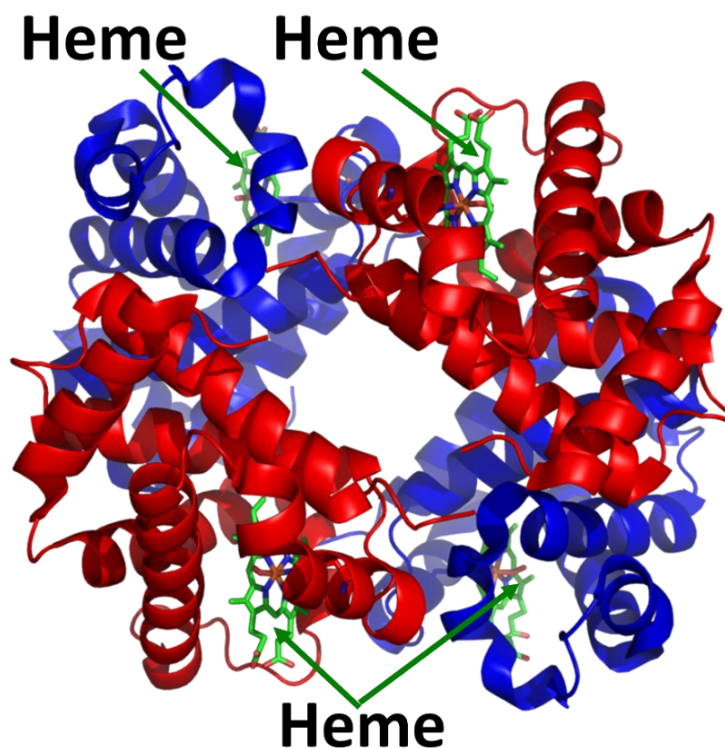


Fig. 3.12: The vertebrate hemoglobin molecule, consisting of four globin subunits (two α - and two β -polypeptides). Each globin is associated with a heme group bound to iron.

Hemoglobins exemplify the role of organic molecules groups in protein function. To be biologically active, each globin polypeptide must associate with *heme* (a cyclic organic molecule) with an iron ion at its center. The iron is what reversibly binds oxygen. All kinds of organisms, from bacteria to plants and animals (and even some anaerobes!) contain hemoglobin variants, each with a heme-like molecule. The organic molecules associated with proteins are called **prosthetic groups**. Other proteins must be bound directly to metal ions (e.g., magnesium, manganese, and cobalt) to be biologically active.



[136-2 Protein Quaternary Structure & Prosthetic Groups](#)



Chemical modifications are posttranslational enzyme-catalyzed events, which are required to make a protein fully functional or to regulate its activity. Many polypeptides are structurally modified by *glycosylation* (e.g., to make membrane *glycoproteins*). Others are *phosphorylated* at one or more specific amino acids in the chain, to regulate their biological activity. These and other modifications account for and enhance the molecular and functional diversity of proteins within and across species.

3.5 Domains, Motifs, and Folds in Protein Structure

The structures of two different proteins in Figure 3.13 share a common domain: the *PH* (Pleckstrin Homology) domain.

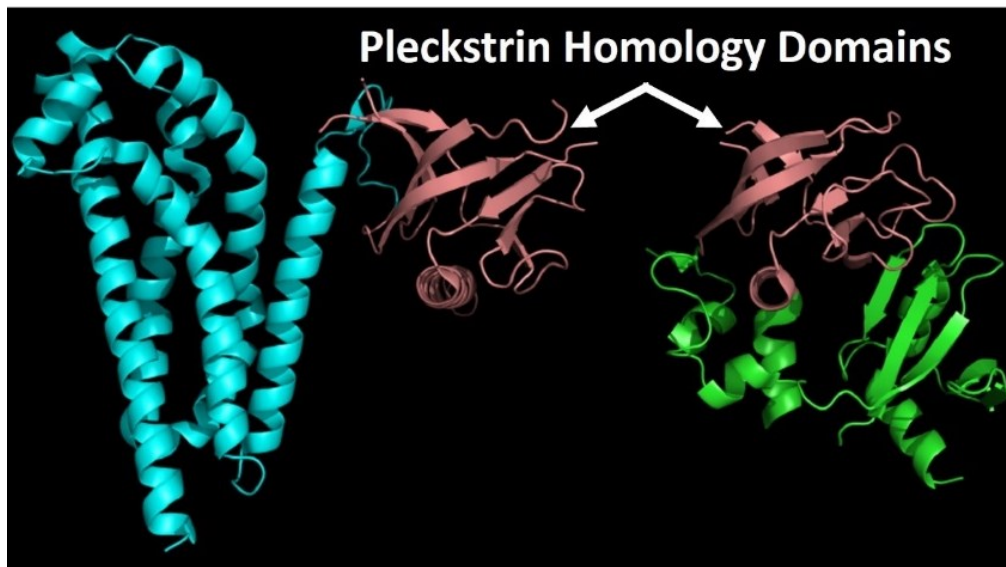


Fig. 3.13: The *pleckstrin homology* (PH) domains shown here in pink (arrows) are examples of common domains in two different proteins. By virtue of these common PH domains, otherwise different proteins can interact with the same cell-signaling factors with roles in coordinating intercellular communication.

These two (and many other) proteins have this domain, allowing them to bind a molecule of *phosphatidyl-inositol triphosphate*, which is generated as part of a common cell-signaling pathway. The implication of this common *domain* is that a cell's signaling pathways allow it to respond to different signals that lead to the same response, albeit under different conditions and probably at different times. Proteins are typically described as consisting of several distinct substructures, discussed here.

A **domain** is an element of the protein's overall structure that is stable and often *folds* independently of the rest of the protein chain. Like the PH domain above, many domains are not unique to the protein products of one gene but instead appear in a variety of proteins. Proteins sharing *more than a few* common domains are encoded by members of evolutionarily related genes comprising **gene families**. Genes for proteins that share only one or a few domains may belong to **gene superfamilies**. Superfamily members can have one function in common, but the rest of their sequences are otherwise unrelated. Domain names often derive from their prominent biological function in the protein they belong to (e.g., the *calcium-binding domain* of *calmodulin*) or from their discoverers (the PH domain). The domain-

swapping that gives rise to gene *families* and *superfamilies* is a natural genetic event. Because protein domains can also be *swapped* on purpose by genetic engineering, we can make *chimeric proteins* with novel functions.



[137-2 Protein Domain Structure and Function](#)

Protein ***motifs*** are small regions of protein 3D structure or of amino-acid sequence shared among different proteins. They are recognizable regions of protein structure that may (or may not) be defined by a unique chemical or biological function.

Supersecondary structure refers to a combination of secondary-structure features, such as *beta-alpha-beta* units or the *helix-turn-helix motif*. They may also be referred to as structural motifs. See ^{3.14}[Supersecondary structure](#) for examples.

A ***protein fold*** refers to a general aspect of protein architecture, like *helix bundle*, *beta-barrel*, *Rossmann fold*, or other "folds" provided in the ^{3.15}[Structural Classification of Proteins](#) database. Look at ^{3.16}[Protein Folds](#) to read more about these structures.

3.6 Proteins, Genes, and Evolution: How Many Proteins Are We?

If evolution did not have to select totally new proteins for each new cellular function, then how many genes does it really take to make an organism? Estimates suggest that it takes from nineteen thousand to twenty-five thousand coding genes to make and to operate a human and all its proteins (Check out Pertea and Salzberg at ^{3.17}[# of genes in a human genome](#) and Abascal, F. et al. at ^{3.18}[~20% of human genes are mysterious!](#)). But our cells, and those of eukaryotes generally, may express as many as one hundred thousand different proteins. In other words, our protein-coding genes seem to be far fewer than the number of proteins we need to make. How is this possible? It appears that there must be better, more efficient ways to evolve new and useful complex cellular tasks than evolving more new genes!

As we already noted, the use of the same twenty amino acids to make proteins in all living things speaks to their early (even prebiotic) selection and to the common ancestry of all living things. Complex conserved domain structures shared among otherwise different proteins imply that evolution of protein function has occurred as much by recombinatorial exchange of DNA segments encoding such substructures as by an accumulation of base substitutions in otherwise redundant genes. Likewise, motifs and folds might also be shared in this way. The protein count also exceeds gene number in part because cells can produce different mRNA variants from the same genes by *alternate splicing*, a process that produces mRNAs that code for *different combinations of substructures from the same gene*! Alternate splicing is discussed in detail in a later chapter. These transcriptional and post-transcriptional phenomena explain eukaryotic protein numbers, but also the conservation of gene and amino-acid sequences across species (e.g., histones and globins), the key testimony to their common ancestry. Along with the synthesis of alternate versions of a single RNA, an ongoing repurposing of useful

regions of protein structure may prove a strategy for producing new proteins without adding new genes to a genome.

3.7 Directed Evolution: Getting Cells to Make New Proteins...for Our Use and Pleasure

Investigators have long speculated that proteins with useful functions could be adapted to human use. Enzyme additives in laundry detergents or spot-removers are already used to digest organic stains. Such enzymes must be extracted from a suitable biological source. But what if we could engineer even better versions of an enzyme? At the molecular level, protein evolution is the natural selection of gene sequences that encode functional polypeptides. This implies that variant “mutant” versions of a gene that encode related polypeptides already exist, from which nature can select one. Thus, changing environmental circumstances might favor one variant protein over another. In nature, mutations that could create polypeptide variants are entirely random. In other words, we humans might wait quite a long time before a *better version* of a protein, say an enzyme, would be available for human use through natural selection. Can we speed up the selective process and more rapidly evolve better, more useful proteins? Yes, we can!

Many industries (e.g., fuel and pharmaceutical industries) have used molecular techniques to create new proteins. It is possible to clone a desired gene, to make targeted mutations in the gene, and then to express them in suitable cells. The expressed proteins can then be screened for example, for mutant enzymes with improved or even novel useful activities. This is a far cry from older techniques that irradiated or otherwise mutagenized cells or organisms, which would then be screened for mutants that looked interesting! Since we can now target mutations to a single specific base within a gene (for example, one that avoids drastic changes in protein folding), it is possible to study the functional effects of very subtle conformational changes in a protein.

The technique, called **directed evolution**, was originally pioneered by *Frances Arnold*, who engineered enzymes that could make renewable fuels (e.g., isobutanol), environmentally friendly pharmaceuticals, and enzymes that function hundreds of times faster and/or at a broader temperature range than their naturally occurring counterparts. For these achievements, Frances Arnold was awarded the Nobel Prize for Chemistry in 2018!

3.8 View Animated 3D Protein Images in the NCBI Database

Viewed by *X-ray diffraction*, we can see that proteins exhibit exquisite diversity. Look at X-ray views of protein tertiary (and even quaternary structures at the Cn3D database of the *National Center for Biotechnology Information* (NCBI). Using this free tool, you can search for any protein in the database. As an example, we'll follow the steps to find human insulin and look at it from all angles. Here are the detailed steps:

1. To access 3D animated images of proteins from the database, click or copy the link below into your browser and hit *Return*: ^{3.20}<https://www.ncbi.nlm.nih.gov/structure?db=Structure>
2. In the *Structure* screen that appears (Fig. 3.14, below), type “normal human insulin” into the search box and click *Search*.

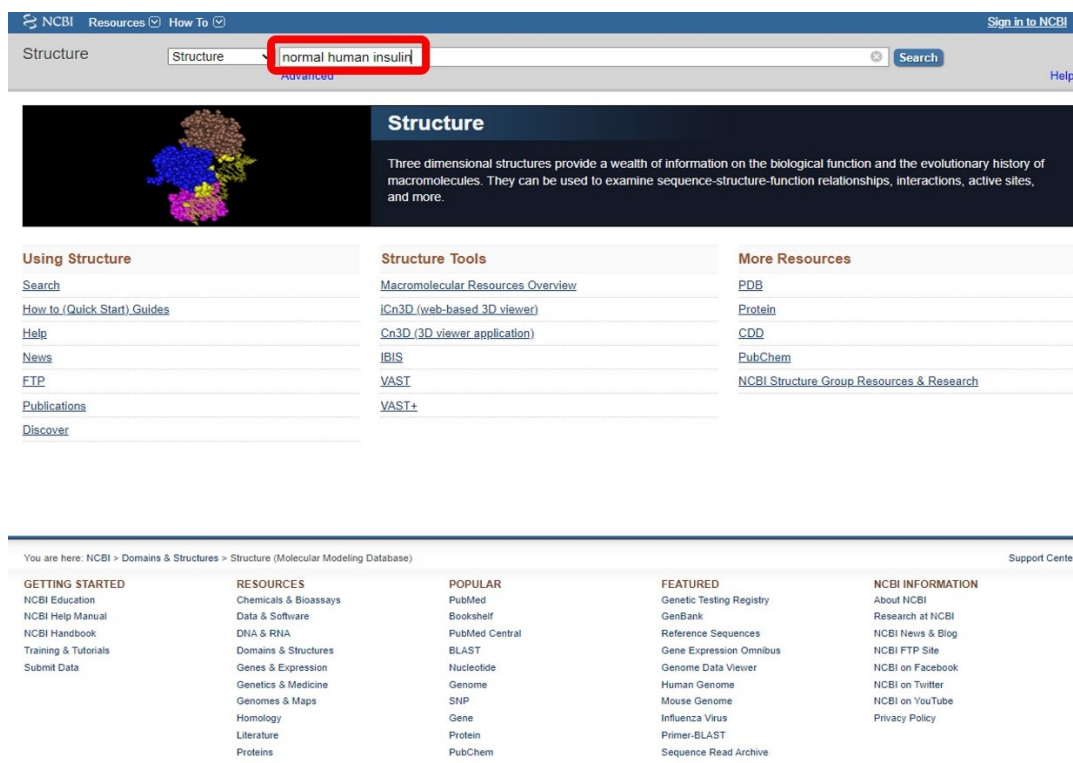


Fig. 3.14: Screenshot of NCBI portal to 3D protein structure database based on X-Ray crystallographs.

- The first batch of X-ray crystallographic insulin structures appears in the *Search results* screen below (Fig. 3.15). Scroll down to *item 4* and click on its URL link (red arrow), or on the image itself.

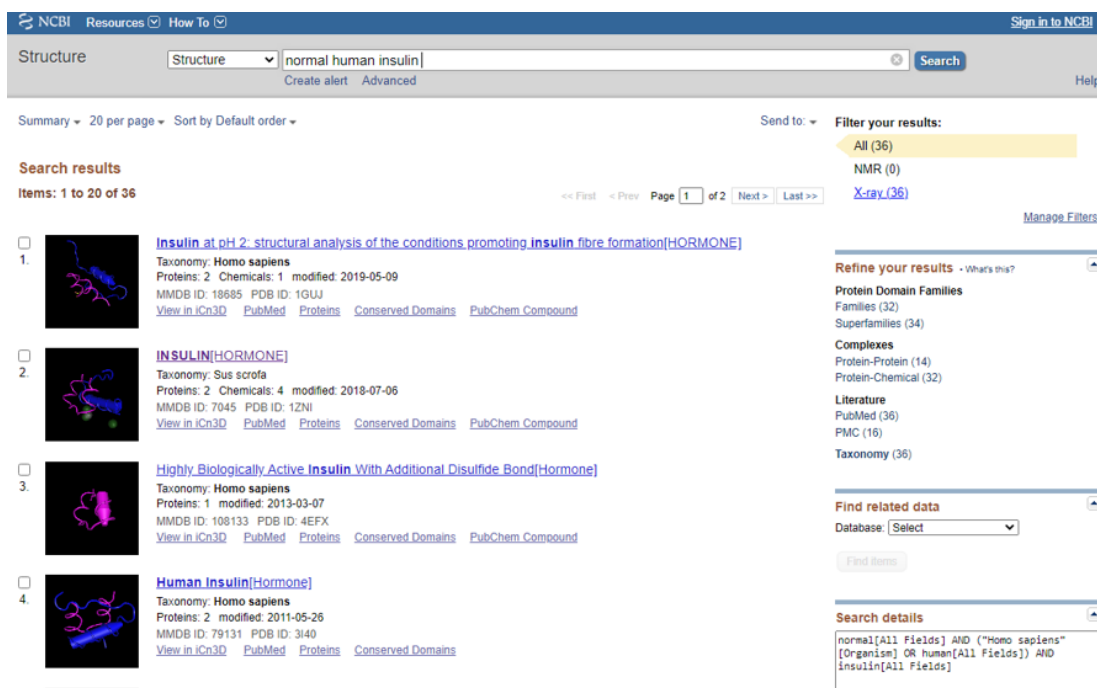


Fig. 3.15: Screenshot of NCBI protein structure database search leading to available macromolecular structures for human insulin (²¹[normal human insulin - Structure - NCBI \(nih.gov\)](https://www.ncbi.nlm.nih.gov/structure/)).

- If the protein does not start rotating in the next screen (Fig. 3.17 below, click *View*, then *Rotate*, then *Autorotate*, then any 'direction.' To control the rotation, click on the molecule and move the cursor to rotate the molecule manually. When you can roll the molecule around, like a beautiful marble in your hand, you'll see its structure and how it might account for its function.

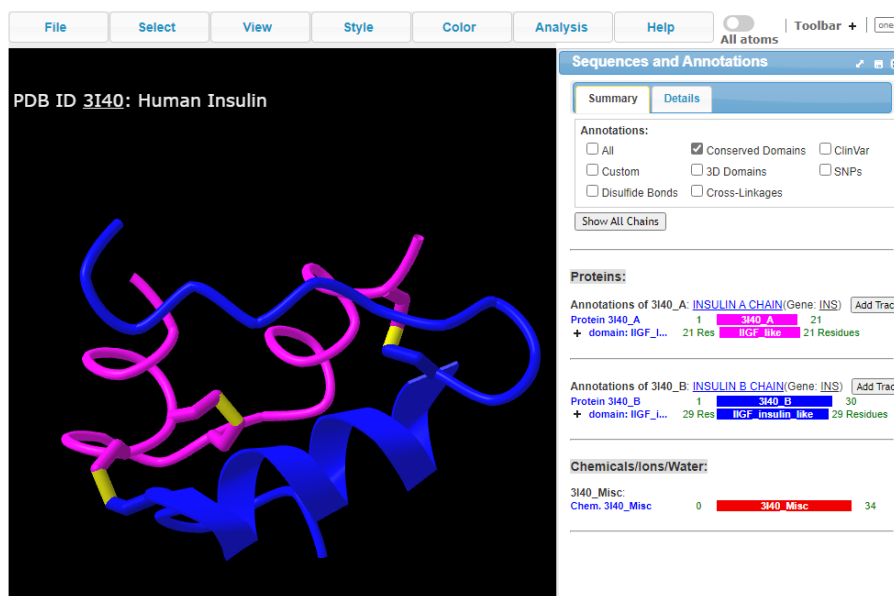


Fig. 3.16: Screenshot of the dynamic 3D X-Ray structure of a human insulin in the NCBI protein structure database ([223I40\(MMDB\) in iCn3D \(nih.gov\)](https://www.ncbi.nlm.nih.gov/3D/3I40))

To rotate the molecule, click **View**, then **Rotate**, then **Autorotation**, then play with the directional options..., and then spin...and enjoy!

Some iText & VOP Key Words and Terms

A β	enzymes	polypeptide backbone
α -carbon	functional groups	primary structure
alpha (a-) helix	gene family	prion
allosteric regulation	gene superfamily	protein folding
Alzheimer's disease	glycosylation	PrP
amino-acid residues	helix-turn-helix motif	quaternary structure
amino end	hemoglobin	random coil
amyloid beta protein	hydrophobic interactions	recombinatorial exchange
amyloid plaques	levels of protein structure	salt bridges
beta barrel	mad cow disease	scales, feathers, fingernails
beta (b-) sheet	multimer	secondary structure
carboxyl end	NCBI Cn3D database	sequence motifs
catalysts	neurofibrillary tangles	sickle cell anemia
chaperones	nucleoskeleton	side chains
configuration	orders of protein structure	structural domain
Creutzfeldt-Jakob disease	Parkinson's disease	structural motif
cytoskeleton	peptide bonds	sulphydryl groups
disulfide bonds	peptide linkages	tau protein
Edman degradation	pleated sheet	

CHAPTER 3 WEB LINKS



3.19



3.20



3.21



3.22



Chapter 4

Bioenergetics

Thermodynamics (Free Energy, Enthalpy, Entropy) Chemical Energy, Open vs Closed Systems

Reminder: For inactive links, google key words/terms for alternative resources.



4.1 Introduction

Three **Laws of Thermodynamics** describe the flow and transfer of energy in the universe:

- **Energy can neither be created nor destroyed.** The rule is also stated as "**universal energy is constant.**"
- **Universal entropy (disorder) is always increasing.**
- **Entropy declines with temperature**—as temperatures approach absolute zero, so goes entropy.

In living systems, we do not have to worry about the third law because the equations for energy exchange in living systems already reflect the temperature dependence of entropy change during reactions. In this chapter we will look at how we came to understand the basic thermodynamic principles and how they apply to living systems. First, we will look at different kinds of energy and at how **redox** reactions govern the flow of energy through living things. Next, we'll look at some simple arithmetic statements of the **laws of thermodynamics** for **closed systems** and at how they apply to chemical reactions conducted under **standard conditions**. Finally, since there is really no such thing as a **closed system**, we look at the energetics of reactions occurring in **open systems**. For an excellent discussion about the application of basic thermodynamic principles to living things, check out Lehninger A. (1971) *Bioenergetics: The Molecular Basis of Biological Energy Transformation*. Benjamin Cummings, San Francisco.

Learning Objectives

When you have mastered the information in this chapter, you should be able to do the following:

1. Explain the difference between *energy transfer* and *energy transduction*.

2. Compare and contrast *potential*, *kinetic* and other kinds of energy (e.g., *mass*, *heat*, *light*).
3. Explain the idea that *chemical equilibrium* is a *dynamic equilibrium*
4. Explain the reciprocal changes in universal *free energy* and *entropy*.
5. Derive the algebraic relationship between *free energy*, *enthalpy*, and *entropy*.
6. Distinguish between *exothermic*, *endothermic*, *exergonic*, and *endergonic* reactions.
7. Predict changes in free energy in a chemical reaction based on changes in the concentrations of reactants and products in *closed systems* as well as *open systems*.
8. Predict how and when an endothermic biochemical reaction will release free energy. (You should be able to do this after working some sample problems of closed system energetics.)
9. Explain how an exergonic reaction occurring in a cell might be energetically unfavorable.
10. Explain and distinguish between the equilibrium and the steady state of reactions.
11. Explain the thermodynamic laws.

4.2 Kinds of Energy

We can easily recognize different kinds of energy around us, like **heat**, **light**, **electricity**, **chemical energy**, **nuclear energy**, **sound**, etc. These different forms of energy are measurable (e.g., calories, joules, volts, decibels, quanta, photons). Even mass is a form of energy; recall Albert Einstein's $E = mc^2$ equation, the *theory of special relativity* (or the *law of relativity*).

The problem in thinking about thermodynamics is that the universe is so big; there are too many kinds of energy to contemplate at once! To simplify, let's imagine only two kinds of energy in the universe: **potential energy** and **kinetic energy**. A helpful example is a dam. The water above the dam has *potential energy*. As the water flows over (or through) the dam, its potential energy is released as *kinetic energy*. In the old days, the kinetic energy of flowing water could be used to power (i.e., turn) a millstone to grind wheat and other grains into flour. These days, water is more likely to flow through a hydroelectric dam, where kinetic energy is converted (*transduced*) to electricity. In this simple view, heat (molecular motion), electricity (a current of electrons), sound (waves), and light (waves or moving "particles") are different forms of kinetic energy. The energy of mass or its position in the universe is potential energy. For example, the energy in a molecule of ATP is potential energy. Physicists talk a lot about potential energy and about the flow and conversion kinetic energy.

Here is an equally simple way to conceptualize energy: it is either **useful** or **useless**. This concept led directly to the arithmetic formulation of the thermodynamic laws. In this utilitarian way of thinking about energy, useless energy is **entropy**, while useful energy can be any of the other forms of energy (potential or kinetic).

Key to understanding bioenergetics is recognizing the differences between closed and open systems in the universe. Biochemical reactions in a test tube are **closed systems** that will reach **equilibrium**. Closed systems are artificial and can only be created in a laboratory, where one can restrict and measure the amount of energy and mass entering or escaping the system. Cells on the other hand (and in fact, every reaction or event in the rest of the universe), are **open systems**. Open systems readily exchange energy and mass with their surroundings.

With this introduction, let's imagine ourselves to be early scientists trying to understand energy flow in the universe by asking how the **laws of thermodynamics** apply to

living systems (*bioenergetics*). We'll see that the *laws* can be mathematically demonstrated precisely because all kinds of energy can be measured and all units of energy (e.g., volts into calories, light quanta into volts, and joules into decibels, mass into photons...) can be interconverted.



[138 Different Kinds of Energy: Chemical Equilibrium](#)



4.3 Deriving Simple Energy Relationships

Thermodynamic laws lead us to equations that help us understand how energy moves (flows) between components of the universe. As we'll discuss, the universe is most easily thought of as a giant closed system within which energy can be transferred and even transduced (i.e., from one form to another) between those components. We can't measure energy flow in or across an entire universe. But we can isolate bits within the universe so that we can quantify energy as it is transferred or transduced within smaller, more manageable closed systems.

4.3.1 Energy in the Universe: The Universe Is a *Closed System*

Consider an event, any event. I think we can agree that when stuff happens, the participants in the happening go from an unstable state to a relatively more stable state. For example, you are carrying a bag of marbles, the bag tears open, the marbles fall to the floor, then roll and spread out, eventually coming to a stop. At that point, the marbles are in a more *stable state* than they were when you were holding the un-ripped bag.

We can all agree that gravity made the marbles fall out of the bag. We might further agree to say that the fallen marbles are in a more stable state than they were in the bag in your hands. That much certainly seems true. If so, we could extend our observation to conclude that the drive to greater stability is what made the marbles fall! In fact, regardless of the force or impetus for any event, the science says that the drive to achieve greater stability is what makes stuff (i.e., events) happen!

This is the essence of the *second law of thermodynamics*: all universal energy-transfer events occur with an increase in stability—that is, an increase in *entropy*. We'll consider the second law and *entropy* in detail shortly.

The tendency of things to go from unstable to more stable is a natural, rational state of affairs, as seen with those marbles on the floor or a messy bedroom with clothes strewn about. Intuitively, mess and disorder are more stable than is a state of order. Of course, marbles dropping or clothing going from folded and hung to wrinkled on the floor releases energy (potential energy) as they fall (kinetic energy). If you don't believe that this release of energy is real, just think of how much energy you will need to pick up the marbles or to refold your clothes (after laundering them of course!).

We can model the flow of energy in the universe in a way that is consistent with the laws of thermodynamics. Since the *first law of thermodynamics* says that *energy can be neither created nor destroyed*, a simple statement of the first law could be the following:

$$E_{\text{universal}} = E_{\text{light}} + E_{\text{heat}} + E_{\text{electrical}} + E_{\text{mass}} + \dots$$

This equation sums up the different kinds of energy in the universe. In Figure 4.1, consider the circle (also called a *Venn diagram*) to be the universe.

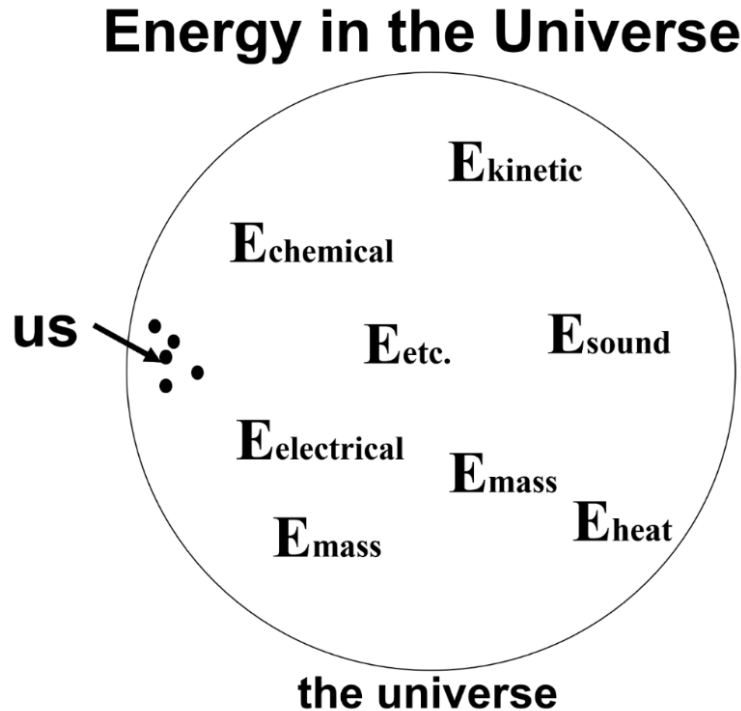


Fig. 4.1: Venn diagram depicting all things in the universe including mass and more familiar kinds of energy. The diagram implies a as a "closed system"; nothing (mass, energy) can enter or exit the universe, consistent with the first law of thermodynamics.

From this we conclude that energy cannot get into or out of the universe, but energy can be transferred within the universe from one place to another or converted from one form to another. It follows then that $E_{\text{universal}}$ is the sum of all kinds of energy in the universe and that this *sum* is a constant—in other words, universal energy is conserved; in the words of the first law, *universal energy is constant*. The following equation expresses this idea; it is a statement of the *first law of thermodynamics*:

$$E_{\text{light}} + E_{\text{heat}} + E_{\text{electrical}} + E_{\text{mass}} + \dots = \text{a constant}$$



[139 First Law of Thermodynamics](#)



In the late nineteenth century, John Venn formalized this visual approach to segregating things and concepts into circles as a way of logically viewing relationships between them. The *Venn diagrams* used in the following figures are simple and help clarify concepts of energy in the universe. For more complex examples showing overlapping components of the universe (i.e., those that share some but not all attributes), google "Venn diagrams." For our purposes, let's stay simple and go with the simpler binary notion of *useful* and *useless* energy.

Our equation shortens to the sum of just two kinds of energy in the universe:

$$E_{\text{universal}} = G_{\text{universal}} + TS_{\text{universal}}$$

In this equation, G is useful energy; it stands for “Gibbs” free energy, named after Josiah W. Gibbs, who is credited with applying the inductive scientific method to the study of thermodynamics. S is useless energy (entropy), and T is absolute temperature (included because of the third law). This equation is also a statement of the *first law*, and it is shown in the revised Venn diagram in Figure 4.2.

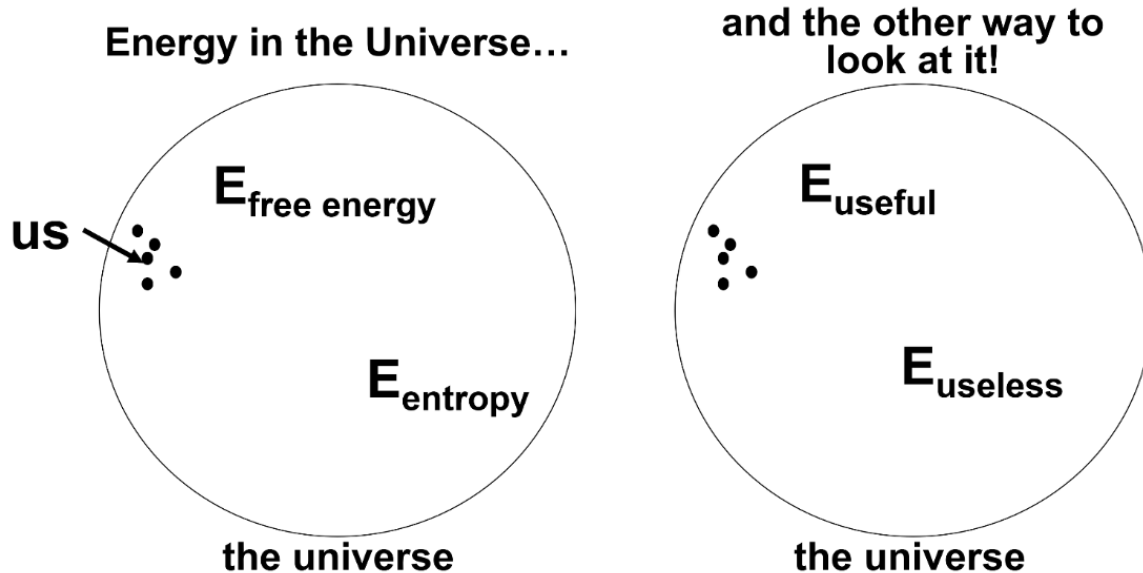


Fig. 4.2: Since even mass is a form of energy, the universe depicted at the left consists of only two components: *entropy* and *free energy* (defined as energy that is available to do work). The universe redrawn at the right just redefines free energy as the *useful* kind, and entropy as *useless* energy (the opposite of useful!). Both models are still consistent with the *first law of thermodynamics*.

In this binary energy model, it follows that as universal entropy increases, free energy in the universe must decrease (as graphed in Figure 4.3).

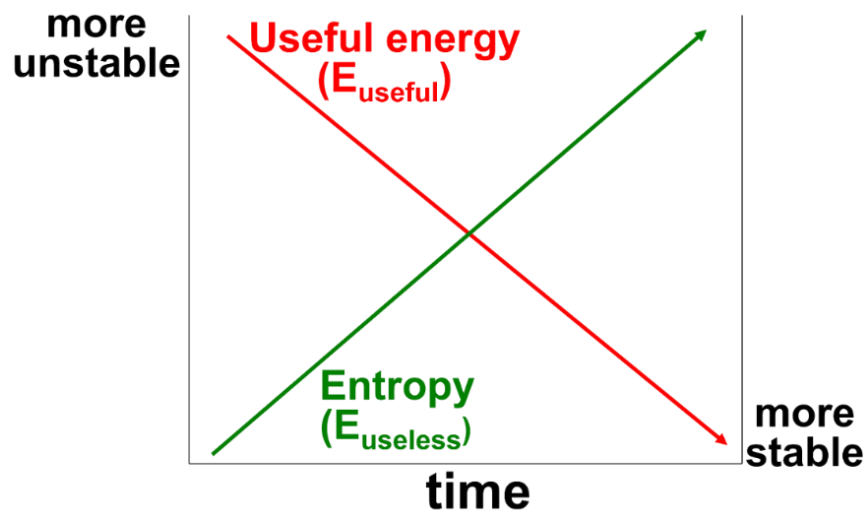


Fig. 4.3: Entropy and free energy have a reciprocal relationship, which is graphed here as it changes over time. If all activities in the universe proceed by increasing entropy, and if the universe consists *only* of free energy and entropy, then as entropy increases, free energy must be decreasing in the universe.

Free or *potentially useful* energy is higher in more ordered and complex and therefore relatively unstable systems. Such unstable, ordered systems will release free energy *spontaneously*. In other words, stuff (i.e., spontaneous reactions) *will* happen!



[140-2 Second Law of Thermodynamics](#)

4.3.2 Energy Is Exchanged Between Systems in the Universe

While the arithmetic statements about changes in energy are useful concepts, the parameters are, of course, not measurable! But if we isolate a bit of the universe, we *can* measure energies and watch energy flow. If we can measure the amount of energy put into or removed from **a system within the universe**, we can write a more useful equation to follow the transfer of energy between a system and its surroundings:

$$\Delta H = \Delta G + T\Delta S$$

In this formula, ΔH is the change (Δ) in enthalpy, (i.e., it is the energy entering/leaving the system in units of heat energy); ΔG is change in free energy; ΔS is change in entropy; and T is the absolute temperature in degrees Kelvin (K).

Heat given off in a reaction (or other event) is often confused with entropy. True, much of the increase in entropy that occurs in living things is indeed in the form of random molecular motion, or heat. But remember that heat can have its uses; not all heat is entropic! Hence, it is more interesting (and accurate!) to think of energy in terms of changes in enthalpy, free energy, and entropy during energy transfers. Thus, we must consider the arithmetic of energy transfers to involve not two, but three terms!

According to the equation $\Delta H = \Delta G + T\Delta S$, interacting systems in our universe would seem to be *closed systems*. Accordingly, energy put into or taken from the system (ΔH) will be exactly balanced by increases or decreases in the other two terms in the equation (ΔG , $T\Delta S$).

Recall that we say a system is closed *not* because it is *really* closed but because we can isolate it well enough to measure energy flow into and out of it. This or any algebraic equation with three variables is useful since, if you know two of the values you can calculate the third! Here is a simple situation to illustrate the point: If I put a liter of water onto a burner and light the flame, the water gets hot. If the temperature of the liter of water rises by 1°C, we know that it has absorbed 1,000 calories (1 kilocalorie, or 1 *food* Calorie) of the heat from the burner.

Since energy interactions depend on physical conditions, such as temperature and air pressure, we need to standardize those conditions when we conduct experiments that measure energy changes in experimentally isolated systems. For more on how this standardization enables the measuring of energy change in chemical reactions (in fact, any energy exchange), go to the following link.



[141 Deriving Closed-System Thermodynamics](#)

Turning to *bioenergetics*, let's apply the equation $\Delta H = \Delta G + T\Delta S$ to the conditions in which chemical reactions occur in cells. Because most life on Earth lives at sea level, where the air pressure is 1 *atm* (atmosphere), typical determinations of ΔH , ΔG , and $T\Delta S$ are made under defined, well-controlled *standard conditions* that include "*unimolar conditions*," an absolute T of 298 °K (25°C), 1 *atm*, and a constant pH of 7.0. The latter is defined as such because the pH inside of a typical cell is close to neutral. Our equation for reactions under *these* standard conditions becomes the following:

$$\Delta H = \Delta G_o + T\Delta S$$

ΔG_o is the *standard free energy change* for a reaction conducted in a *closed system* under standard conditions. ΔH and ΔS are still the enthalpy and entropy changes, but they are determined under standard conditions. So, what are those "*unimolar conditions*"? To meet these conditions, you would measure the calories released by burning (oxidizing) a small amount, say 180 *mg* of glucose. Then just multiply the number of calories released by 1,000 to get ΔH , which equates to the number of calories that would have been released by burning 180 *gm* (i.e., the molecular mass, or a whole mole) of the stuff.

Now we are ready to consider examples of how we determine the energetics of reactions.

4.3.3 How is Enthalpy Change (ΔH) Determined?

ΔH (the heat released or absorbed) in a chemical reaction can easily be determined by conducting the reaction under standard conditions in a *bomb calorimeter*, which is essentially an inner chamber surrounded by an outer chamber filled with water (Figure 4.4).

Measuring ΔH for a Chemical Reaction in a Calorimeter

Reaction will release or absorb heat, measurable by change in temperature of water in the calorimeter water 'jacket'

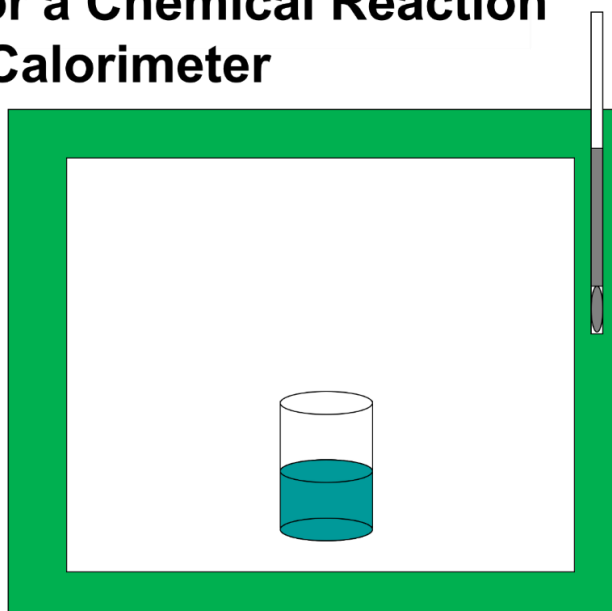


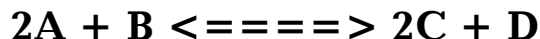
Fig. 4.4: Basic design of a *bomb calorimeter*, with an inner chamber for conducting reactions and an outer chamber containing water with a thermometer to measure temperature change (heat absorption or release, or ΔH) during the reaction.

Food manufacturers determine the **Calorie** content of food by *bomb calorimetry*. As a reaction takes place in the beaker, it will either release or absorb heat, either heating or cooling the water in the calorimeter jacket, as measured by the thermometer. A closed system reaction that releases heat as it reaches *equilibrium* is defined as **exothermic**; the ΔH for an *exothermic* reaction is negative. For example, a package says that a chocolate bar has 90 Calories. This means that burning the bar will generate 90 kilocalories (Kcal) of heat as measured in the calorimeter. Recall that 1 Calorie (with a capital C) is equal to 1,000 calories, (1 Kcal). One calorie (cal; lowercase) is the energy needed to raise the temperature of 1 gram of water by one degree Celsius.

You are probably most familiar with reactions that release heat, but some chemical reactions actually absorb heat. Take the common hospital cold pack, for example. Squeeze it to get it going, toss it into a calorimeter and watch the temperature in the calorimeter drop as the pack absorbs heat from its surroundings! Such reactions are called **endothermic**, with a positive ΔH . OK, we can measure one of the energy parameters, but we will need to know at least one other, either ΔG or ΔS before the equation $\Delta H = \Delta G + T\Delta S$ becomes useful.

4.3.4 How Is Standard Free Energy Change (ΔG) Determined?

As it turns out, ΔG (the standard free energy change) is *directly proportional* to the concentrations of the reactants and the products of a reaction conducted to completion (i.e., **equilibrium**) under standard conditions. Therefore, to determine ΔG , we need to know or to be able to measure the concentration of reactants and reaction products before and after a chemical reaction (i.e., when the reaction reaches *equilibrium*). Take the following generic chemical reaction:



The **Boltzmann equation** relates ΔG to equilibrium concentrations of A, B, C, and D:

$$\Delta G = -RT \ln K_{eq} = RT \ln \frac{[C]^2[D]}{[A]^2[B]}$$

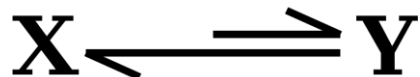
In this equation, **R** is the gas constant (1.806 cal/mol-deg); **T** is 298 °K; and **K_{eq}** is the **equilibrium constant**. As you can see, the K_{eq} for the reaction is the ratio of the product of the concentrations of the products (raised to their stoichiometric powers) to the product of the concentrations of the reactants (raised to *their* stoichiometric powers).

Once you know (or have determined) the equilibrium concentrations of reactants and products in a chemical reaction, you can use this equation to calculate the ΔG for a reaction. Remember, if the ΔG is a negative number, the reaction is defined as **exergonic**, i.e., reactions release free energy. Reactions that release free energy are also referred to as **spontaneous reactions**. If the ΔG is a positive number, the reaction absorbs free energy. Defined as **endergonic**, such *uphill* reactions are *not* spontaneous.



1.3.5 Working an Example Using These Equations for Closed Systems

Consider the following reaction:



This reaction is *endergonic*, but if you do not know [X] or [Y] (and therefore don't know the ΔG_o), how can you tell? If you are given [X] and [Y], you can also do the math. Assume we measure the concentrations of reactants and products for this reaction at equilibrium, with the following results:

$$[\text{X}] = 2.5 \text{ Kcal/mol and } [\text{Y}] = 500 \text{ cal/mol}$$

Try using the Boltzmann equation (in section 4.3.4) to calculate the standard free energy for this reaction. What is the **Keq** for this reaction? What is the ΔG_o for the reaction? If you did not come up with a Keq of 0.2 and an absolute value for the standard free energy for $|\Delta G_o|$ of 866.2 Kcal/mol, recalculate or collaborate with a classmate. Hint: make sure that you convert the units in your equation so that they are all the same (e.g., calories to Kcal)! Based on the calculated value of ΔG_o , is this reaction in fact endergonic? Now if you conduct the reaction in a bomb calorimeter, you find that it proceeds to equilibrium with a $\Delta H = -750$ Kcal/mol. What kind of reaction is this? Together with the *enthalpy change*, it is now possible to use the equation $\Delta H = \Delta G_o + T\Delta S$ to calculate an absolute value for the entropy change: $|\Delta S| = 116.2$ cal/mol-deg for the reaction. At equilibrium, the reaction proceeded with an increase in entropy under standard conditions.

If you still did not get the correct answer, recalculate, or collaborate with a classmate.



[143 Determining \$\Delta S\$ in Closed Systems](#)



4.4 Summary: The Properties of Closed Systems

First, let's reiterate that there is no such thing as a closed system, unless of course the universe is one! What we call a closed system is simply one for which we can measure the energy entering and exiting it and within which we can measure energy transfers and transductions (changes from one kind of energy to another). Features of closed systems can be defined by their properties:

- Closed systems are experimentally *defined by an investigator*.
- Defined *standard conditions* apply.
- Energy entering or leaving the system is *measurable*.
- Reactions reach *equilibrium* regardless of reaction rate.
- Product and reactant concentrations at equilibrium are *constant* (hence the Keq).
- Measured energy transfers/transductions are *constant* for a set of defined conditions.

Now let's turn our attention to open systems, since that's what cells are!

4.5 Open Systems and Actual Free Energy Change

Cells are *open systems* that are constantly exchanging mass and energy with their environment; they *never reach equilibrium*. In addition, diverse organisms live under very different atmospheric conditions and maintain different body temperatures (e.g., your cat has a higher body temperature than you do!). Clearly, the conditions under which cells conduct their biochemical reactions are decidedly *nonstandard*. However, while *open systems* do not reach equilibrium, they *do* achieve a *steady state* in which the rate of input of energy and matter is equal to the rate of output of energy and matter. Think of a biochemical pathway like glycolysis. If a cell's energy needs are constant, the pathway will reach a *steady state*. Of course, a cell's need for energy (as ATP) can change as energy needs change. If it does, then the steady state of ATP production will change to meet the needs of the cell.

Later we will discuss just how energy flows through living things, from sunlight into the chemical energy of nutrient molecules, into energy-rich fuels like ATP, and finally into the performance of all manner of cellular work. For now, let's characterize open systems by their properties:

- Open systems *exchange energy and mass with their surroundings*.
- Open systems *never reach equilibrium*
- They *achieve steady state* where the energy input rate = output rate.
- The *steady state can change*.
- In open systems, *endergonic reactions can be energetically favorable* (spontaneous reactions), and *exergonic reactions can become energetically unfavorable*.

Fortunately, there is an equation to determine free energy changes in open systems. For our chemical reaction $2A + B \rightleftharpoons 2C + D$, this equation would be the following:

$$\Delta G' = \Delta G_o + RT \ln \frac{[C]_{ss}^2 [D]_{ss}}{[A]_{ss}^2 [B]_{ss}}$$

$\Delta G'$ is the *actual free energy change* for a reaction in an open system; ΔG_o is the standard free energy change for the same reaction under standard conditions in a closed system; R is again the gas constant (1.806 cal/mol-deg); and T is the absolute temperature in which the reaction actually occurs. The *ss* subscripts designate reactant and product concentrations measured under *steady-state* conditions. You can see here that this equation states a relationship between ΔG_o and $\Delta G'$. So, to determine the actual free energy of a biochemical reaction in a cell or any living tissue, all you need to know are the ΔG_o for the reaction, the steady-state concentrations of reaction components in the cells/tissues, and the absolute T under which the reactions are occurring.



Elsewhere, we will use the reactions of the glycolytic pathway to exemplify the properties, as well as the energetics of open and closed systems. At that time, pay careful attention to the application of the terminology of energetics in describing energy flow in closed vs open systems.

Some iText & VOP Key Words and Terms

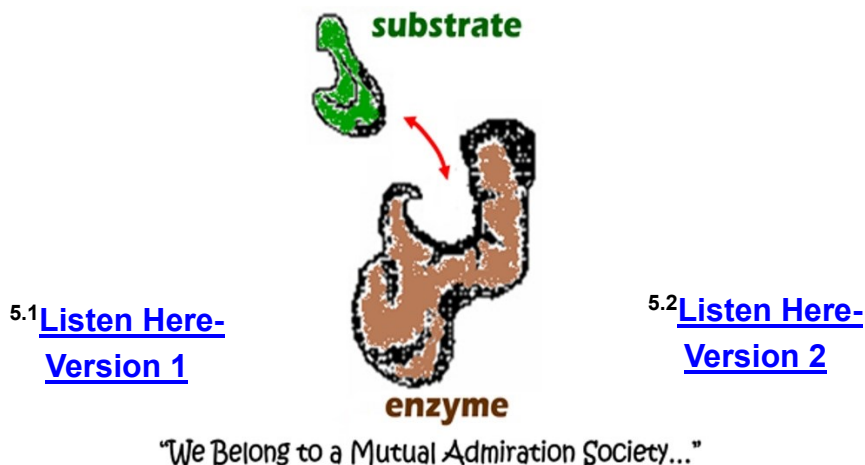
actual free energy	endothermic	law of conservation
ATP	energy	laws of thermodynamics
bioenergetics	energy transduction	light
Boltzmann equation	energy transfer	mass
calories	enthalpy	open-system properties
calorimeter	entropy	open systems
chemical energy	equilibrium constant	order vs entropy
chemical equilibrium	exergonic	standard conditions
closed systems	exothermic	standard free energy
decibels	free energy	steady state
$e = mc^2$	gas constant	useful energy
electricity	Gibbs free energy	useless energy
endergonic	K _{eq}	volts

Chapter 5

Enzyme Catalysis and Kinetics

Mechanism of enzyme catalysis, induced fit, activation energy, understanding enzyme kinetics

Reminder: For inactive links, google key words/terms for alternative resources.



5.1 Introduction

By definition, all catalysts accelerate chemical reactions, including enzymes. But enzymes and inorganic catalysts differ in important ways (Table 5.1).

Table 5.1
Enzymes vs Inorganic Catalysts

Inorganic Catalysts	Enzymes
e.g., Ni, Pt, Ag, etc.	e.g., pepsin, trypsin, ATP synthase, ribonuclease, etc.
increase rxn rate	increase rxn rate
unchanged at end of rxn	unchanged at end of rxn
non-specific	highly specific
rigid, inflexible	flexible - can undergo allosteric change...
cannot be regulated	can be regulated

In this chapter, we look at the properties and mechanism of action of enzymes. These include allosteric change (*induced fit*, *enzyme regulation*), energetic events (changes in *activation energy*), and how enzymes work in open and closed (experimental) systems.

Most **enzymes** are proteins; a few are RNAs. All enzymes are long flexible polymers that fold into intricate shapes, able to recognize and tightly bind specific target molecules. In contrast, inorganic catalysts are rigid, weakly binding many molecules and catalyzing many random reactions. The specificity of an enzyme lies in the structure and flexibility of its **active site**. We will see that the active site of an enzyme undergoes conformational change during catalysis.

The structural flexibility of enzymes also explains their ability to respond to metabolites in a cell, metabolites which indicate the cell's biochemical status. When such *regulatory* metabolites bind to an enzyme, the resulting conformational changes can speed up or slow down the catalytic rate of the reaction—a phenomenon called **allosteric regulation**. As you might imagine, changing the rate of a biochemical reaction can change the rate of an entire biochemical pathway—and ultimately the steady-state concentrations of products and reactants in that pathway.

To understand the importance of allosteric regulation, we'll look at how we measure the speed of enzyme catalysis. As we consider the classic early twentieth-century enzyme kinetic studies of Leonor Michaelis and Maud Menten, we'll focus on the significance of the ***K_m*** and ***V_{max}*** values that they derived from their data. But before we begin our discussion, remember that chemical reactions are intrinsically reversible. The action of catalysts, either organic or inorganic, depends on this concept of reversibility.

Finally, let's give a nod to recent human ingenuity that enabled enzyme action to turn an *extracellular profit*! You can now find enzymes in household cleaning products like spot removers and detergents, where they function to digest and to remove stains caused by fats and pigmented proteins. Enzymes added to meat tenderizers also digest (hydrolyze) animal proteins down to smaller peptides. Enzymes can even clean a clogged drain!

Learning Objectives

When you have mastered the information in this chapter, you should be able to do the following:

1. Describe how the flexibility of proteins and RNAs makes them ideal biological catalysts.
2. Compare and contrast the properties of *inorganic* and *organic* catalysts.
3. Explain why *catalysts do not change equilibrium concentrations* of a reaction conducted in a closed system.
4. Compare the activation energies of catalyzed and uncatalyzed reactions and explain the roles of *allosteric effectors* in enzymatic reactions.
5. Discuss how allosteric sites interact with an enzyme's active site and explain the concept of the rate-limiting reaction in a biochemical pathway.
6. Write simple rate equations for chemical reactions.
7. Write possible chemical rate equations for catalyzed reactions.
8. Distinguish between *V_{max}* and *K_m* in the Michaelis-Menten kinetics equation.
9. State what *V_{max}* and *K_m* say about the progress of an enzyme-catalyzed reaction.
10. Interpret enzyme kinetics and the *progress of an enzyme-catalyzed reaction* from this data.
11. More accurately identify Leonor Michaelis and Maud Menten!

5.2 Enzymes and the Mechanisms of Enzyme Catalysis

Studies by George W. Beadle and Edward L. Tatum correlating mutations with enzyme deficiencies in *Neurospora crassa* (bread mold) and *Drosophila melanogaster* led them to propose the **one-gene/one-enzyme hypothesis** in 1941. By 1958, they shared the Nobel Prize in Physiology or Medicine for this work. Their original hypothesis had already morphed twice—first into the **one-gene/one-protein**, and then into the **one-gene/one-polypeptide** hypothesis..., helping to launch the age of molecular biology.

The subsequent discovery of RNA catalysts came as quite a surprise! The revelation of RNA catalysts, dubbed **ribozymes**, earned Sidney Altman and Thomas Cech a Nobel Prize in Chemistry in 1989. *Ribozymes* are now known to catalyze RNA splicing (the removal of unwanted regions of a precursor RNA). They also catalyze a step in the protein synthesis (translation) by ribosomes. In fact, almost no biochemical reactions or pathways exist that are not directly the result of enzyme catalysis, from the digestion of nutrients in your mouth, stomach, and small intestines to pretty much every chemical reaction inside your cells (check out Kornberg A. 1989. *Never a Dull Enzyme*. Ann. Rev. Biochem. 58:1-30). The focus in this chapter is on the long history of *protein enzyme catalysis*. But as you study, you may recognize that the mechanisms of enzyme catalysis described here involve common essential features seen in all biocatalysts.

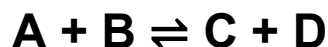
Most enzymes are soluble inside or outside cells, while a few are part of membranes or other cellular structures. In all cases, they bind to soluble **substrates** (the reactants in enzyme-catalyzed reactions). The large size and exquisite diversity of protein structures make enzymes highly specific catalysts. As already noted, the specificity of an enzyme results from the shape of its **active** site, which is dependent on the 3D arrangement of amino acids in and around the region. The *substrates* of a catalyzed biochemical reaction are bound to and held in place on the enzyme while rapid bond rearrangements take place. The flexibility of enzymes allows enzymes to change in shape at the active site during catalysis. The flexibility of enzymes also enables small metabolites in cells to interact with an enzyme, change its shape and thereby changing its catalytic rate. The latter phenomenon enables **allosteric regulation**, allowing cells to control the rates and even the direction of many biochemical reactions and pathways. As we will see, enzymes may be bound to *prosthetic groups* or ions, which contribute to the shape and activity of the enzyme.

Any understanding the mechanism of catalysis must also include knowledge about the energetics of catalyzed reactions. We'll see that enzymes lower the **activation energy** of a chemical reaction, and that *activation energy* is an inherent *energy barrier* to the reaction. Finally, we look at the energetics of enzyme action. We will also describe the action of biological catalysis in terms biochemical pathways and the structural features of enzymes (e.g., active-site and overall conformation and the *affinities* of an enzyme for its substrates), as well as free energy changes that occur during catalysis. Of course, structural and energy considerations of enzyme catalysis are related, as we will see in more detail in a later chapter.



5.2.1 Structural Considerations of Catalysis

From a chemistry course, you may recall that the rate of an uncatalyzed reaction is dependent on the concentration of the reactants in solution. This is the *law of mass action*, recognized in the nineteenth century. Look at this simple reaction:



The *law of mass action* makes two key assumptions:

1. At any given time following the start of the reaction, the rate of product formation is proportional to the concentrations of the reactants and products ([A], [B], [C], and [D] in this case).
2. Chemical reactions in the laboratory (i.e., a "closed system") eventually reach equilibrium, at which point the net rate of formation of reaction products is zero. In other words, the forward and reverse reactions occur at the same rate.

There are no products (i.e., C and D) at the start of the reaction written above. As there are no products yet, the reaction rate should be directly proportional only to the concentration of the reactants. The *law of mass action* predicts that any chemical reaction will occur its fastest rate when reactants (e.g., A and B above) are first mixed—that is, when A and B at their highest concentrations. This is because there are more reactant molecules in solution and a greater likelihood that they will collide in an orientation that allows the bond rearrangements necessary for the reaction to occur.

Of course, chemical rates decline as reactant concentrations diminish products accumulate over time. But even as the rate of formation of C and D slows down, their rising levels of should also influence the rate of their own production. Remember, all chemical reactions are inherently reversible; those rising concentrations of products will begin to push the reaction in reverse to form A and B, further slowing down the net accumulation of C and D.

The chemical rate equations that you may recall from a chemistry course in fact treat all chemical reactions as reversible (elsewhere we'll address the concept that some reactions in cells are biologically irreversible!). Chemical rate equations enable the determination of reaction rates for our sample reaction. Here is an equation for the rate of formation of the products C & D:

$$\mathbf{\text{rate of formation of C and D} = k_1[A][B] - k_{-1}[C][D]}$$

In this equation, $k_1[A][B]$ is the rate of the forward reaction and $k_{-1}[C][D]$ is the rate of the reverse reaction. This equation recognizes that the reaction is reversible. The equation states that the net reaction rate is equal to the *rate of the forward reaction* ($k_1[A][B]$) minus the *rate of the back reaction* ($k_{-1}[C][D]$). The equation is valid (applicable) at any time during the reaction. k_1 and k_{-1} are *rate constants* for the forward and reverse reactions, respectively.

So how do catalysts work? Catalysts increase chemical reaction rates by bringing reactants (now called substrates) together more rapidly than they would encounter each other based just on random molecular motion in solution. This is possible because catalysts have an **affinity** for their substrates. In the case of inorganic catalysts, relatively weak, generic forces account for the affinity of reactants and inorganic catalysts. Thus, a metallic catalyst (e.g., silver or platinum) attracts molecules with an appropriate (usually a charge) configuration. If the attraction (**affinity**) is sufficient, the metal will hold reactants in place long enough to catalyze the bond rearrangements of a chemical reaction. In contrast to inorganic catalysts, enzymes have evolved highly specific shapes with physical-chemical properties. As a result, typical enzymes only attract substrates for specific reactions, and do so with high affinities. The **lock-and-key** mechanism was the first attempt to explain enzyme-substrate specificity (Figure 5.1, below).

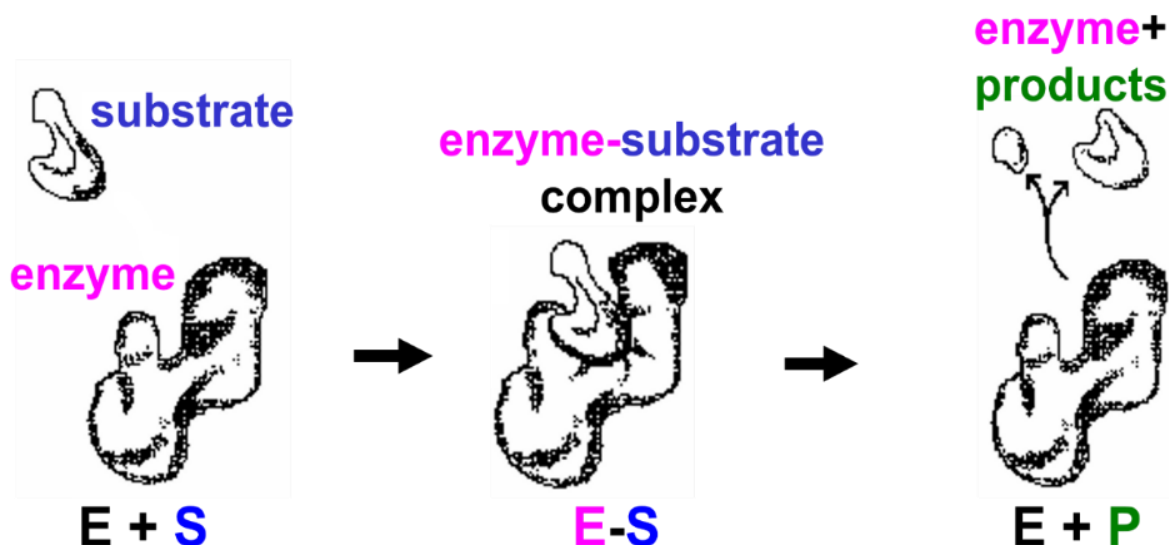


Fig. 5.1: In the lock-and-key model, reactants i.e., substrates (S) are 'keys' that fit in enzyme (E) 'locks': (1, left), E and S attract each other; (2, middle) S enters the catalytic site on E like a key in a lock, forming an E-S complex (middle); (3, right) catalysis occurs, E and P separate.

The active site of an enzyme has an exquisite, selective affinity for its substrate(s). This affinity is many times greater than those of inorganic catalysts for generic reactants. The result is that enzymes are more efficient and faster than inorganic catalysts. According to this model, the affinity of an enzyme for a particular substrate engages the substrate "key" in the tumblers (i.e., in the active site) of the enzyme's "lock." Thus engaged, the substrate(s) would undergo the bond rearrangements specific for the catalyzed reaction to generate products and to regenerate an unchanged enzyme.

However, X-ray crystallography of enzyme-substrate interactions revealed that the active sites of enzymes change shape during catalysis. This **allosteric change** suggested the **induced-fit** mechanism of enzyme action. In this model, enzyme-substrate affinity causes the substrate to bind to the enzyme surface. Once bound, the enzyme undergoes an allosteric change, drawing the substrate(s) more tightly into the active site and catalyzing the reaction. Of course, after the reaction products detach from the enzyme, it returns to its original shape. (modeled in Figure 5.2, below).

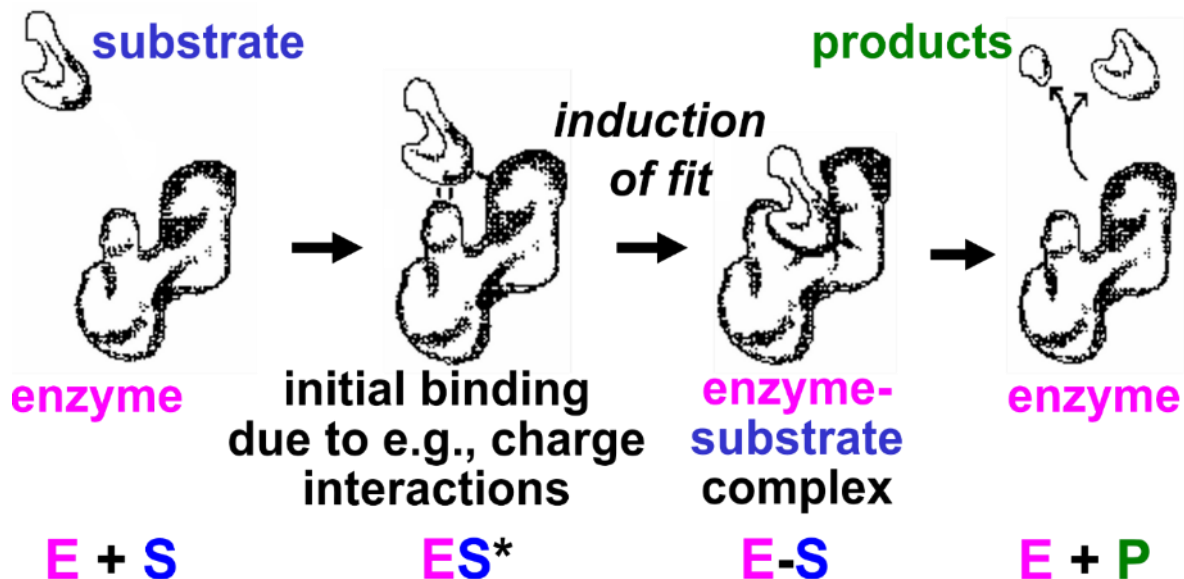


Fig. 5.2: In the *induced-fit* model, E and S attract each other initially forming a short-lived ES^* , inducing an allosteric (shape) change leading to a tight fit of S in the catalytic site of E to form an E-S complex. Catalysis occurs and E and P separate.



[146-2 Induced-Fit Mechanism of Enzyme Action](#)

5.2.2 Energetic Considerations of Catalysis

Consider the random motion of substrates in solution, which only occasionally encounter one another. They even more rarely bump into one another in just the right orientation to cause a reaction. This explains why adding more reactants or increasing the temperature of a reaction can speed it up: it increases the number of random as well as productive molecular collisions. Unlike molecules and reactions in a test tube, living organisms do not have these options for increasing reaction rates. But they do have enzymes! All catalysts work by lowering the **activation energy** (E_a) for a reaction, thereby increasing the rate of the reaction. Activation energy is essentially a barrier that makes it difficult for interacting substrates to come together to actually undergo a biochemical reaction.

Inorganic catalytic surfaces attract reactants where catalysis can occur. The attractions are weak compared to those of enzymes and their substrates. An enzyme's active site very strongly attracts otherwise randomly distributed substrates, making enzyme catalysis faster than inorganic catalysis. Again, cells cannot use inorganic catalysts, most of which are insoluble and would attract reactants indiscriminately—not a good way for cells to control metabolism! The advent of enzymes with their specificity and high rate of catalysis was a key event in *chemical evolution* and was required for the origins of life. As we saw, allosteric change during the “induction of fit” enables specific catalysis. In fact, a catalyzed reaction will be faster than the same reaction catalyzed by a piece of metal, and of course much faster (millions of times faster!) than the uncatalyzed reaction. The energetics of catalysis helps to

explain why. Let's look at the energetics of a simple reaction in which A and B are converted to C and D (Figure 5.3).

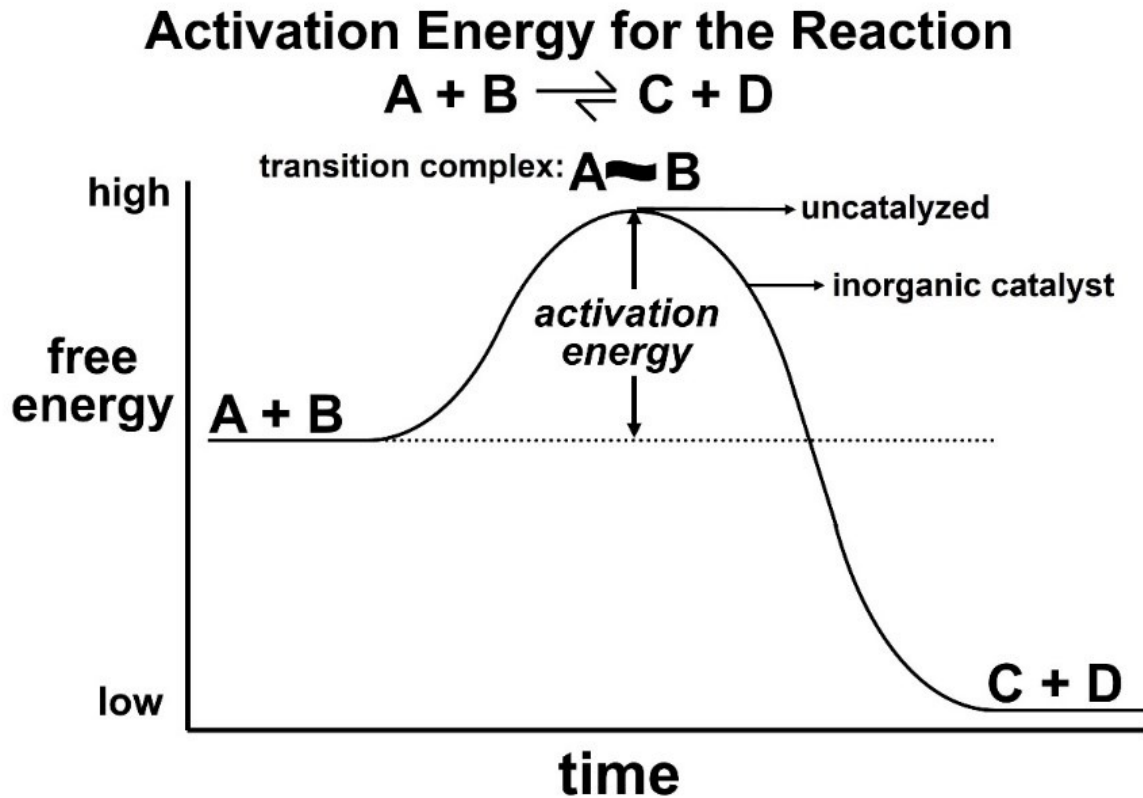


Fig. 5.3: This graph follows the free energy change over time as reactants A and B react. The *activation-energy* peak in the middle of the curve defines that *free energy barrier* which A and B must overcome before much C and D can be made. The *barrier* is due mainly to thermal motion of A and B molecules that rarely encounter each other in relatively dilute solutions. Enzymes are even more efficient than inorganic catalysts in lowering the activation-energy barrier to a reaction.

Conducted in a closed system, enzyme-catalyzed reactions rapidly reach equilibrium. Like all catalysts, enzymes are not consumed by the reactions they catalyze, nor do they alter equilibrium concentrations of reactants and products of these reactions. One estimate is that as few as 4,000 biochemical reactions are catalyzed in a given cell at a given time. Another estimate is that there are 20,000-25,000 genes in the human genome. Many of these encode enzymes. What are these genes and their protein products good for if so few biochemical reactions are needed by an average cell? Is this in fact an underestimate?



[147 Enzyme Activation Energy](#)

5.3 Enzyme Regulation

We noted that some enzymes are regulated, which just means that there are factors in the cell that can slow down or speed up their rate of catalysis. In this way the cell can respond quickly

to the metabolic needs reflected by the intracellular levels of these factors. Factors that slow down catalysis are *inhibitors*. Those that speed up catalysis are called *activators*.

In addition to responding to intracellular molecular indicators of the metabolic status of the cell, enzymes may also be inhibited by drugs, poisons, or changes in the chemical milieu (e.g., pH). Since cellular reactions occur as part of biochemical pathways, regulating a single enzyme can affect an entire pathway (e.g., Figure 5.4).

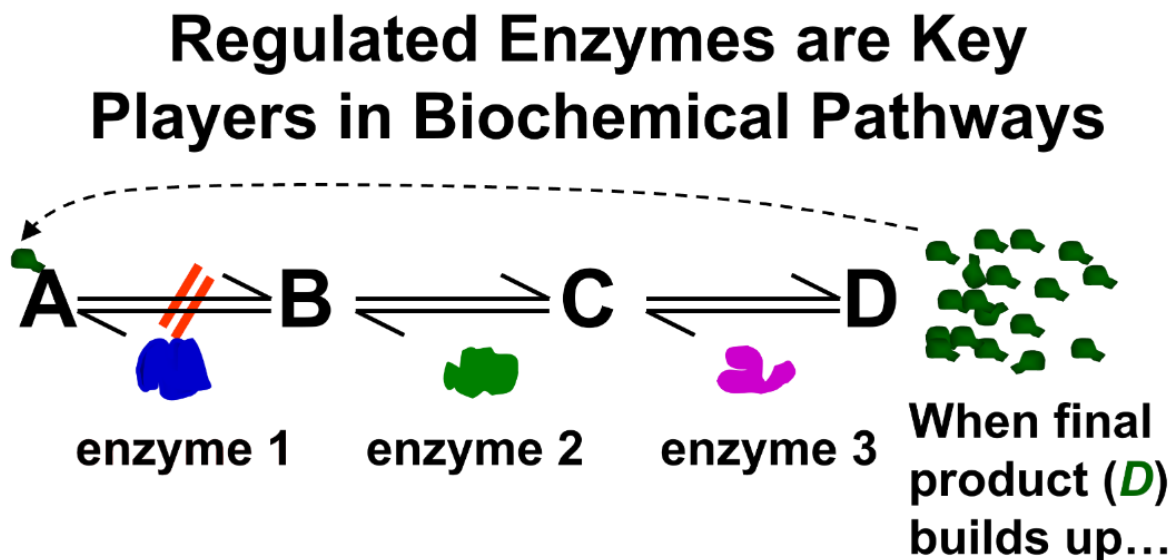


Fig. 5.4: Reaction products in a biochemical pathway can alter the rate of synthesis of the pathway's final products. In this generic pathway, the end-product (D) accumulates (far right). At some point, excess E binds to the allosteric regulatory sites on enzyme 1 at the left, blocking the formation of B, an intermediate product. This kind of *feed-back inhibition* prevents a wasteful overproduction of D.

This biochemical pathway will produce substance **E**. Under normal conditions, another series of metabolic reactions would consume **E**. However, if the cell no longer needs so much **E**, it will accumulate in the cell. If there is an excess of **E** in the cell, some of the excess might bind to one of the enzymes. In the pathway shown, **E** binds to *enzyme 1*. This binding causes an allosteric change in the enzyme, inhibiting catalysis and slowing down the entire pathway.

In this example of *allosteric regulation*, the inhibitory regulation of **enzyme 1** evolved to control the rate of production of substance **E**. This common mode of *allosteric regulation* is called *feedback inhibition*. Enzymes can be regulated precisely because they can be *bent out of shape* (or into shape for that matter!). Some small metabolites are chemical information when they accumulate in cells and can communicate cellular metabolic status. The result is a decrease or increase in enzyme activities to achieve an appropriate cellular response.

Whether an activator or an inhibitor of enzyme catalysis, regulatory molecules typically bind to enzymes at *allosteric regulatory* sites, causing local conformational changes in the enzyme, which are transmitted to the active site. *Enzyme inhibition* would occur if a change in shape reduced the affinity of the enzyme for a substrate—or if it reduced the rate of bond rearrangements after the substrate had already entered the active site. *Enzyme activation*

would occur if the allosteric effect were to increase this affinity and/or catalytic rate. The mechanism of allosteric regulation of enzyme activity is shown in Figure 5.5.

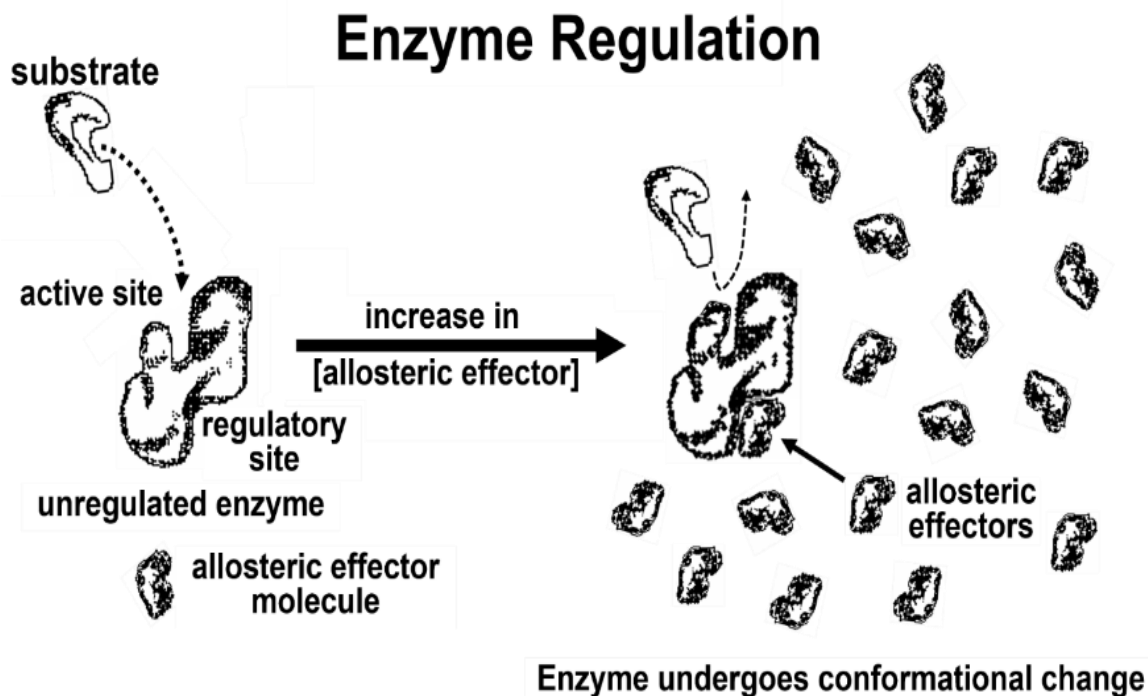


Fig. 5.5: Levels of small metabolic chemicals (*metabolites*) can reflect and control a cell's metabolic status. Such metabolites, called *allosteric effectors*, have an affinity for *regulatory sites* on an enzyme (shown at the left), causing conformational (shape) changes in its *active site* (seen at the right). Allosteric regulation can reduce or increase the affinity of enzyme and substrate, either inhibiting catalysis, as suggested here, or stimulating it.



[148-2 Allosteric Regulation of Enzyme Activity](#)



5.4 Enzyme Kinetics

Studying enzyme kinetics not only tells us about the catalytic properties of a particular enzyme but can also reveal important properties of biochemical pathways. Thus, one can determine a standard *rate-limiting reaction* under a given set of conditions by comparing kinetic data for each enzyme in a biochemical pathway. But what else can we do with kinetic data?

5.4.1 Why Study Enzyme Kinetics?

Apart from their value in teaching us how a given enzyme actually works, kinetic data has considerable clinical value. For example, let's assume that we know the kinetics of every enzyme in the biochemical pathway ending in the synthesis of a liver metabolite in a healthy person. Also assume we know what the *rate-limiting* (slowest) enzyme in the pathway is. Consider the following scenario: A patient presents with higher-than-normal blood levels of **X**. Is this because the normal *rate-limiting* reaction is no longer rate-limiting in the patient? If so, what is new rate-limiting enzyme?

One can ask similar questions of an alternate scenario in which a patient is producing too little of the metabolite. A cellular biochemical might deviate from "normal" levels for a variety of reasons:

- ***Viral and bacterial infection or environmental poisons:*** These can interfere with a specific reaction in a metabolic pathway; remedies depend on this information!
- ***Chronic illness resulting from mutational enzyme deficiencies:*** knowing the kinetics of enzymes in a pathway suspected to be involved in the illness could lead to treatments, including medications designed to enhance or to inhibit (as appropriate) enzyme activity.
- ***Genetic illness tied to metabolic deficiency:*** If a specific enzyme is the culprit, investigation of a pre- and/or postnatal course of treatment might be possible (e.g., medication or perhaps even gene therapy).
- ***Lifestyle changes and choices:*** These might include eating habits, usually remediated by a change in diet.
- ***Lifestyle changes brought on by circumstance rather than choice:*** These are changes due to aging. An all-too-common example is the onset of type 2 diabetes. This can be treated with medication and/or delayed by switching to a low-carb diet favoring hormonal changes that would improve proper sugar metabolism.

Knowing the rate-limiting reaction(s) in a biochemical pathway can identify regulated enzymes. This allows research to focus on the search for remedies to correct a metabolic imbalance. As noted, ribozymes are RNA molecules that catalyze biochemical reactions; their kinetics can also be analyzed and classified. We will consider how enzymes are regulated later, when we discuss glycolysis, a biochemical pathway that most living things use to extract energy from nutrients. For now, let's look at an overview of *experimental* enzyme kinetics.

5.4.2 How We Determine Enzyme Kinetics and Interpret Kinetic Data

In enzyme kinetic studies, the enzyme is considered to be a reactant, albeit one that is regenerated by the end of the reaction. The reaction begins when substrate is added to the enzyme. In enzyme kinetic studies, the concentration of the enzyme is held constant while reaction rates are measured, after different amounts of substrate are added. As a consequence, all catalyzed reactions will reflect *saturation* of the enzyme at high concentration of substrate. This is the basis of saturation kinetics illustrated below in Figure 5.6.

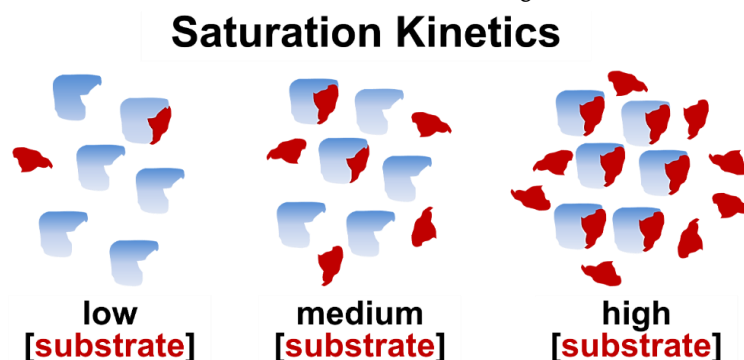
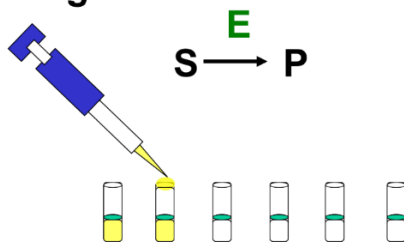


Fig. 5.6: Enzyme saturation occurs when all available enzymes in solution are bound to substrate(s). Saturation kinetics experiments measure catalytic rates at different substrate(s) concentrations. At high substrate concentrations the active sites of all the enzymes are occupied, i.e., enzyme is saturated with substrate (far right). Under these conditions the reaction occurs at its fastest rate.

In the illustration, the active sites on all the enzyme molecules are bound to substrate molecules at a high substrate concentration. Under these conditions, a catalyzed reaction is proceeding at its fastest. Let's generate some kinetic data to see saturation in action. The experiment illustrated in Figure 5.7 will determine the kinetics of the conversion of substrate (S) to product (P) by an enzyme (E).

Determining the Kinetics of the Reaction



1. Hold **[E]** constant while varying **[S]**
2. Measure the **initial rate** of **[P]** formation at each **[S]**.

Fig. 5.7: The test tubes shown in this experiment each contain the same concentration of enzyme. After adding different amounts of substrate(s) to each tube, rates of product formation near the start of the reaction are measured to determine the kinetics of the enzyme-catalyzed reaction.

A series of reaction tubes are set up, each containing the same concentrations of E ([E]) but different concentrations of S ([S]). The concentration of P ([P]) produced at different times (beginning just after the start of the reaction in each tube) is plotted to determine the *initial rate* of P formation for each [S] tested. Figure 5.8 below is such a plot.

Plot of the Rate of an Enzyme-Catalyzed Reaction at Different Substrate (S) Concentrations

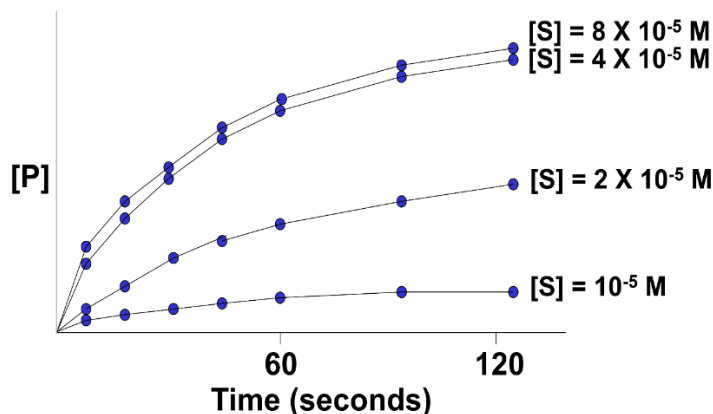


Fig. 5.8: This graph plots the rate of product (P) formation at different substrate concentrations ([S]) while holding the enzyme concentration ([E]) constant, as described in the experimental protocol in Fig. 5.7.

In this hypothetical example, the rates of the reactions (amounts of P made over time) do not increase beyond an [S] higher than 4×10^{-5} M. The upper curves thus represent the maximal rate of the reaction at the experimental concentration of enzyme. We say that the maximal reaction rate occurs at **saturation**.



Next, we can estimate the initial reaction rate (v_0) at each substrate concentration by plotting the slope of the first few time points through the *origin* of each curve in the graph. Consider the graph of the initial reaction rates estimated in this way in Figure 5.9 (below).

Finding the Initial Reaction Rate of an Enzyme-Catalyzed Reaction at Different Substrate (S) Concentrations

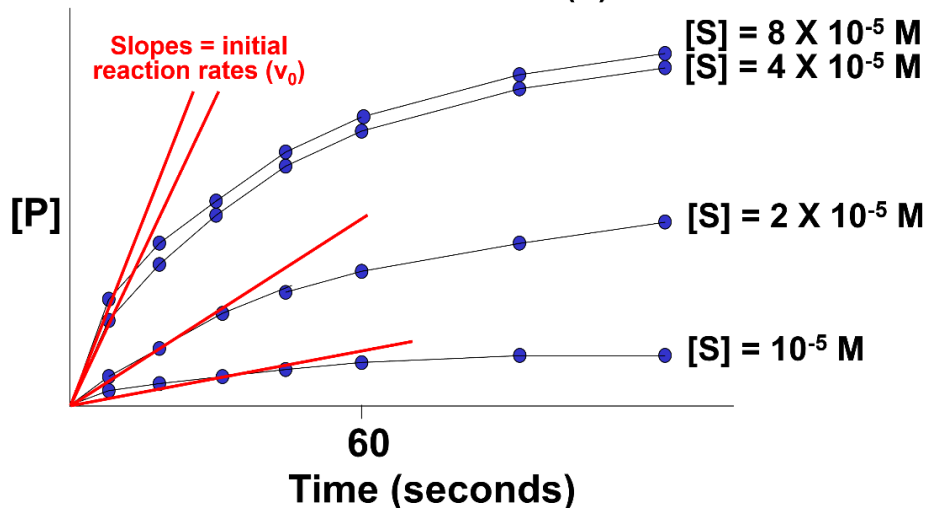


Fig. 5.9: This graph shows the slopes (straight lines at the left) of the initial reaction rates taken from the curves in the previous graph (Fig. 5.8). From the graph we can say that the enzyme became saturated somewhere between the two highest concentrations of substrate.

The slope of each straight line is the v_0 for the reaction at a different $[S]$ value near the very beginning of the reaction, when $[S]$ is high and $[P]$ is vanishingly low. Next, we plot these rates (slopes, or v_0 values) against the different $[S]$ values in the experiment to get the curve of the reaction kinetics in Figure 5.10.

Initial Enzyme Reaction Rates and the Rectangular Hyperbola

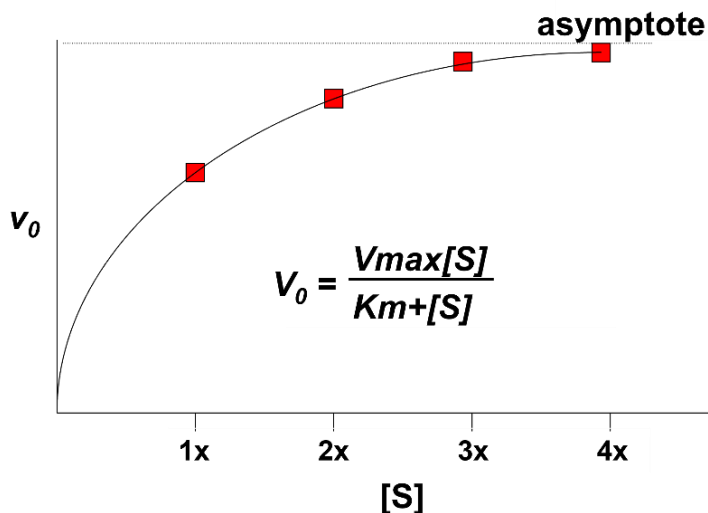


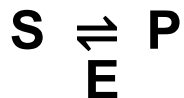
Fig. 5.10: This graph plots the initial reaction rates (slopes, or v_0) for the reactions plotted in Figure 5.9. The curve is approaching a maximum v_0 (the asymptote at the right). The formula shown for this curve describes a *rectangular hyperbola*.

This is an example of Michaelis-Menten kinetics, which is common to many enzymes. Named after the biochemists who realized that the curve described *rectangular hyperbola*. Put another way, the equation mathematically describes the mechanism of catalysis of the enzyme! The following equation mathematically describes a rectangular hyperbola:

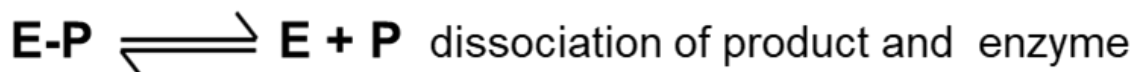
$$y = \frac{xa}{x+b}$$

You might be asked to understand the derivation of the Michaelis-Menten equation in a biochemistry course. You might even be asked to do the derivation yourself! We won't make you do that but suffice it to say that Michaelis and Menten started with some *simple assumptions* about how an enzyme-catalyzed reaction would proceed and then wrote reasonable chemical and rate equations for those reactions. The goal here is to understand those assumptions, to see how the kinetic data support those assumptions, and to realize what this tells us about how enzymes really work.

Here is one way to write the chemical equation for a simple reaction in which an enzyme (E) catalyzes the conversion of substrate (S) to product (P):



Michaelis and Menten rationalized that this reaction might actually proceed in three steps. In each step, enzyme E is treated as a reactant in the conversion of S to P. The resulting chemical equations are shown below:



Reasoning that the middle reaction (the conversion of E and S to E and P) would be the fastest one and therefore would not be the *rate-limiting reaction of catalysis*, they only considered the first and third reactions to be relevant in determining the overall kinetics of product formation. Then they wrote the following rate equations for just these two chemical reactions (as one would in an introductory chemistry course):

$$V_{\text{E-S formation}} = k_1[\text{E}][\text{S}] - k_{-1}[\text{E-S}]$$

$$V_{\text{P formation}} = k_2[\text{E-S}] - k_{-2}[\text{E}][\text{P}]$$

Both of these equations describe a straight line. Therefore, neither describe the observed hyperbolic reaction kinetics. Finally, if you solve one of the equations for a term common to

both (e.g., E-S) and then plug your solution into the other equation, your combined equation still describes a straight line—again, not the expected rectangular hyperbola. To arrive at a chemical rate equation consistent with a rectangular hyperbola, Michaelis and Menten had to make several assumptions, including those made by G. E. Briggs and J. B. S. Haldane, about how E, S, and P would behave in a catalyzed reaction. Those assumptions allowed them to rewrite each equation, then to combine and to rewrite them into a single mathematical equation—one that did indeed describe a rectangular hyperbola. Here are Briggs and Haldane's assumptions:

1. $[S] \gg [E]$ at the start and during the “steady state”
2. $[P] \ll [S]$ at the start of a reaction.
3. All E is bound to S at the start of the reaction.
4. $[E]_{\text{total}} = [E]_{\text{free}} + [E-S]$ at all times.

We've already seen the equation that Michaelis and Menten derived from these assumptions, which is now known as the Michaelis-Menten equation:

$$V_0 = \frac{V_{\text{max}}[S]}{K_m + [S]}$$

The take-home message here is that the assumptions about an enzyme-catalyzed reaction are a good approximation of how the reaction proceeds over time. Michaelis and Menten defined V_{max} and K_m as key kinetic factors in enzyme-catalyzed reactions. In the generic example of substrate conversion to product, we saw that increasing $[S]$ results in a higher rate of product formation because there is a higher rate of encounters between enzyme and substrate molecules. At some point, however, increasing $[S]$ does not increase the initial reaction rate any further. Instead, v_0 *asymptotically* approaches a theoretical maximum for the reaction, defined as **V_{max}** , the *maximum initial rate*. As we have already seen, V_{max} occurs when all available enzyme active sites are saturated (occupied by substrate). At this point, the intrinsic catalytic rate determines the *turnover rate* of the enzyme. The substrate concentration at which the reaction rate has reached $\frac{1}{2} V_{\text{max}}$ is defined as **K_m** (the *Michaelis-Menten constant*), which is a ratio of the rate constants derived when rate equations for E-S and P formation (above) are rewritten and then combined into the single equation that describes a rectangular hyperbola.



[150 Graphing Enzyme Kinetic Data](#)

To recapitulate, the two most important kinetic properties of an enzyme are the following:

1. The rate at which the enzyme becomes saturated with a particular substrate, which is related to the reaction's K_m

2. The maximum rate of the catalyzed reaction, which is described by the reaction's V_{max}

These properties suggest how an enzyme might behave under cellular conditions and can show how the enzyme should respond to allosteric regulation by natural inhibitory or activating factors, as well as how it might respond to poisons or other noxious chemicals.

For clear, detailed explanations of enzyme catalytic mechanisms, check out Jencks WP 1987, *Catalysis in Chemistry and Enzymology* Mineola, NY, Courier Dover Publications. You can also find more details of how kinetic equations are derived (an important step in understanding how the enzyme works) in any good biochemistry textbook, or you can check out the *Michaelis-Menten Kinetics* entry in the ^{5.3}[All About Enzymes](#) Wikipedia link.

Some iText & VOP Key Words and Terms

activation energy	enzyme	Michaelis-Menten constant
active site	enzyme activation	Michaelis-Menten kinetics
allosteric change	enzyme inhibition	rate-limiting reaction
allosteric regulation	enzyme kinetics	ribozyme
allosteric site	enzyme regulation	saturation kinetics
biochemical pathway	induced fit	substrate specificity
catalytic RNAs	inorganic catalyst	substrates
conformation	K _m	V _{max}

CHAPTER 5 WEB LINKS

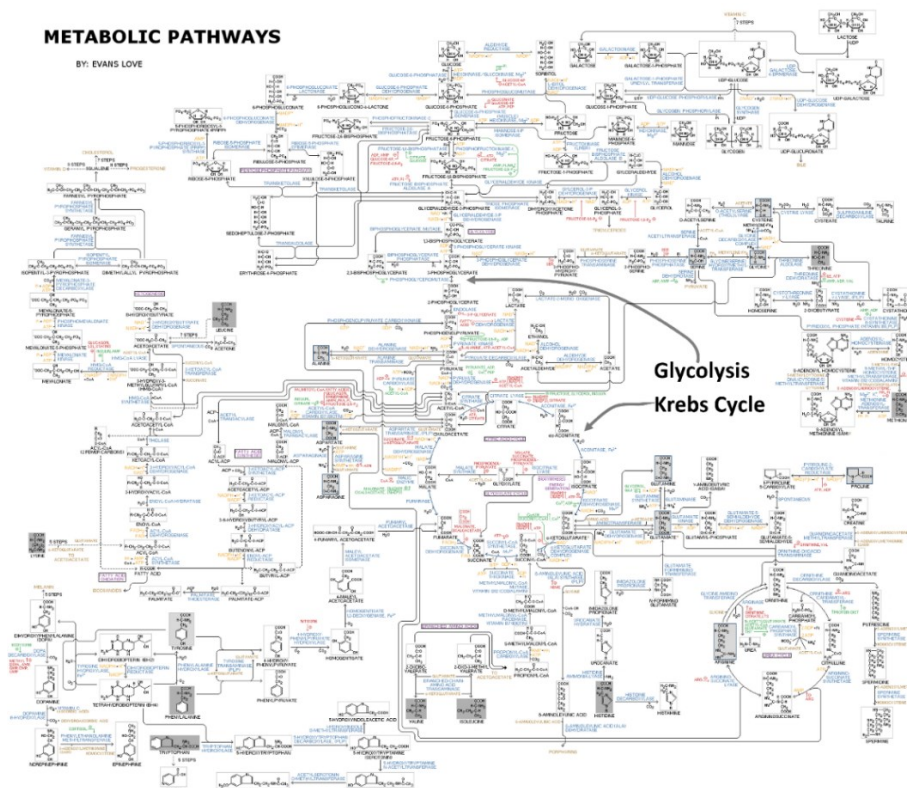


Chapter 6

Glycolysis, the Krebs Cycle, and the Atkins Diet

Glycolysis, gluconeogenesis, the Krebs cycle—getting energy from food; enzyme regulation; how cells capture free energy (bioenergetics); liver cells glucose metabolism; diets to fool your body

Reminder: For inactive links, google key words/terms for alternative resources.



This (in fact, any!) metabolic pathway chart highlights the centrality of the Krebs (citric-acid) cycle to virtually all biochemical pathways. See a sizable version of this chart at [6.1 Human Metabolic Pathways](#)

6.1 Introduction

We used to get free metabolic pathway wall charts like this one from vendors of biochemical reagents. The big picture is correct (as noted, a sizable version is available at the link below the chart), but the charts may be out of date in small details of metabolism. In this chapter, we'll zoom in on the middle of the chart, encompassing glycolysis and the Krebs cycle (named for Hand Krebs, its discoverer) to see how thermodynamic laws apply to chemical reactions. We've looked at the principles governing *thermodynamics* (the flow of energy in the universe) and *bioenergetics* (energy flow in living systems). We saw evidence that energy is exchanged between components in the universe, but that it can be *neither created nor destroyed*. That makes the universe a *closed system*, a conclusion codified as the *first law of thermodynamics*. Personally, I find the idea of a closed universe troubling since there is no escape from it—that is, until I remind myself that the universe is a big place, and I am only a small part of a small system that you can define for yourself: the solar system, planet Earth, the country you pledge allegiance to, your city or village, your school, your farm, or your homestead!

You may derive comfort from the realization that you can move from one system to another and even exchange goods and services between them. This is a metaphor for energy flow between systems in the universe. We also saw that the *first law* applies to **closed systems within the universe**, and that there are no closed systems in the universe! Any system in the universe is open, always exchanging energy and mass with neighboring systems. What we mean by the term *closed system* is that we can define and *isolate* some small part of the universe and then *measure* any energy that this isolated system gives up to its environment or takes in from it. The simplest demonstration of the *first law* in action is the *bomb calorimeter*, which measures heat released or absorbed during a chemical reaction.

A second thermodynamic concept says that energy flows from one place to another only *when it can*. In the vernacular, we say that **energy flows downhill**. Anything that happens in the universe (a galaxy moving through space, a planet rotating, you getting out of bed in the morning, coffee perking you up, your cells burning sugar, DNA replicating...) does so because energy flows downhill. We saw that any happening or event in the universe, however large or small, must be **spontaneous**. That is, it occurs with a release of **free energy**. Remember, *spontaneous* means "by itself" and not necessarily "instantaneous" or "fast"! Finally, we noted that when enzymes catalyze biochemical reactions in a closed system, the reactions still reach equilibrium, despite the higher rate of the catalyzed reaction. What does this tell you about the energetics of catalyzed reactions in *closed systems*?

With this brief reminder about energy flow and what enzymes do, we'll look at how our cells capture **nutrient free energy** (the chemical energy in foods), a topic that will include examples of the energetics of closed systems that reach equilibrium and open systems that don't! First, we tackle **glycolysis**, an **anaerobic fermentation** pathway for generating chemical energy from glucose, as well as the first of several **aerobic** pathways of **respiration**. We'll see that most of the energy from glycolysis and respiration is captured in molecules of ATP, the universal energy currency of life, used by cells to...live! Then we look at **gluconeogenesis**, a regulated reversal of glycolysis. We ask when, where, and why we would want to make rather than burn glucose. Finally, we begin a discussion of respiration with a look at the **Krebs cycle**. The complete respiratory pathway can be summarized by the following equation:



The **standard free energy change** for this reaction (ΔG°) is about -687 Kcal/mol. This is the maximum amount of nutrient free energy that is (at least in theory) available from the complete respiration of a mole of glucose. Given a cost of about 7.3 Kcal to make each mole of ATP (*adenosine triphosphate*), how many moles of ATP might a cell produce after burning a mole of glucose? We'll figure this out here.

Learning Objectives

When you have mastered the information in this chapter, you should be able to do the following:

1. Explain the difference between fermentation and respiratory glycolysis and the role of redox reactions in both processes.
2. Calculate, compare, and contrast ΔG° and $\Delta G'$ for the same reaction, and explain any differences in free energy in open and closed systems.

- Describe and explain the major events of the first stage of glycolysis, and then trace the free energy changes through the formation of G-3-P.
- Describe and explain the major events of the second stage of glycolysis and trace the free energy changes through the formation of pyruvate and lactic acid.
- State the role of *redox reactions* in glycolysis and fermentation.
- Compare and contrast glucose (i.e., carbohydrates in general), ATP, NADH, and FADH₂ as *high-energy* molecules.
- Explain why only a few cell types in the human body conduct *gluconeogenesis*.
- Explain how gluconeogenesis, an energetically unfavorable pathway, can occur at all.
- Explain why the *Atkins diet* (and similar low carbohydrate/high protein diets) works and speculate on any downsides.
- Explain the *supercatalyst* concept and speculate on why a *supercatalyst* (e.g., the *Krebs cycle*) would have evolved.
- Explain the role of high-energy linkages and electron carriers in the Krebs cycle.
- In terms of energetics and biochemical reactions, compare *phosphate ester linkages* in ATP and GTP with the *thioester linkages* in *acetyl-S-CoA* and *succinyl-S-CoA*.
- Speculate on why the *E. coli* *Krebs cycle* generates GTP rather than ATP molecules.

6.2 Glycolysis: A Key Pathway in Energy Flow through Life

One of the properties of life is that living things require energy. The pathways of energy flow through life are shown in Figure 6.1.

Free Energy Flow Through Life

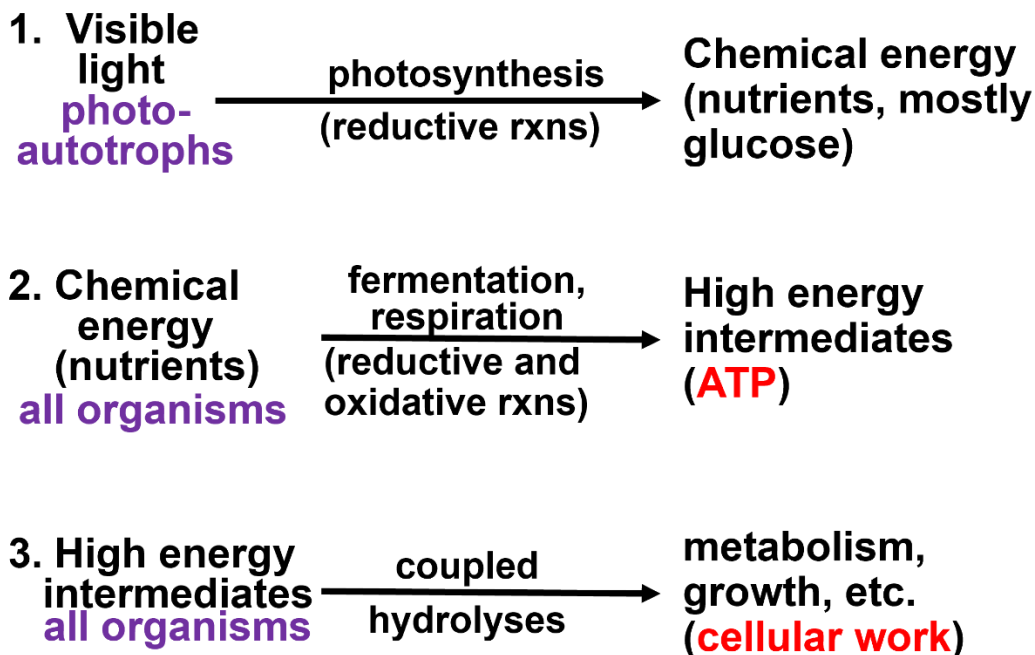


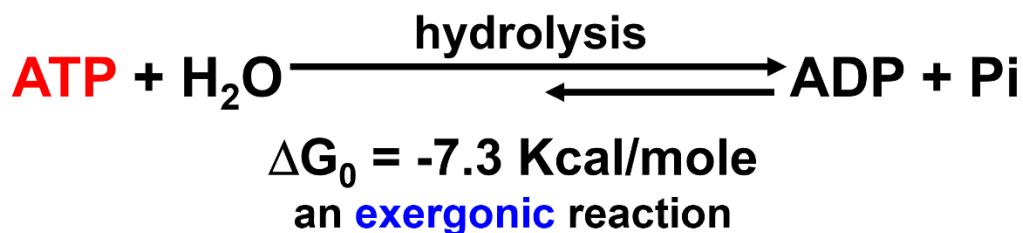
Fig. 6.1: Three paths (top, middle and bottom, respectively) follow of the *flow of free energy* through life, from visible light to chemical energy (e.g., photosynthesis of glucose) to the high-energy intermediates like ATP (fermentation or respiration), and finally to the work of a cell (growth and metabolism).

To begin with, recall that the most common intracellular *energy currency* with which living things “pay” for cellular work is **ATP**. The energy to make ATP on planet Earth ultimately comes from the sun via *photosynthesis*. Recall that light energy fuels the formation of glucose and O₂ from CO₂ and water in green plants, algae, cyanobacteria, and a few other bacteria. Photosynthesis even produces some ATP directly, but not enough to fuel all cellular and organismic growth and metabolism. In fact, all cells, even plant cells, use **fermentation** and/or **respiration** (*anaerobic* or *aerobic* processes, respectively) to capture nutrient free energy (mostly) as ATP.

ATP is called a **high-energy intermediate** because its hydrolysis releases a large amount of free energy. In the condensation reactions that make ATP, it takes about 7.3 Kcal of free energy to link a phosphate to ADP in a *phosphate ester* linkage.

Having captured nutrient free energy in a form that cells can use, ATP hydrolysis then releases that free energy to fuel cellular work. Cellular work includes bending cilia, whipping flagella, contracting muscles, transmitting neural information, building polymers from monomers, and more. The free energy needed to make ATP in animal cells comes exclusively from nutrients (sugars, fats, and proteins). As noted, plants get free energy directly from sunlight, but they mobilize nutrient free energy, which they make in much the same way as the rest of us get it from what we eat! The energetics of ATP hydrolysis and synthesis are summarized in Figure 6.2.

Energetics of **ATP** Hydrolysis and Synthesis



therefore,

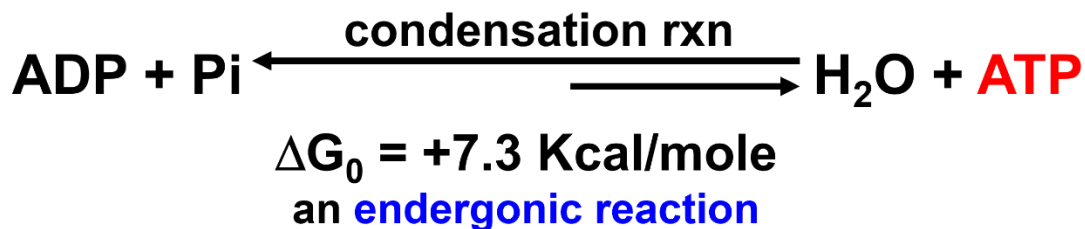
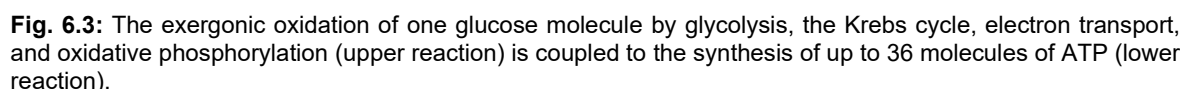


Fig. 6.2: For cells, the flow of free energy involves hydrolyzing ATP (the main chemical energy currency of life on earth) to power the cellular work (upper chemical equation) and harnessing free energy from the environment to make more of it (lower chemical equation)!

Cells Make ATP using nutrient free energy



As we will see, glycolysis is an evolutionarily conserved biochemical fermentation used by all organisms to capture a small amount of nutrient free energy. Check out Fothergill-Gilmore LA [(1986) *The evolution of the glycolytic pathway*. Trends Biochem. Sci. 11:47-51] for more detail. The glycolytic pathway occurs in the cytosol of cells, where it breaks down each molecule of glucose ($C_6H_{12}O_6$) into two molecules of **pyruvic acid (pyruvate)**: $CH_3COCOOH$). This occurs in two stages, capturing nutrient free energy in two ATP molecules per glucose molecule that enters the pathway. Figure 6.4 (below) gives an overview of glycolysis, highlighting its *two stages*.

Stage 1 of glycolysis consumes ATP. Phosphates are transferred first from ATP to glucose and then to fructose-6-phosphate, reactions catalyzed by **hexokinase** and **phosphofructokinase**, respectively. So, these *Stage-1* phosphorylations *consume* free energy. Later, in *Stage 2* of glycolysis, nutrient free energy is captured in ATP and **NADH** (reduced nicotinamide adenine dinucleotide). NADH forms in **redox reactions**, in which **NAD⁺** is reduced as some metabolite is oxidized. In *Stage 2*, it is *glyceraldehyde-3-phosphate* that is oxidized—but more on that later! To summarize, by the end of glycolysis, a single starting glucose molecule has been split into two molecules of **pyruvate** while four molecules of ATP and two molecules of NADH have been produced.

The diagram illustrates the metabolic pathway of glycolysis, starting from glucose and ending with pyruvate. It is divided into two main stages:

- Stage 1:** This stage involves the conversion of glucose into two 3-carbon molecules (3C).
 - Glucose enters the cell from the outside (out) through the cell membrane and moves into the cell (in).
 - Glucose is phosphorylated to glucose-6-phosphate by the enzyme **hexokinase**, consuming **ATP**.
 - Glucose-6-phosphate is converted to fructose-6-phosphate by the enzyme **glucose-6-phosphate isomerase**.
 - Fructose-6-phosphate is phosphorylated to fructose-1,6-diphosphate by the enzyme **phosphofructokinase**, also consuming **ATP**.
 - Fructose-1,6-diphosphate is cleaved by **fructose diphosphate aldolase** into dihydroxyacetone phosphate and glyceraldehyde-3-phosphate (2x).
- Stage 2:** This stage involves the conversion of glyceraldehyde-3-phosphate into pyruvate, producing **ATP** and oxidizing **CH₂O**.
 - Glyceraldehyde-3-phosphate is converted to 1,3-bisphosphoglycerate by **glyceraldehyde phosphate dehydrogenase**. This step is coupled with a **redox reaction** where **2NAD⁺ + 2P_i** are converted to **2NADH**.
 - 1,3-bisphosphoglycerate is converted to 3-phosphoglycerate by **phosphoglycerate kinase**, producing **2ATP**.
 - 3-phosphoglycerate is converted to 2-phosphoglycerate by **phosphoglycerate mutase**.
 - 2-phosphoglycerate is converted to phosphoenolpyruvate by the enzyme **enolase**.
 - Phosphoenolpyruvate is converted to pyruvate by **pyruvate kinase**, producing **2ATP**.

The final products are **2 pyruvate** molecules. The diagram also indicates that **ATP is consumed** in Stage 1 and **ATP is made, CH₂O oxidized** in Stage 2.

Pyruvate will be metabolized either anaerobically or aerobically. The alternate fates of pyruvate are summarized in Figure 6.5.

The diagram illustrates the two main types of cellular respiration:

- Anaerobic metabolism:** This pathway occurs in the **cytoplasm** and involves **fermentation**. It starts with **NADH** being converted back to **NAD⁺**. The end products are **lactate, alcohol, succinate, etc.**
- Aerobic metabolism:** This pathway involves **respiration** and occurs within the **mitochondrion**. The end products are **CO₂, H₂O... in:**

Fig. 6.5: Alternate Fates of Pyruvate: *fermentation*, the *anaerobic reduction* to alcohol or other end-products..., or *respiration*, the *aerobic oxidation*, using oxygen as a final electron acceptor, resulting in H₂O and CO₂ production).



151 Overview of Glycolysis

111



A familiar anaerobic glycolytic pathway is the production of alcohol by yeast in the absence of oxygen. Another is the production of lactic acid by skeletal muscle during strenuous exercise, which leads to the *muscle fatigue* you might have experienced after an especially vigorous workout. Muscle fatigue is due to a buildup of lactic acid in the muscle cells, which under these conditions can't oxidize pyruvate and instead reduces it to lactate. Other cell types produce different fermentative end products, while still capturing free energy in two ATPs per starting glucose.

We will also consider **gluconeogenesis**, a pathway that essentially reverses the glycolysis and results in glucose synthesis. *Gluconeogenesis* may occur under normal conditions as well as during high-protein/low-carb diets and during fasting or starvation. Next, we learn that **cellular respiration**—the aerobic oxidation of pyruvate after **incomplete glycolysis**—takes place in the mitochondria of eukaryotic cells. We'll see the role of the Krebs cycle (also called the **TCA**, or **Tri-Carboxylic Acid** cycle) in the complete oxidation of pyruvate and why the oxidation takes a cycle. We will take a moment to look at the experiments of Hans Krebs that revealed this cycle that starts a **respiratory** pathway that oxidizes glucose to CO_2 and H_2O , leaving no carbohydrates behind. As we look at the reactions of glycolysis and the Krebs cycle, watch for redox reactions in both pathways.

6.3 Some Details of Glycolysis

Here we focus on the enzyme-catalyzed reactions and free energy transfers between pathway components, looking at the energetics and enzymatic features of each reaction.

6.3.1 Glycolysis, Stage 1

Reaction 1: In the first reaction of glycolysis, **hexokinase** rapidly phosphorylates glucose entering the cell, forming **glucose-6-phosphate (G-6-P)**. Figure 6.6 (below) shows that the overall reaction is **exergonic**. The **standard free energy change** for the reaction is -4 Kcal per mole of G-6-P synthesized.

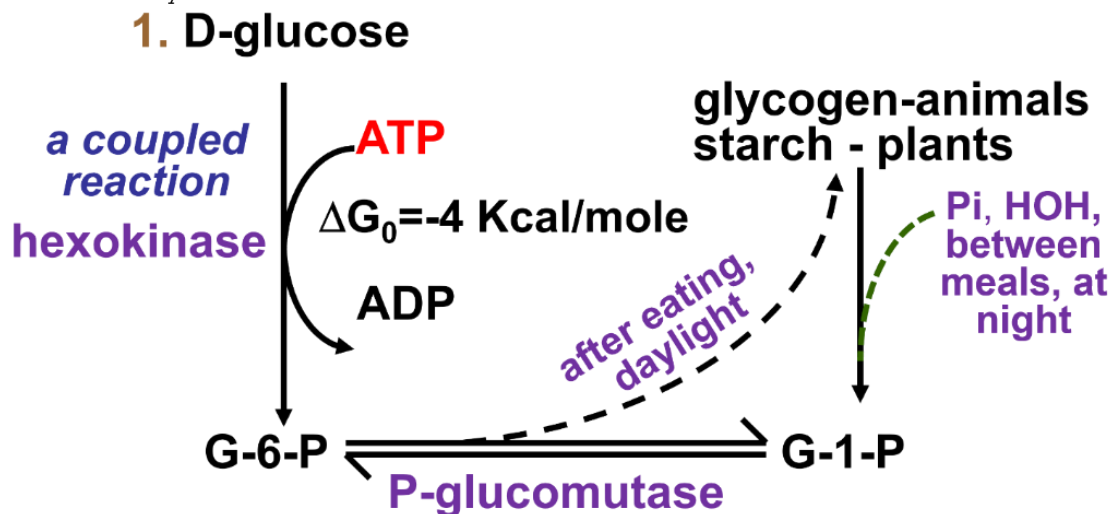


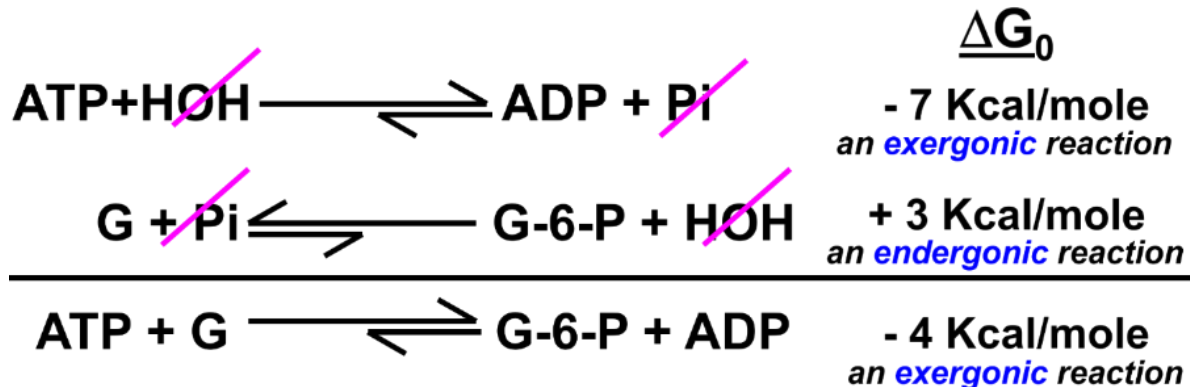
Fig. 6.6: In reaction 1, phosphorylation of glucose to make glucose-6-phosphate (G-6-P) consumes a molecule of ATP. If cellular energy needs are being met, G-6-P will be polymerized to make storage polysaccharides. In animals, G-6-P can be retrieved by glycogen hydrolysis to G-1-P by *glycogen phosphorylase* when the cells require nutrient energy; at that time, the G-6-P will resume glycolysis.

The hexokinase reaction is a *coupled reaction*, in which *phosphorylation* of glucose is coupled to ATP hydrolysis. The free energy of ATP hydrolysis (an energetically favorable reaction) fuels glucose phosphorylation (an energetically *unfavorable* reaction). The reaction is also *biologically irreversible*, as shown by the single vertical arrow.

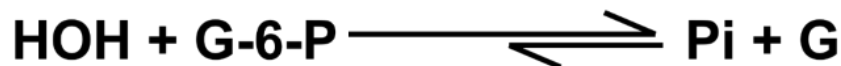
Excess dietary glucose can be stored in most cells (especially liver and kidney cells) as *glycogen*—a highly branched polymer of glucose monomers. In green algae and plants, glucose made by photosynthesis is stored as polymers of starch. When glucose is necessary for energy, glycogen and starch hydrolysis form glucose-1-phosphate (**G-1-P**), which is then converted to **G-6-P**. Let's look at the energetics, i.e., the flow of free energy in the hexokinase-catalyzed reaction. This reaction can be seen as the sum of two reactions in Figure 6.7.

Hexokinase - Energetics

Under standard conditions: **Glucose phosphorylation can be seen as the sum of two reactions:**



What is ΔG_0 for:



G-6-P phosphatase

Fig. 6.7: Free energy flow (exchange) for reaction 1 of glycolysis, and for the removal (hydrolysis) of the phosphate from G-6-P. The hexokinase reaction is the sum of 2 reactions in which the exergonic hydrolysis of ATP is coupled to the endergonic phosphorylation of glucose to G-6-P. What do you think the ΔG_0 is for the hydrolysis of G-6-P, shown in the reaction at the bottom of the slide? Can you figure out the standard free energy change for the hydrolysis of G-6-P?

Recall that ATP hydrolysis is an *exergonic reaction*, releasing about 7 (~7) Kcal/mol (rounding down!) in a closed system under standard conditions. The condensation reaction of glucose phosphorylation occurs with a ΔG_0 of +3 Kcal/mol. Under standard conditions, this is an *endergonic* reaction. Summing up the free energy changes of the two reactions, we can calculate the overall ΔG_0 of -4 Kcal/mol for the coupled reaction under standard conditions in a closed system.

The reactions in Figure 6.7 are written as if they are reversible. However, we said that the overall coupled reaction is *biologically irreversible*. Why the contradiction? To explain, we say that an enzyme-catalyzed reaction is biologically irreversible when its products have a relatively low affinity for the enzyme's active site, making catalysis of the reverse reaction very inefficient. While enzymes catalyzing biologically irreversible reactions don't facilitate the return of products back into reactants, they are often allosterically regulated. This is the case for hexokinase. Imagine a cell that slows its consumption of G-6-P because its energy needs are being met. As a result, cellular G-6-P levels rise. As you might expect, the hexokinase reaction slows down so that the cell doesn't unnecessarily consume a precious nutrient free energy resource. The *allosteric regulation* of hexokinase is shown below in Figure 6.8.

Hexokinase - *Enzymatics*:

- ***Biologically irreversible*: enzyme can't readily catalyze reverse reaction.**

Value to the organism? once in cell, G can't leave;
glucose transporter doesn't recognize G-6-P

- ***Allosteric regulation* by G-6-P (inhibition)**

Value to the organism? cells keep what they need,
share what they don't

Fig. 6.8: Enzymatics of the hexokinase reaction: this enzyme catalyzes a *biologically irreversible* reaction and is allosterically regulated. This ensures that glucose entering a cell cannot leave it as G-6-P. Hexokinase is one of several 'allosterically regulated' glycolytic enzymes whose regulation allows some cells to slow the reaction and divert glucose to other cells.

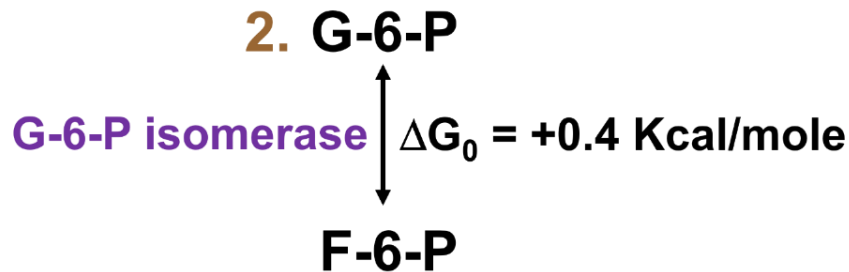
As G-6-P concentrations rise in the cell, excess G-6-P binds to an allosteric site on hexokinase. The resulting conformational change in the enzyme is then transferred to the active site, inhibiting the glucose phosphorylation reaction. The inhibition is reversible: when G-6-P levels decline in the cell, it comes off of the enzyme, the allosteric change is reversed, and uninhibited reaction rates resume.



[152-2 Glycolysis Stage 1, Reaction 1](#)



Reaction 2: In this slightly endergonic and reversible reaction, *isomerase* catalyzes the isomerization of **G-6-P** to *fructose-6-P* (**F-6-P**). The reaction is shown below in Figure 6.9.



Energetics: mildly endergonic,
freely reversible, unregulated

Enzymatics: isomerases catalyze
isomer interconversions.

Fig. 6.9: Reaction 2 of glycolysis is the isomerization of G-6-P to F-6-P by an isomerase (fructose is a structural isomer of glucose). The reaction is endergonic and reversible.

Reaction 3: In this biologically irreversible reaction, 6-P-fructokinase catalyzes the phosphorylation of F-6-P to make *fructose 1,6-diphosphate* (F1,6-diP). In this *coupled reaction*, ATP again provides the second phosphate. The overall reaction is written as the sum of two reactions in Figure 6.10.

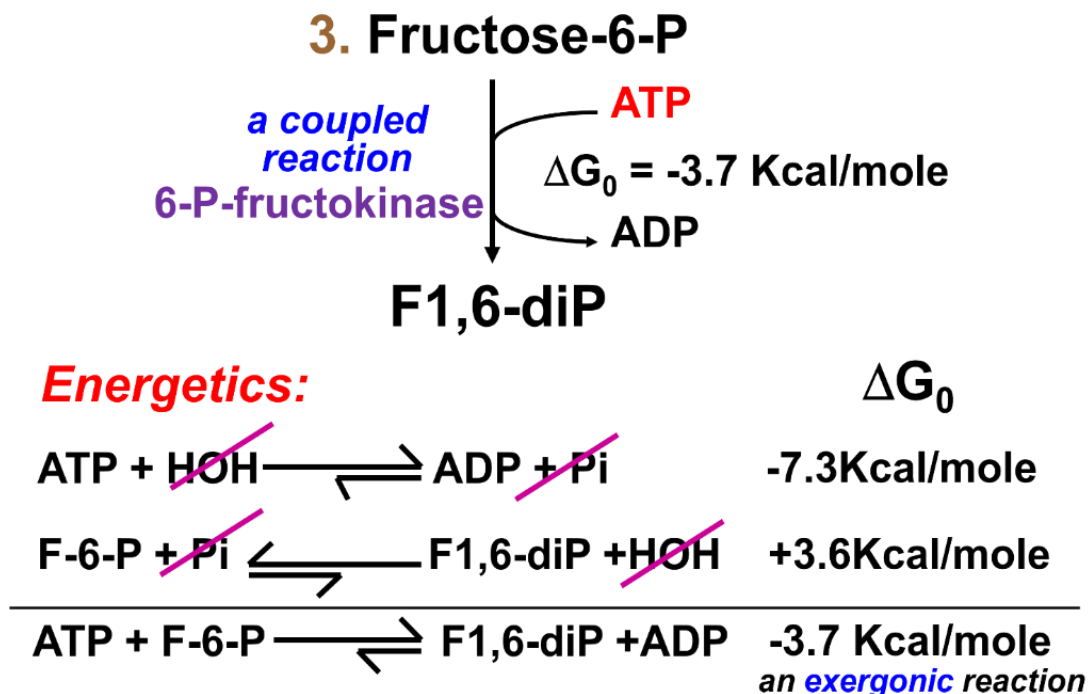


Fig. 6.10: In reaction 3 of glycolysis, a kinase catalyzes the phosphorylation of F-6-P to F1,6-diP in a biologically irreversible reaction, consuming molecule of ATP.

Like the hexokinase reaction, the 6-P-fructokinase reaction is a coupled, exergonic, and allosterically regulated reaction. Multiple **allosteric effectors**, including ATP, ADP, AMP, and long-chain fatty acids, regulate this enzyme.

Reactions 4 and 5: These are the last two reactions of the first stage of glycolysis. In *reaction 4*, F1,6-diP (still a 6-C sugar) is reversibly split into *dihydroxyacetone phosphate (DHAP)* and *glyceraldehyde-3-phosphate (G-3-P)*, each 3-C molecules. In *reaction 5* (also reversible), DHAP is converted into another G-3-P. Both reactions are summarized below in Figure 6.11.

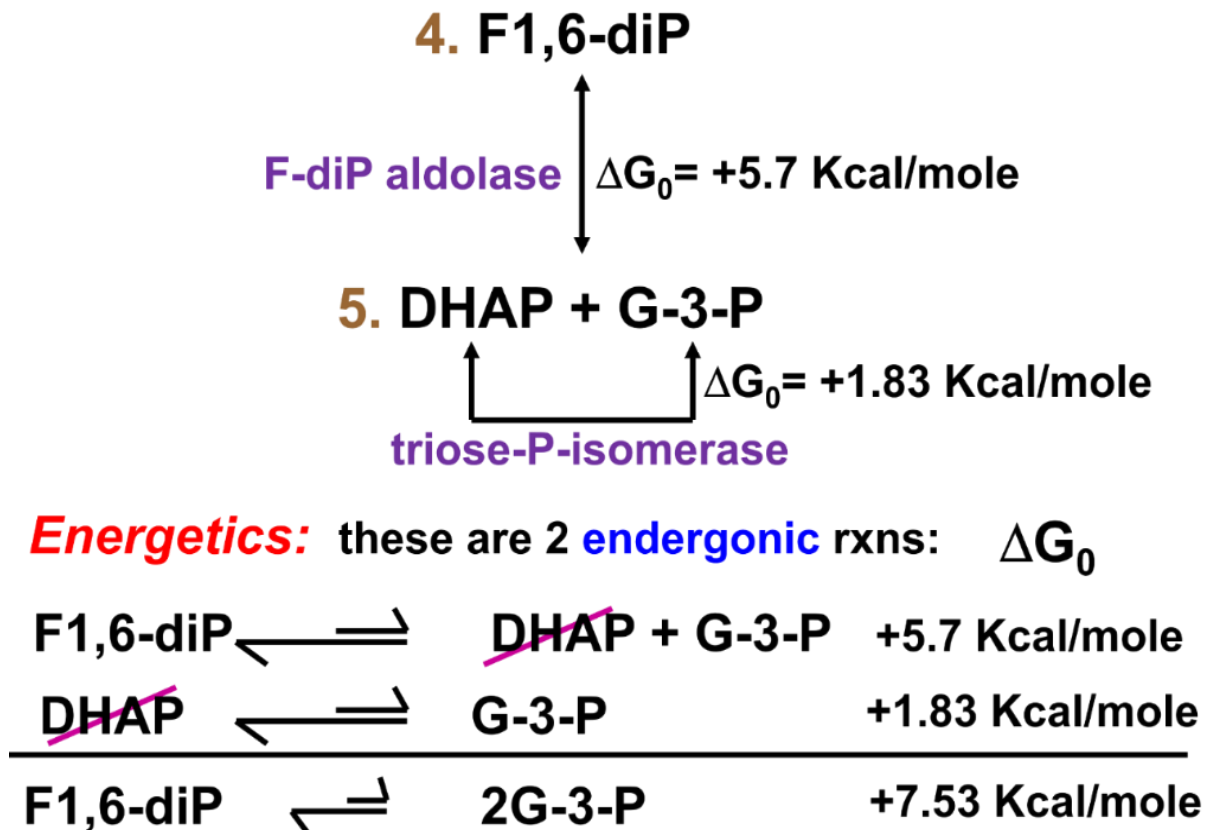


Fig. 6.11: In reaction 4, F1,6-diP is split into glyceraldehyde-3-P (G-3-P) and dihydroxyacetone phosphate (DHAP) by an aldolase. In reaction 5, DHAP is isomerized to G-3-P. Both reactions are endergonic, consuming free energy. These two reactions mark the end of Stage 1 of glycolysis.

The net result is the formation of two molecules of G-3-P at the end of the reactions of *Stage 1*. The enzymes *F-diP aldolase* and *triose-P-isomerase* both catalyze freely reversible reactions. Also, both reactions proceed with a positive free energy change and are therefore *endergonic*. The sum of the free energy changes for splitting F1,6-diP into two G-3-Ps is a whopping +7.5 Kcal/mol, a very energetically unfavorable process.

In summary, by the end of *Stage 1* of glycolysis, two ATP molecules have been consumed, and one 6-C carbohydrate has been split into two 3-C carbohydrates. We have also seen two biologically irreversible and allosterically regulated enzymes.



[153 Glycolysis Stage 1, Reactions 2-5](#)

6.3.2 Glycolysis, Stage 2

We will follow just one of the two molecules of G-3-P generated by the end of *Stage 1* of glycolysis but remember that both proceed through *Stage 2* (the remainder) of glycolysis.

Reaction 6: In this redox reaction, G-3-P is oxidized to *1,3-diphosphoglyceric acid (1,3-diPG)*, and NAD^+ is reduced to NADH. The reaction catalyzed by the *G-3-P dehydrogenase* enzyme, as shown in Figure 6.12.

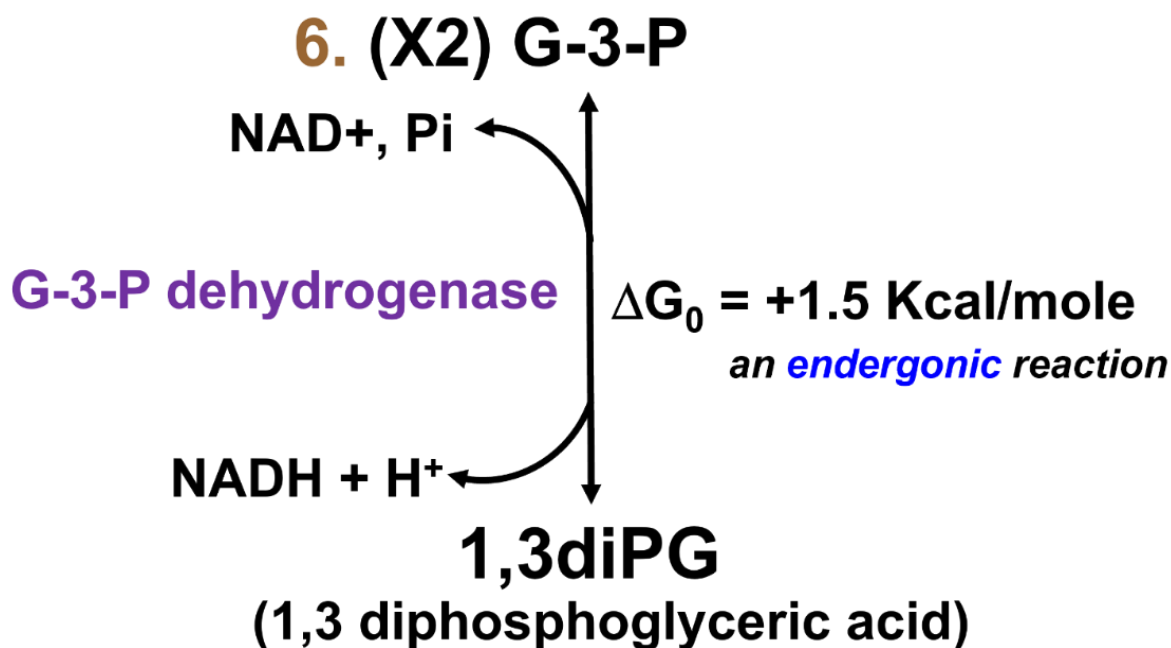


Fig. 6.12: Reaction 6, a redox reaction catalyzed by G-3-P dehydrogenase, is freely reversible. hydride (H^-) ions are transferred to NAD^+ from G-3-P to make NADH and 1,3 diphosphoglyceric acid (1,3diPG).

This *freely reversible endergonic* redox reaction, removes a hydrogen molecule (H_2) from G-3-P, leaving behind phosphoglyceric acid. This short-lived oxidation intermediate is phosphorylated to make *1,3 diphosphoglyceric acid (1,3diPG)*. At the same time, the hydrogen molecule is split into a hydride ion (H^-) and a proton (H^+). The H^- ions reduce NAD^+ to **NADH**, leaving the protons behind in solution. Remember that all of this is happening in the active site of the same enzyme, and remind yourself of what is oxidized and what is reduced here!

Even though it catalyzes a reversible reaction, *G-3-P dehydrogenase* is allosterically regulated. However, in contrast to the regulation of hexokinase, that of G-3-P dehydrogenase is more complicated! The regulator is NAD^+ , and the mechanism of allosteric regulation is called **negative cooperativity**. It turns out that the higher the $[\text{NAD}^+]$ in the cell, the lower the affinity of the enzyme for more NAD^+ and the faster the reaction in the cell! The mechanism is discussed at the following link.



[154 Glycolysis Stage 2, Reaction 6](#)



Reaction 7: This reaction, catalyzed by *phosphoglycerate kinase*, is freely reversible and exergonic (Figure 6.13), yielding ATP and 3-phosphoglyceric acid (3PG).

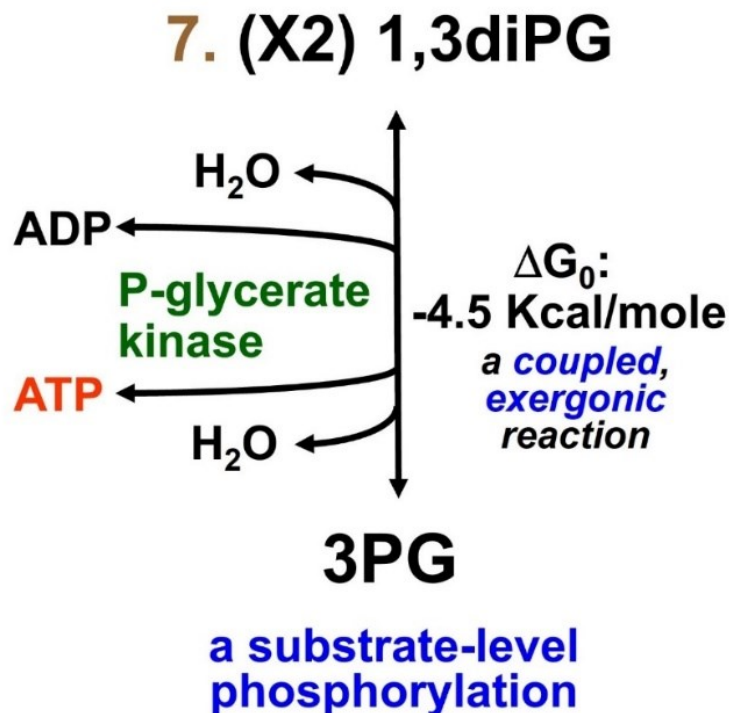


Fig. 6.13: Reaction 7—a reversible, *exergonic, coupled* reaction generates ATP using free energy released by the hydrolysis of one of the phosphates from 1,3diPG. The mechanism of ATP synthesis here is called *substrate-level phosphorylation*.

In glycolysis, catalysis of phosphate group transfer between molecules by **kinases** is called ***substrate-level phosphorylation***, one of the ways of phosphorylating ADP to make ATP. In this *coupled reaction*, the free energy released by hydrolyzing a phosphate from 1,3diPG is used to make ATP. Remember that this reaction occurs twice per starting molecule of glucose, so that two ATPs have been synthesized during this reaction of glycolysis. We call 1,3diPG a ***very high-energy phosphate compound***.

Reaction 8: This freely reversible endergonic reaction transfers the phosphate from the number 3-C of 3PG to the number 2-C (Figure 6.14). *Mutases* like *phosphoglyceromutase* are enzymes that catalyze the transfer of functional groups within a molecule.

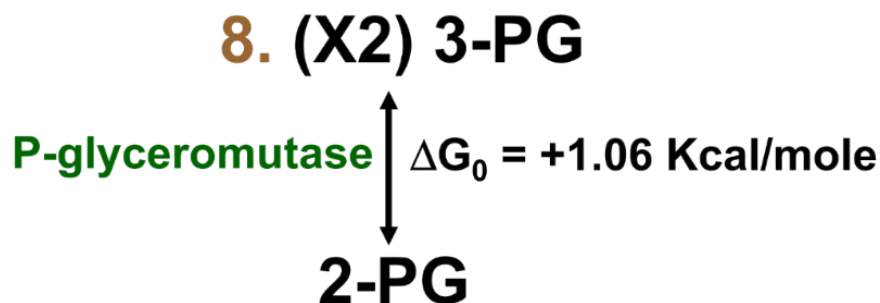


Fig. 6.14: Reaction 8 is a reversible, *endergonic* reaction catalyzed by a mutase. Mutases catalyze the transfer of a chemical group from one part of a molecule to another. Here, 3-PG is converted to 2-PG.

Reaction 9: In this reaction (Figure 6.15), *enolase* catalyzes the conversion of 2-PG to phosphoenolpyruvate (PEP).

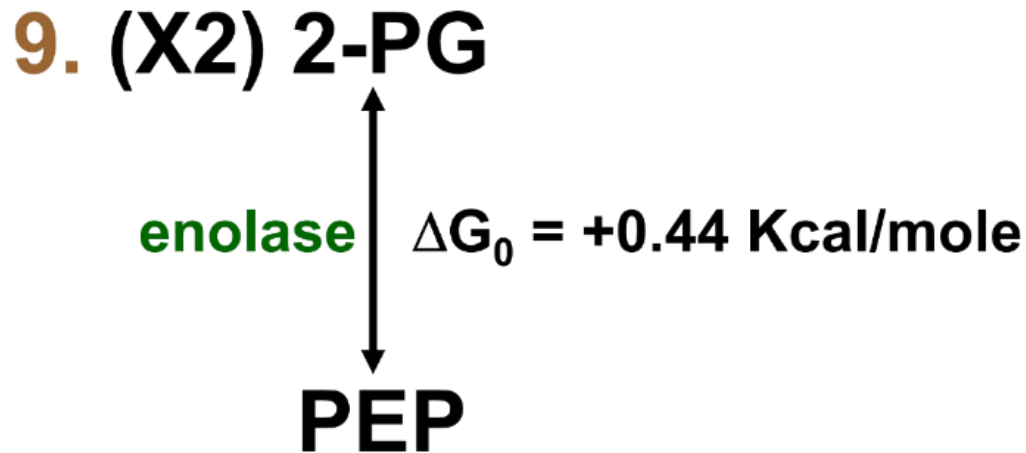


Fig. 6.15: Reaction 9 is a reversible *endergonic* reaction. An *enolase* catalyzes the conversion of 2-PG to phosphoenolpyruvate (PEP).

Reaction 10: This reaction results in the formation of a molecule of *pyruvic acid* (*pyruvate*), illustrated in Figure 6.16.

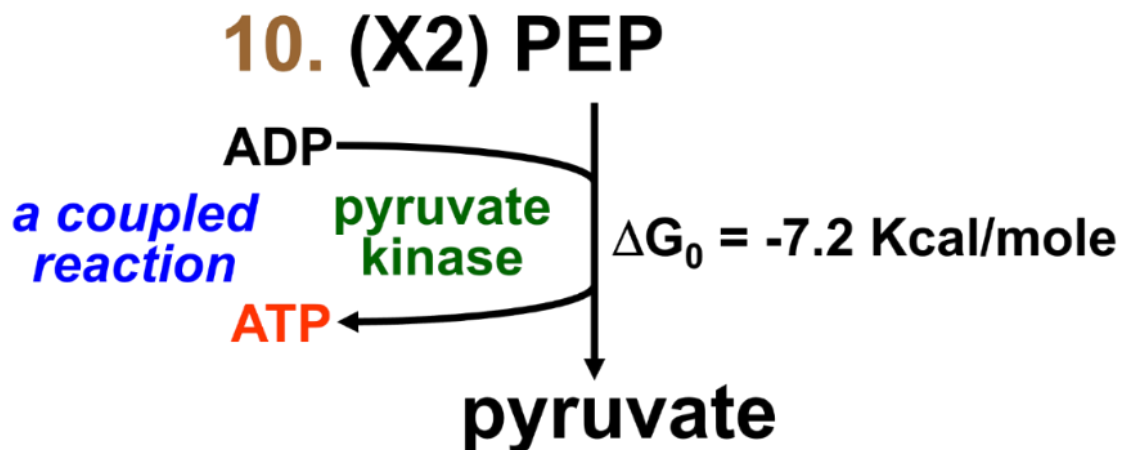


Fig. 6.16: Reaction 10 is biologically irreversible. In a coupled reaction, pyruvate kinase catalyzes a highly exergonic transfer of a phosphate on PEP to ADP to make ATP.

Remember again that two pyruvates are produced per starting glucose molecule. The enzyme *pyruvate kinase* couples the *biologically irreversible*, exergonic hydrolysis of a phosphate from PEP and the transfer of that phosphate to ADP in a *coupled reaction*. The reaction produces PEP, another *very high-energy* phosphate compound. Pyruvate kinase is allosterically regulated by increases in cellular levels of ATP, citric acid, long-chain fatty acids, F1,6-diP—or even PEP, one of its own substrates.



[155-2 Glycolysis Stage 2, Reactions 7–10](#)



As we have seen, there are alternate fates of pyruvate, the product of incomplete glycolysis. One is the aerobic mitochondrial oxidation of pyruvate, following *incomplete glycolysis*. The other is an anaerobic **fermentation**, or *complete glycolysis*, in which pyruvate is reduced to one or another end product. Recall that muscle fatigue results when skeletal muscle uses anaerobic fermentation to get the free energy required for vigorous exercise by reducing pyruvate to **lactic acid**. It is the accumulation of lactic acid in skeletal-muscle cells that causes muscle fatigue. The enzyme LDH (*lactate dehydrogenase*) that catalyzes the reduction of pyruvate to lactate is regulated—but not allosterically! Instead, different types of muscle tissues regulate LDH by making different versions of the enzyme. Go to the following link for a more detailed explanation.



[156 Fermentation: Regulation of Pyruvate Reduction Is NOT Allosteric!](#)

6.4 A Chemical and Energy Balance Sheet for Glycolysis

Compare the balance sheets for *complete* glycolysis (fermentation) to lactic acid and for *incomplete* (aerobic) glycolysis, showing chemical products and energy transfers (Figure 6.17).

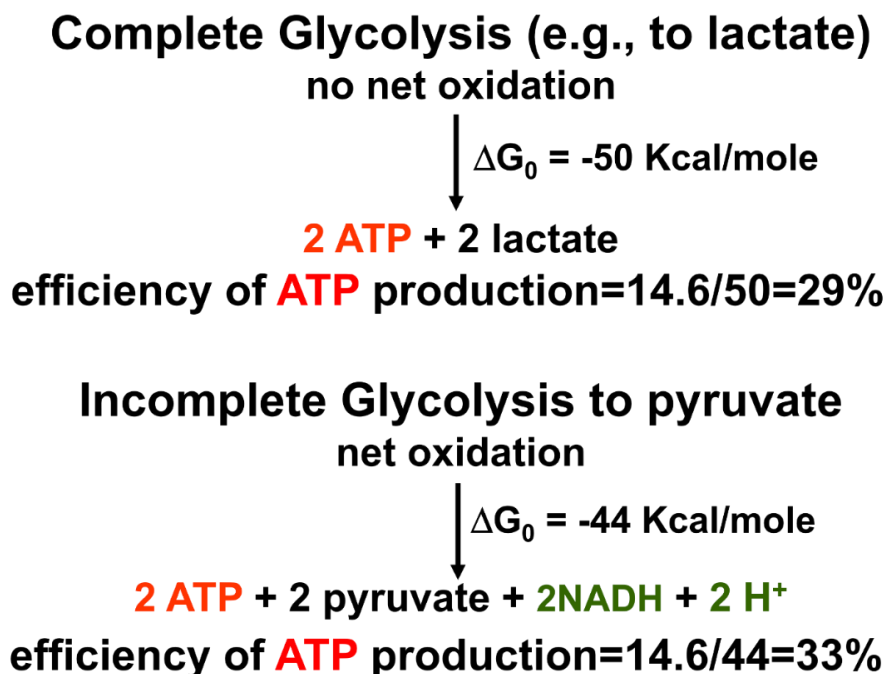


Fig. 6.17: The free energy and ATP yields of complete glycolysis (fermentation) and incomplete glycolysis (respiration) result from net exergonic pathways. The percentages, which represent the efficiency of ATP production, are based on the ratios of the free energy captured as ATP to the free energy released by the different pathways for metabolizing glucose. From the data, incomplete glycolysis is a more efficient way to extract nutrient free energy.

There are two reactions in *Stage 2* of glycolysis that each yield a molecule of ATP, and each occurs twice per starting glucose molecule. Stage 2 of glycolysis thus produces four ATP molecules per glucose. Since *Stage 2* consumed two ATPs, the net yield of chemical energy for

the cell by the end of glycolysis is two ATPs, whether complete to lactate or incomplete to pyruvate! Because anaerobic cells can't make use of oxygen, they have to settle for the measly few (15) kilocalories' worth of ATP that they get from a fermentation. Since there are 687 kilocalories potentially available from the complete combustion of a mole of glucose, there is a lot of nutrient free energy left on the table, to be captured during the rest of respiration (i.e., pyruvate oxidation).



[157-2 Balance Sheet of Glycolysis](#)



Remember also that the only redox reaction in aerobic glycolysis is in Stage 2. This is the oxidation of G-3-P, a 3-C glycolytic intermediate. Now check out the redox reaction of a fermentation pathway. Since pyruvate, also a 3-C intermediate, was reduced, there has been *no net oxidation of glucose* (i.e., glycolytic intermediates) in complete glycolysis.

By this time, you will have realized that glycolysis is a *net* energetically favorable (*downhill, spontaneous*) pathway in a closed system, with an overall negative ΔG_o . Under normal circumstances, glycolysis is also spontaneous in most of our cells, driven of course, by a constant need for energy to do cellular work. Thus, you would expect that the actual free energy, or $\Delta G'$, of glycolysis, is also normally negative. In fact, glycolysis in actively respiring cells proceeds with a release of more free energy than it would in a closed system. In other words, the $\Delta G'$ for glycolysis in active cells is more negative than the ΔG_o of glycolysis! Feel free to investigate the truth of this statement on your own.

Before we discuss the aerobic fate of pyruvate, let's take look at ***gluconeogenesis***, the Atkins diet, and at not-so-normal circumstances when glycolysis essentially goes in reverse (at least in a few cell types). Under these conditions, glycolysis is energetically unfavorable, and even the otherwise-exergonic reactions of glycolysis (those with a negative ΔG_o) will proceed with a negative $\Delta G'$.

6.5 Gluconeogenesis

In well-fed animals, most cells can store some of glucose as glycogen, which they break down as needed to retrieve nutrient energy as G-6-P. Glycogen hydrolysis (***glycogenolysis***) produces G-1-P, which is converted to G-6-P as we saw at the beginning of *Stage 1* of glycolysis. But glycogen in most cells is quickly used up between meals. Therefore, most cells depend on a different, external source of carbohydrates (i.e., glucose) other than diet. Those sources are liver (and to a lesser extent kidney) that can store large amounts of glycogen after meals. In continual feeders (cows and other ruminants), glycogenolysis is ongoing.

Glycogenolysis by liver cells supplies glucose to the blood for up to six to eight hours between meals in *intermittent feeders* (like us), to be distributed as needed to all cells of the body. Thus, you can expect to use up liver and kidney glycogen reserves after a good night's sleep, a period of intense exercise, or any prolonged period of low carbohydrate intake (fasting or starvation). Under these circumstances, animals use ***gluconeogenesis*** (literally, *new glucose synthesis*) in liver and kidney cells to provide systemic glucose to nourish other cells.

In healthy individuals, the hormones glucagon and insulin regulate blood *glucose homeostasis*, depending on cellular glucose (energy) needs. These hormones raise or lower blood glucose levels to protect the organism from *hypoglycemia* (low blood sugar) and *hyperglycemia* (high blood sugar), respectively. The gluconeogenic pathway produces glucose from carbohydrate and non-carbohydrate precursors that include pyruvate, lactate, glycerol, and *gluconeogenic amino acids*. The latter are amino acids that can be converted to alanine. The reactions of glycolysis and gluconeogenesis are shown side-by-side in Figure 6.18.

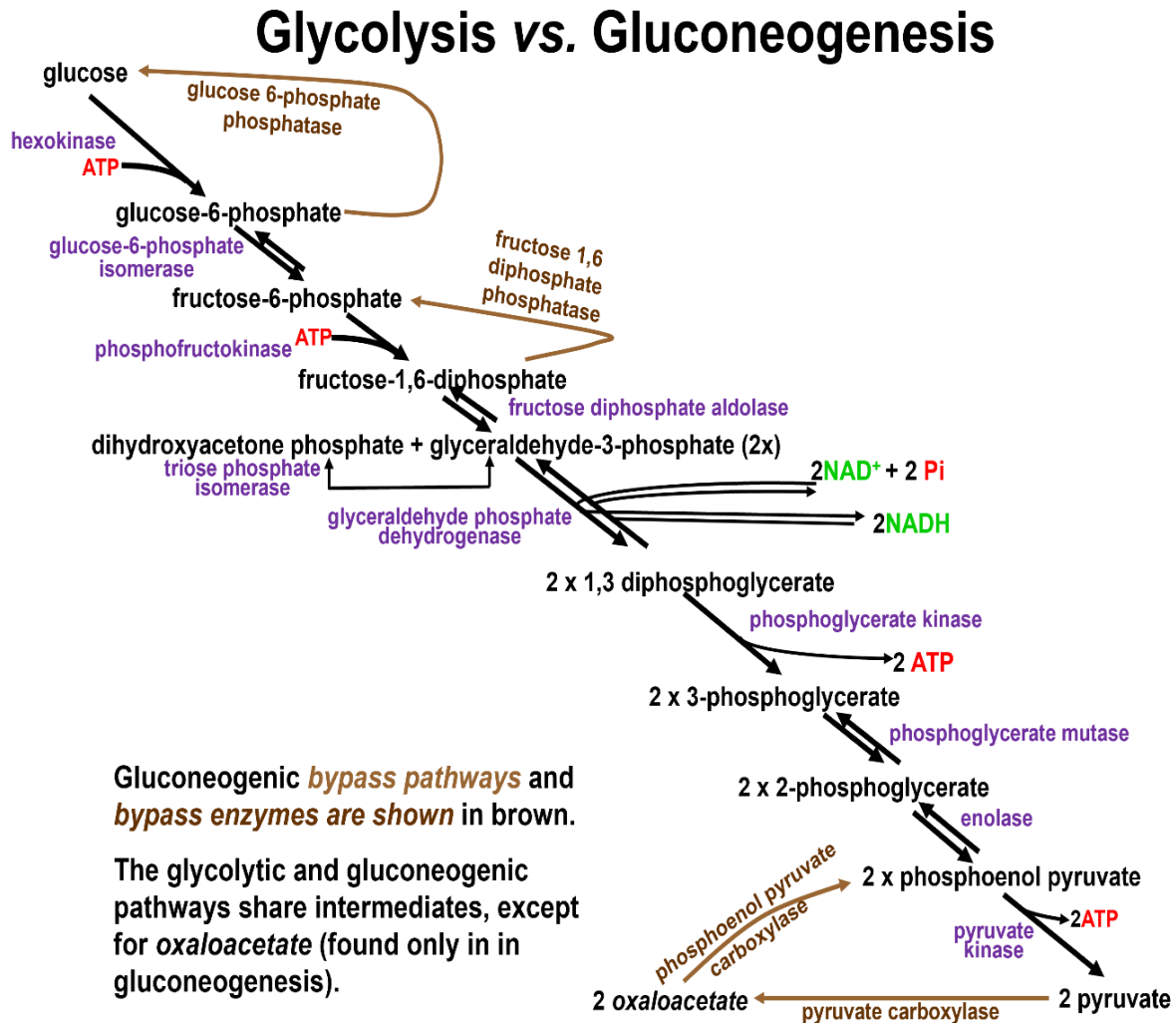


Fig. 6.18: Comparison of incomplete glycolysis to gluconeogenesis. The two pathways are essentially the reverse of one another, except for the *bypass enzymes* in gluconeogenesis (shown in green), which are required to get around biologically irreversible enzymes of glycolysis.

In Figure 6.18, look for *bypass* reactions that are catalyzed by *carboxylases* and *phosphatases* and at the glycolytic reactions that function in reverse during gluconeogenesis. Glycolysis overall is an exergonic pathway; therefore, gluconeogenesis must be an endergonic pathway. In fact, the free energy released by glycolysis generates a net of two ATPs from two pyruvates, while synthesizing glucose from two pyruvates during gluconeogenesis, costs four ATPs and two GTPs!

Likewise, gluconeogenesis is only possible if the bypass enzymes are present. These are necessary to get around the three biologically irreversible reactions of glycolysis. Except for the *bypass reactions*, gluconeogenesis is essentially a reversal of glycolysis. As drawn in the pathways in Figure 6.18, glycolysis and gluconeogenesis would seem to be cyclic. In fact, this apparent cycle was recognized by Carl and Gerty Cori, who shared the 1947 Nobel Prize for Medicine or Physiology with Bernardo Houssay for discovering how glycogen is broken down in muscle cells (and in fact in most cells) to pyruvate to be reconverted to glucose in liver cells. Named after the Coris, The **Cori cycle** (Figure 6.19, below) recognizes the interdependence of liver and muscle in glucose breakdown and re-synthesis.

The Cori Cycle Connecting Glycolysis and Gluconeogenesis

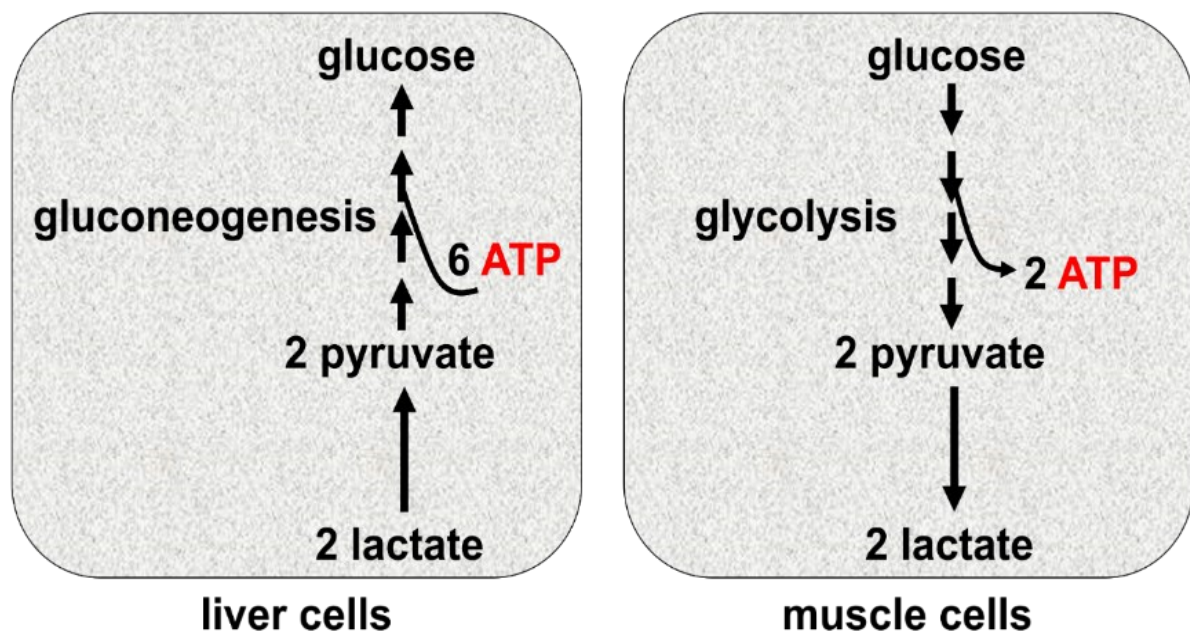


Fig. 6.19: The *Cori cycle* reveals the relationship between glycolysis and gluconeogenesis. Lactic acid produced by complete glycolysis in active skeletal muscle goes to the liver, where it could be converted to pyruvate and then to glucose.

Despite the free energy requirement of gluconeogenic reactions in a closed system under standard conditions, gluconeogenesis in liver and kidney cells is energetically favorable..., because cells are open systems. Pyruvate accumulation in liver cells and a rapid release of new glucose into the blood drive the gluconeogenesis forward, synthesizing glucose with a negative $\Delta G'$, a decline in actual free energy. Of course, glycolysis and gluconeogenesis are not simultaneous. Which pathways operate in which cells is tightly controlled. Glycolysis is the norm in all cell types, even liver and kidney. However, the cessation of glycolysis in favor of gluconeogenesis in the latter cells is under hormonal control (Figure 6.20, below).

Key in turning on liver gluconeogenesis is the role of glucocorticoid hormones. What causes the secretion of glucocorticoids? A long night's sleep, fasting and in the extreme, and starvation are forms of *stress*. Stress responses start in the *hypothalamic-pituitary axis*.

Different stressors cause the *hypothalamus* to secrete a *neurohormone* which stimulates the release of *ACTH* (*adrenocorticotrophic hormone*) from the *pituitary gland*.

ACTH then stimulates the release of cortisone and other glucocorticoids from the cortex (outer layer) of the adrenal glands. As the name glucocorticoid suggests, these hormones participate in the regulation of glucose metabolism.

Hormonal Control of Gluconeogenesis

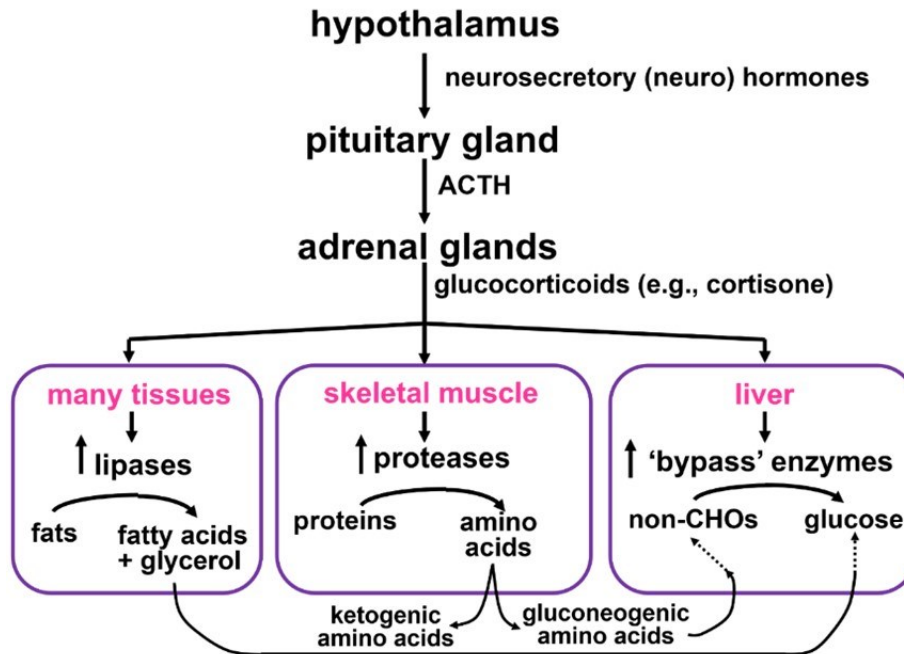


Fig. 6.20: Hormones of the *hypothalamic-pituitary axis* control gluconeogenesis. A need for glucose-derived energy stimulates hypothalamic hormones, which in turn stimulate the pituitary gland to release ACTH (adrenocorticotrophic hormone). ACTH then stimulates the adrenal glands to release glucocorticoid hormones stimulate many tissues to control gluconeogenic glucose production, as well as the use of alternate nutrient fuels by many cells in the body.

Here is what happens at times of low blood sugar (e.g., when carbohydrate intake is low):

1. Glucocorticoids stimulate the synthesis of gluconeogenic bypass enzymes in liver cells.
2. Glucocorticoids stimulate *protease* synthesis in skeletal muscle, causing hydrolysis of the peptide bonds between amino acids. Gluconeogenic amino acids circulate to the liver where they are converted to pyruvate, a major precursor of gluconeogenesis. Some amino acids are ketogenic and are converted to Acetyl-S-CoA, a precursor to *ketone bodies*.
3. Glucocorticoids stimulate increased levels of enzymes including *lipases* which catalyze hydrolysis of the ester linkages in triglycerides (fat) in adipose and other cells. This generates *fatty acids* and *glycerol*.
4. Glycerol circulates to liver cells where it is taken up and converted into to G-3-P, augmenting gluconeogenesis. Fatty acids circulate to liver cells where they are oxidized to Acetyl-S-CoA that is then converted to ketone bodies. and released to the circulation.
5. Most cells switch from glycolysis to fatty acid oxidation an alternate energy source when glucose is limiting. Heart and brain cells depend on glucose for energy, but under extreme conditions (prolonged fasting, starvation), brain cells can use ketone bodies as an energy source of last resort.

Thus, the essential roles of glucocorticoids include the following:

1. Enabling most cells to oxidize fats (fatty acids) for energy
2. Allowing brain cells to use gluconeogenic glucose for energy, and in the extreme, ketone bodies as an alternate energy source
3. Allowing cardiac muscle to use gluconeogenic glucose as its energy source

It's a pity that we humans can't use fatty acids as gluconeogenic substrates! Plants and some lower animals have a *glyoxylate cycle* pathway that can convert fatty-acid oxidation products directly into gluconeogenic carbohydrate substrates. Lacking this pathway, we (and higher animals in general) cannot convert fats to carbohydrates, in spite of the fact that we can all too easily convert the latter to the former! For us, when the gluconeogenic response is inadequate to the task, the body can resort to ketogenic fat metabolism. Think of this as a *last resort*, leading to the production of ketone bodies and the "acetone breath" in long-term fasting or people with severe eating disorders (e.g., *anorexia nervosa*).

The dark side of the limits of gluconeogenic metabolism is prolonged starvation, which will eventually overwhelm the gluconeogenic response. You see this in reports from third-world regions where many people may suffer starvation due to drought, other natural disasters, or war. The spindly arms and legs of starving children result from muscle wasting as the body tries to provide the glucose necessary for survival.

6.6 The Atkins Diet and Gluconeogenesis

You may know the *Atkins Diet* as an ultra-low-carb diet, one of several *ketogenic diets*. Glucocorticoid hormones released on such a diet trick the body into maintaining a constant gluconeogenic state. Glucose production by the gluconeogenic liver is limited and reserved for brain and heart cells; the rest of the cells in our bodies switch to burning fats, hence the weight loss. Carried to an extreme, restriction of carbohydrate intake results in high blood levels of ketones and the *acetone breath* also encountered during extreme malnutrition. Discredited some years ago, the Atkins and similar 'ketogenic' diets (e.g., Paleo, South Beach) are now back in favor.

Low-carb diets are important in the control of any diabetic disease. Older folks with type-2 (adult-onset) diabetes can control their disease with a low-carb diet, with exercise, and with medication. For example, metformin is a drug that blocks gluconeogenesis to reduce glucose synthesis from gluconeogenic substrates, at the same time as it stimulates cellular receptors to take up available glucose. Look at Hundal RS et al. (2000. *Mechanism by Which Metformin Reduces Glucose Production in Type 2 Diabetes*. Diabetes 49: 2063–9) for more details on how *metformin* works. Given the prevalence of adult obesity and type 2 diabetes in the United States, it's likely that someone you know is taking *metformin* or another similar medication! Unfortunately, a high rate of obesity in children may lead to type 2 diabetes in youngsters in the U.S. Of course, the treatment of last resort insulin, the pancreatic hormone that increases cellular glucose uptake.



6.7 The Krebs/TCA/Citric acid cycle

The glycolytic reactions through fermentative reactions to pyruvate produce ATP anaerobically. The evolution of respiration (the aerobic use of oxygen to efficiently burn nutrient fuels) had to wait until photosynthesis created the oxygenic atmosphere we live in now. Read more about the source of our oxygenic atmosphere in Dismukes GC et al. [(2001) *The origin of atmospheric oxygen on Earth: the innovation of oxygenic photosynthesis*. Proc. Nat. Acad. Sci. USA 98:2170-2175].

The **Krebs cycle** is the earliest pathway of eukaryotic mitochondrial oxygenic respiration. The Krebs cycle no doubt evolved a few reactions at a time, perhaps in the beginning, as a means of protecting anaerobic cells from the "poisonous" effects of oxygen as photosynthetic organisms spread. Along the way, natural selection elaborated the aerobic Krebs cycle, electron transport and oxidative phosphorylation metabolic pathways we see today. Whatever their initial utility, these reactions were an adaptive response to the increase in oxygen in the Earth's atmosphere. As we've seen, respiration is a much more efficient pathway than glycolysis is for extracting chemical energy from nutrients. Animals rely on it, but even plants and photosynthetic algae use the respiratory pathway when sunlight is not available! Here we focus on oxidative reactions in the mitochondrion, beginning with pyruvate oxidation and continuing to the redox reactions of the Krebs cycle.

In mitochondria, *pyruvate dehydrogenase* catalyzes pyruvate oxidation to **Acetyl-S-Coenzyme A (Ac-S-CoA)**. The Krebs cycle then completely oxidizes Ac-S-CoA. These mitochondrial redox reactions generate CO₂ and a lot of reduced electron carriers (NADH, FADH₂). The free energy released in these redox reactions is coupled to the synthesis of only one ATP per oxidized pyruvate (i.e., two per the glucose we started with). In fact, it is the NADH and FADH₂ molecules that have captured most of the free energy in the original glucose molecules. The entry of pyruvate from glycolysis into the mitochondrion and its oxidation are summarized in Figure 6.21.

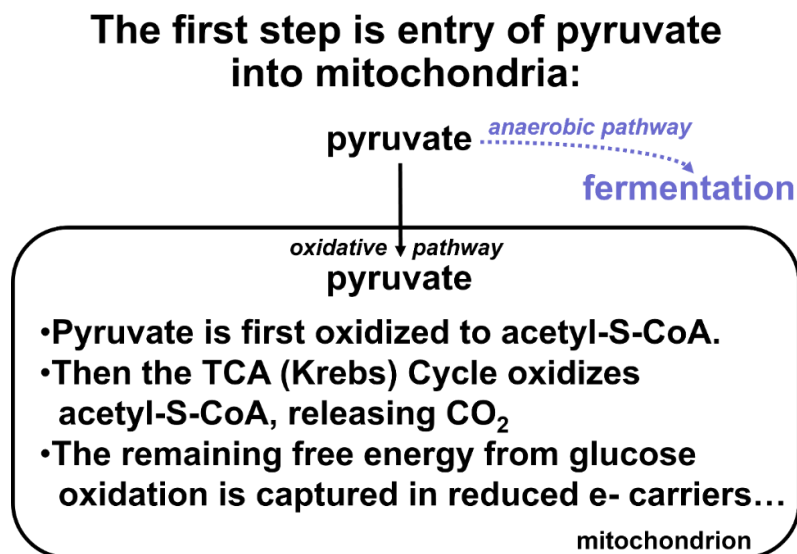
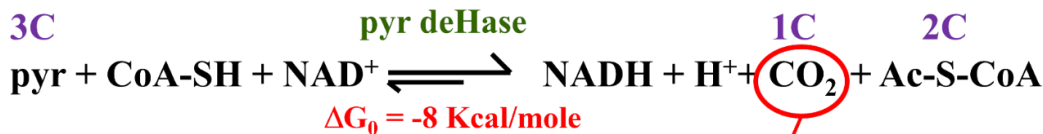


Fig. 6.21: Pyruvate enters the mitochondrion to be oxidized to acetyl-S-Coenzyme A (acetyl-S-CoA).

Pyruvate oxidation by pyruvate dehydrogenase converts a 3-C carbohydrate into a 2-C acetate molecule, releasing a molecule of CO₂. In this highly exergonic reaction, a reduced coenzyme A (CoA-SH) forms a *high-energy thioester* (the *-S- linkage*) with acetate to make **Ac-S-CoA**. In this reaction, NAD⁺ is reduced to NADH a CO₂ molecule is produced, along with the Ac-S-CoA. Figure 6.22 shows the complete redox reaction.



First of the CO₂ molecules resulting from glucose respiration!

Fig. 6.22: Pyruvate dehydrogenase catalyzes pyruvate oxidation to Ac-S-CoA, releasing a molecule of CO₂ and reducing NAD⁺ to NADH.

The **Krebs cycle** (illustrated in Figure 6.23 below) functions during respiration to oxidize Ac-S-CoA and to reduce NAD⁺ and FAD to NADH and FADH₂, respectively.

Overview of the Citric Acid (Krebs) Cycle

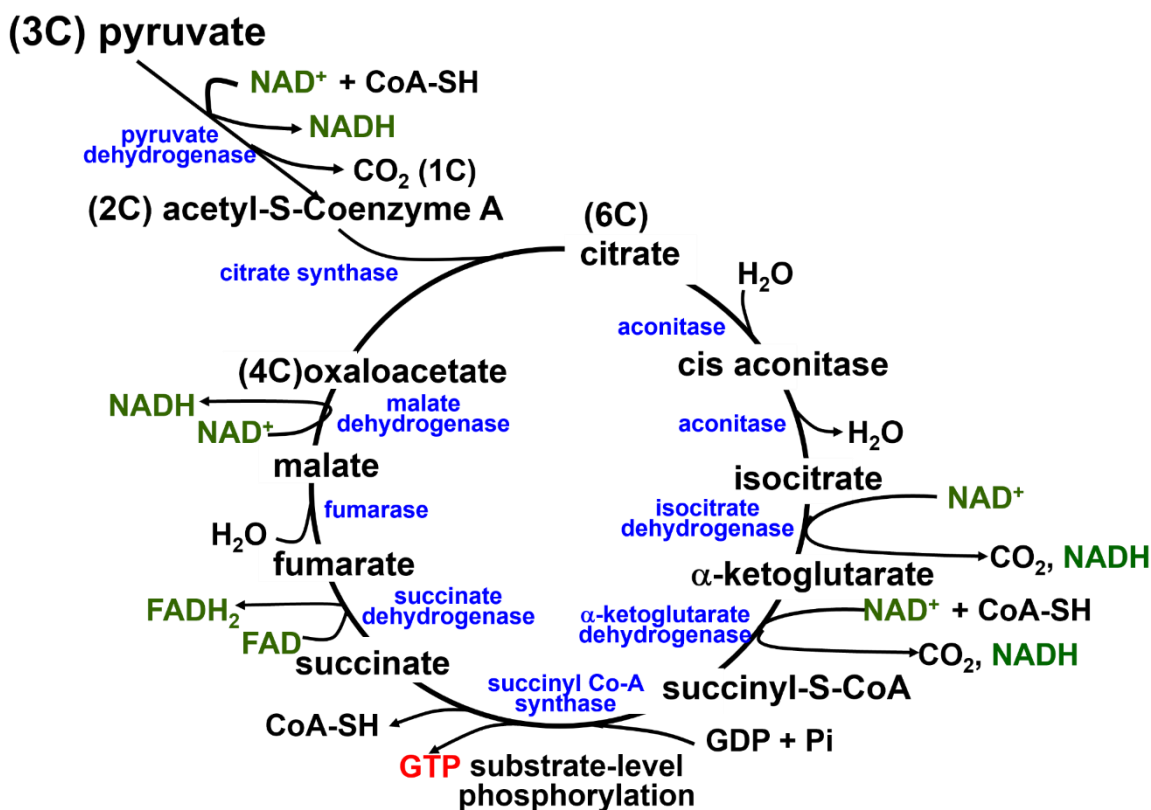


Fig. 6.23: Krebs Cycle highlights: The first reaction is the condensation of Ac-S-CoA and oxaloacetate (OAA), forming citric acid (citrate). Four reactions in the cycle are redox reactions which create reduced electron carriers (NADH, FADH₂); one reaction makes GTP by substrate-level phosphorylation.

Overall, the cycle is exergonic. Intermediates of the Krebs cycle also function in amino acid metabolism and interconversions. All aerobic organisms alive today share the Krebs cycle. This is consistent with its spread early in the evolution of our oxygenic environment. Because of the central role of Krebs cycle intermediates in other biochemical pathways, parts of the pathway may even have predated the complete respiratory pathway. This centrality of the Krebs cycle to cellular metabolism is emphasized in the biochemical pathways chart shown at the top of this chapter.

After the oxidation of pyruvate in mitochondria, the resulting Ac-S-CoA enters the Krebs cycle in the first reaction, condensing with oxaloacetate (OAA) to form citric acid, or citrate. Citrate is a tricarboxylic acid (TCA), for which the cycle was first named. There are four redox reactions in the Krebs cycle. As we discuss the cycle, look for the accumulation of reduced electron carriers (FADH_2 , NADH) and a small amount of ATP synthesis by substrate-level phosphorylation. Also follow the carbons in pyruvate into CO_2 . The following checklist will help you understand the events of the cycle:

1. Find the two molecules of CO_2 produced in the Krebs cycle itself.
2. Find GTP (which quickly transfers its phosphate to ADP to make ATP). Note that in bacteria, ATP is made directly at this step.
3. Count all the reduced electron carriers (NADH , FADH_2) in the biochemical pathway. Each will "carry" a pair of electrons into the mitochondria. If you include the electrons on each of the NADH molecules made in glycolysis, how many electrons have been removed from cytoplasmic glucose for during its complete oxidation in mitochondria?

Remember that glycolysis produces two pyruvates per glucose and thus two molecules of Ac-S-CoA. Thus, the Krebs cycle turns twice per glucose that enters the glycolytic pathway. The *high-energy thioester linkages* formed in the Krebs cycle fuel ATP synthesis as well as the condensation of oxaloacetate and acetate to form citrate in the first reaction. Each NADH carries about 50 Kcal of the 687 Kcal of free energy originally available in a mole of glucose; each FADH_2 carries about 45 Kcal of this free energy. This energy will fuel ATP production during electron transport and oxidative phosphorylation.



[159-2 Highlights of the Krebs Cycle](#)



Finally, the story of the discovery of the Krebs cycle is at least as interesting as the cycle itself! Albert von Szent-Györgyi initially discovered several organic acid oxidation reactions that were thought at the time to be part of a linear pathway of nutrient oxidation, and for which he won a Nobel Prize in 1937. Hans Krebs then performed the elegant experiments showing that the reactions were part of a cyclic pathway. He proposed (correctly!) that the cycle would be a *supercatalyst* that would catalyze the oxidation of yet another organic acid. Some of the experiments are described by Krebs and his coworkers in their classic paper: Krebs HA, et al. [(1938) *The formation of citric and α -ketoglutaric acids in the mammalian body*. Biochem. J. 32: 113–117].

Hans Krebs and Fritz Lipmann shared the 1953 Nobel Prize in Physiology or Medicine. Krebs was recognized for his discovery of the TCA cycle, which more commonly carries his name. Lipmann was recognized for proposing ATP as the mediator between food

(nutrient) energy and intracellular work energy and for discovering the reactions that oxidize pyruvate and synthesize Ac-S-CoA, bridging the Krebs cycle and oxidative phosphorylation (to be considered in the next chapter).



160 Discovery of the Krebs Cycle

You can read Krebs' review of his own research in Krebs HA [(1970) *The history of the tricarboxylic acid cycle*. Perspect. Biol. Med. 14:154-170]. For a classic read on how Krebs described his supercatalyst suggestion, check out ^{6.2}[Citrate and \$\alpha\$ -Ketoglutarate Formation in Mammals](#). For more about the life of Lipmann, check the brief Nobel note on the ^{6.3}[Fritz Lipmann Biography](#).

Some iText & VOP Key Words and Terms

Acetyl-S-coenzyme A (Ac-S-CoA)	free energy	oxidation, reduction
ADP, ATP, GDP, GTP	Free energy capture	phosphatase enzymes
aerobic	fructose	phosphate-ester linkage
anaerobic	G, G6P, F6P, F1,6-diP	redox reactions
Atkins diet	gluconeogenesis	respiration
biochemical pathways	gluconeogenic amino acids	SDH (succinate dehydrogenase)
energetics	glycolysis	spontaneous reaction
bypass reactions, enzymes	glyoxylate cycle	stage 1
C ₆ H ₁₂ O ₆ (glucose)	High-energy bond (linkage)	stage 2
cells as open systems	High-energy molecules	standard conditions
Cori cycle	isomerase enzymes	steady state
dehydrogenase enzymes	kinase enzymes	stoichiometry of glycolysis
DHAP, G3P, 1,3diPG, 3PG, 2PG, PEP, Pyr	Krebs (TCA, citric acid) cycle	substrate-level phosphorylation
diabetes	metabolic effects of low-carb diets	Succinyl-S-CoA
energetics of glycolysis	metformin	supercatalyst
energy flow in cells	mitochondria	synthase enzymes
equilibrium	mutase enzymes	thioester linkage
FAD (oxidized nicotinamide adenine dinucleotide)	NAD ⁺ (oxidized nicotinamide adenine dinucleotide)	$\Delta G'$ (actual free energy change)
FADH ₂ (reduced flavin adenine dinucleotide)	NADH (reduced nicotinamide adenine dinucleotide)	ΔG^o (standard free energy change)
fermentation	nutrients	

CHAPTER 6 WEB LINKS



6.1



6.2



6.3

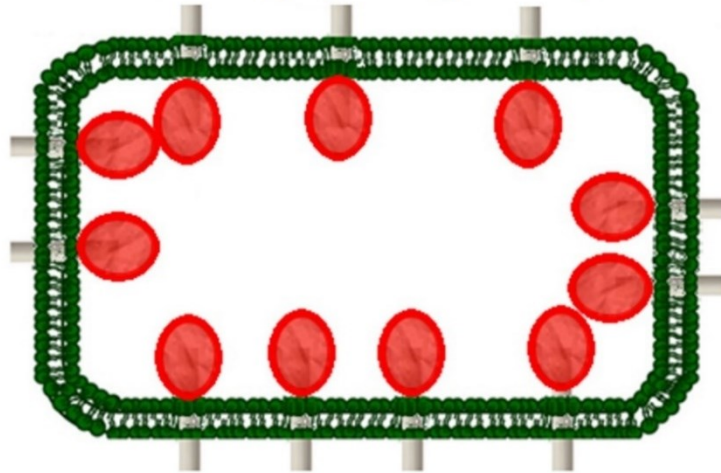
Chapter 7

Electron Transport, Oxidative Phosphorylation, Photosynthesis

Electron transport, oxidative phosphorylation, and protein motors make ATP in mitochondria; H₂O is split, CO₂ fixation by "light" and "dark" reactions makes glucose in chloroplasts

Reminder: For inactive links, google key words/terms for alternative resources.

The lollipop energy factory



will make you shake and dance!

[Listen Here!](#) or use the QR code above.

7.1 Introduction

We saw that glycolysis generates two pyruvates per glucose, and that subsequent oxidation of each pyruvate generates two Ac-S-CoA molecules. In aerobic cells, oxidation of each Ac-S-CoA by the *Krebs cycle* captures about 4% (30/687 Kcal) available in a mole of glucose—in just two ATP molecules; not much for all that biochemical effort! But a total of twenty-four H⁺ ions (protons) in glucose have captured in *reduced electron carriers* (NADH, FADH₂) in *redox reactions*. Here we'll see how oxidation of NADH and FADH₂ by an **electron transport system** releases free energy that fuels the creation of a **proton (H⁺) gradient**. Then we'll see how dissipation of this gradient powers **oxidative phosphorylation** and ATP synthesis. *Electron transport* and *oxidative phosphorylation* account for the cellular capture most nutrient free energy in ATP. We'll contrast oxidative phosphorylation with glycolytic and Krebs cycle *substrate-level phosphorylation* and look at an energy balance sheet for respiration. Then we'll look at cellular free energy capture from alternate nutrients. Finally, we look at **photosynthesis**, and then compare ATP production in what are essentially, opposite pathways.

Learning Objectives

When you have mastered the information in this chapter, you should be able to do the following:

1. Explain the *centrality* of the Krebs cycle to aerobic metabolism (cellular respiration).
2. Identify *sources of electrons in redox reactions* leading to and within the Krebs cycle.
3. Draw the *path of electrons* from the Krebs cycle to the electron transport chain to O₂.

- Trace the *evolution of the electron transport chain* from its location on an aerobic bacterial membrane to its location in eukaryotic cells.
- List the expected properties of a *proton gate* and a *proton pump*.
- Interpret experiments* involving redox reactions, ATP synthesis, and ATP hydrolysis conducted with intact mitochondria and separated mitochondrial membranes.
- Distinguish between the *pH*, *proton (H⁺)* and *electrical gradients* established by electron transport.
- Explain the *chemiosmotic mechanism of ATP synthesis* and contrast it with *substrate-level phosphorylation*.
- Discuss the evolution of electron transport in respiration and in photosynthesis and each.
- Trace the different paths that electrons can take in photosynthesis and explain the circumstances that would result in the different paths.
- Explain the similar (or identical) biochemical intermediates in respiration and photosynthesis.

7.2 Electron Transport Chains

All cells use an ***electron transport chain (ETC)*** to oxidize substrates (Krebs cycle products) in *exergonic* reactions. The flow of electrons from reduced substrates through an *ETC* is like the movement of electrons between the poles of a battery. In the case of the battery, the electron flow releases free energy to power motors, lights, cell phones.... During mitochondrial respiration, reduced electron carriers (NADH, FADH₂) are oxidized, and electrons flow down an *ETC* to molecular oxygen (O₂) to make water. In plants and other photosynthetic organisms, electron flow down an *ETC* oxidizes NADPH (nicotinamide adenine dinucleotide phosphate, a phosphorylated version of NADH), eventually reducing CO₂ to sugars.

In both respiration and photosynthesis, the oxidation of energy-rich reduced electron carriers releases ***free energy***. This free energy of electron flow is coupled to the active transport of protons (H⁺ ions) across a membrane by ***proton pumps***. The proton pumps create a chemical (proton) gradient and consequently a pH and electrical gradient. In shorthand, we say that the free energy once in reduced substrates is now in an ***electrochemical gradient***. The proton gradient (in fact, any gradient) is a source of potential (i.e., free) energy. The proton gradients of respiration and of photosynthesis will be used to make ATP by ***oxidative phosphorylation*** and by ***photophosphorylation***, respectively.

The ***Chemiosmotic Mechanism*** explains how the electrochemical gradient forms and how it stores free energy, and how its free energy ends up in ATP. For his insight on this mechanism, Peter Mitchell won the Nobel Prize in Chemistry in 1978. Read Mitchell's original proposal of the *chemiosmosis model* of mitochondrial ATP synthesis in Mitchell P (1961) *Coupling of phosphorylation to electron and hydrogen transfer by a chemiosmotic type of mechanism*. Nature 191:144-148.

7.3 Electron Transport in Respiration

Here we focus on the details of respiration as it occurs in the mitochondria of eukaryotic cells. The end products of electron transport are NAD⁺, FAD, water, and protons. The protons end up outside the mitochondrial matrix because they are pumped across the cristal membrane

using the free energy of electron transport. *Electron transport* and *oxidative phosphorylation* are summarized in Figure 7.1.

Mitochondrial Electron Transport and Oxidative Phosphorylation

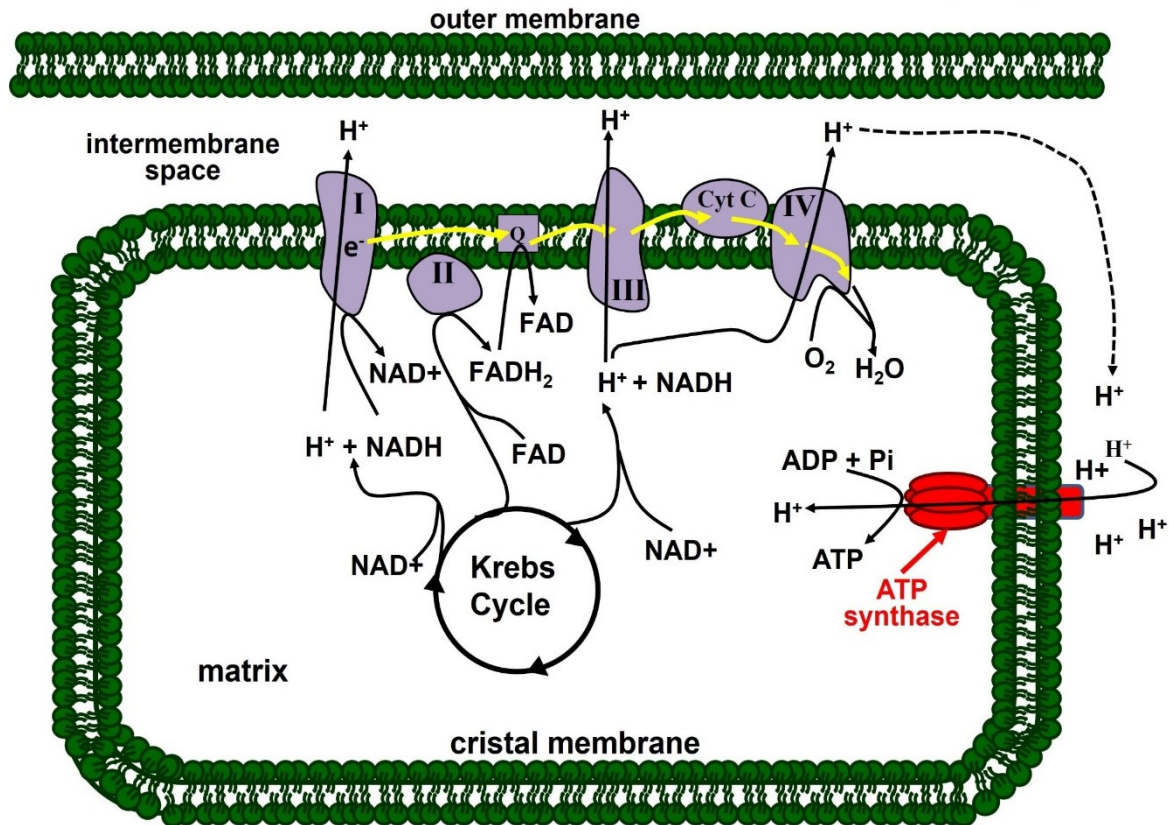


Fig. 7.1: Electron transport and oxidative phosphorylation on the cristal membrane: NADH and FADH₂ oxidation feeds electrons into electron transport, releasing free energy to powers proton pumps that force H⁺ ions (i.e., protons) out of the mitochondrion. The resulting pH (i.e., H⁺) gradient fuels ATP synthesis when protons flow into the mitochondrial matrix through a regulated cristal membrane ATP synthase (red).

Roman-numbered protein complexes, along with coenzyme Q ("Q") and cytochrome C ("Cyt C") constitute the ETC. The role of the respiratory ETC is to oxidize NADH or FADH₂ to NAD⁺ and FAD, respectively. Electrons from these reduced electron carriers are transferred from one ETC complex to the next. At the end of the chain, electrons, protons, and oxygen unite in complex IV to make water. Under standard conditions in a closed system, electron transport is downhill, with an overall release of free energy (a negative ΔG_0) at equilibrium.



[161 Electron Transport Oxidizes Reduced Electron Carriers](#)



[162-2 Finding the Free Energy of Electron Transport](#)



[163-2 Separating Electron Transport from Oxidative Phosphorylation](#)



In the illustration above, we can see three sites in the respiratory ETC that function as H^+ pumps. At these sites, the negative change in free energy of electron transfer is large and coupled to the action of a *proton pump*, which forces protons out of the mitochondrial matrix and across the cristal membrane. Because the outer mitochondrial membrane is freely permeable to protons, the gradient created is, in effect, between the cytoplasm and the mitochondrial matrix. Protons flow back into the mitochondrial matrix through lollipop-shaped *ATP synthase* complexes on the cristal membrane, and this flow releases the gradient free (potential) energy that will be harnessed to make ATP by oxidative phosphorylation.



[164-2 Proton Pumps Store Free Energy of the ETC in Proton Gradients](#)



7.4 Oxidative Phosphorylation in Respiration

Oxidative phosphorylation is the mechanism by which ATP captures the free energy in the mitochondrial proton gradient. Most of the ATP made in aerobic organisms is made by oxidative phosphorylation, and not by substrate-level phosphorylation (the mechanism of ATP synthesis in reactions of glycolysis or the Krebs cycle). Some oxidative biochemistry may have existed early in evolution. After photosynthetic organisms appeared, the response of anaerobic organisms to the toxicity of rising environmental oxygen levels may have been to coopt early oxidative biochemistries oxygen detoxification. The later elaboration of respiratory metabolism was undoubtedly selected because it turned out to be significantly more efficient at making ATP than were the anaerobic fermentations (such as "complete" glycolysis). In other words, oxidative phosphorylation is more efficient than substrate-level phosphorylation.

Summarizing electron travel down the electron-transport chain fuels the three proton pumps that establish a proton gradient across the *cristal* membrane that stores free energy. Oxidative phosphorylation then allows controlled diffusion of protons back into the mitochondrial matrix through cristal membrane\ ATP synthases, fueling ATP production. We say that the proton gradient has a **proton motive force**. We also recognize the proton gradient as a pH gradient as well as an electrical gradient (i.e., the difference in electric potential).

The use of this proton motive force to make ATP is regulated. Conditions in the cell control when the energy stored in this gradient will be released to make ATP. The switch that allows protons to flow across the cristal membrane to relieve the proton gradient is a mitochondrial cristal membrane *ATP synthase*, a tiny, complex enzymatic protein motor.

When this "switch" is open, protons *pumped* out of the mitochondrial matrix during electron transport can flow back into the matrix through the *ATP synthase*. For a clear discussion of this complex enzyme, see P. D. Boyer (1997) *The ATP synthase – a splendid molecular machine*. Ann. Rev. Biochem. 66:717-749.

The *splendid molecular machine* was discovered when isolated mitochondria were themselves fractionated. When a membrane fraction (*Fraction 1*) of the organelle was shown to hydrolyze ATP, the putative enzyme responsible was designated an F1 ATPase. Intact mitochondria cannot hydrolyze ATP! In further experiments, vesicles formed from isolated

mitochondrial membranes were exposed to high pH in the presence of ADP and inorganic phosphate, resulting in ATP synthesis. Hence, F1 ATPase was redesignated the mitochondrial ATP synthase (or F1 ATP synthase). The capture of free energy of protons flowing through this lollipop-shaped ATP synthase by facilitated diffusion is shown in Figure 7.2.

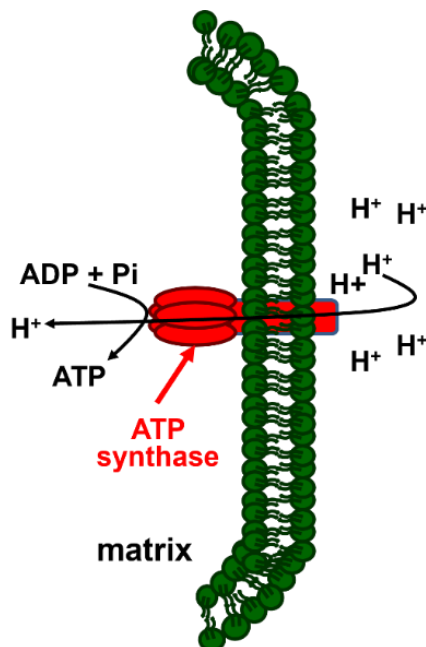


Fig. 7.2: The passive diffusion of protons through the cristal membrane ATP synthase relieves the proton gradient, releasing free energy that fuels ATP synthesis in the mitochondrial matrix.

If the three ETC sites in the cristal membrane that actively transport protons are **proton pumps**, then the cristal membrane ATP synthase complexes function as regulated **proton gates** which catalyze ATP synthesis as protons flow through. For their discovery of the details of ATP synthase function, P. D. Boyer and J. E. Walker shared the Nobel Prize in Chemistry in 1997.



[165-2 Proton Gates Capture Proton Gradient Free Energy as ATP](#)

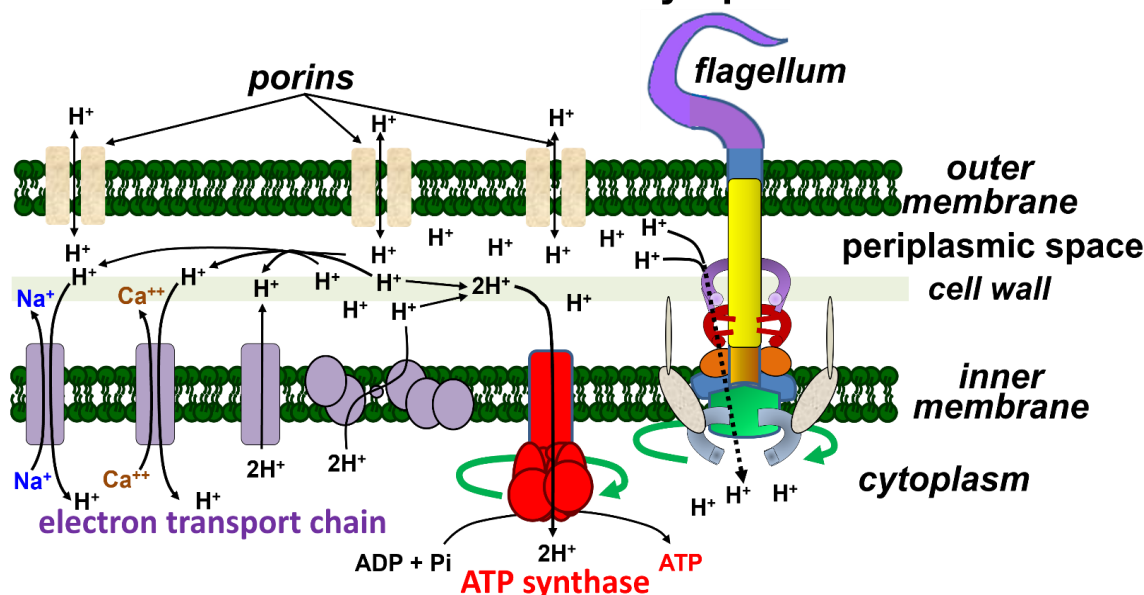
The ratio of ATP to ADP concentrations regulates proton flow through the ATP synthase gates. A high ATP/ADP ratio in the mitochondrial matrix indicates that the cell does not need more ATP and closes the proton gate so that the proton gradient cannot be relieved. On the other hand, a low ATP/ADP ratio in the matrix means that the cell is hydrolyzing a lot of ATP and that the cell needs more. Then the proton gate opens, and protons flow through cristal membrane ATP synthases back into the matrix, going down the concentration gradient. As they flow, they release free energy to power a protein motor in the enzyme, which in turn activates ATP synthesis. Just as we did for glycolysis, we can count the ATPs and see how much free energy we get from aerobic respiration, (i.e., the complete oxidation of glucose). You can see this in the following link.



[166 A Balance Sheet for Respiration](#)

If the *endosymbiotic theory* is correct, then aerobic bacteria are the evolutionary ancestor to mitochondria, and the bacterial cell membrane should be the site an ETC and a *chemiosmotic mechanism* of ATP generation, much like that in mitochondria. This is in fact the case. And proton gradients not only power molecular motors linked to ATP synthesis but are also linked directly to the spinning of a bacterial flagellum (yet another "splendid molecular machine," Figure 7.3, below).

Bacterial electron transport chain pumps protons and cations from cytoplasm:



Relief of the resulting ion gradient powers ATP synthase and flagellar motor.

Fig. 7.3: A bacterial electron transport system can pump protons and other cations (Na⁺, Ca⁺⁺) from the cytoplasm into the periplasmic space, at the left on the inner (cell) membrane. Relief of the proton gradient can power an ATP synthase or motility by fueling a spinning flagellum (at the right). In both cases, proton flow into the cell powers complex protein motors.

Electron transport in the cell membrane creates the gradient, in some cases co-transporting a proton and another cation (Na⁺, Ca⁺⁺). Relief of the proton gradient typically powers both ATP synthesis and the flagellum, though recent studies indicate that relief of the Na⁺ gradient can also spin the flagellum in some species of bacteria.

7.5 Photosynthesis

Plants have evolved different photosynthetic adaptations, called *C3*, *CAM* and *C4*. All are essentially the reverse of respiration. We'll consider each in turn later. Chemically, photosynthesis is the reverse reaction of respiration. Compare the following chemical reactions of respiration and photosynthesis:

1. $\text{C}_6\text{H}_{12}\text{O}_6 + 6\text{O}_2 \rightleftharpoons 6\text{CO}_2 + 6\text{H}_2\text{O} \quad (\Delta G_o = -687 \text{ Kcal/mole})$
2. $6\text{CO}_2 + 6\text{H}_2\text{O} \rightleftharpoons \text{C}_6\text{H}_{12}\text{O}_6 + 6\text{O}_2 \quad (\Delta G_o = +687 \text{ Kcal/mole})$

If respiration (reaction 1) is a complete oxidation of glucose to H_2O and CO_2 , then photosynthesis (reaction 2) is a reduction of CO_2 by electrons from H_2O to make glucose. Thus, photosynthesis is an endergonic reaction pathway. During photosynthesis, sunlight (i.e., visible light) fuels the reduction of CO_2 (Figure 7.4).

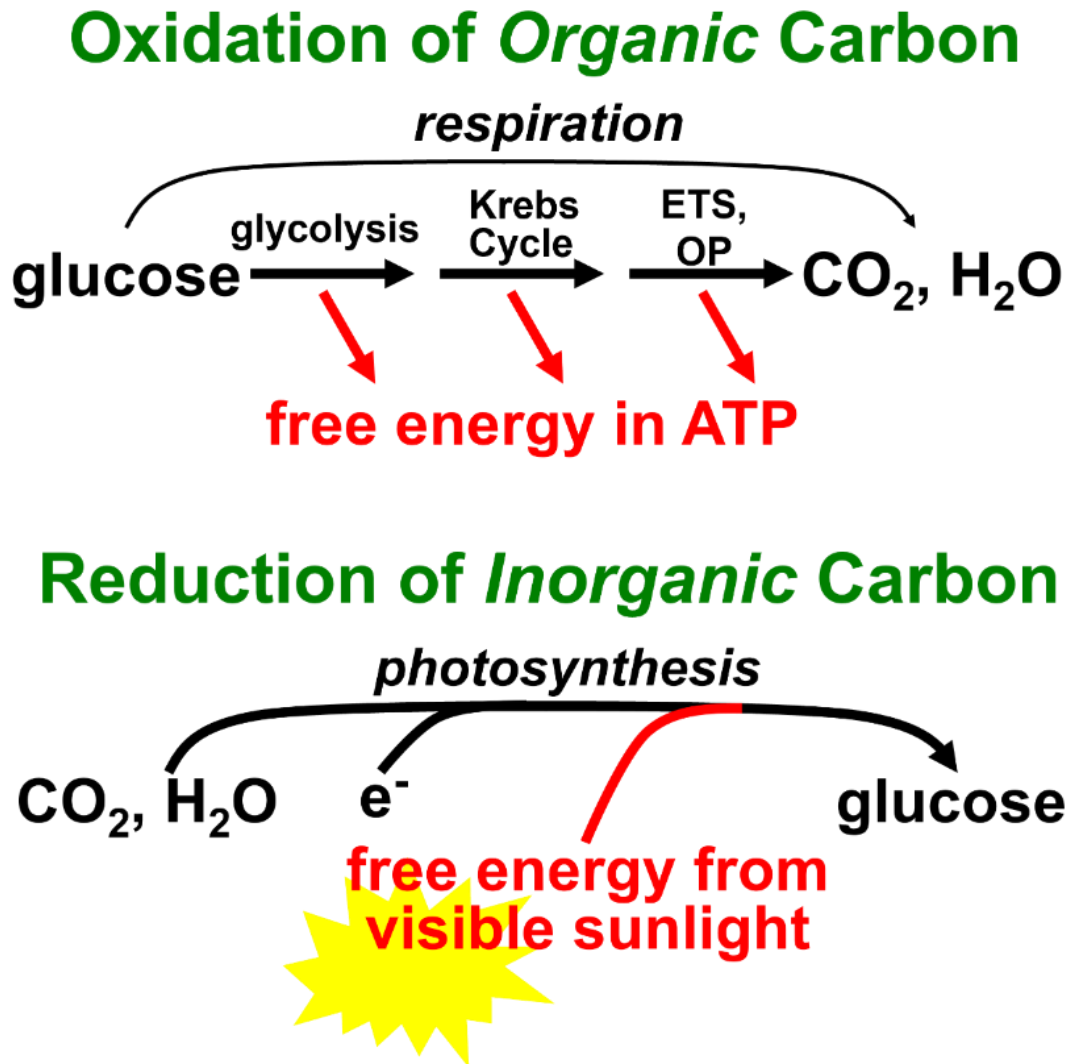


Fig. 7.4: The flow of carbon through life. Oxidation of organic carbon (e.g., glucose) releases free energy to make ATP. Free energy from sunlight fuels photosynthesis, turning CO_2 back into organic carbon (glucose).

Photosynthesis began in the absence of oxygen; it preceded oxygenic respiration on Earth. But after it evolved, it led to increasing levels of atmospheric oxygen that resulted in the selection of oxygenic respiratory pathways (the Krebs cycle, electron transport, and oxidative phosphorylation). We will see here that photosynthesis and respiration both have electron-transport-ATP-synthesizing systems that share similar features. This suggests that both pathways share a common evolutionary ancestry. Elsewhere, we will consider what a common ancestral system might have looked like. Two biochemical pathways make up photosynthesis:

- **Light-dependent reactions** use visible light energy to remove electrons from water to reduce electron carriers, to pump protons, and to make ATP.
- **Light-independent reactions** use ATP to transfer electrons from the reduced electron carriers to CO_2 to synthesize glucose.

These light-dependent and light-independent pathways of photosynthesis are summarized below in Figure. 7.5.

Photosynthesis has light-dependent and light-independent components:

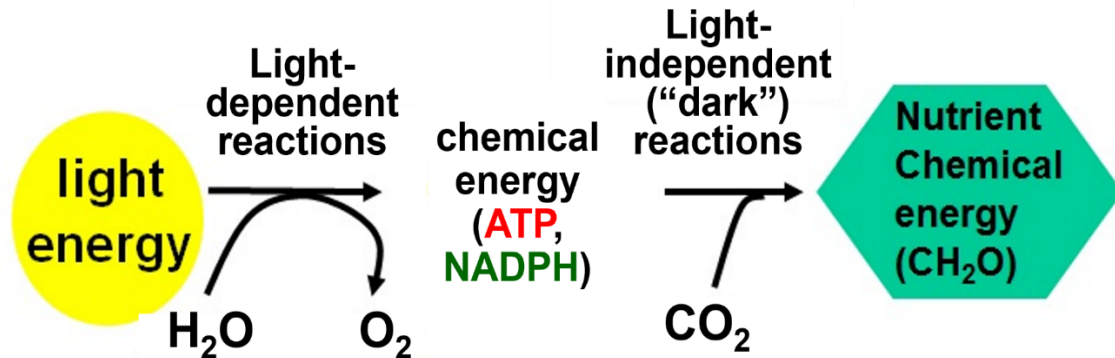


Fig. 7.5: The *light-dependent* ("light") reactions of photosynthesis "split" water, releasing oxygen and protons, and capturing solar free energy as chemical energy. The *light-independent* ("dark") reactions use this chemical free energy to "fix" CO_2 into organic molecules (nutrients).

7.5.1 The Light-Dependent Reactions

Colored substances contain **pigments** which reflect the colors that we see, and which simultaneously absorb all the other colors of visible light. Early studies asked which plant pigments absorbed the light that allow (or as we say now, "support") photosynthesis. The abundant chlorophyll pigment we see in plant tissues is actually two separate green pigments: *chlorophyll a* and *chlorophyll b*. One might therefore predict that light absorbed by either or both chlorophylls will support photosynthesis. Figure 7.6 illustrates the experiment that tested this hypothesis.

Photosynthesis Uses Visible Light

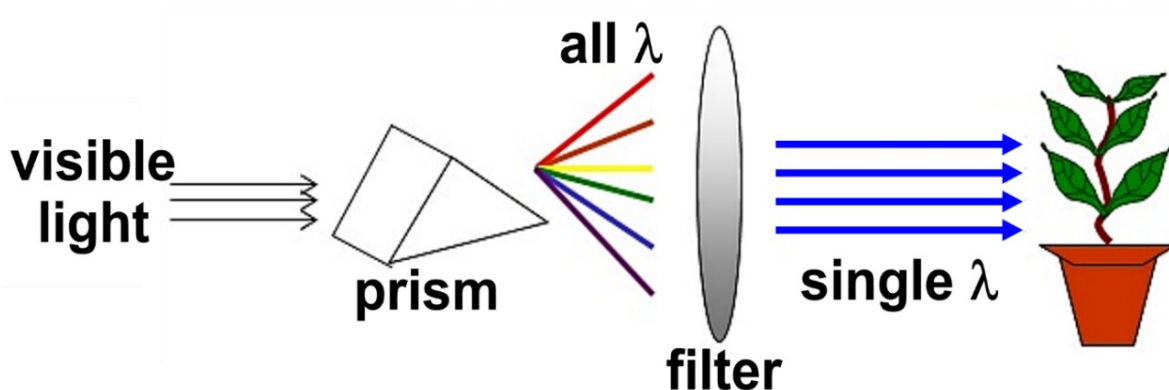


Fig. 7.6: This experiment asks which wavelengths of light support photosynthesis. Visible light is separated into different colors (wavelengths) by a prism (left), and then passed through filters (center) that block all but one wavelength of light (right) to see which wavelengths support plant growth and carbohydrate synthesis.

The *action spectrum* of photosynthesis (Figure 7.7, below) plots the results of this experiment.

An Action Spectrum of Photosynthesis

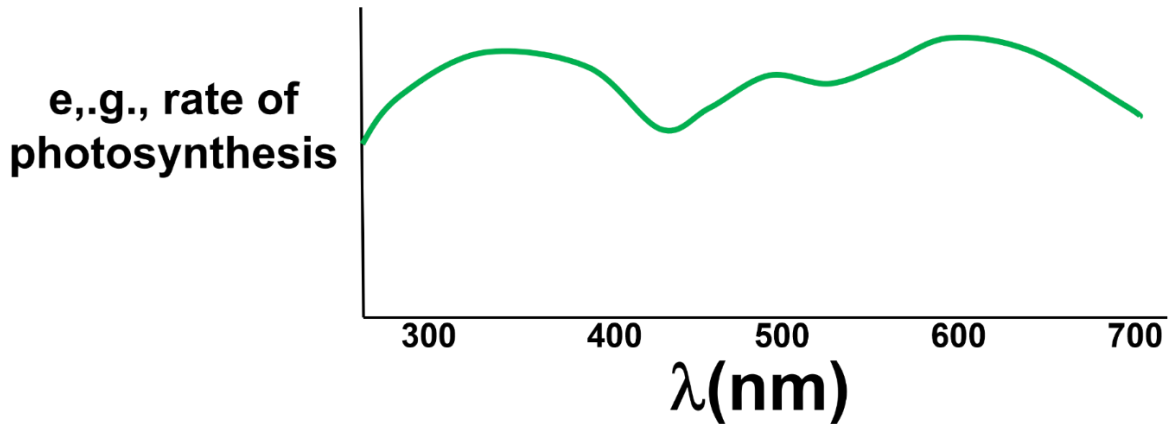


Fig. 7.7: The action spectrum of photosynthesis plots the rate of photosynthesis (e.g., glucose synthesis) of a plant illuminated by different wavelengths of light. The graph shows that a wide range of visible-light wavelengths support photosynthesis.

The spectrum shows that all wavelengths of visible light energy can support photosynthesis. In addition, other experiments revealed that radiation other than visible light (e.g., ultraviolet and infrared light) do not support photosynthesis. One can conclude that chlorophylls are likely not the only pigments to support photosynthesis.

Chlorophylls are easily purified from leaves. The graph in Figure 7.8 shows an average *absorbance spectrum* of chlorophylls. The absorbance of *chlorophyll a* and *chlorophyll b* are slightly different, but they center at wavelengths of 450 nm and 675 nm.

Absorbance Spectrum of Chlorophylls

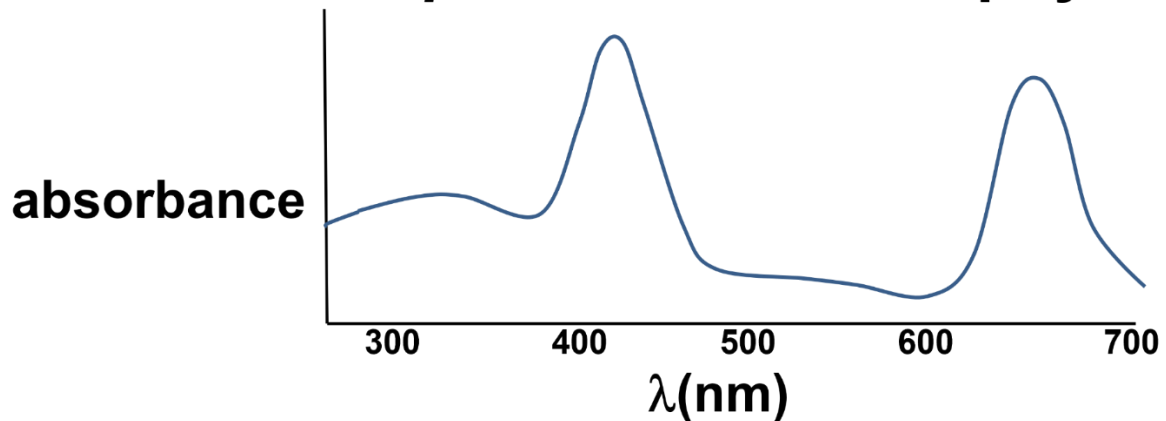


Fig. 7.8: This graph plots the absorbance spectrum of purified chlorophylls, showing two main peaks of absorbance, suggesting the hypothesis that different pigments must be absorbing other wavelengths of light that support photosynthesis.

Comparing the action and absorbance spectra, we can conclude that the chlorophylls support photosynthesis, but that chlorophylls alone do not account for the *action spectrum* of photosynthesis!

In fact, other pigments absorbing light elsewhere in the visible spectrum also support photosynthesis. Of course, we knew that these other pigments were present in leaves and other photosynthetic plant tissues because we could see many of them as fall colors. All of these pigments (including chlorophylls) are found in **chloroplasts**, the organelles that conduct photosynthesis in plants. Take another look at the structure of chloroplasts in the electron micrographs in Figure 7.9.

Low & High Magnification EM Cross-Sections Through a Chloroplast

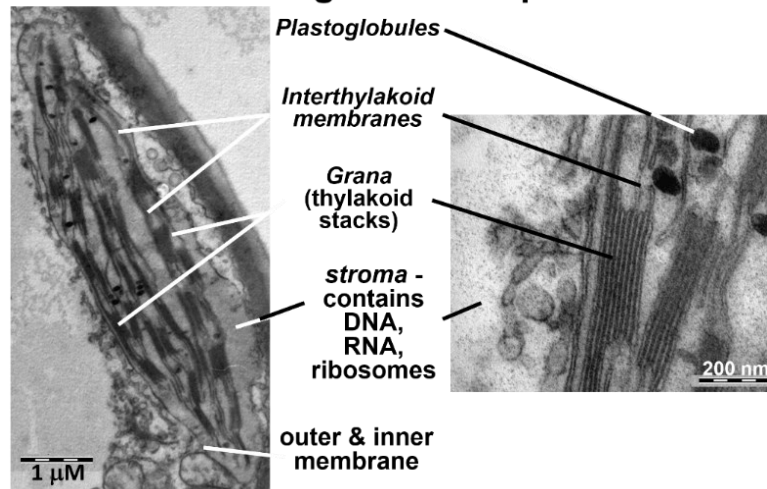


Fig. 7.9: A low-power transmission micrograph of a chloroplast (left) shows grana (thylakoid stacks). A high-power micrograph (right) shows a few grana in detail, revealing thylakoid membranes in a granum and the interthylakoid membranes linking grana.

Figure 7.10 shows that the visible light *absorbance spectra* of the *chlorophylls*, *carotenoids* and possibly other pigments coincide with the photosynthetic *action spectrum*. This implies that the absorption of light by those pigments is responsible for photosynthesis.

Plant Pigment *Action* vs. *Absorbance Spectra*

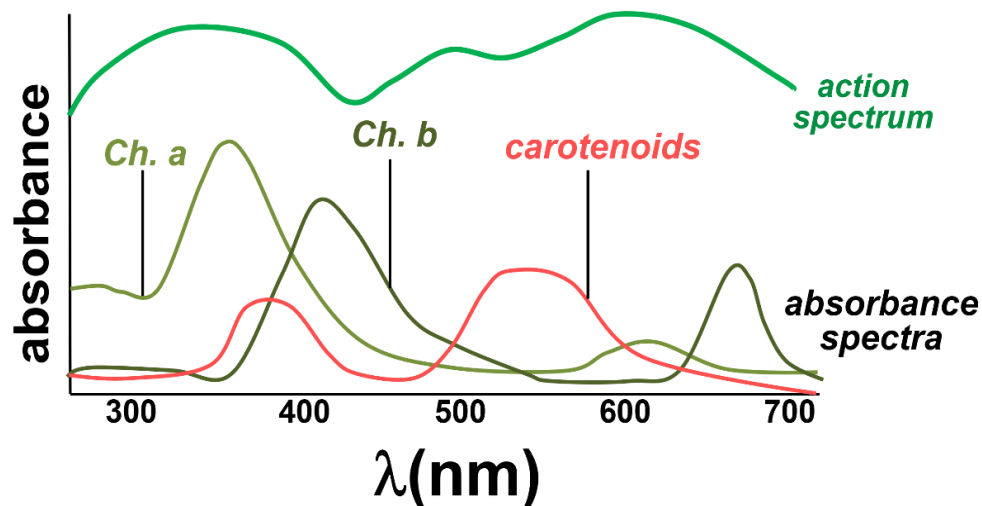


Fig. 7.10: Here a photosynthetic action spectrum is shown above the absorbance spectra of different plant pigments. The absorbance peaks of different chloroplast pigments lie under different regions of the action spectrum, consistent with the hypothesis that multiple pigments support photosynthesis.

In fact, *chlorophyll a*, *chlorophyll b*, carotenoids, and other *accessory pigments* participate in capturing light energy for photosynthesis. Two pigment clusters, called *reaction centers*, capture light energy as part of *photosystem 1 (PSI)* and *photosystem 2 (PSII)* on *thylakoid membranes* of chloroplasts.

Johann Deisenhofer, Robert Huber and Hartmut Michel first determined the 3D structure of a bacterial reaction center. Then they unraveled the relationship between the structure of the proteins in the center and the membrane in which they were embedded. For this, they shared the 1988 Nobel Prize in Chemistry.

The activities of PSI are animated at ^{7,4}[Photosystem I Action](#). You can watch as photon of light excites electron (e^-) pairs from PSI pigments, which then transfer their energy from pigment to pigment, ultimately to *chlorophyll a P700*. The impact of the electron pair excites a pair of electrons from *chlorophyll a P700*, and that pair is captured by a *PSI e^- acceptor*, an event referred to as *charge separation*. Next, the reduced PSI acceptor is oxidized as electrons move down a short ETC, eventually reducing NADP^+ to NADPH. Electrons on NADPH will eventually be used to reduce CO_2 to a carbohydrate. So far, so good! But that leaves an electron deficit in PSI. The Z-Scheme illustrated in Figure 7.11 (below) follows electrons *taken from water* (absorbed through roots) into PSII, which will replace those electrons missing from PSI.

The Z-Scheme: Electrons Missing from PSI Replaced With Electrons from PSII

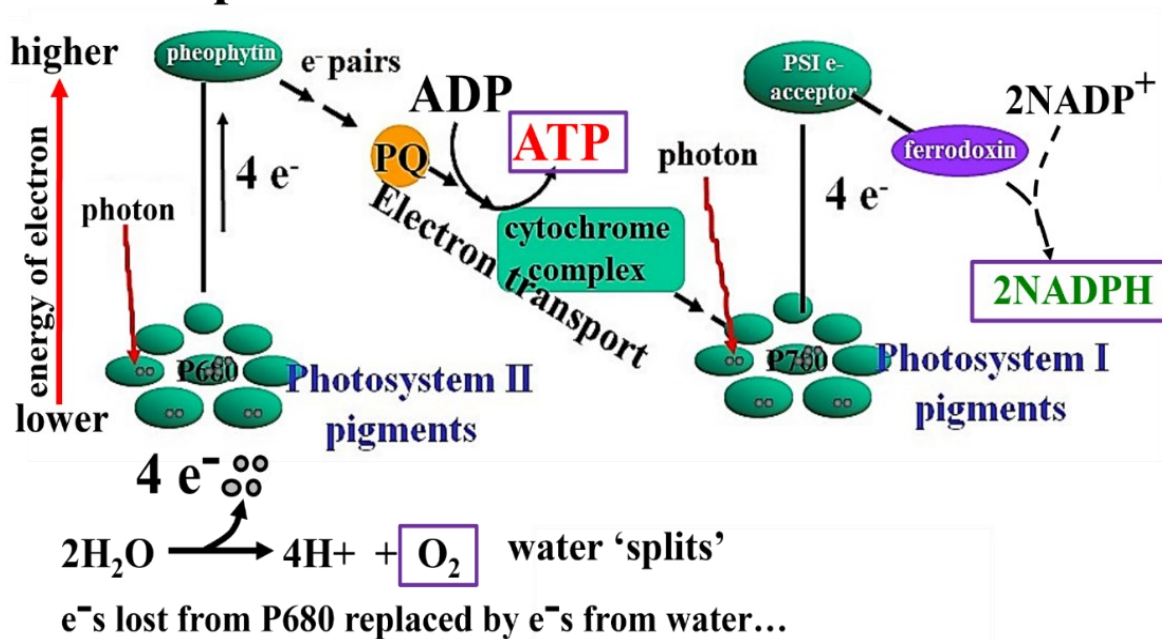


Fig. 7.11: In the “Z-Scheme” of photosynthesis, pigments in PSI and PSII absorb light energy that excites electrons captured by electron acceptors. Electrons excited from PSI reduce NADPH, the starting point of the light-independent reactions (right). Electrons excited out of PSII come from splitting water, releasing oxygen (left). These electrons will flow down a photosynthetic electron transport chain to replace electrons lost from PSI. Along the way, free energy from those excited electrons fuels ATP synthesis (middle of the illustration).

Let's summarize electron flow through the Z-scheme. Light excites an e^- pair from the **P680** form of *chlorophyll a* in PSII. A PSII electron acceptor in the thylakoid membrane, identified as **pheophytin**, captures these electrons, another act of photosynthetic *charge separation*. An electron transport chain oxidizes the pheophytin, transferring e^- pairs down to PSI.

Some of the free energy released then pumps protons from the *stroma* into the luminal space of the thylakoid membranes. The gradient free energy fuels ATP synthesis as protons flow back into the stroma through a chloroplast **ATP synthase**. The video at ^{7.5}[Z-Scheme Action](#) animates the entire Z-Scheme, showing first how PSI electrons reduce NADP^+ and then how PSII electrons replace missing PSI electrons, making ATP along the way. The oxygen released by splitting water ends up in the atmosphere.

The goal of photosynthesis is to make and to store glucose, so the photosynthesizing cells get ATP to power glucose production by **photophosphorylation** during electron transport via the Z-scheme. However, **cyclic photophosphorylation** can occur when the cell's need for ATP exceeds the capacity of the Z-scheme to supply it. Cyclic photophosphorylation is illustrated in Figure 7.12.

Energy from light can be used to make ATP directly, without splitting water: *Cyclic Photophosphorylation*

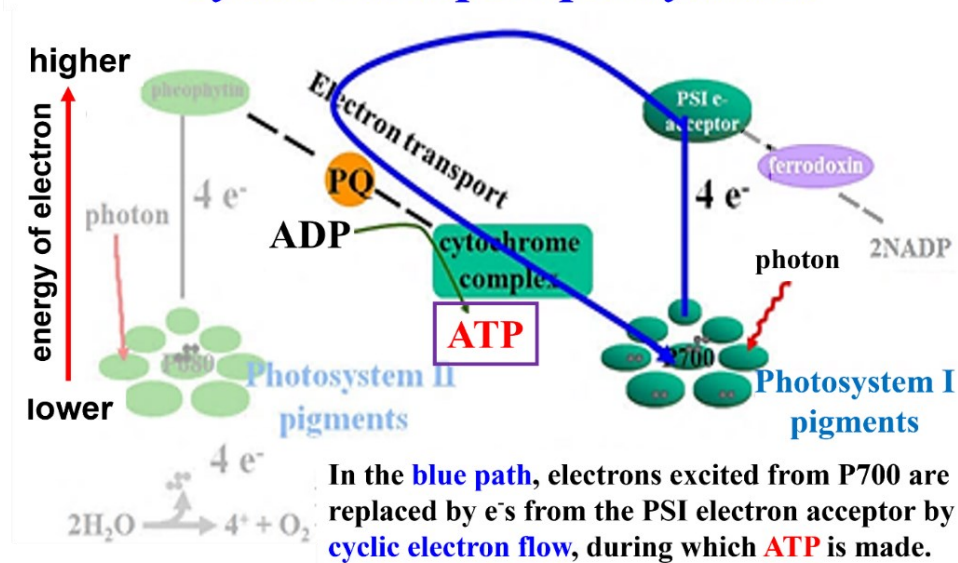


Fig. 7.12: In *Cyclic Photophosphorylation* electrons excited from PSI don't reduce NADP^+ , PSII is not active, water is not split, and the excited PSI electrons are captured by the photosynthetic electron transport chain, to be returned to PSI. ATP made by this cycle is used for cellular work when there is no need or capacity to make more sugar.

Cyclic Photophosphorylation is a variation on the light-dependent reactions, a kind of time-out to make ATP without reducing NADP^+ and thus, without reducing CO_2 to make sugar). The cycle simply takes excited electrons to the PSI electron acceptor, and instead of sending them to NADP^+ , it deposits them on PC (*plastocyanin*) in the electron transport chain between PSII and PSI. These electrons then flow down the "long line" of the Z, right back to PSI, releasing their free energy to make ATP. In light, the electrons just keep going up and

around, hence the name *Cyclic Photophosphorylation*. The path of electrons shown in Figure 7.12 is also animated at ^{7,6}[Cyclic Photophosphorylation Activity](#).

7.5.2 The Light-Independent (“Dark”) Reactions of Photosynthesis

As we have seen, the light-dependent reactions of photosynthesis require light energy and water to generate O₂, ATP, and NADPH. In the *light-independent* (or “*dark*”) reactions of chloroplasts, the ATP and NADPH will provide free energy and electrons (respectively) for carbon fixation (the reduction of CO₂ to make carbohydrates). There are three main pathways for the so-called “dark reactions.”

7.5.2.a The Calvin-Cycle C₃ Pathway

The Calvin cycle is the most common light-independent reaction pathway. In 1950 Melvin Calvin and his colleagues (James Bassham and Andrew Benson) described the complete cycle from CO₂ to glucose (Figure 7.13). They followed the radioactive ¹⁴C isotope of carbon into ¹⁴CO₂, then into glucose, and then into carbohydrates in photosynthesizing plants. For his work, Calvin received the Nobel Prize in Chemistry in 1961.

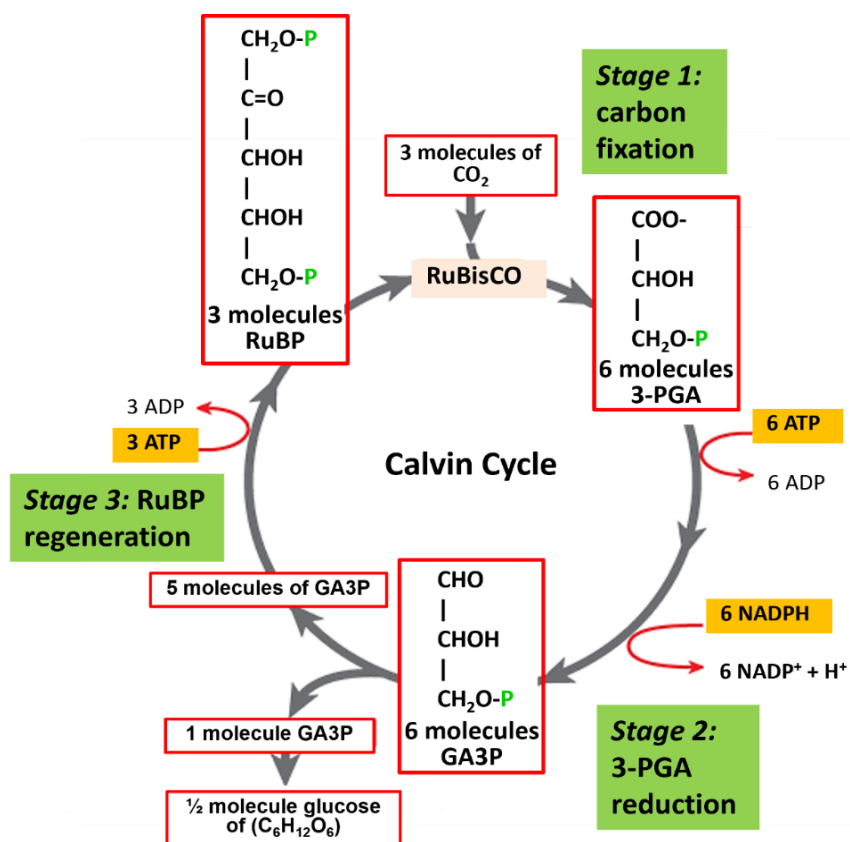


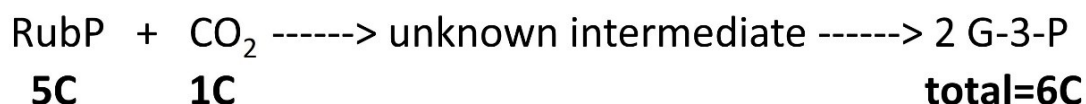
Fig. 7.13: In the Calvin cycle in C₃ (i.e., most) plants, three ribulose biphosphate (RuBP) molecules are fixed to three CO₂ molecules, catalyzed by *RuBisCO*. The 3 resulting 6-C molecules are split into six 3-C carbohydrates that then make six glyceraldehyde-3-phosphate molecules (G3P, or GA3P in the illustration). Five of these go on to regenerate three RuBP molecules; the sixth G3P waits for the Calvin cycle to repeat, producing another G3P. Two G3Ps from two turns of the cycle are substrates for glucose synthesis.

Check the animation at ^{7.7}[Calvin Cycle Action](#). Each carbon dioxide entering the Calvin cycle is "fixed" to a 5-C *ribulose biphosphate* molecule (RuBP), catalyzed by the enzyme ***RuBP carboxylase-oxygenase***, or ***RUBISCO*** for short. Seemingly, the expected 6-C molecule quickly splits into two 3-C molecules since it has not been found as an intermediate to date! The first detectable products are two molecules of 3PGA (3-phosphoglyceric acid). Each 3PGA is in turn reduced to glyceraldehyde-3-phosphate (GA3P in the illustration, or G-3-P). Some of the Calvin cycle intermediates should look familiar: the cycle regenerates the RuBP *and* the essential intermediates of glucose production. Figure 7.14 (below) may help with the arithmetic of the Calvin cycle.

Calvin Cycle Arithmetic!

How to make one molecule of glucose and regenerate 5 molecules of RubP

The basic reaction:



Making glucose and regenerating RubP:

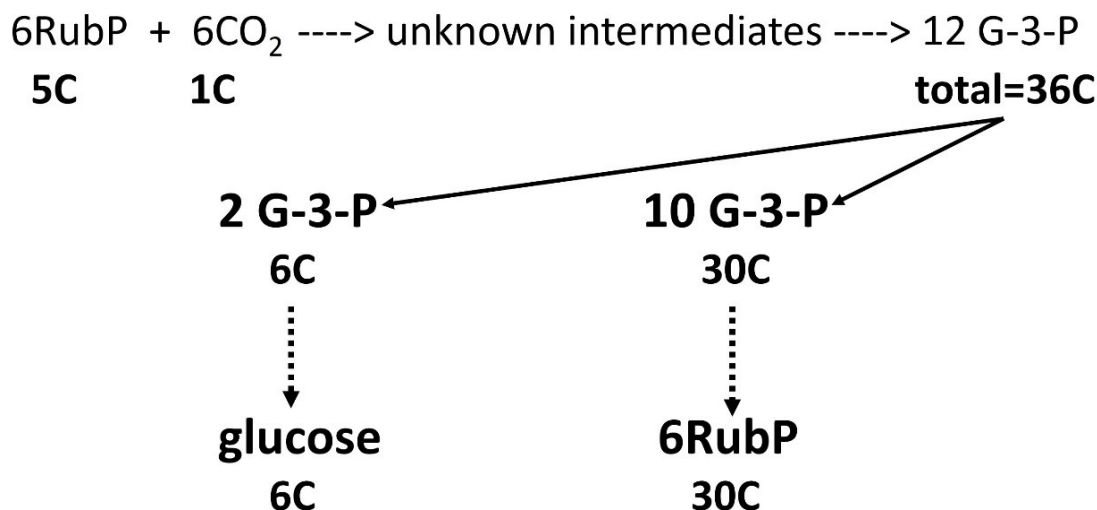


Fig. 7.14: Visualizing a turn of the Calvin Cycle can be difficult. This presentation of Calvin Cycle arithmetic shows how 6 turns of the cycle accounts for making one glucose and regenerating five RuBP molecules.

Perhaps the easiest way to see this is to imagine the cycle going around six times, fixing six molecules of CO₂, as shown in the illustration. In fact, a photosynthetic cell must fix six CO₂ carbons onto six RuBP molecules to generate a single 6-C glucose. This leaves thirty carbons on the table (as G-3-P molecules) with which to make six 5-C RuBP molecules.

There are times that even plants in temperate environments suffer prolonged hot, dry spells. Perhaps you've seen lawns grow slowly and turn brown after a dry heat wave in summer, only to grow and re-green when rains return. The **C3** plants can resort to

photorespiration during drought and dry weather, closing their stomata to conserve water. Under these conditions, CO₂ can't get into the leaves—and O₂ can't get out! As CO₂ levels drop and O₂ levels rise in photosynthetic cells, the Calvin cycle slows down. Instead of fixing CO₂, **RUBISCO** now catalyzes "O₂ fixation" using its *oxygenase* activity. The combination of RuBP with O₂ splits RuBP into 3-C and 2-C molecules: **3-phosphoglyceric acid** (3-PG) and **phosphoglycolate**, respectively (Figure 7.15).

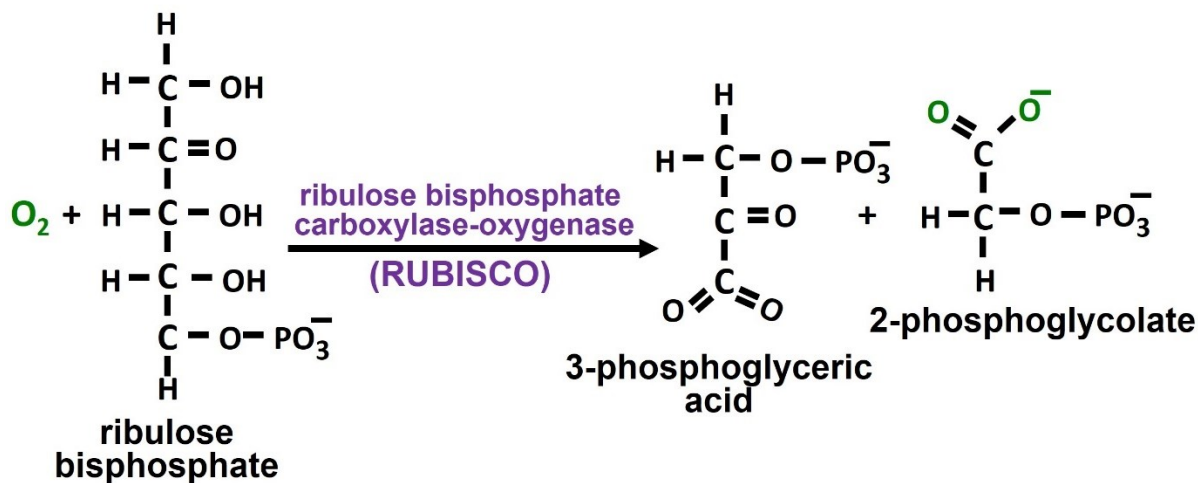


Fig. 7.15: *Photorespiration* is a C₃-plant strategy to conserve water during a dry spell. Under these conditions, RuBP carboxylase-oxygenase uses its *oxygenase* activity to, in effect, 'fix oxygen', by splitting ribulose biphosphate (at the left) into 3-phosphoglyceric acid and 2-phosphoglycolate (at the right).

Under these conditions, not only does photorespiration result in only one three-carbon carbohydrate (compared to two in the Calvin cycle), but the phosphoglycolate produced is *cytotoxic* (not healthy for cells!). It is necessary to remove the phosphate from phosphoglycolate and to then metabolize the remaining *glycolic acid*; this costs energy. Therefore, photorespiration can only be sustained for a short time.

On the other hand, plants that have adapted to live in hot, arid environments all the time have evolved one of two alternate pathways: the **CAM** (**Crassulacean Acid Metabolism**) and the **C₄** pathway. Each is an alternative to the **C₃** pathway of carbon fixation.

7.5.2.b The CAM Photosynthetic Pathway

Crassulacean acid metabolism (CAM) was discovered in the *Crassulaceae* family, which contains succulents like sedum (a common ground cover), cactuses, jade plants, and some orchids. This pathway was selected in evolution to allow plants to conserve water, especially during the high daytime temperatures. The CAM pathway is shown below in Figure 7.16.

Stomata in chlorenchyma (**mesophyll** leaf cells) close during the day to minimize water loss, which would result from **transpiration**. The stomata open at night, allowing plant tissues to take up CO₂. CAM plants fix CO₂ by combining it with **PEP** (*phosphoenolpyruvate*) to make **oxaloacetate** (**OAA**). This eventually produces **malic acid** that is stored in plant cell vacuoles. By day, stored malic acid retrieved from the vacuoles splits into pyruvate and CO₂. The CO₂ then enters chloroplasts and joins the Calvin cycle to make glucose and the starches.

Crassulacean Acid Metabolism (CAM) Photosynthesis: Adaptations to Hot Arid Environments

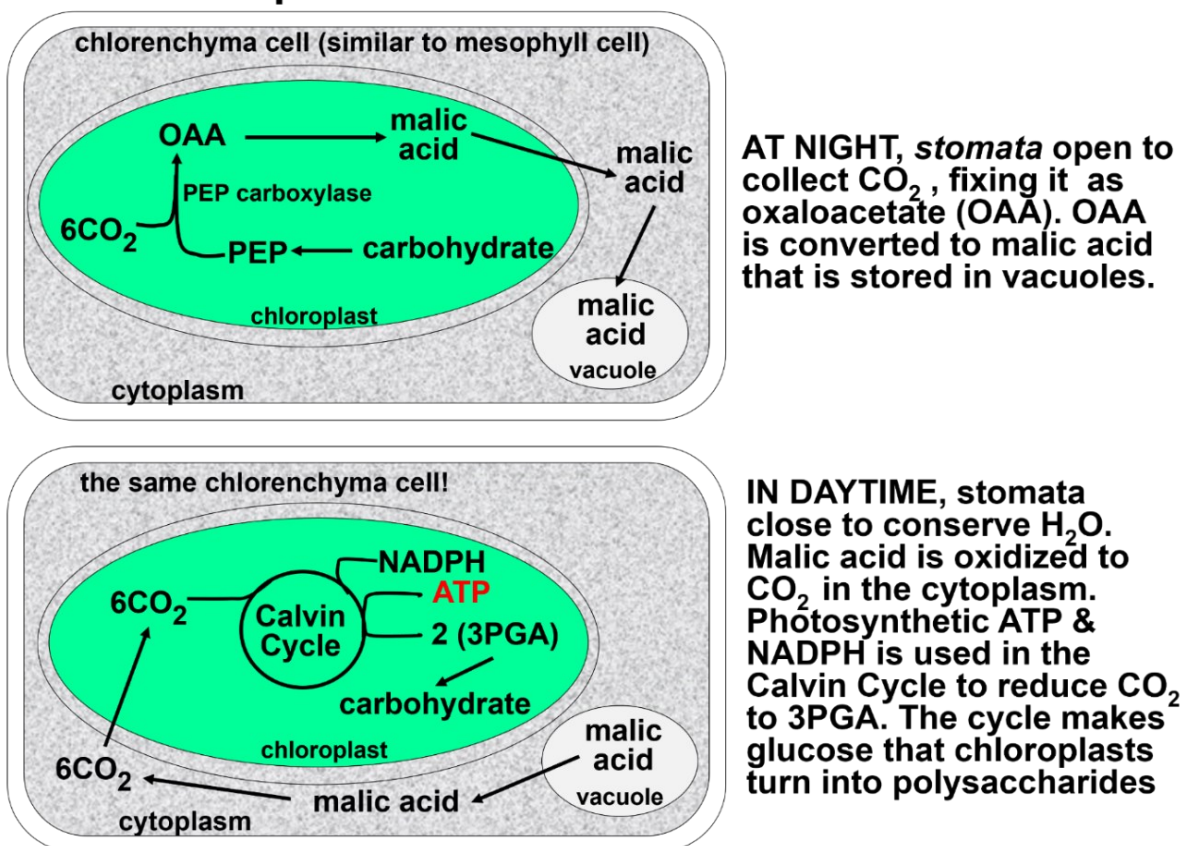


Fig. 7.16: Crassulacean Acid Metabolism evolve to help some plants conserve water during high daytime temperatures. Chlorenchymal cell stomata open at night, fixing CO_2 to oxaloacetate (OAA) which is converted to malic acid to be stored. Stomata close at daybreak to conserve water while stored malic acid releases CO_2 to be refixed to make glucose.

In sum, CAM plant *mesophyll* cells do the following:

- They open stomata at night to collect, to fix, and to store CO_2 as an organic acid.
- They close stomata to conserve water in the daytime.
- They re-fix the stored CO_2 as carbohydrates, using the NADPH and ATP from the light reactions the next the day.

7.5.2.c The *C4* Photosynthetic Pathway

The end-product of the *C4 pathway* is the malic acid (4 carbons). As in the CAM pathway, phosphoenolpyruvate (PEP) carboxylase catalyzes carbon fixation, converting PEP to oxaloacetate (OAA) and OAA to malate. But the pathways differ in what next happens to malate (malic acid). Figure 7.17 (below) shows the role of malate in the *C4* pathway.

C4 Photosynthesis: Adaptations to Hot Tropical Environments

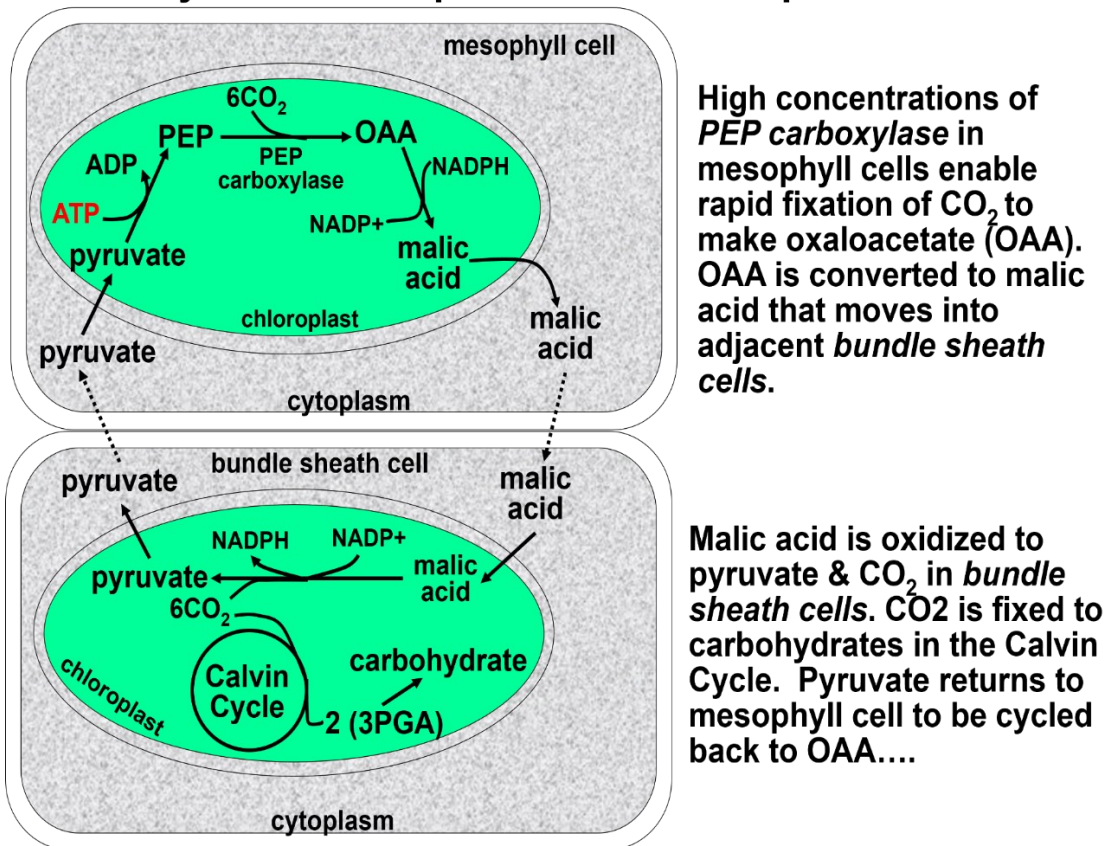


Fig. 7.17: Some plants that spend a lot of time in hot conditions have adapted to use the C4 photosynthetic pathway. In these plants the “dark” reactions are shared by mesophyll and bundle sheath cells, allowing efficient photosynthesis during the day, even hot ones! Stomata open part of day in mesophyll cells, CO_2 enter and is fixed to PEP that is converted to OAA, then malate. The malate moves to bundle sheath cells, where it releases CO_2 to be re-fixed by the Calvin cycle.

C4 metabolism diverges from the CAM pathway after malate formation. PEP-carboxylase catalysis is rapid in **C4** plants, in part because malic acid does not accumulate in the *mesophyll* cells. Instead, malate is rapidly transferred from the mesophyll cells into adjacent **bundle sheath** cells, where it enters chloroplasts. As a result, the **C4** plants can keep their stomata open to capture CO_2 for part of the day (unlike CAM plants), but they must close them for at least part of the day to conserve water. The four-carbon malic acid is oxidized to pyruvate (three carbons) in the bundle-sheath-cell chloroplasts. The CO_2 that is released then enters the Calvin cycle to be rapidly fixed by RUBISCO. Of course, the result is that, compared to **C3** photosynthetic metabolism, C4 plant photosynthesis results in a more efficient use of water and faster carbon fixation, properties most useful under dry, hot conditions. Corn is perhaps the best-known **C4** plant! By the way, we don't think of corn as tropical plant! Why not?

7.6 On the Origins and Evolution of Respiration and Photosynthesis

In one scenario, abundant chemical energy on our once-cooling planet favored the formation of cells that could capture free energy from energy-rich nutrients in the absence of any O_2 . We initially thought that these first cells would extract nutrient free energy by non-oxidative,

fermentation pathways. But they would have been voracious, quickly depleting environmental nutrient resources. Of course, this didn't happen!

It was the evolution of autotrophy that saved heterotrophic life from an early extinction! Why? Because autotrophs create organic molecules by extracting free energy from inorganic molecules or from light. An alternative scenario gaining traction suggests that the first cells may respired but using something other than oxygen as a final electron acceptor. In this scenario (considered in detail elsewhere), non-oxygenic “oxidative” chemistries came first, followed by the evolution of anoxic fermentative chemistries, then by photosynthesis, and finally by oxygenic respiratory pathways. In either scenario, we can assume that photosynthesis existed before oxygenic respiration.

We also assume that oxygenic photoautotrophs that capture free energy from light would become the most abundant autotrophs, if for no other reason than sunlight is always available (during the day!), and of course, O₂ is now abundant in the air! Early photoautotrophs were likely the ancestors of today's cyanobacteria, and in fact, a phylogenetic study of many genes—including genes in plastids and in nuclei that encode plastid proteins—along with a wealth of cyanobacterial genome data, suggests a common ancestry of freshwater cyanobacteria and eukaryotic chloroplasts (Ponce-Toledo, R.I. et al., 2017, *An Early-Branching Freshwater Cyanobacterium at the Origin of Plastids*. Current Biology 27:386-391).

Finally, think about the origins of respiratory metabolism and the endosymbiotic origins of mitochondria. How did cellular respiration co-opt the photosynthetic electron transport reactions that captured the electrons from H₂O needed to reduce CO₂ and then direct those reactions to the task of burning sugars back to H₂O and CO₂? As photosynthetic organisms emerged and atmospheric O₂ increased, higher O₂ levels would become toxic to most living things. Still, some autotrophic cells must have had a preexisting genetic potential to conduct detoxifying respiratory chemistry. These would have been facultative aerobes able to switch from photosynthesis to respiration as O₂ levels rose. Today's *purple non-sulfur bacteria* (e.g., *Rhodobacter sphaeroides*) are just such facultative aerobes! Perhaps we aerobes descend from the ancestors of *Rhodobacter*-like cells that survived and spread from localized environments, where small amounts of O₂ threatened their otherwise strictly anaerobic neighbors. Is it possible that the endosymbiotic critter that became the first mitochondrion in a eukaryotic cell was not just any aerobic bacterium, but a purple photosynthetic bacterium?

Some iText & VOP Key Words and Terms

active transport of protons	energy efficiency of glucose metabolism	PEP carboxylase
ATP synthase	energy flow in glycolysis	pH gradient
bacterial flagellum	energy flow in Krebs cycle	pheophytin
C4 photosynthesis	F1 ATPase	photosynthesis
Calvin cycle	FAD	photosystems
CAM photosynthesis	FADH ₂	proton gate
carotene	light-dependent reactions	proton gradient
chemiosmotic mechanism	light-independent reactions	proton pump
chlorophyll a	malic acid	PSI electron acceptor
chlorophyll b	mitochondria	PSII electron acceptor
complexes 1(I), 2(II), 3(III), 4(IV)	molecular motor	redox reactions

Crassulaceae	NAD ⁺	reduced electron carriers
cristal membrane	NADH	RUBISCO
cyclic photophosphorylation	NADP ⁺	RuBP
cytochromes	NADPH	splitting water
dark reactions	outer membrane	stoichiometry of glycolysis
electrochemical gradient	oxidative phosphorylation	stoichiometry of the Krebs cycle
electron transport system (chain)	P ₆₈₀	substrate-level phosphorylation
endosymbiotic theory	P ₇₀₀	Z-scheme

CHAPTER 7 WEB LINKS

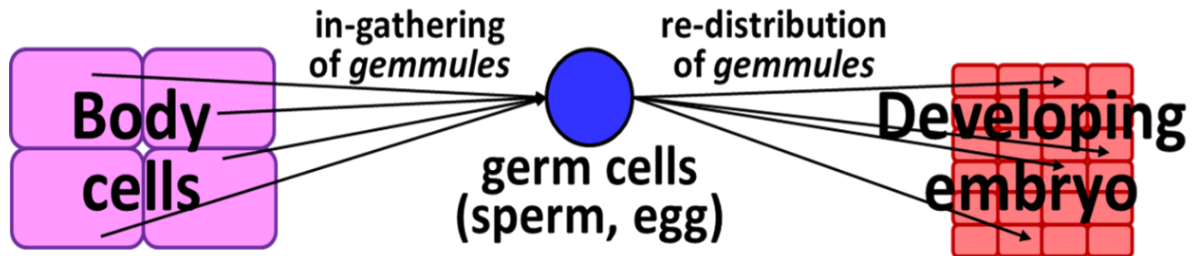


Chapter 8

DNA, Chromosomes and Chromatin

The “stuff of genes”; the double helix; chromosomes and chromatin;
naked, circular bacterial DNA

Reminder: For inactive *links*, google key words/terms for alternative resources.



Darwin proposes *pangenesis* of *gemmules* to explain natural selection

8.1 Introduction

In 1868, seeking a source for the hereditary variation to support evolution by natural selection, Charles Darwin came up with ***pangenesis*** (from the Greek *pan*, “all”; *genetikos*, “origins”) in which hereditary particles that he called *gemmules* (from Greek, meaning “small buds”) are gathered from all cells of the body cells, to be funneled into *germ cells* and later redistributed to body cells during embryogenesis.

In the 1890s, Hugo de Vries, at first unaware of Mendel's work independently rediscovered the Mendelian laws of heredity in plants, adopting and repurposing Darwin's term *pangenes* for Mendel's heritable factors. One of three scientists credited with the rediscovery of Mendel's work, he also coined the term ***mutation*** for novel traits that suddenly appeared in his plants. In 1905, William Bateson first defined ***genetics*** as the study of inheritance and in 1909, Wilhelm Johannsen, a Danish botanist, shortened *pangene* to ***gene***, also coining the terms ***genotype*** and ***phenotype*** to distinguish genes from the visible traits they determine. Finally, we knew by the start of the twentieth century that chromosome shapes were species-specific, that each species had a characteristic ***karyotype*** (from the Greek *karyo*, “nuclear”). Feel free to search for more historical details. This brief etymological history brings us to the topic of this chapter – the stuff of genes!

Here we look at classic experiments that led to our understanding that genes are composed of ***DNA***. We already knew that genes are on chromosomes (*chromo*: colored; *soma*: body). Early twentieth-century ***gene mapping*** had shown the relative location (***locus***) of genes on the linear chromosomes of eukaryotes. We'll see that eukaryotic chromosomes are highly ***condensed*** structures composed of DNA and protein, visible only during ***mitosis*** or ***meiosis***. During the much longer ***interphase*** portion of the eukaryotic ***cell cycle***, chromosomes decondense to ***chromatin***, a less-organized form of protein-associated DNA in the nucleus. In this chapter, we'll learn that eukaryotic chromatin is the gatekeeper of *gene activity*, a situation quite different from the case in bacterial cells. Unlike eukaryotic cells, bacteria contain a tiny amount of more simply organized DNA. We'll learn that bacterial *gene (mutation) mapping* revealed the bacterial “chromosome” to be a closed, naked, and circular DNA double helix.

Such knowledge of DNA structure and organization is essential to our understanding of how and when cells turn genes on and off. When we look at the *gene regulation* in an upcoming chapter, remember this and that the genetic content of eukaryotic as well as of prokaryotic organisms is species-specific.

Learning Objectives

When you have mastered the information in this chapter, you should be able to do the following:

1. Summarize the evidence that led to acceptance that genes are made of DNA.
2. Discuss how Chargaff's DNA *base ratios* support DNA as the "stuff of genes."
3. Interpret the findings of Griffith, of Avery et al. and of Hershey and Chase in their historical context.
4. Outline and explain how Watson and Crick built their model of a DNA double helix.
5. Distinguish between conservative, semiconservative, and dispersive replication.
6. Describe and/or draw the progress of a *viral infection*.
7. Trace the fate of $^{35}\text{SO}_4$ (*sulfate*) into proteins that are synthesized in cultured bacteria.
8. Distinguish between chromatin and chromosomes in terms of DNA organization and speculate on how this organization impacts replication.
9. List some different uses of karyotypes.
10. Compare and contrast euchromatin and heterochromatin structure and function.
11. Outline an experiment to purify *histone H1* from chromatin.
12. Formulate a hypothesis to explain why chromatin is found only in eukaryotes.
13. Describe the roles of different histones in nucleosome structure.
14. Explain the role of Hfr strains in mapping genes in *E. coli*.
15. Explain the chemical rationale of using different salt concentrations to extract 10 nm nucleosome filaments vs 30 nm solenoid structures from chromatin.

8.2 The Stuff of Genes

Mendel presented his Laws of Inheritance in 1865 but they were not widely understood—probably because they included a strong dose of math and statistics at a time when quantitative biology was not in vogue. By the late 1800s we knew that all eukaryotic cells contain a nucleus, shown by histological staining to contain mainly proteins and nucleic acids; the nucleus as the site genetic information was gaining traction. As the 19th century was ending and the 20th was beginning, the discovery paired chromosomes emerging from nuclei during mitosis and the rediscovery of Mendel's work on paired hereditary factors, set off an explosion of genetic studies that led to the understanding that any organism's observable traits are based on patterns of inheritance of paired genetic factors housed in the nuclei of cells. In honor of Gregor Mendel, we call this seminal insight *Mendelian Genetics*.

By 1901, Albrecht had discovered the four DNA bases (*adenine, thymine, cytosine, and guanine*) and uracil, the substitute for thymine in RNA, earning him the 1910 Nobel Prize in Physiology or Medicine. Yet for many years, DNA was thought to be just a small molecule made up of only its four bases. How could such a simple molecule account for the inheritance of so many different physical and physiological traits? Recall Beadle Tatum and Lederberg's ***one-gene/one-enzyme*** hypothesis, recognizing the inheritance of enzyme activities according to Mendelian rules. After morphing first into the ***one-gene/one-protein***, and then into the ***one-***

gene/one-polypeptide hypothesis, this relationship between genes and polypeptides supported proteins as the stuff of genes, and an enduring prejudice against DNA as the genetic material, despite early and growing evidence to the contrary.

Compared to DNA, only chains of up to twenty different amino acids in polypeptides and proteins seemed to have had the potential for enough structural diversity to account for the growing number of heritable traits in each organism. Proteins were arguably the most likely candidates for the molecules of inheritance. The experiments you'll read about here began around the start of World War I and lasted until just after World War II. The work of Phoebus Levene over this period revealed that DNA was no mere tetramer of deoxynucleotides but was in fact a long polymer (check it out at ^{8.1}[Long Nucleotide Polymers](#)). This led to the clever experiments that eventually forced the conclusion that DNA, not protein, was the genetic molecule, despite it being composed of just four monomeric units. Finally, we will look at the classic work of Watson, Crick, Franklin, and Wilkins, which revealed the structure of the genetic molecule.

8.2.1 Griffith's Experiment

Fred Neufeld, a German bacteriologist studying pneumococcal bacteria in the early 1900s, discovered three immunologically different strains of *Streptococcus pneumoniae* (types I, II and III). The virulent strain (type III) was responsible for much of the mortality during the **Spanish flu** (influenza) *pandemic* of 1918–20. This pandemic killed between twenty million and one hundred million people, commonly because of bacterial pneumonia, a *superinfection* by *Streptococcus pneumoniae*. Viral infection may weaken the immune defense against *S. pneumoniae*, and a recent study suggests how: the virus causes bronchial and alveolar capillary leakage increasing susceptibility of the lungs to bacterial infection (see ^{8.2}[Influenza Causes Bronchoalveolar Capillary Leakage](#)).

In the 1920s, Frederick Griffith was working with virulent **wild-type** (type III) and **benign** (type II) strains of *S. pneumoniae*. The two strains are easy to tell apart in petri dishes because the virulent strain grew as morphologically *smooth* (**S**) colonies, while the benign strain formed *rough* (**R**) colonies. It turns out that the *S* cells are coated with a polysaccharide (mucus) capsule, which makes their colonies appear smooth, and The *R* cell colonies look rough (i.e., don't glisten) because they lack the polysaccharide coating. Griffith knew that injecting mice with *S* cells killed them within about a day, and that injecting them with nonlethal *R* cells, on the other hand, caused no harm. He decided to test the hypothesis that exposure of mice to the *R* strain of *S. pneumoniae* might immunize them against lethal infection by subsequent exposure to *S* cells. To test his hypothesis, Griffith injected mice with *R* cells. Sometime later, he injected them with *S* cells. Unfortunately, the attempt to immunize the mice against *S. pneumoniae* was unsuccessful: both the control mice injected with *S* cells, and the experimental mice (that received the *R*-strain cells first and then *S* cells) all died in short order! As expected, mice injected only with *R* cells survived. Griffith also checked the blood of his mice for the presence of bacterial cells by growing the cells from the blood in petri dishes; he made the following observations:

- Mice injected with benign **R**-strain cells survived, and after he plated blood from the mice on a nutrient medium, no bacterial cells grew.

- Mice injected with S-strain cells died, and bacterial cell colonies grew from their blood. Griffith's experiments and results, published in 1928, are summarized in Figure 8.1.

Griffith's Discovery of Bacterial *Transformation*

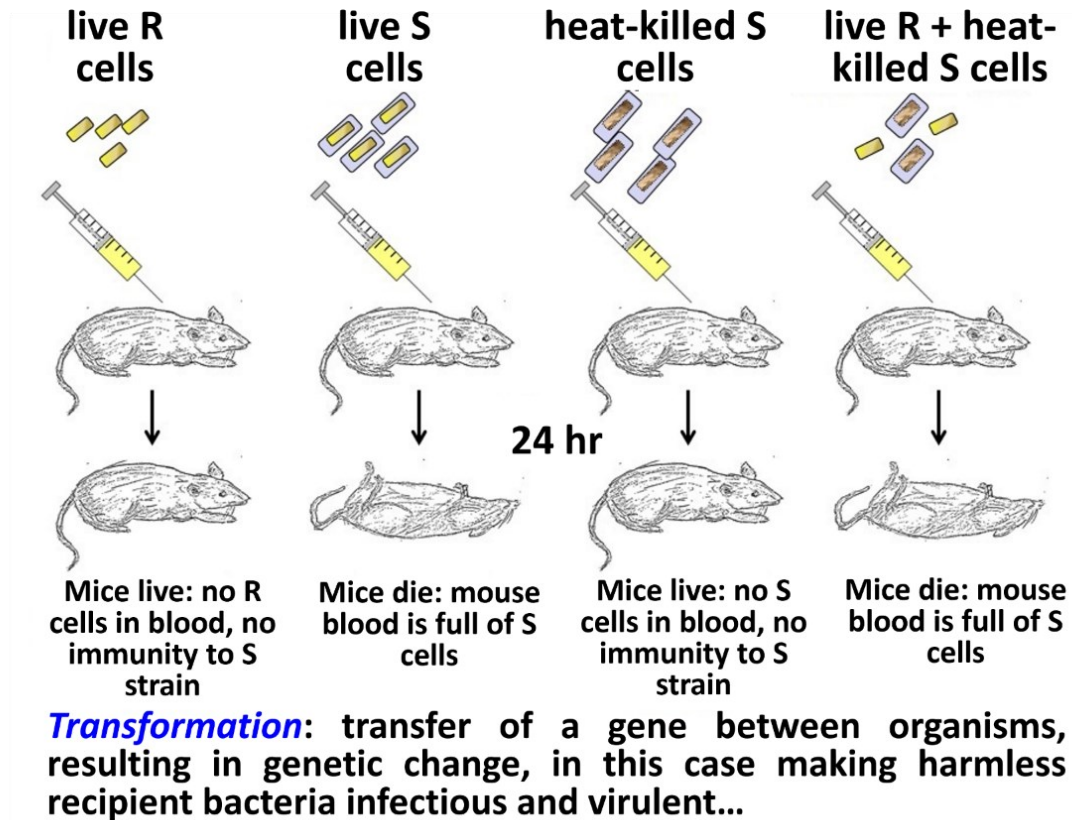


Fig. 8.1: Griffith did the experiments shown here, hoping to make antibodies against streptococcal pneumonia. Across the top are his experiments, with 2 control experiments at the left. The 2 key experiments are at the right. Neither of these yielded the desired antibodies. Instead, he discovered a chemical from dead cells that genetically transformed live bacteria (the experiment at the far right). He called the chemical a "transforming principle", what we now know as DNA.

Griffith then performed two other experiments, also shown in the illustration:

- He injected mice with heat-killed S cells; these mice survived. Blood from these mice contained no bacterial cells. This was expected, since heating the S cells would likely have the same effect as pasteurization has on bacteria in milk!
- Griffith also injected mice with a *mixture of live R cells and heat-killed S cells*, in the hope that the combination might induce immunity in the mouse, whereas injecting the R cells alone had failed. You can imagine his surprise when, far from being immunized against S-cell infection, the injected mice died, and abundant S cells had accumulated in their blood.

Griffith realized that something important had happened in his experiments: in the mixture of live R cells and heat-killed S cells, something released from the dead S cells had transformed some R cells. Griffith named this "something" the **transforming principle**—a molecule in the debris of dead S cells that could sometimes be acquired by a few live R cells and turn them into virulent S cells.

8.2.2 The Avery-MacLeod-McCarty Experiment

While Griffith didn't know the chemical identity of his *transforming principle*, his experiments led to studies that proved DNA to be **the stuff of genes**. With improved molecular purification techniques developed in the 1930s, Oswald Avery, Colin MacLeod, and Maclyn McCarty transformed R cells in vitro (that is, without the help of a mouse!). They purified heat-killed S-cell DNA, proteins, carbohydrates, and lipids and separately tested the transforming ability of each molecular component on R cells in a test tube. Since only the **DNA fraction** of the dead S cells could cause bacterial transformation, the three scientists concluded that DNA must be the *transforming principle*. Figure 8.2 summarizes their experiments. But despite the evidence, DNA was not yet readily accepted as the stuff of genes.

Avery et al. ask: What causes transformation? (What is the **transforming principle**)?

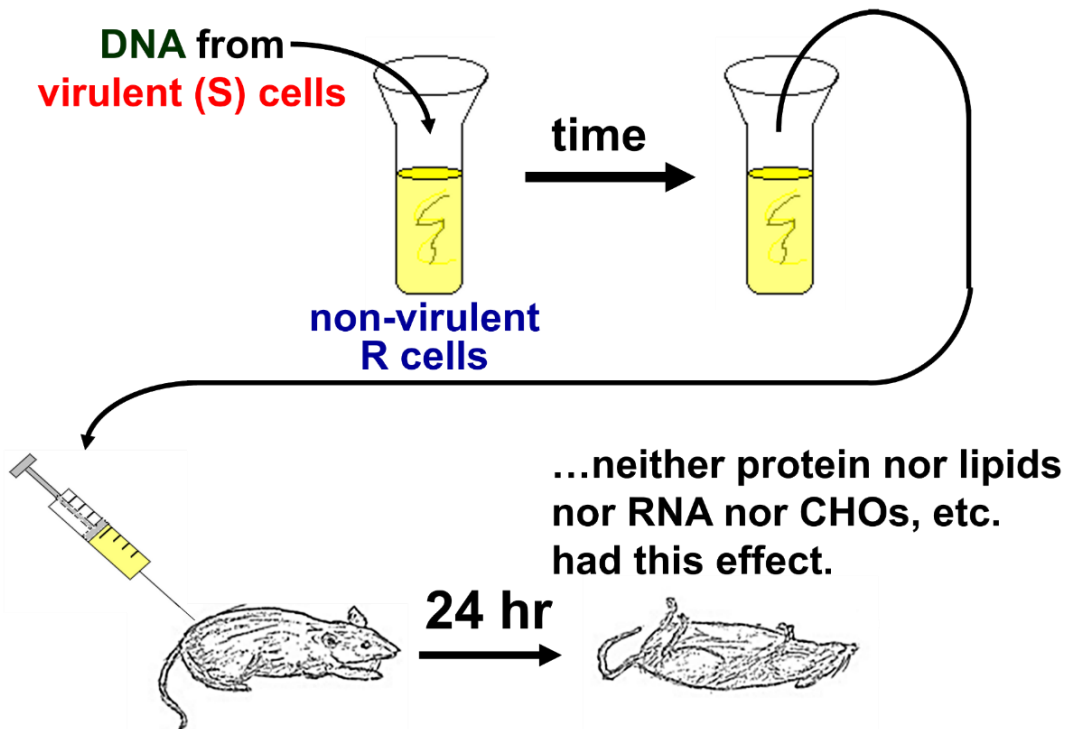


Fig. 8.2: The experiments of Avery et al. demonstrated for the first time that the chemical “stuff of genes” (i.e., of inheritance) is DNA.

The sticking point was that DNA is composed of only the four nucleotides: monophosphates of adenosine, cytidine, guanosine, and thymidine. Scientists knew that DNA was a large polymer, but they still thought of it as a simple molecule—a polymer made up of repeating sequences of the four bases (e.g., -A-G-C-T-A-G-C-T-A-G-C-T-A-G-C-T-A-G-C-T...).

It seemed that only the endless possible combinations of the twenty amino acids in proteins seemed to promise the biological specificity necessary to account for an organism's many genetic traits. Lacking structural diversity, DNA was explained as a mere scaffold for protein genes.

To adapt Marshall McLuhan's famous statement: "The medium is the message" (i.e., airwaves do not merely *convey* but *are* the message), many scientists still believed that proteins were the medium of genetic information *as well as* the functional message itself. The reluctance of influential scientists of the day to accept DNA as a *transforming principle* deprived its discoverers of the Nobel Prize stature they deserved. After new evidence made further resistance to that idea untenable, even the Nobel Committee admitted its error in failing to award a Nobel Prize for the discoveries of Avery, MacLeod, and McCarty. It took the experiments of Alfred Hershey and Martha Chase finally put to rest any notion that proteins were genes.



[167-2 Transformation In and Out of Mice—Griffith, McCarty, et al.](#)

8.2.3 The Hershey-Chase Experiment

Earlier we learned that viruses cannot live an independent life! A bacterial virus, called a bacteriophage (or just phage for short), consists of DNA inside a protein capsule. The life cycle of bacterial viruses is shown in Figure 8.3.

Structure & Life Cycle of a *Bacteriophage* (phage, or bacterial virus)

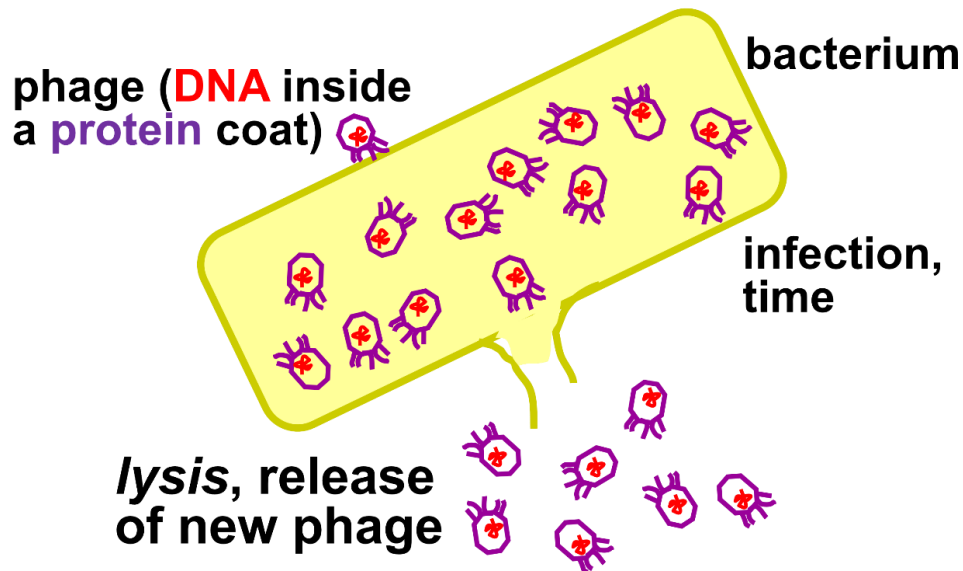


Fig. 8.3: Life cycle of a bacteriophage: the phage's coat remains attached to a cell after infection, but the chemicals inside the phage enter the infected cell.

Phage are inert particles until they bind to and infect bacterial cells. Phage added to a bacterial culture can be seen attached to bacterial surfaces with an electron microscope. Investigators found that they could detach phage particles from bacteria simply by agitation in a blender (like one you might have in your kitchen). Centrifugation then separated the bacterial cells in a pellet at the bottom of the centrifuge tube, leaving the detached phage particles in the supernatant.

By adding phage to bacteria, then detaching them from the bacteria at different times, investigators were able to determine how long the phage had to remain attached before the bacteria would become infected. It turned out that pelleted cells that had been attached to phage for short times would survive and reproduce when resuspended in a growth medium, but pelleted cells left attached to phage for longer times would become infected. When centrifugally separated from the detached phage particles and resuspended in fresh medium, these infected cells would go on to lyse, releasing new phage. Therefore, the transfer of genetic information for virulence (i.e., infectivity) from the phage to cells took some time. Furthermore, the viral genetic material responsible for infection was apparently no longer associated with the phage capsule, which could be recovered from the centrifugal supernatant.

Alfred Hershey and Martha Chase designed an experiment to determine whether the DNA that was enclosed by the viral *protein capsule* or the capsule protein itself caused a phage to infect the bacterium. In the experiment, they separately grew *Escherichia coli* cells infected with **T2 bacteriophage** in the presence of either ^{32}P or ^{35}S (radioactive isotopes of phosphorous and sulfur, respectively). The result was to generate phage that contained either radioactive DNA or radioactive proteins, but not both (recall that only DNA contains phosphorous and only proteins contain sulfur). They then separately infected fresh *E. coli* cells with either ^{32}P - or ^{35}S -labeled radioactive phage. Their experiment is illustrated in Figure 8.4.

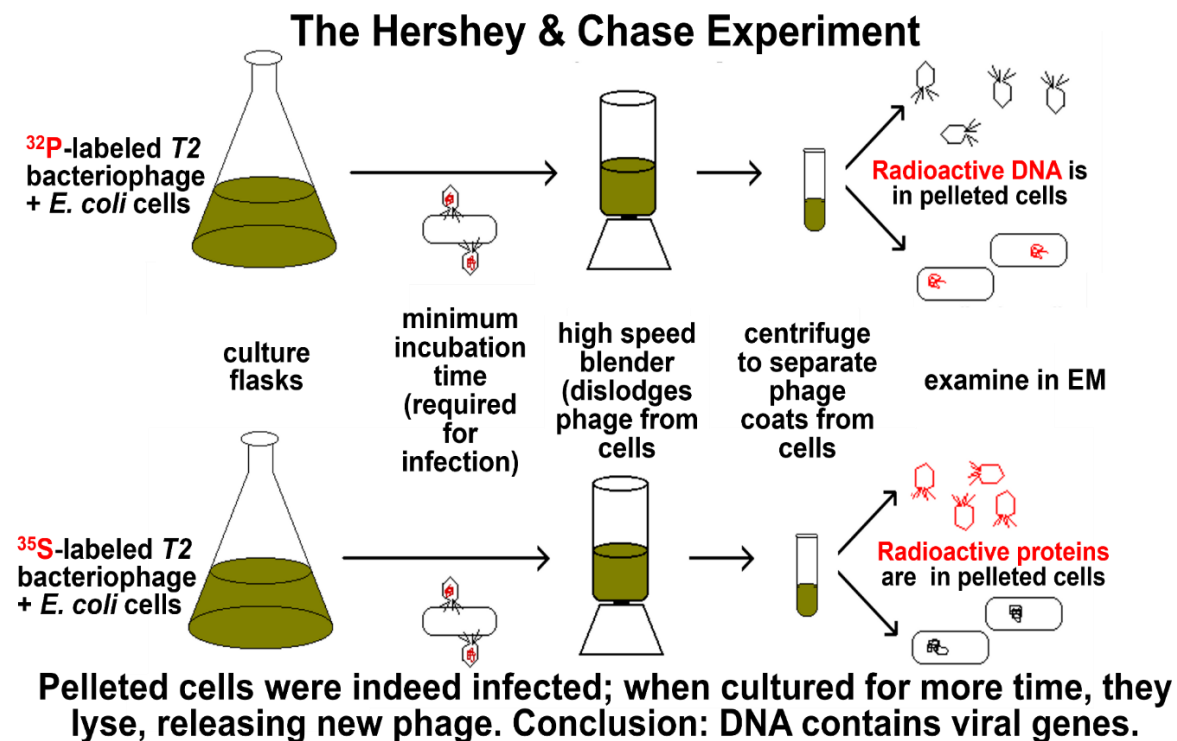


Fig. 8.4: Alfred Hershey and Martha Chase demonstrated that the chemical stuff of viral genes is DNA by showing that cells exposed to phage for the short time needed to become infected contained only the DNA from the virus, and not the viral proteins.

Phage and cells were incubated with either ^{32}P or ^{35}S just long enough to allow infection. Some of each culture was allowed to go on and lyse to prove that the cells were infected. The remainder of each mixture was sent to the blender. After centrifugation of each blend, the pellets and supernatants were examined to see where the radioactive proteins or

DNA had gone. From the results, the ^{32}P always ended up in the pellet of bacterial cells, while the ^{35}S was found in the phage remnant in the supernatant. Hershey and Chase concluded that the genetic material of bacterial viruses was DNA, not protein, and that—just as Avery's group had suggested—DNA was the bacterial transforming principle. Given the earlier resistance to “simple” DNA being the genetic material, Hershey and Chase used cautious language in framing their conclusions. They need not have; all subsequent experiments confirmed that DNA was the genetic material. Concurrent with these confirmations were experiments demonstrating that DNA might not be (indeed, was not) such a simple, uncomplicated molecule!

For their final contributions to pinning down DNA as the “stuff of genes,” Hershey shared the 1969 Nobel Prize in Physiology or Medicine with Max Delbrück and Salvador E. Luria. So why didn't Martha Chase get a share of the recognition in the Nobel Prize, given that she was Hershey's sole coauthor on the paper presenting their findings? You may well ask!



[168 Hershey and Chase: Viral Genes Are in Viral DNA](#)

In a later chapter you will read about our benevolent (even crucial) *microbiomes* and the phenomenon of *horizontal gene transfer* (essentially, what Griffith discovered). Under the rubric *not only in the lab*, you can read about the role of horizontal gene transfer how gut bacteria can become harmful in ^{8.3} [A link to the full paper is in *Nature Communications* at *Not Only In the Lab: Gut Bacteria Breaking Bad*.](#)

8.3 DNA Structure

By 1878 a substance in the pus of wounded soldiers derived from cell nuclei (called *nuclein*) was shown to be composed of five bases (the familiar ones of DNA and RNA). The four bases known to make up DNA (as part of nucleotides) were thought to be connected through the phosphate groups, in short repeating chains of four nucleotides. By the 1940s, we knew that DNA was a long polymer. Nevertheless, it was still considered too simple to account for genes. After the Hershey and Chase experiments, only a few holdouts would not accept DNA as the genetic material. When the scientific community finally accepted DNA as the *stuff of genes*, the next questions were the following:

- What did DNA look like?
- How did its structure account for its ability to encode and reproduce life? While the four phosphate-linked nucleotide composition of DNA was known for some time, it became mandatory to explain how such a “simple molecule” could inform the thousands of proteins necessary for life. The answer to this question lay (at least in part) in an understanding of the physical structure of DNA, made possible by the advent of **X-ray crystallography**.

8.3.1 X-Ray Crystallography and the Beginnings of Molecular Biology

If a substance can be crystallized, the crystal will diffract X-rays at angles revealing regular (repeating) structures of the crystal. William Astbury demonstrated that high molecular-weight

DNA had just such a regular structure. His *crystallographs* suggested DNA to be a linear polymer of stacked bases (nucleotides), each separated from the next by 0.34 nm. Astbury is also remembered for coining the term "*molecular biology*" to describe his studies. The term now covers all aspects of biomolecular structure, encompassing all of the topics covered in this book!

In an irony of history, the Russian biologist Nikolai Konstantinovich Koltsov had already intuited in 1927 that the basis for the genetic transfer of traits would be a "giant hereditary molecule" made up of "two mirror strands that would replicate in a semiconservative fashion using each strand as a template" (^{8.4}[The Remarkable Nikolai Koltsov](#) - scroll down to read the biography). This was a pretty fantastic inference if you think about it, since it was proposed long before Watson, Crick, and their colleagues worked out the structure of the DNA double-helix!

8.3.2 Wilkins, Franklin, Watson, and Crick—DNA Structure Revealed

Maurice Wilkins, an English biophysicist at King's College, London, had obtained samples of a highly purified, very high molecular weight DNA "gel" (or "fibers") that he obtained from a Swiss colleague, Rudolph Signer. Wilkins' first X-ray diffraction studies with these fibers yielded good results. After joining Wilkin's lab, Rosalind Franklin and her assistant, Raymond Gosling, produced even higher resolution X-ray diffraction images of Signer's fibers, among which was her most famous (and definitive) crystallograph, *Photo 51* (Figure 8.5).

Photo 51

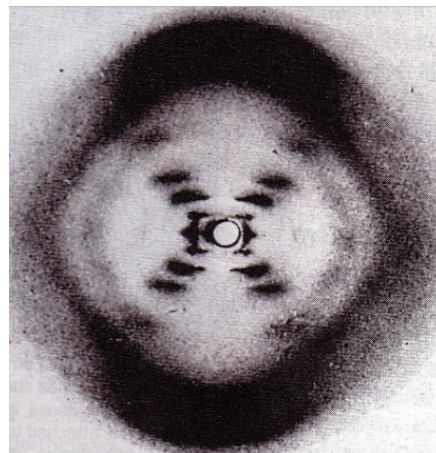


Fig. 8.5: *Photo 51*, the X-ray crystallograph of DNA taken by Rosalind Franklin, revealed three crucial molecular dimensions that ultimately led to the double-helical DNA model.

This image confirmed Astbury's **0.34 nm** repeated dimension and revealed two more numbers, **3.4 nm** and **2 nm**, reflecting additional repeat structures in the DNA crystal. These numbers and other evidence suggested to Franklin that DNA's negatively charged phosphates were on the outer surface of what was likely a helix. At some point, Wilkins shared *Photo 51* with James Watson and Francis Crick, who instantly realized its significance, used it along with other data to build DNA models out of nuts, bolts, and tubes. Their models eventually revealed DNA to be a pair of **antiparallel, complementary** strands of nucleic acid polymers — shades of Koltsov's mirror-image macromolecules! Each strand is a string of nucleotides linked

by *phosphodiester linkages*, the two strands held together in a double helix by complementary H-bond interactions. Let's look at the evidence for these conclusions. As we do, refer to the two illustrations of the double helix below, in Figure 8.6.

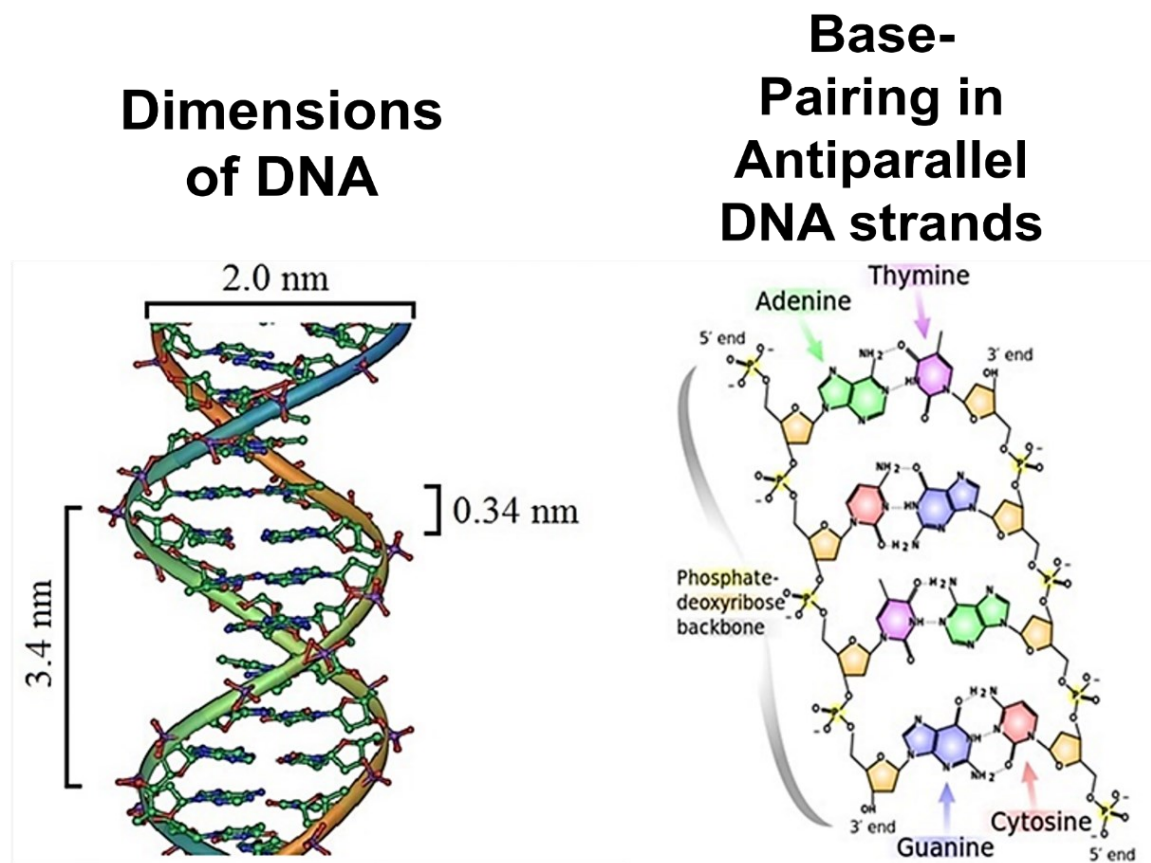


Fig. 8.6: The three molecular dimensions from Franklin's Photo 51 could be best explained by the proposition of a DNA double helix (left). Model-building by Watson and Crick revealed the antiparallel structure of the double helix (right).

Recalling that Astbury's 0.34 nm dimension was the *distance between successive nucleotides* in a DNA strand, Watson and Crick surmised that the 3.4 nm repeat was a structurally meaningful tenfold multiple of Astbury's number. When they began building their DNA models, Watson and Crick confirmed from the bond angles connecting the nucleotides, that the DNA strands were forming a helix, from which they suggested that the 3.4 nm repeat was detecting the *pitch* of the helix (i.e., the length of one complete turn of the helix). This meant that there were ten bases per turn of the helix. They further suggested that the 2.0 nm number might reflect the diameter of the helix. When their scale models of a single-stranded DNA helix predicted a helical diameter much less than 2.0 nm, they were able to confirm an intertwined double-stranded DNA in their model (their *double helix*) that more nearly met the requirement for a 2.0 nm diameter. In building this double helix, Watson and Crick realized that bases in opposing strands would indeed come together within the double helix to form H-bonds, holding the helices together. However, for their double helix to have a constant diameter of 2.0 nm, they also realized that the smaller **pyrimidine** bases, **thymine** (T) and **cytosine** (C), would have to H-bond to the larger **purine** bases, **adenine** (A) and **guanine** (G), and that neither A-G nor C-T pairing could account for a uniform 2.0 nm diameter along the length of the double helix.

And now to the question of how a “simple” DNA molecule could have the structural diversity needed to encode thousands of different polypeptides and proteins. In early studies, purified *E. coli* DNA was chemically hydrolyzed down to nucleotide monomers. The hydrolytic products contained nearly equal amounts of each base, reinforcing the notion that DNA was that simple molecule that could not encode genes. But Watson and Crick had private access to data from Erwin Chargaff, who had determined the base composition of DNA isolated from different species, including *E. coli*. He found that the base composition of DNA from different species was not always *equimolar*, meaning that for some species, the DNA was not composed of equal amounts of each of the four bases (see some of this data in Table 8.1).

Table 8.1

Composition of DNA from Different Organisms				
base		human	yeast	fly
purine	A	27%	21%	15%
pyrimidine	C	23%	32%	35%
purine	G	23%	32%	35%
pyrimidine	T	27%	21%	15%

The mere fact that Chargaff had shown that DNA from some species could have base compositions that deviated from equimolarity put to rest the argument that DNA had to be a very simple sequence. Finally, it was safe to accept the obvious, namely that DNA sequences could vary almost infinitely and could indeed be *the stuff of genes*. Even more convincing was his data showing unique patterns of base ratios. Although base compositions could vary between species, the **A/T** and **G/C ratios** were always 1 for every species; likewise, the ratio of **(A+C)/(G+T)** and **(A+G)/(C+T)**. From this information, Watson and Crick inferred that **A** (a purine) would H-bond with **T** (a pyrimidine), and that **G** (a purine) would H-bond with **C** (a pyrimidine) in the double helix. Building their model with this new information, they also confirmed that H-bonding between the complementary bases would be maximal only if the two DNA strands were **antiparallel**, leading to the most stable structure of the double helix. Watson and Crick published their paper *A Structure for Deoxyribose Nucleic Acid* in 1953 in *Nature* (read an annotated version of this seminal article at ^{8.5}[The Classic 1953 Watson-Crick](#)). Their article is also famous for predicting a *semiconservative* mechanism of replication, something that had been predicted by Koltsov twenty-six years earlier, albeit based on intuition... and much less evidence! Watson, Crick, and Wilkins shared a Nobel Prize in 1962 for their work on DNA structure. Franklin was never nominated for her work on DNA structure, and since she died in 1958, her chances ended since Nobel prizes are not awarded posthumously. In the years between 1953 and 1958, Franklin had left King's College for Cambridge where she studied *Tobacco Mosaic Virus*, an RNA virus. By X-ray crystallography, she showed that the RNA was inside the virus, interwoven with the viral coat proteins. Posthumously she *did* get delayed recognition, including a named building at King's College and later, a whole Chicago university. There remain questions about why Franklin did not get more timely credit for her role in the work, but a BBC documentary at ^{8.6}<https://www.youtube.com/watch?v=FpZQvDFQpmg> sheds some light in the controversy.



[169-2 Unraveling the Structure of DNA](#)

8.3.3 Meselson and Stahl's Experiment—Replication is Semiconservative

Confirmation of Watson and Crick's suggestion of semiconservative replication came from Matthew Meselson and Franklin Stahl's very elegant experiment, which tested the three possible models of replication (Figure 8.7)

Models of Replication

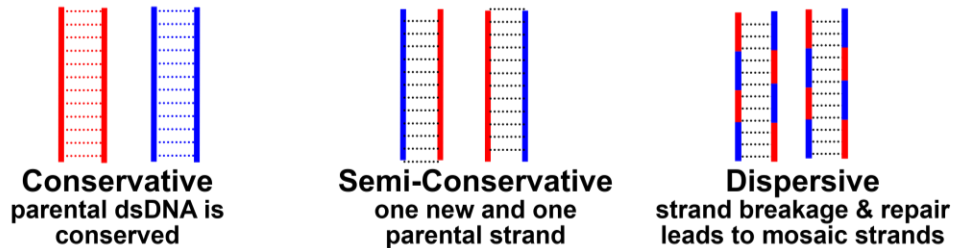


Fig. 8.7: Three possible modes of DNA replication could be imagined.

In their experiment, *E. coli* cells were grown in a medium containing ^{15}N , a "heavy" nitrogen isotope. After many generations, the DNA in all of the cells had become labeled with the heavy isotope. At that point, the ^{15}N -tagged cells were placed back into a medium containing the more common, "light" ^{14}N isotope and were allowed to grow for exactly one generation. Figure 8.8 shows Meselson and Stahl's predictions for their experiment. Meselson and Stahl knew that ^{14}N -labeled and ^{15}N -labeled DNA would *form separate bands* after centrifugation on *CsCl* (cesium chloride) density gradients.

Meselson & Stahl Experiment & Predictions

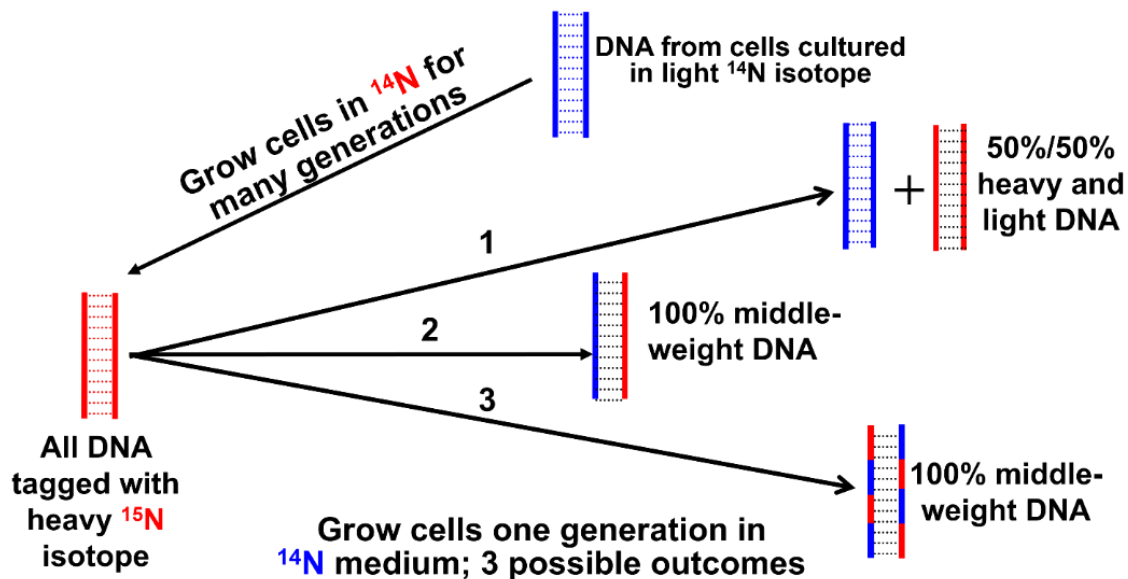


Fig. 8.8: Meselson and Stahl predicted experimental outcomes based on possible modes of replication.

They tested their predictions by purifying and centrifuging the DNA from the ^{15}N -labeled cells grown in the ^{14}N medium for one generation. They found that this DNA formed a one band with a density between that of ^{15}N -labeled DNA and ^{14}N -labeled DNA, eliminating a *conservative* DNA replication model (possibility #1), as Watson and Crick had also predicted.

That left two possibilities: replication was either semiconservative (possibility #2) or dispersive (possibility #3). The dispersive model was *eliminated* when DNA isolated from cells that had been grown for a second generation on ^{14}N were shown to contain two bands of DNA on the CsCl density gradients.



[170 Replication is Semiconservative](#)



8.4 Chromosomes

It was clear by the twentieth century that eukaryotic nuclei were somehow involved in inheritance and further, that the behavior of paired chromosomes in meiosis mimicked the behavior of Mendel's heritable factors. Hence, the *chromosome theory of heredity*, which posits that chromosomes contain genes. Let's take a brief look at how we came to know this. As with much of genetics, it begins with Mendel! He was unable to account for deviations from his *Law of Independent Assortment*—e.g., an unexpected green pea plant among the many yellow pea plants expected from a genetic cross. But, in the early 1900s, genetic studies of *Drosophila melanogaster* in Thomas Hunt Morgan's lab showed that Mendel's misbehaving pea plant traits were due to recombination (crossing over) between bits of homologous chromosomes during meiosis. Furthermore, these crossings over were frequent. In fact, the recombination frequency of a given pair of genes was reproducible, effectively a constant. This suggested that such gene pairs were linked on the same chromosome, consistent with *chromosome theory*. For this work, Morgan earned the 1933 Nobel Prize for Physiology or Medicine. He further suggested that genes far apart on a chromosome would cross over more often than those closer together. Alfred H. Sturtevant (one of Morgan's students) tested the possibility that recombination frequencies might be used to map the position (*locus*) of genes on chromosomes. He confirmed this by generating the first maps of *D. melanogaster*'s four chromosomes! See ^{8.7}[D. melanogaster Gene Mapping](#) (or I. Lobo & K. Shaw, 2008; T. H. Morgan, Genetic recombination, and gene mapping. *Nature Educ.* 1:205) for more details.) Let's look at a typical chromosome (Figure 8.9).

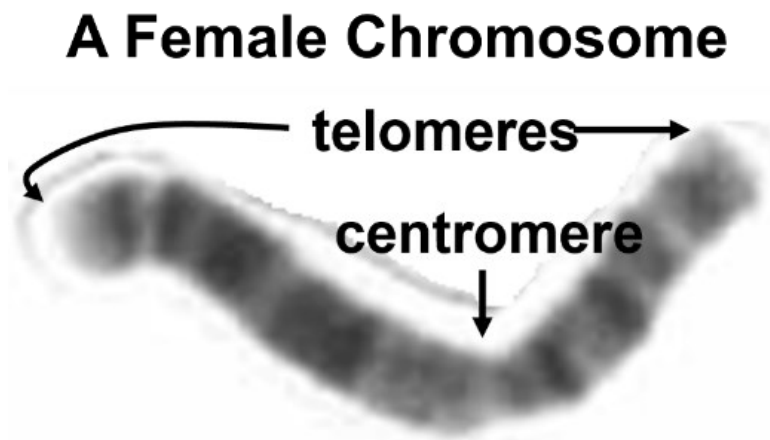
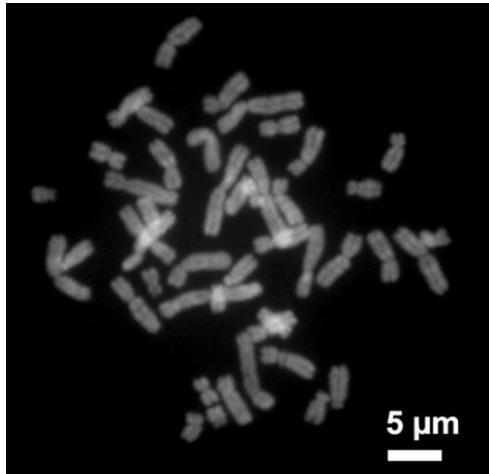


Fig. 8.9: A female chromosome stained for banding and defining a *centromere* and two *telomeres*.

Late nineteenth-century microscopists saw chromosomes condense a dispersed state as the nucleus broke down during mitosis or meiosis. They were visible as they moved to opposite

poles of the cell during cell division. Look at chromatids separating in mitotic anaphase in the computer-colored micrograph at ^{8.8}[Chromatid Separation in Anaphase](#).

It's possible to distinguish one chromosome from another by **karyotyping**. Cells in metaphase of mitosis placed under pressure will burst and the chromosomes spread out. Figure 8.10 shows such a chromosomal spread.



Human Chromosomes

Mitotic chromosomes are spread by bursting cells in mitosis and staining to reveal banding and morphological variation along their lengths.

Fig. 8.10: Spread of human mitotic cell chromosomes.

Such spreads showed that chromosomes come in matched pairs, again parallel to Mendel's paired heritable factors. By the early 1900s, the number, sizes, and shapes of chromosomes were known to be species-specific. Cutting apart micrographs like the ones in Figures 8.10 and 8.11, and then pairing the chromosomes by their morphology, generates a **karyotype**. Paired human homologs are easily identified in the modern colorized micrograph (Figure 8.11).

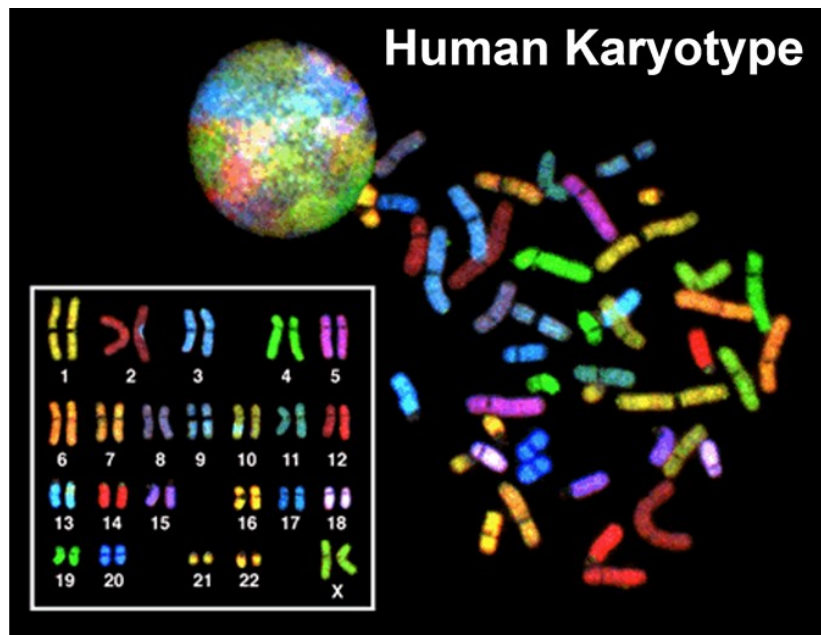


Fig. 8.11: Computer-colored human chromosome spread (right) and **karyotype** (inset at left). An intact nucleus is seen above the karyotype.

As seen in cells undergoing mitosis, all dividing human cells contain twenty-three pairs of homologous chromosomes. The karyotype in Figure 8.11 is from a female; note the pair of homologous sex (X) chromosomes (lower right of the inset). X and Y chromosomes in males are not truly homologous. Chromosomes in the original spread and in the aligned karyotype were stained with fluorescent antibodies against chromosome-specific DNA sequences to "light up" the different chromosomes.



[171-2 DNA, Chromosomes, Karyotypes, & Gene Maps](#)

8.5 Genes and Chromatin in Eukaryotes

Chromosomes and chromatin are a uniquely eukaryotic association of DNA with proteins of different types and amounts. Bacterial DNA (as well as prokaryotic DNA generally) is relatively "naked"—that is, not visibly associated with protein.

The electron micrograph of an *interphase* cell (Figure 8.12) reveals that the chromatin can itself exist in various states of condensation.

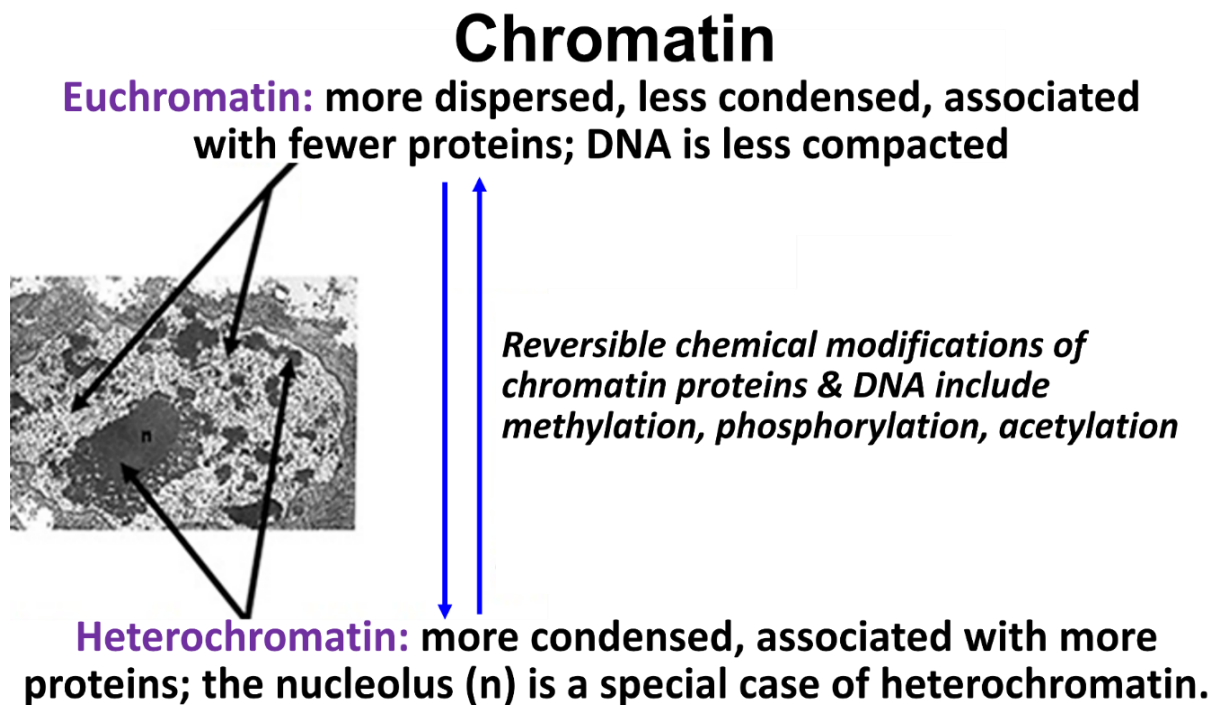


Fig. 8.12: Transmission micrograph of *euchromatin* and *heterochromatin* in the nucleus.

Chromatin is maximally condensed during mitosis to form chromosomes. During interphase, chromatin exists in more- or less-condensed forms, called **heterochromatin** and **euchromatin**, respectively. Transition between these chromatin forms involves changes in the amounts and types of proteins that can bind to the chromatin during gene regulation (i.e., when genes are turned on or off). Experiments (to be described later) showed that active genes tend to be in the more dispersed euchromatin, where enzymes of replication and transcription

have easier access to the DNA. Transcriptionally *inactive* genes are heterochromatic, obscured by additional chromatin proteins present in heterochromatin.

We can define three levels of chromatin organization in general terms:

1. DNA is wrapped around histone proteins (*nucleosomes*) like "beads on a string."
2. Multiple nucleosomes are coiled (condensed) into 30 nm fiber (solenoid) structures.
3. The 30 nm fibers are packed in higher order to form the familiar metaphase chromosome.

These aspects of chromatin structure were determined by gentle disruption of the nuclear envelope of nuclei, followed by salt extraction of extracted chromatin. Salt extraction dissociates most of the proteins from the chromatin. The results of a low-salt extraction are shown in Figure 8.13.

A Chromatin Extraction

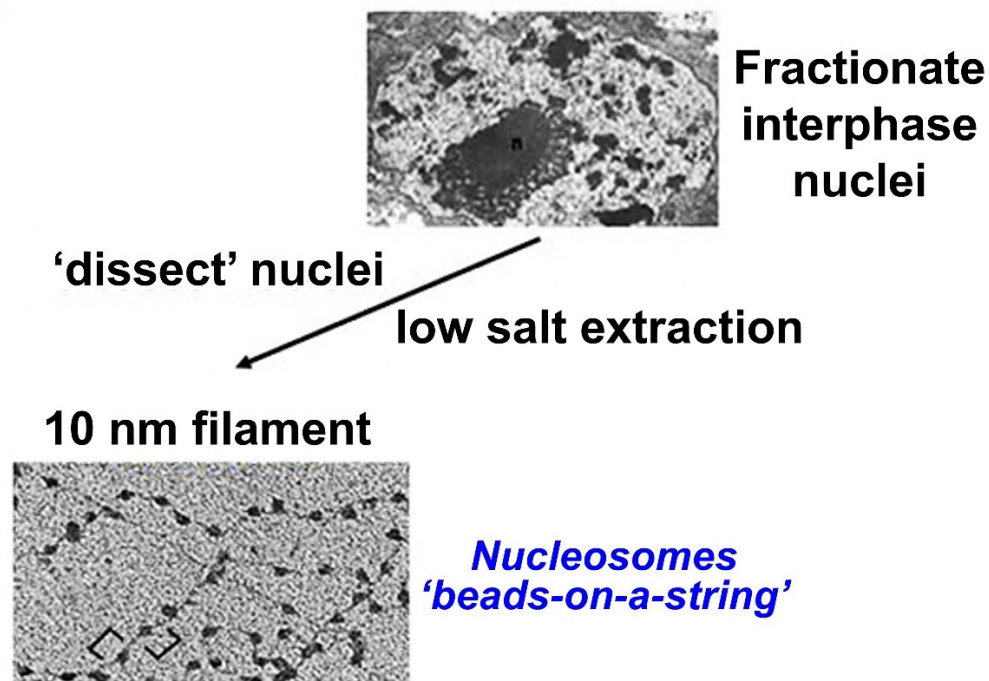


Fig. 8.13: Low-salt fractionation of interphase nuclei yields 10 nm nucleosome *beads on a string*.

When the low-salt extract is centrifuged and the pellet is resuspended, the remaining chromatin looks like a 10 nm filament of attached nucleosomes, or *beads on a string*. DNA-wrapped *nucleosomes* are the beads, which are in turn linked by uniform lengths of the metaphorical DNA "string."

After a high-salt chromatin extraction, the structure visible in the electron microscope is the 30 nm *solenoid*, the coil of nucleosomes, themselves a polymer of histone proteins encoiled (wrapped up in) DNA, as modeled below in Figure 8.14. The effect of the higher [salt] is to change the chromatin conformation by re-aggregating nucleosomes with non-histone nuclear proteins. As shown in the illustration, increasing the salt concentration of an already-extracted nucleosome preparation will cause the "necklace" to re-fold into the 30 nm solenoid structure.

Alternate Chromatin Extractions

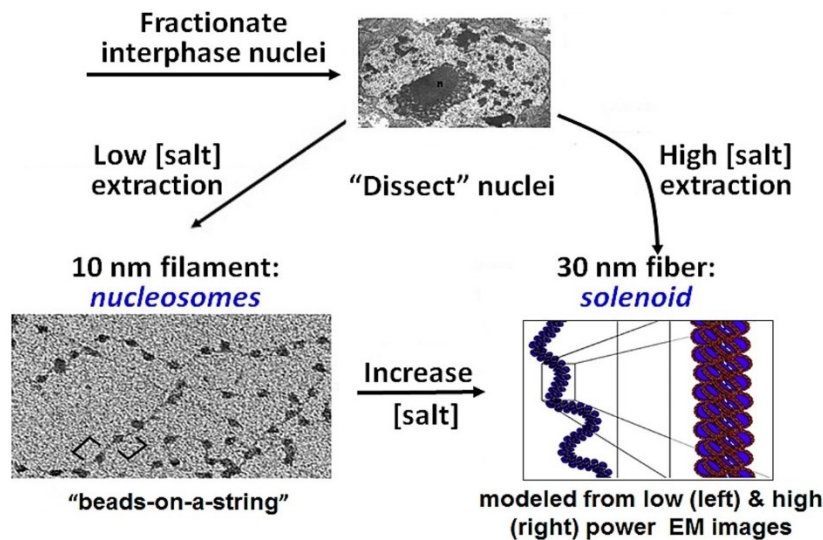


Fig. 8.14: High-salt chromatin extraction from nuclei, or high-salt treatment of 10 nm filaments selectively removes chromatin proteins, yielding 30 nm *solenoid* structures, essentially coils of 10 nm filaments. The solenoid can also form by adding salt to isolated 10nm filament extracts.

In fact, there are at least five levels (*orders*) of chromatin structure. These are illustrated below in Figure 8.15.

Organization of DNA in Chromatin

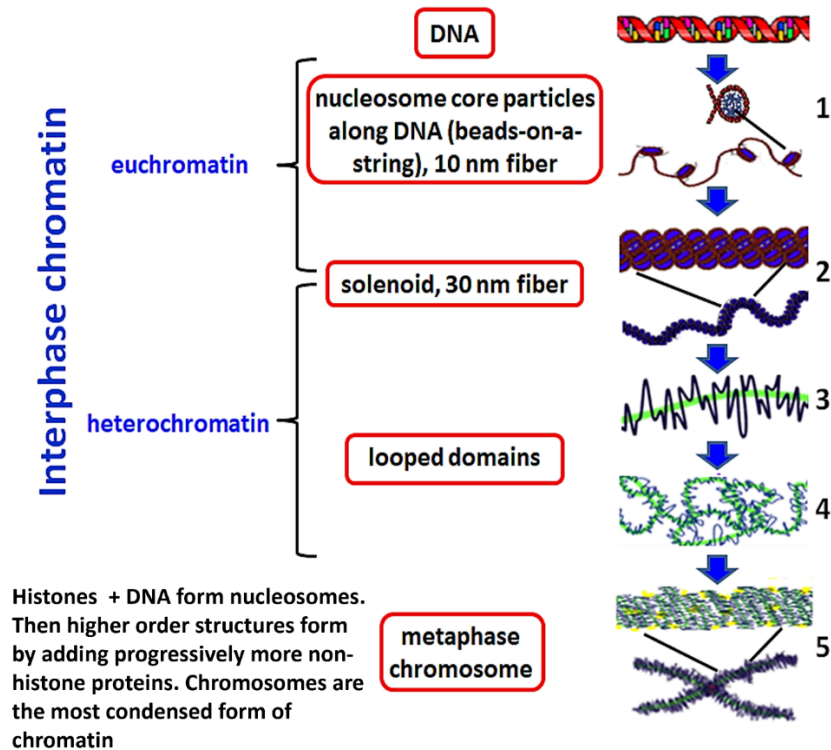


Fig. 8.15: Five different levels (orders) of chromatin structure.

The first two orders, as we have discussed, are the *string of beads* (#1) and the *solenoid* (#2). Other extraction protocols reveal other orders of chromatin structure (#3 and #4). Metaphase chromosomes of mitosis (#5) are the “highest order,” or most condensed form of chromatin.

The results of deoxyribonuclease (**DNase I**) digestion of a beads-on-a-string extract are shown in Figure 8.16. **DNase I** briefly digests nucleosome “necklaces,” degrading DNA between the “beads” and leaving behind shortened, different length 10 nm filaments. After a longer digestion, only single beads (nucleosomes) remain, bound to small amounts of DNA.

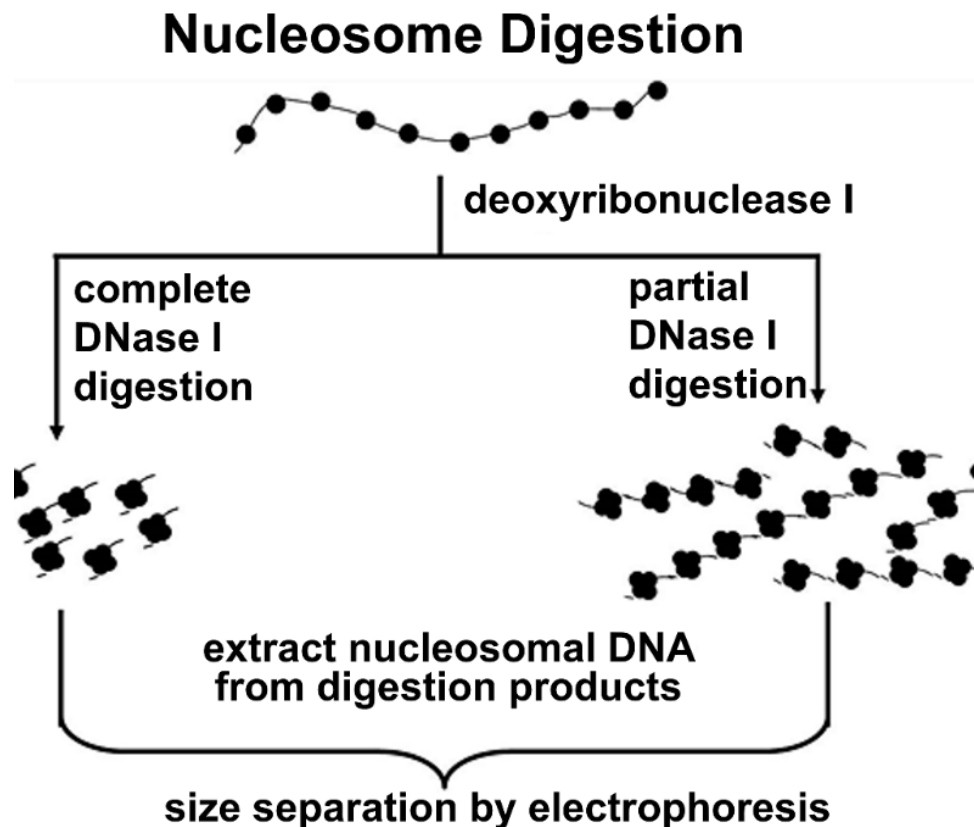


Fig. 8.16: Short digestion times with DNase I leave behind shortened 10 nm filaments (or *beads on a string*). Longer digestion releases single nucleosomes associated with some DNA.

Roger Kornberg, one son of Nobel laureate Arthur Kornberg (discoverer of the first DNA polymerase enzyme of replication—see the next chapter), participated in the discovery and characterization of nucleosomes while still a postdoc! He discovered that each nucleosome consisted of histone proteins and about two hundred base pairs of DNA. Here are some details about his experiments:

After DNA extracts of a brief DNase I digest of nucleosome ‘beads,’ the nucleosomes remained. But about 80 nucleotides of DNA had disappeared, suggesting that *nucleosomes* are separated by a DNA “linker” of about eighty base pairs. DNA extracted from nucleosomes themselves was about 147 base pairs long, a closer estimate of the DNA wrapped around the proteins of the nucleosome. Five *histone* proteins could be isolated from nucleosomes and separated by electrophoresis; the results are illustrated at the left in Figure 8.17 (below).

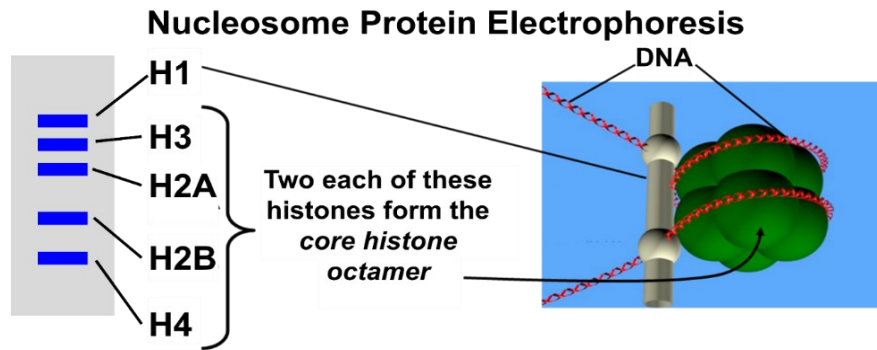


Fig. 8.17: Electrophoretic separation of five histones extracted from 10 nm filaments (left) and the role of histones in the nucleosome.

Histones are basic proteins, with many *lysines* and *arginines*, whose positively charged side chains enable them to bind to the acidic negatively charged, acidic *phosphodiester backbone* of double helical DNA. The DNA wraps around octamers, each of which is composed of two out of four histones (H3, H2A, H3B, H4), to form the *nucleosome* (Figure 8.17, right). Histone H1 is a linker between nucleosomes. About a gram of histones associates with each gram of DNA.



[172-2 Nucleosomes: DNA and Protein](#)



[173-2 Chromatin Structure: Dissecting Chromatin](#)

As you might guess, acid extraction of chromatin should selectively remove the basic histone proteins. An electron micrograph of the chromatin remnant after an acid extraction of metaphase chromosomes is shown below in Figure 8.18.

Acid extraction removes histones from metaphase chromosomes, leaving behind non-histone proteins:

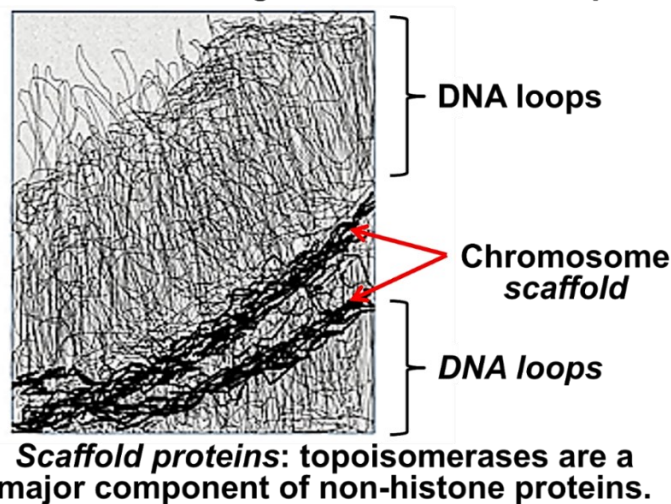


Fig. 8.18: Acid extraction of chromatin removes histones, leaving nonhistone proteins behind. (See the original micrograph at ^{8.9}[Acid Extract of Chromatin](#).)

DNA freed of the regularly spaced histone-based nucleosomes loops out and away from the long axis of the chromatin, leaving behind an association of DNA with non-histone proteins. The dark material along this axis is a protein scaffold, most of which is made up of *topoisomerase*, an enzyme that prevents DNA from breaking apart under the strain of replication, as we will see in detail later.

8.6 Structure and Organization of Bacterial DNA... and Bacterial Sex

Sexual reproduction allows compatible genders (think *male* and *female*) to share genes, a strategy that increases species diversity. It turns out that bacteria and other single-celled organisms can also share genes and spread diversity with the help of plasmids. Plasmids are circular, extrachromosomal bits of DNA that contain their own genes (often, antibiotic-resistance genes), and which replicate themselves in bacteria. We'll close this chapter with a look at sex (*E. coli* style!) and plasmid-assisted gene-mapping experiments showing linearly arranged genes on a circular bacterial DNA molecule (the bacterial "chromosome").

E. coli sex begins when F^+ and F^- cells meet. These "opposite" mating type cells can share DNA during **conjugation**. F^+ cells contain the extrachromosomal **fertility plasmid**, or **F plasmid**, which is separate from the *E. coli* chromosome. The *F plasmid* has genes which encode **sex pili** on F^+ cells, as well as factors needed to form a "mating bridge," or **conjugation tube**. Figure 8.19 (below) illustrates the behavior of the *F plasmid* during conjugation.

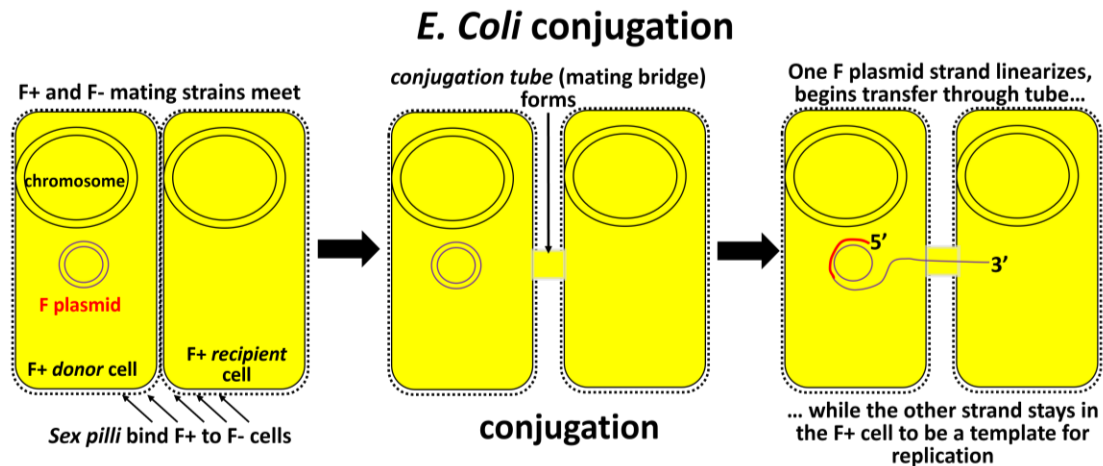


Fig. 8.19: Bacterial *conjugation* (sex in bacteria): *F plasmids* (Fertility plasmids) transfer chromosomal DNA from *E. coli* F^+ mating strain cells to F^- strain cells, leading to genetic diversity.

When an F^+ (*donor*) cell meets an F^- (*recipient*) cell, donor cell sex pili initiate recognition. A conjugation tube then forms, linking the cytoplasm of both cells. After the donor F^+ cell nicks one strand of the *F plasmid* DNA, the nicked strand enters the conjugation tube, rolling into the F^- cell. Both the "traveling" DNA strand and the intact DNA circle remaining in the donor cell replicate (shown in red in Figure 8.19). *E. coli* conjugation can have different outcomes:

1. One of two semi-conservatively replicated *F plasmids* stays in the donor cell, and the other ends up in the recipient cell, making the recipient cell a new F^+ *donor* cell!
2. The *F plasmid* inserts itself into recipient cell's chromosomal DNA, typically at a specific site in the DNA where there is sufficient sequence similarity between the plasmid and the chromosomal DNA. This allows insertion by *recombination*. With the fertility-factor genes

now part of its chromosome, the recipient cell becomes an *Hfr* (*High-frequency recombination*) cell, and it will produce *Hfr-strain* progeny cells.

In both outcomes the mating type of the recipient F^- cell becomes either an F^+ or an *Hfr* cell that can initiate mating with other F^- cells as illustrated below in Figure 8.20.

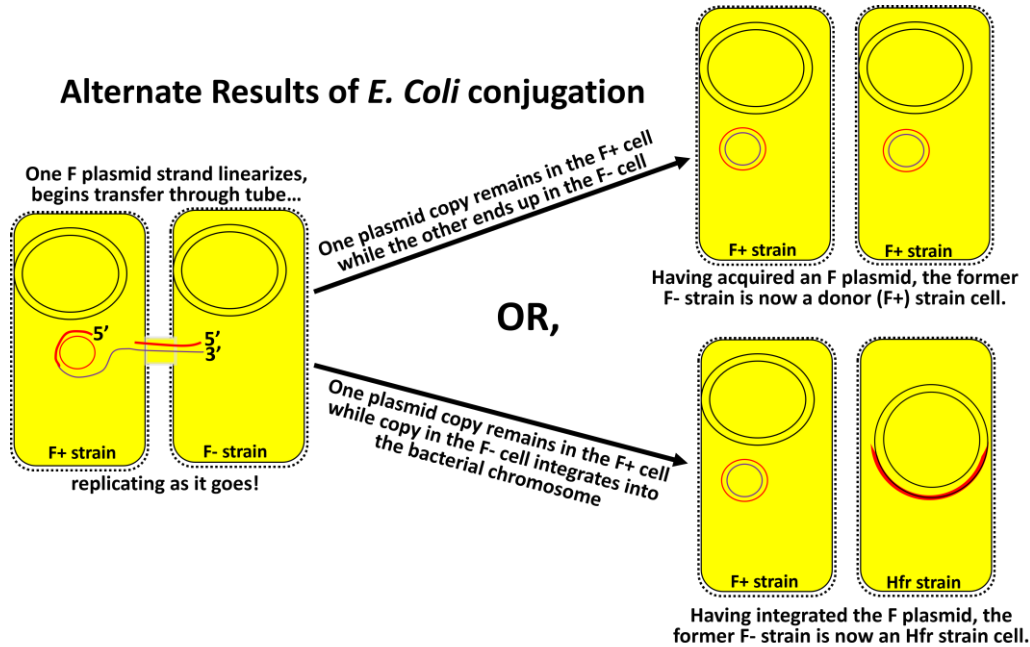


Fig. 8.20: After conjugal F plasmid transfer, F^- recipients can become F^+ or, if the plasmid integrates into the recipient genome, it will become an *Hfr* cell.

Hfr cells readily express their integrated F-plasmid genes, and like F^+ cells, they develop sex pili and form a conjugation tube with an F^- cell. One strand of the bacterial chromosomal DNA will be nicked at the original insertion site of the F plasmid. The next events parallel the replicative transfer of an F plasmid during F^+/F^- conjugation, except that only part of the *Hfr* donor chromosomal DNA is transferred, as seen in Figure 8.21 (below).

***E. Coli* Hfr Strain Conjugation with F^- Strain**

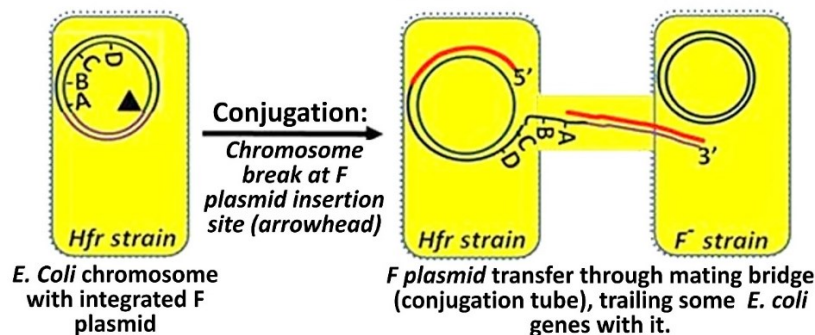


Fig. 8.21: During conjugation with *Hfr-donor* strain cells, a portion of the donor's chromosomal DNA follows the excised plasmid across the conjugation bridge into the F^- recipient cell.

In the illustration, the F plasmid (red) has inserted itself *in front of* an **A** gene so that when it enters the conjugation tube, it brings along several *E. coli* chromosomal genes. Because of the size of the bacterial chromosome, only a few bacterial genes enter the recipient gene before transfer is aborted. But in the brief time of DNA transfer, at least some genes did get into the

recipient F^- strain where they can be expressed. Here is the outline of an experiment that allowed bacterial genes on a circular DNA chromosome to be mapped:

1. Hfr cells containing functional **A**, **B**, **C**, and **D** genes were mated with recipient cells containing mutants of the A, B, C, or D genes.
2. Conjugation was mechanically disrupted at different times after the formation of a conjugation tube.
3. Recipient cells from each of the disrupted conjugations were then grown into culture and analyzed for specific gene function.

A chromosome map based on such experiments generated the map shown below in Figure 8.22:

Specific F Plasmid Insertion Sites in *E. coli* Chromosomes

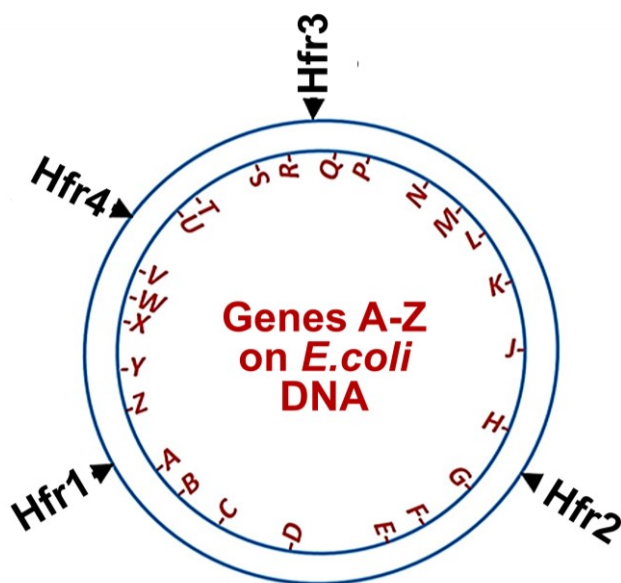


Fig. 8.22: Map of the *E. coli* chromosome (its genome!) based on conjugation of F^- recipient cells with different strains of Hfr donor cells (indicated as Hfr 1, 2, 3...): the different Hfr donors are cells in from which Hfr plasmids have integrated at different places on a recipient cell's chromosome. When the Hfr strains transfer DNA to new F^- cells, they bring along different regions of the donor cell chromosome (and thus different genes) into their conjugated recipient cells.

Let's look in detail at how this map could have been generated. First, a recipient cell with a *mutant A* gene would have acquired a wild-type A gene (and therefore A-gene function) after only a short time before conjugation disruption. Progressively longer conjugation times (measured in separate experiments) were required to transfer genes B, then C, then D to the recipient cell. Thus, the order of these genes on the bacterial chromosome was **-A-B-C-D-**. In fact, the timing of conjugation that led to F^- mutants acquiring a functional gene from the Hfr strain was so refined that it was possible to determine not only the gene locus but also the size (length) of the genes! Thus, the time to transfer a complete gene to an F^- cell reflects the length of the gene and the linear order of genes on bacterial DNA. Recall that genes mapped along the eukaryotic chromosomes had already implied a linear order of the genes. However, little was known about eukaryotic chromosome structure at the time, and the role of DNA as the "stuff of genes" was not yet appreciated. These bacterial mating experiments demonstrated for the first time that genes are linearly arranged not just along a chromosome but also along

the DNA molecule. Over time, many bacterial genes were mapped all along the *E. coli* chromosome by isolating many different Hfr strains in which an F plasmid had been inserted into different sites around the DNA circle. As shown below in Figure. 8.22. These Hfr strains were mated to F⁻ bacteria, each with mutations in one or another known bacterial gene. As in the original "ABCD" experiment, the order of many genes was determined and even shown to be linked at a greater or lesser distance to those ABCD genes and each other. Using these Hfr strains in conjugation experiments, it was shown that in fact, the different Hfr cells transferred different genes into the recipient cells in the order implied by the chromosome map illustrated in Figure 8.22. What's more, when the experiment was done with *Hfr4* (in this generic diagram), the order of genes transferred after longer times of conjugation was found to be -V-W-X-Y-Z-A-B-.... The conclusion here is that an *E. coli* DNA molecule (i.e., its "chromosome") is a closed circle! We will see visual evidence of this fact in the next chapter, with some discussion of how this evidence informed our understanding of DNA replication.

8.7 Phage Can Integrate Their DNA into the Bacterial Chromosome

We have seen how a phage can infect bacterial cells and co-opt the host's bacterial metabolic machinery to reproduce itself, eventually lysing the bacterium and releasing the phage. This **lytic** activity of infection is one of two alternate phage life cycles. **Lysogeny** is the other life-cycle pathway for bacteriophage. Lysogeny typically begins when the phage DNA is incorporated (i.e., is spliced) into the host bacterium's genome. There it will replicate along with the host's chromosome during rounds of *binary fission*. Upon excision, usually when the bacterium is under stress (e.g., nutrient-deprivation), the phage DNA can reenter the lytic pathway, reproducing and lysing the host cell. During excision, phage DNA can pick up bacterial DNA (i.e., genes) that will then be transferred to a new host cell in a subsequent infection. The transfer of genes from one bacterial cell to another in this way is called **bacterial transduction**. Esther and Joshua Lederberg studied the lysogenic activity of lambda phage (λ phage) in *E. coli*, showing that—like conjugation—bacterial transduction by λ phage results in **lateral gene transfer** between bacteria. In general, lateral gene transfer increases genetic diversity in bacteria, in fact in many prokaryotes and eukaryotes. We will see more examples of lateral gene transfer in later chapters. The events of lysogenic phage infection are shown below in Figure 8.23.

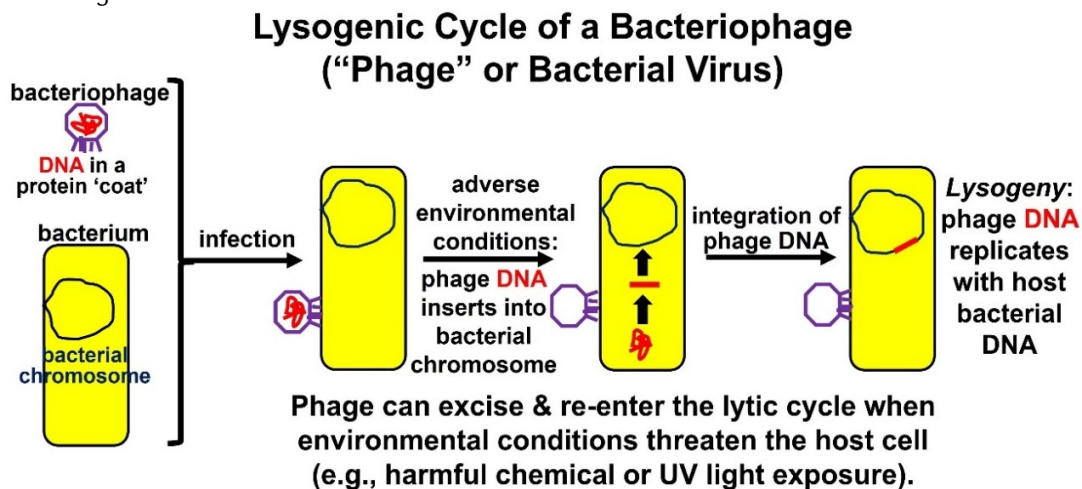


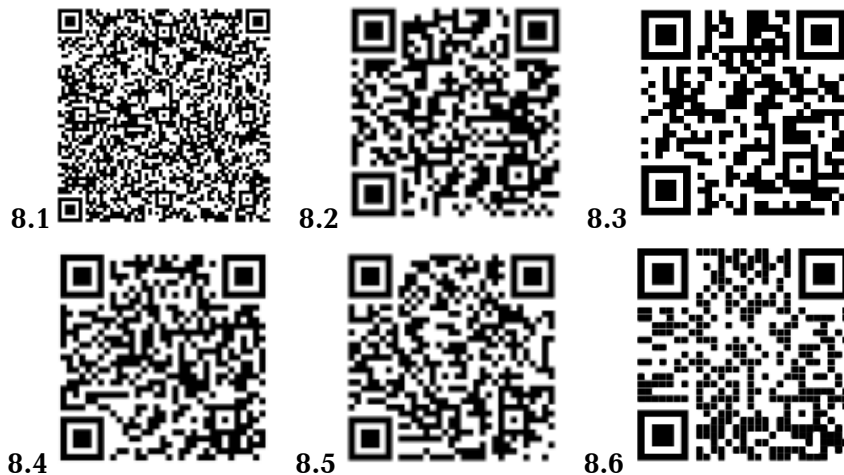
Fig. 8.23: The lysogenic life cycle of phage.

Recall that Joshua Lederberg shared a Nobel Prize with George Beadle and Edward Tatum for proposing the one-gene/one-enzyme hypothesis (eventually the one-gene/one-polypeptide relationship). Esther Lederberg had studied earlier with Beadle, Tatum, and her husband at Stanford University in 1944, and many credit her with helping her mentors win that 1958 Nobel Prize. In her own right, she held adjunct or directorial positions at the University of Wisconsin–Madison and at Stanford University, where her other seminal research included the discovery of lambda phage (1950), the invention of the *replica-plating* technique (1951), the demonstration of lambda-phage transduction in *E. coli*, and the discovery of the *fertility factor* in *E. coli*. See ^{8,10}[Esther Lederberg's Accomplishments](#) to explore why she was never a faculty member or a Nobel laureate.

Some iText & VOP Key Words and Terms

10 nm filament (fiber)	double helix	mutations
30 nm solenoid (fiber)	euchromatin	nonhistone proteins
5'-to-3' replication	F and Hfr plasmids	nucleosomes
antiparallel DNA strands	F– strain	recipient cell
bacterial conjugation	F+ strain	replication
base ratios	fertility plasmid	<i>S. pneumoniae</i> type 2-R
beads-on-a-string	heterochromatin	<i>S. pneumoniae</i> type 3-S
chromatin	Hfr strain	semiconservative replication
chromosomes	histone octamer	sex pili
conjugation tube	histone proteins	solenoid fiber
conservative replication	influenza	spindle fibers
deoxyribonuclease	karyotype	transforming principle
discontinuous replication	levels of chromatin packing	X & Y chromosomes
dispersive replication	mating bridge	X-ray crystallography
DNA	metaphase chromatin	X-ray diffraction
donor cell	mitosis & meiosis	

CHAPTER 8 WEB LINKS





Chapter 9

Details of DNA Replication and Repair

Replicons, replication origins and forks; bidirectional replication; many enzymes of replication; Okazaki fragments; mechanisms, actions, and enzymes of DNA repair; when DNA repair fails

Reminder: For inactive links, google key words/terms for alternative resources.

Elongation leading to



a circular but tasty double helix

9.1 Introduction

Replication begins at one or more **origins of replication** along DNA. **Helicase enzymes** catalyze unwinding of the double helix, creating *replicating bubbles*, or **replicons**, with **replication forks** at either end. Making a new DNA strand starts with making an RNA **primer** using RNA nucleotides and **primase** enzymes. DNA nucleotides are added to the 3' ends of primers by a **DNA polymerase**. Later, other DNA polymerases catalyze removal of RNA primers and the replacement of the hydrolyzed ribonucleotides with new deoxyribonucleotides. Finally, **DNA ligases** stitch together the fragments of new DNA synthesized at the replication forks. This complex mechanism is common to the replication of "naked" prokaryotic DNA *and* of chromatin-encased eukaryotic DNA and must therefore have arisen early in the evolution of replication biochemistry. In this chapter, we'll see details of replication, including differences between prokaryotic and eukaryotic replication that arise because of differences in DNA packing. As with any process with many moving parts, replication is error prone, so we'll also see how the overall **fidelity** of replication relies on **DNA repair** mechanisms that target specific kinds of replication mistakes, or **mutations**. Lest we think that uncorrected errors in replication are always bad, we'll learn that they usually *do not* have bad outcomes. Instead, that they leave behind the mutations necessary for *natural selection* and the *evolution of diversity*.

Learning Objectives

When you have mastered the information in this chapter, you should be able to do the following:

1. Explain how Cairns interpreted his theta (θ) images.
2. Compare and contrast the activities of enzymes required for replication.

3. Describe the order of events at an origin of replication and at each *replication fork*.
4. Compare *unidirectional* and *bidirectional* DNA synthesis from an origin of replication.
5. Outline the basic synthesis and proofreading functions of *DNA polymerase I* and explain why only this one of the several DNA polymerases can proofread newly synthesized DNA.
6. Identify the major players and their roles in the initiation of replication.
7. Explain how Reiji Okazaki's experimental results were not entirely consistent with how both strands of DNA replicate.
8. List the major molecular players (enzymes, etc.) that elongate a growing DNA strand.
9. List the nonenzymatic players in replication and describe their functions.
10. Describe how the structure of *telomerase* enables proper replication.
11. Compare the activities of topoisomerases 1 and 2.
12. Explain the reasoning behind the hypothesis of *processive replication*.
13. Compare and contrast the impacts of germline and somatic mutations.
14. Describe common forms of DNA damage.
15. List enzymes of replication that were adapted to tasks of DNA repair.
16. Explain why a DNA glycosylase is useful in DNA repair.
17. Explain the connection between "breast cancer genes" and DNA repair.

9.2 DNA Replication

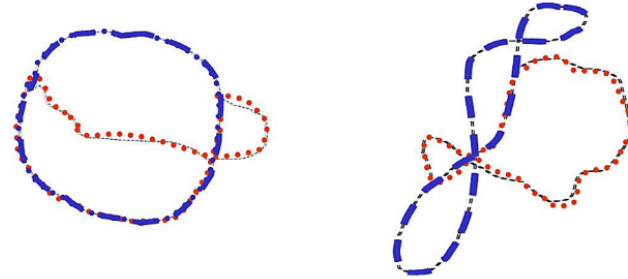
As we've seen, single DNA strands have directionality, with a 5' nucleotide phosphate end and a 3' deoxyribose hydroxyl end. This is also true for circular bacterial chromosomes; you can see this if the circle is broken! Because the strands of the double helix are *antiparallel*, the 5' end of one strand aligns with the 3' end of the other at both ends of the double helix. In looking at the complementary pairing of bases of deoxynucleotides in the double helix, Watson and Crick quickly realized that the base sequence of one strand of DNA could serve as a template to make a new complementary strand. As we'll see, this structure raised questions about DNA replication. Finding answers began with experiments that visualized the replication process.

9.2.1 Visualizing Replication and Replication Forks

Recall the phenomenon of bacterial conjugation showing that the bacterial chromosome (i.e., genome) is circular. In 1963, John Cairns confirmed this fact by direct visualization of bacterial DNA. He cultured *E. coli* cells for long periods on ³H-thymidine (³H-T), a radioactive precursor to DNA, to make their entire cellular DNA radioactive. He then disrupted the cells gently to minimize damage to the DNA. The released DNA was allowed to settle and to adhere to membranes. Cairns placed a sensitive film over the membrane, allowed time for the radiation to expose the film and then developed the films (*autoradiographs*).

When he examined the autoradiographs in the electron microscope, Cairns saw tracks of silver grains, the same kind of silver in a photographic emulsion that make up an image on film in old-fashioned photography. A facsimile of Cairns' results is illustrated in Fig. 9.1 below. He measured the length of the "silver" tracks on his autoradiographs, which most often consisted of three different closed loops, or circles (see the right-hand drawing in the figure). The circumferences of two of these circles were always equal, and their lengths, measured in nanometers, corresponded closely to a predicted DNA content of a single, non-dividing cell. Cairns therefore interpreted these images to be bacterial DNA in the process of replication.

Facsimile of Silver Tracks Seen by Cairns in His Electron Microscope Autoradiographs



Circumferences of complete circles (red dots or blue dashes) were always 1.36 nm, consistent with 4×10^6 nt (2×10^6 base pairs), and with the 0.0044 pg DNA per bacterial cell.

Fig. 9.1: Cairns cultured *E. coli* cells in radioactive (^3H -) thymidine to let the cells to make radioactive ('hot') DNA. He extracted the 'hot' DNA and spread it on electron microscope (EM) 'grids', coating them with radiation-sensitive *emulsion*. After exposing the grids he saw 'silver tracks' in the EM, like those drawn here. In each image, Cairns could always trace at least 2 equal length (circumference) *intersecting circles*, consistent with the known DNA content per bacterial cell.

Cairns called his replicating chromosomes "*theta images*" because they resembled the Greek letter *theta* (θ). From many θ images, he arranged a sequence (shown in Figure 9.2) to illustrate his inference that replication starts at a single *origin of replication* on the bacterial chromosome, before proceeding around the circle at replication forks to completion.

Cairn's Interpretation of 'Theta' (θ) Images

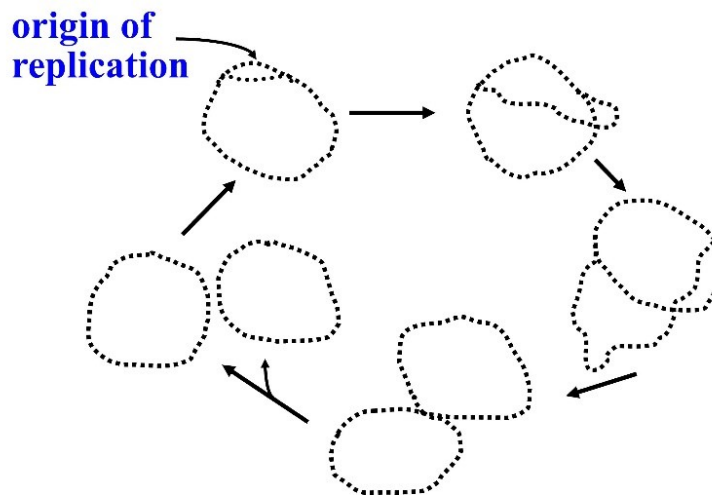


Fig. 9.2: Many of Cairns' silver tracks resembled the Greek letter theta, so called them theta images. He could order his theta images to suggest the progress of chromosomal replication, as drawn here.



11.2.2 Visualizing Bidirectional Replication

David Prescott demonstrated **bidirectional replication**, which began at a replication origin but replicated DNA in *opposite directions* at not one but two **replication forks** (Figure 9.3).

Demonstration of Bidirectional Replication in *E. coli*

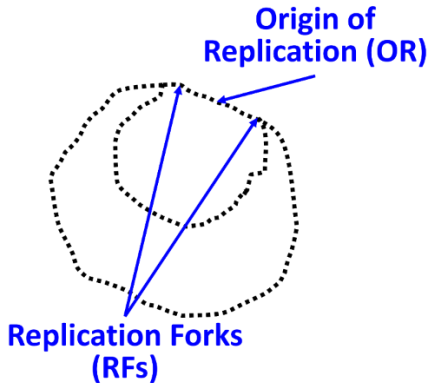


Fig. 9.3: Theta images suggested that DNA replication happens where silver tracks intersect, but at one or both intersections? D. Prescott showed that in the *E. coli* circular chromosome, the intersections were replication forks (RFs) where DNA was unwinding and replicating in opposite directions from a single origin of replication (OR). The *E. coli* OR was later shown to be a specific sequence.

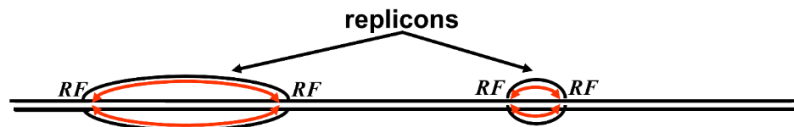


[176 Semiconservative Bidirectional Replication from Two RFs](#)



A typical rate of DNA synthesis is about 2×10^6 base pairs per hour, just about the size of the *E. coli* genome! A typical eukaryotic cell nucleus contains thousands of times more DNA than a bacterium but divides every fifteen to twenty hours. Even a small eukaryotic chromosome can contain hundreds or thousands of times as much DNA as a bacterium. Thus, eukaryotic cells can't afford to replicate DNA at a bacterial rate! Eukaryotes solved this problem *not* by evolving faster replication biochemistry, but by synthesizing their DNA bidirectionally from multiple origins of replication, creating multiple **replicons**. These grow, joining other growing replicons along each linear chromosome, as suggested in Figure 9.4.

Replicon Growth & Replication in Eukaryotes



Newly replicated daughter strands from semiconservative replication are in red. Multiple origins of replication form many unsynchronized replicons, ensuring on-time completion of replication in the S phase of the cell cycle.

Fig. 9.4: Replication in linear eukaryotic chromosomes begins at many ORs and proceeds bidirectionally, forming 'replicons' that eventually merge.



[177 Multiple Replicons in Eukaryotes](#)



Before we consider the biochemical events at replication forks in detail, let's look at the role of DNA-polymerase enzymes in the process.

9.3 DNA Polymerases Catalyze Replication

Before considering what happens at replication forks in detail, let's look at the role of **DNA polymerases** in the process. The first DNA polymerase enzyme was discovered in *E. coli* by Arthur Kornberg, for which he received the 1959 Nobel Prize in Chemistry. However, the rate of catalysis of new DNA, at least in vitro, was too slow to account for the in vivo rate of *E. coli* replication. It was Thomas Kornberg, one of Arthur's sons, who later found two additional, faster-acting DNA polymerases. (We already met the older Kornberg brother, Roger!)

All DNA polymerases require a template strand against which to synthesize the new complementary strand. In successive condensation reactions, all DNA polymerases catalyze the addition of nucleotides to the 3' end of the growing DNA strand. Finally, all DNA polymerases also have the odd property that they can only add to a preexisting strand of nucleic acid, raising the question of where the "preexisting" strand comes from! The polymerases catalyze the formation of a phosphodiester linkage between the end of a growing strand and an incoming nucleotide. The latter must complement a nucleotide in the template strand. Energy for the formation of the phosphodiester linkage comes in part from the hydrolysis of two phosphates (**pyrophosphate**) from the incoming nucleotide during the reaction. While replication requires the participation of many nuclear proteins in both prokaryotes and eukaryotes, polymerases perform the basic steps of replication (Figure 9.5).

DNA Polymerases Catalyze 5'-to-3' Replication

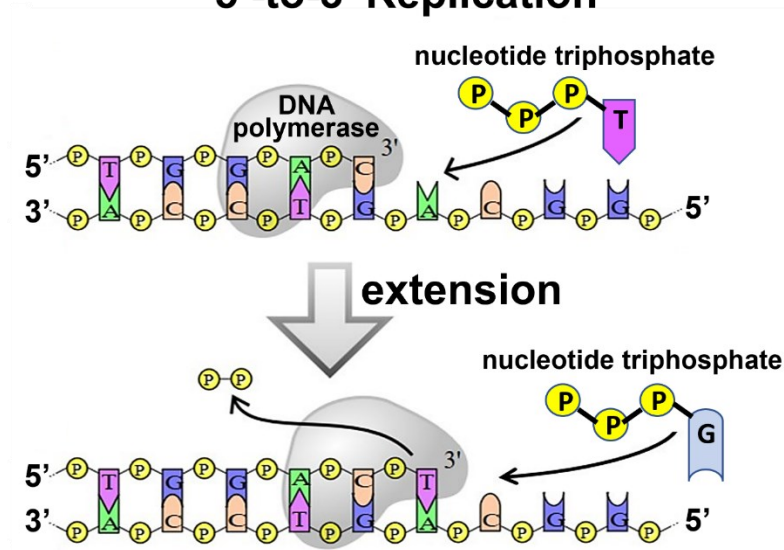


Fig. 9.5: DNA polymerases require a template strand and use deoxynucleotide triphosphate precursors (upper illustration) to catalyze the replication of new DNA in the 5' to 3' direction, shown here extending the new DNA by one nucleotide (lower illustration).



Although DNA polymerases replicate DNA with high fidelity and make as few errors as one per 10^7 nucleotides, mistakes do occur. The proofreading ability of some DNA polymerases corrects many of these mistakes. The polymerase can sense a mismatched base pair, slow down, and then catalyze repeated hydrolyses of nucleotides until it reaches the mismatched base pair. Figure 9.6 illustrates this basic proofreading by a DNA polymerase.

Some DNA Polymerases Can Proofread and Correct Errors:

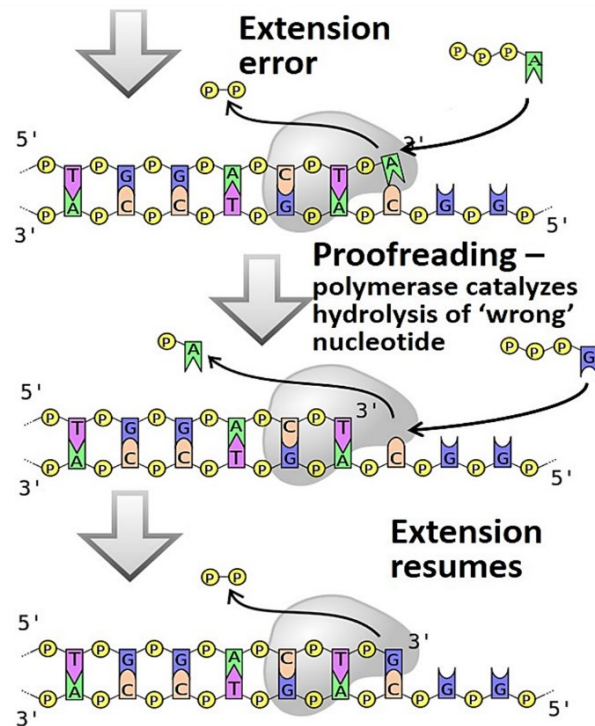


Fig. 9.6: If during replication, a wrong nucleotide is inserted in a growing DNA (upper illustration), some DNA polymerases will detect and proofread the incorrect nucleotide and replace it with the correct one (middle illustration). Replication will then resume (lower illustration).

After mismatch repair, the DNA polymerase resumes its forward movement at replication forks. Of course, not all mistakes are caught by this or other repair mechanisms (see section 9.5, "DNA Repair"). While mutations in eukaryotic germline cells that escape correction can cause genetic diseases, most of these replication errors are the very mutations that fuel evolution. Without mutations in germline cells (i.e., egg and sperm cells), there would be no mutations and no evolution, and without evolution, life itself would have reached a quick dead end! On the other hand, replication mistakes can generate mutations in somatic cells. If these somatic mutations escape correction, they can have serious consequences, including the generation of tumors and cancers.

9.4 The Process of Replication

As noted, DNA replication is a sequence of repeated condensation (dehydration synthesis) reactions linking nucleotide monomers into a DNA polymer. Replication, like all biological polymerizations, proceeds in three enzymatically catalyzed and coordinated steps: **initiation**, **elongation**, and **termination**.

9.4.1 Initiation

As we have seen, DNA synthesis starts at one or more origins of replication. These are DNA sequences targeted by *initiation proteins* (Figure 9.7).

Initiation of Replication in *E. coli*

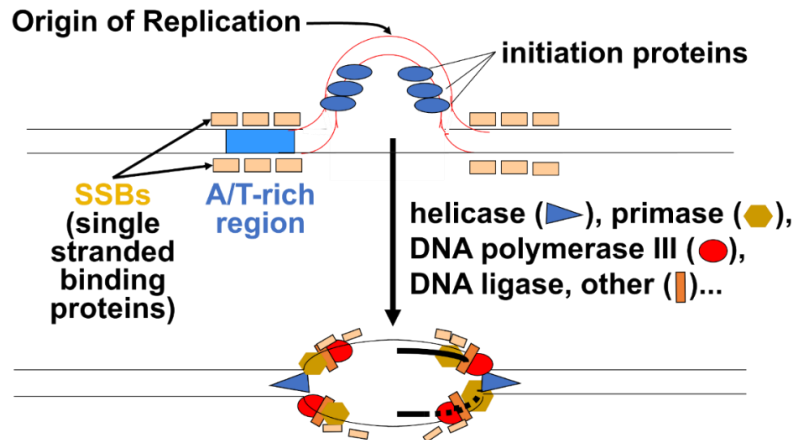


Fig. 9.7: In *E. coli*, replication starts when initiation proteins and SSBs (Single-Stranded Binding proteins) bind to the DNA at the origin of replication (OR), bending the DNA (upper illustration). Helicase begins unwinding the DNA as DNA polymerase and other enzymes of replication bind to the DNA to begin replication (lower illustration).

After these proteins break the hydrogen (H-) bonds at the origin of replication, the DNA double helix is progressively unzipped in both directions (i.e., by *bidirectional replication*). The separated DNA strands serve as templates for new DNA synthesis. Sequences at replication origins that bind to initiation proteins tend to be rich in adenine and thymine bases. This is because A-T base pairs have two H-bonds, which require less energy to break than the three H-bonds that hold G-C pairs together. Once initiation proteins loosen H-bonds at a replication origin, **DNA helicase** uses the energy of ATP hydrolysis to further unwind the double helix. DNA polymerase III is the main enzyme that then elongates new DNA. Once initiated, a replication bubble (replicon) forms as repeated cycles of elongation proceed at opposite replication forks.



Recall that new nucleotides are only added to the free 3' hydroxyl end of a preexisting nucleic acid strand. Since no known DNA polymerase can start synthesizing new DNA strands from scratch, this is a problem! It turns out that DNA polymerases require a **primer**, a nucleic acid strand onto which they can add nucleotides. So, what is that primer, and where does it come from? *RNA polymerases* catalyze RNA synthesis from in a 5'-to-3' direction and are the only polymerases that grow a new nucleic acid strand from the first base. Therefore, it was suggested that *primers* for replication might actually be RNA. Discovery of short stretches of RNA at the 5' end of Okazaki fragments confirmed the notion of RNA primers. We now know that cells use a special RNA polymerase called **primase**, to make RNA primers against DNA

templates. Replication from the 3' end of a primer implies that DNA polymerases can add deoxynucleotides to the 3' end of the RNA. We'll see in the next section that the requirement for RNA primers is nowhere more evident than in the events at a replication fork.

9.4.2 Elongation

Looking at elongation at one replication fork, we see another problem: one of the two new DNA strands will grow continuously toward the replication fork as the double helix unwinds. But what about the other strand? Either this other strand must grow in pieces in the opposite direction, or it must wait to begin synthesis until the double helix is fully unwound. The problem is illustrated in Figure 9.8.

If one new DNA strand at an RF is synthesized *non-stop*, what about the DNA against the opposite template?

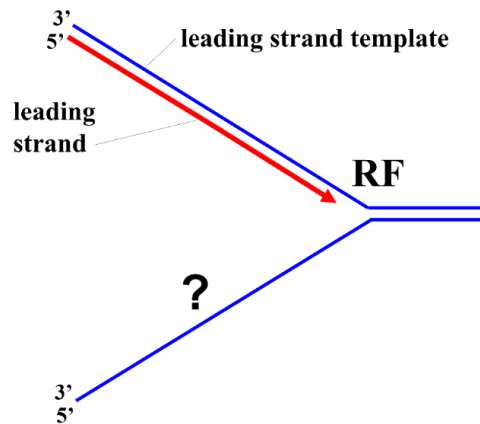


Fig. 9.8: In replication, the 5'-to-3' strand elongation catalyzed by all DNA polymerases presents a problem at the RF: only one new DNA strand (the leading strand) can be made continuously along its parental template strand. How does replication progress along the opposite template strand?

If one strand of DNA must be replicated in fragments, then those fragments would have to be stitched (i.e., ligated) together, as hypothesized in Figure 9.9.

Hypothesis: One of the new DNA strands at an RF is synthesized *in pieces*

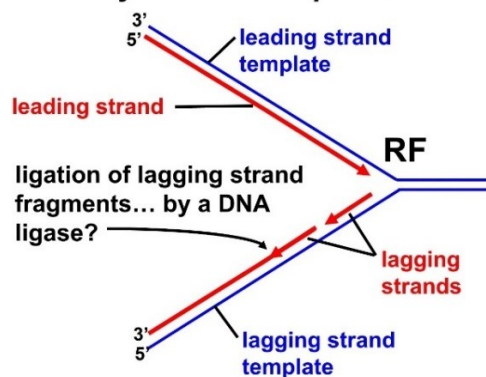
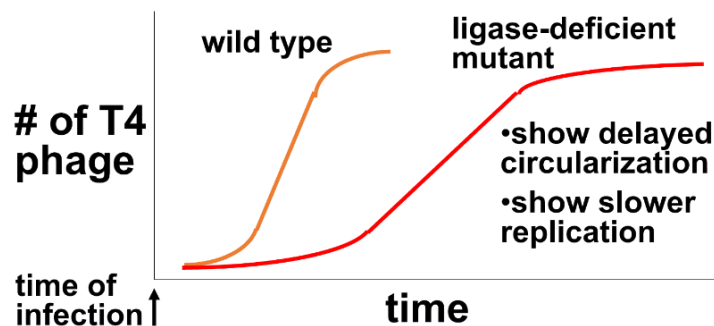


Fig. 9.9: The hypothesis proposed here is that during the elongation of DNA strands, at least one DNA strand at a replication fork (the lower, so-called lagging strand) must be synthesized discontinuously, i.e., in pieces. Each new "piece" would begin with an RNA primer. And these pieces would have to be correctly stitched together into a continuous DNA strand.

According to this hypothesis, a new **leading strand** of DNA grows (is lengthened) continuously by sequential addition of nucleotides to its 3' end, against its **leading-strand template**. The other strand, however, would be made in pieces that would be joined in phosphodiester linkages in a subsequent reaction (discontinuous replication). Because joining these new DNA fragments should take extra time, this new DNA is called the **lagging strand**, making its template the **lagging-strand template**.

Reiji Okazaki and his colleagues were studying infections of slow-growing mutants of T4 phage in *E. coli* host cells. A DNA ligase enzyme was already known to catalyze the circularization of the linear phage DNA molecules that were being replicated in infected host cells. Okazaki's hypothesis was that a deficient DNA ligase in the mutant phage not only slowed down the circularization of the replicating T4 phage DNA but would also slow the joining phage DNA fragments replicated against at least one of the two template DNA strands. The investigators compared the growth rates of wild-type and mutant T4 phage and demonstrated that slow growth of the mutant phage was due to a deficient **DNA ligase** enzyme. (Figure 9.10).

Okazaki studied slow-growing ligase deficient mutants of T4 phage:



Perhaps, reasoned Okazaki, mutant phage would be slow to ligate fragments of DNA that were replicated in pieces...

Fig. 9.10: Wild-type and mutant T4 bacteriophage growth curves are compared. Slow growth of the T4 mutant phage in infected cells (at the right) is due to a slow-acting DNA ligase enzyme encoded by a mutant T4 gene. R. Okazaki suggested that slower growth was due to inefficient ligation (elongation) of lagging strand DNA, as well as the slow circularization of the linear T4 DNA that enters host cells during infection.

When they tested the hypothesis, they found that short DNA fragments did indeed accumulate in *E. coli* cells infected with ligase-deficient mutants but not in cells infected with wild type phage. The lagging strand fragments are now called **Okazaki fragments**. You can check out Reiji Okazaki's original research at ^{9.1}[1968 Okazaki article](#).



[180-2 Okazaki Experiments: Solving a Problem at an RF](#)



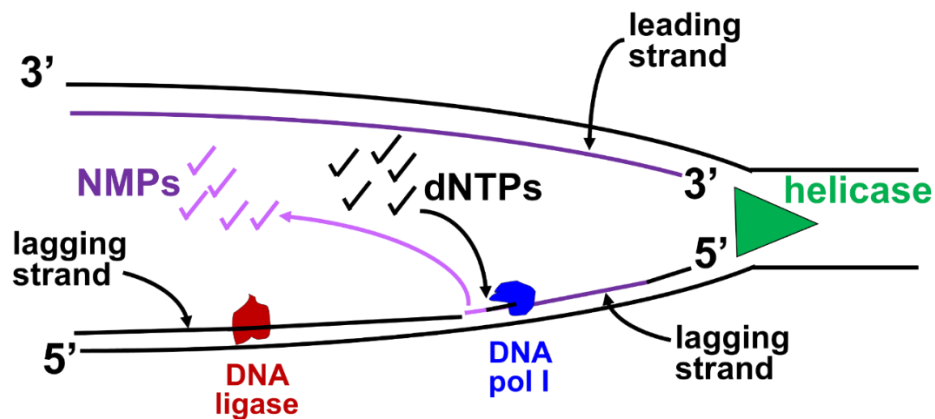
[181 Okazaki Fragments Are Made Beginning with RNA Primers](#)



Each Okazaki fragment would have to begin with a 5' RNA primer, creating yet another dilemma! The RNA primer must be replaced with deoxynucleotides before the fragments could be stitched together—a process that indeed happens (Figure 9.11, below). Removal of RNA primer nucleotides from Okazaki fragments requires the action of ***DNA polymerase I***.

This is the slow-acting DNA polymerase first characterized by Arthur Kornberg. DNA polymerase I has the unique ability to catalyze hydrolysis of the phosphodiester linkages between the RNA (or DNA) nucleotides and the 5' end of a nucleic acid strand.

Removal of Okazaki Fragment Primers: a form of DNA Repair



- **DNA Pol 1** binds to 5' end of the primer
- **DNA Pol 1** adds dNMPs 3' end of newer Okazaki fragment as it hydrolyzes NMPs 5' to 3' from previous Okazaki fragment.
- **DNA ligase** 'stitches' Okazaki fragments together.

Fig. 9.11: Steps in the synthesis of DNA against the lagging template strand.

Another enzyme, *flap endonuclease 1 (FEN1)* plays a role in removing "flaps" of nucleic acid from the 5' ends of the fragments. These flaps are often displaced by polymerase as the enzyme replaces the replication primer. At the same time as the RNA nucleotides are removed, DNA polymerase I catalyzes their replacement by appropriate deoxynucleotides.

Finally, when a fragment is composed entirely of DNA, ***DNA ligase*** links it to the rest of the already-assembled lagging-strand DNA. Because of its 5' *exonuclease* activity (not found in other DNA polymerases), DNA polymerase I also plays unique roles in DNA repair (discussed further in section 9.5).

As Cairns suggested and as others had demonstrated, replication proceeds in two directions from the origin and forms a replicon with two replication forks (RFs). Each RF has a primase associated with replicating Okazaki fragments along lagging strand templates. Figure 9.12 below illustrates the requirement for primases at replication forks.

Bi-Directional Replication Requires **Primase** at both Replication Forks

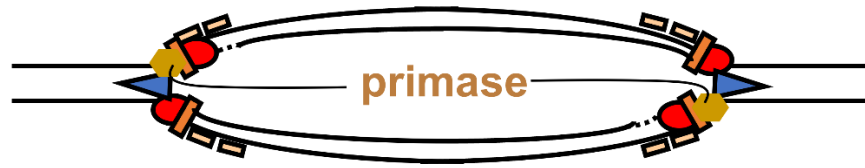


Fig. 9.12: As replication proceeds, all the enzymes (primase, DNA polymerase III, DNA polymerase I, DNA ligase, helicase) and other proteins required for replication are assembled in each replication fork (RF) of replicons. Note that primase is present at both RFs, but only on the lagging strands.

Now we can ask what happens when replicons reach the ends of linear chromosomes in eukaryotes.



[182 Replication Elongation in *E. coli*](#)

9.4.3 Termination

In prokaryotes, replication is complete when two replication forks meet after replicating their portion of the circular DNA molecule. In eukaryotes, many replicons fuse to become larger replicons, eventually reaching the ends of the chromosomes. And now there is still another problem, illustrated in Figure 9.13!

Problem at the End of a Chromosome

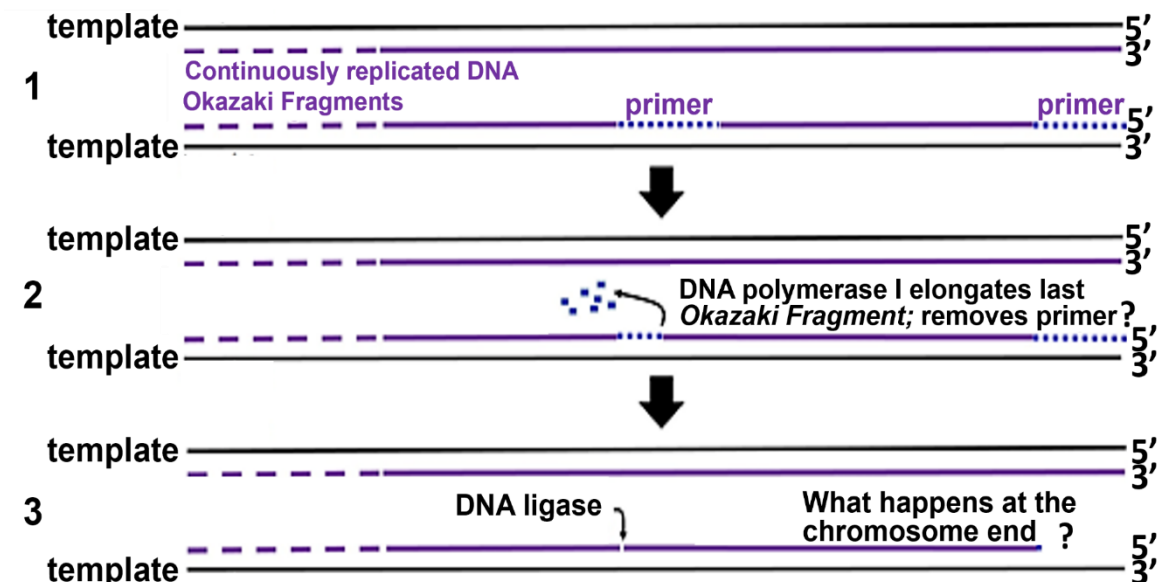


Fig. 9.13: Problems arising in lagging strand replication when replication reaches the ends (telomeres) of linear chromosomal DNA. After a final Okazaki strand is primed and replicated (panel 1) the primer is removed from the penultimate fragment (panel 2). Then primer on the last strand is removed (panel 3). But then, there would be no DNA strand 3' end at which to add replacement DNA nucleotides. This would cause chromosomal shortening at each cell division.

When a replicon nears the end of a double-stranded DNA molecule (i.e., the end of a chromosome), the new continuously synthesized strand stops when it reaches the 5' end of its template DNA, and the primer is removed, having completed its replication. But what about lagging-strand replication? The illustration shows primer removal from an Okazaki fragment primed near the end of the chromosome, and replacement with DNA nucleotides catalyzed by DNA polymerase I. The question marks above the DNA point to a dilemma: if a final Okazaki fragment is primed and synthesized, DNA polymerase I has no free 3' end to begin RNA nucleotide replacement with DNA nucleotides. The problem then would be that every time a cell replicates, at least one strand of new DNA would get shorter. Of course, this would not do—and doesn't happen! Eukaryotic replication undergoes a *termination* process that extends the length of one of the two strands using the enzyme *telomerase*, as illustrated in Figure 9.14.

Telomerase: a ribonucleoprotein (RNP)



1. primer removal (DNA Pol 1)
2. binding of telomerase
3. extension of template strand by telomerase: synthesis of **telomere repeat sequences**
4. telomerase shifts to right
5. more extension of (3' end of) template strand by telomerase
6. Priming, synthesis and ligation of 'extra' Okazaki fragment from extended template strand...

Fig. 9.14: Following removal of the primer from a telomeric Okazaki fragment, the *ribonucleoprotein* enzyme *telomerase* prevents chromosome-shortening. Its RNA serves as a template to generate repeats at the 3' telomeric end of lagging-strand DNA. When the extended repeated sequences on the 3' end of the template DNA are long enough, they then serve as templates for new primer and Okazaki fragment synthesis, maintaining chromosome length.

Telomerase consists of several proteins and an RNA molecule. From the drawing, the RNA component serves as a template for 5'→3' extension of the problematic DNA strand. The protein with the requisite reverse transcriptase activity is called *Telomerase Reverse Transcriptase*, or *TERT*. The *Telomerase RNA Component* is called *TERC*. Carol Greider, Jack Szostak, and Elizabeth Blackburn shared the 2009 Nobel Prize in Physiology or Medicine for discovering telomerase.



[183 Telomere Replication Prevents Chromosome Shortening](#)



We know now that differentiated, non-dividing cells no longer produce the telomerase enzyme but that telomerase genes are still active in dividing cells (e.g., stem cells and cancer cells), which contain abundant telomerase.

9.4.4 Is Replication *Processive*?

Drawings of replicons and replication forks suggest separate events on each DNA strand. Yet events at replication forks seem to be coordinated. Thus, replication may be ***processive***, meaning both new DNA strands are replicated in the same direction at the same time, smoothing out the process. How might this be possible? The drawing in Figure 9.15 shows how lagging-strand template DNA bends so that it points in the same direction as the leading strand at the replication fork.

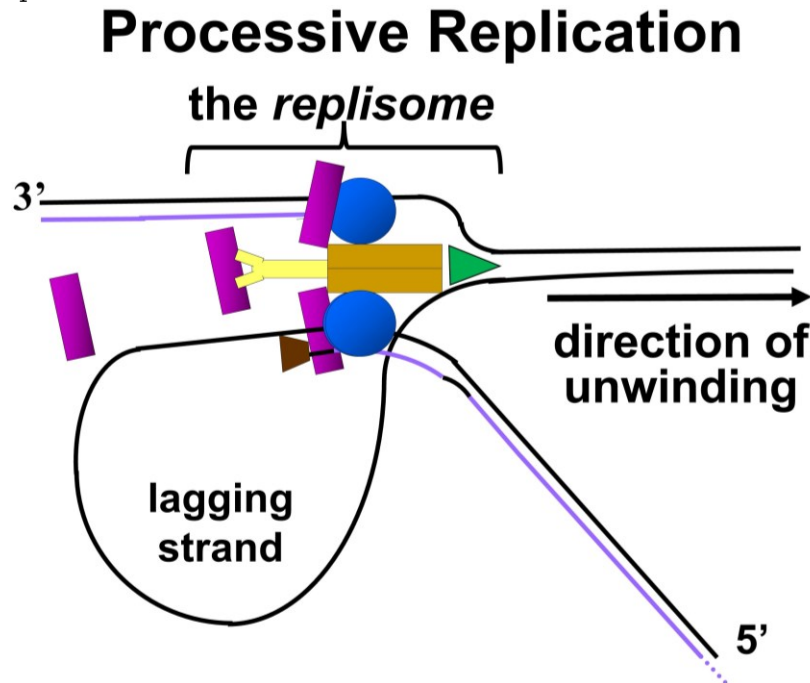


Fig. 9.15: The Processive Replication hypothesis unites all components for replicating both DNA strands in a *replisome*. In this model, the lagging strand template is "looped backwards" so the direction of replication new DNA strands occurs in the same direction.



[184 Processive Replication](#)



The ***replisome*** structure cartooned at the replication fork consists of *clamp proteins*, primase, helicase, DNA polymerase, and single-stranded binding proteins, among others. Now, newer techniques of visualizing replication by real-time fluorescence videography have called the processive model into question, suggesting that the replication process is anything but smooth! Is the observed jerky movement of DNA elongation in the video an artifact? Or is the model of smooth, coordinated replication integrated at a replisome no longer valid? And if not, must coordination of replication be defined and achieved in some other way? Or, finally, are lagging-strand and leading-strand replication simply just *not* coordinated? Check out the video yourself at ^{9.2}[Real-Time Fluorescent Replication Video](#).

9.4.5 One More Problem with Replication

Cairns recorded many images of *E. coli* of the sort shown in Figure 9.16.

DNA Changes Shape (*Topology*) During Replication

Unwinding the double helix causes *supercoiling*
(1 twist for each turn of the helix unwound).

from Cairns'
images:

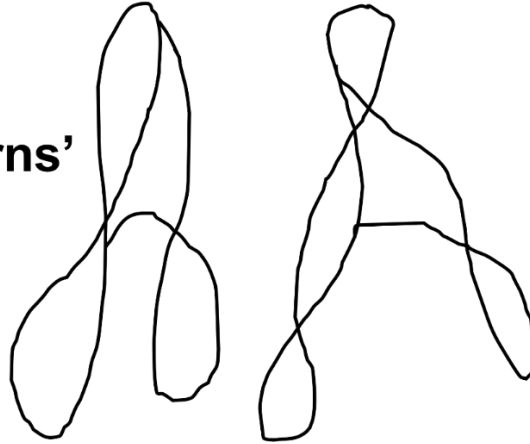


Fig. 9.16: Unwinding a circular DNA molecule or any eukaryotic double helix (rigidly associated with chromosomal proteins) causes the DNA to form supercoils, twisting and coiling in on itself. Cairns saw such supercoils in *E. coli* chromosome, as in this illustration.

The coiled, twisted appearance of the replicating circles was interpreted to be a natural consequence of the helically intertwined strands of DNA pulling apart (as is the case with intertwined strands of any material). As the strands continued to unwind, the DNA would twist into a *supercoil* of DNA. Too much unwinding would cause the phosphodiester linkages in the DNA to rupture, fragmenting the DNA. Obviously, this does not happen.

Experiments were devised to demonstrate supercoiling and to test hypotheses explaining how cells would *relax* the supercoils during replication. Testing these hypotheses revealed the *topoisomerase* enzymes. These enzymes bind and hold on to DNA, catalyze the hydrolysis of phosphodiester linkages, control the unwinding of the double helix, and finally catalyze the re-formation of the phosphodiester linkages. It is important to note that the topoisomerases are not part of a replisome and can act far from a replication fork, probably responding to tensions anywhere in the supercoiled DNA. Recall that topoisomerases comprise much of the protein lying along eukaryotic chromatin.



[185-2 Topoisomerases Relieve Supercoiling During Replication](#)



We have considered most of the molecular players in replication. Key replication proteins and their functions are listed in Table 9.2 below, taken from ^{9.3}[DNA Replication](#).

Table 9.2

ENZYME	FUNCTION IN DNA REPLICATION
DNA helicase	A helix destabilizing protein that unwinds a double helix at replication forks
DNA polymerase	Builds new double-stranded DNA, adding deoxy-nucleotides 5'-3'; some can proofread and correct errors
DNA clamp protein	Prevents DNA pol III from separating from the parent template strand.
Single-strand binding proteins (SSBs)	Keep unwound DNA strands at replication forks from re-annealing during replication
topoisomerases	Relax supercoiled DNA caused by DNA unwinding during replication
DNA gyrases	A specific kind of topoisomerase
DNA ligase	Joins Okazaki Fragments to growing DNA strands during replication
primase	Initiates replication using nucleotides to synthesize an RNA primer required for DNA polymerases to then add deoxynucleotides
telomerase	Enzyme that adds repetitive DNA sequences to telomeric DNA to maintain the length of eukaryotic chromosomal DNA

9.5 DNA Repair

We generally accept the notion that replication faithfully duplicates the genetic material, occasional making mistakes (i.e., mutating). At the same time, evolution would not be possible without those mistakes, those mutations, and mutation is not possible without at least some adverse consequences. Because we see the results of mutation as disease, the word *mutation* in common parlance (and even among scientists) anticipates dire consequences. But mutations (changes in DNA sequence) are a requirement of *life*. In fact, mutations occur frequently between generations. Most are inconsequential, and many are corrected by one or another mechanism of DNA repair.

9.5.1 Germline vs Somatic Mutations: A Balance Between Mutation and Evolution

Germline mutations are heritable. When present in one, but especially in both alleles of a gene, such mutations can result in genetic disease (e.g., Tay-Sachs disease, cystic fibrosis, hemophilia, and sickle cell anemia). Rather than causing disease, some germline mutations may increase the *likelihood* of becoming ill (e.g., mutations of the *BRCA2* gene greatly increase a woman's odds of getting 195, cancer).

Somatic mutations in actively dividing cells might result in benign "cysts" or malignant tumors (i.e., cancer). Other somatic mutations may play a role in dementia (Alzheimer's disease) or in some neuropathologies (e.g., along the autism spectrum). Since the complex chemistry of replication is subject to an inherently high rate of error, cells have evolved systems of DNA repair, so that they may survive high mutation rates. As we saw, DNA polymerases themselves can proofread, so incorrectly inserted bases can be quickly removed and replaced.

Beyond this, multiple mechanisms have evolved to repair mismatched base pairs and other kinds of damaged DNA that escape early detection. How often and where DNA damage occurs is random, as is which damage will be repaired and which will escape to become a mutation. For those suffering the awful consequences of unrepaired mutation, the balance between retained and repaired DNA damage is, to say the least, imperfect. However, evolution and the continuance of life itself rely on this balance.

9.5.2 What Causes DNA Damage?

The simplest damage to DNA is a point mutation, the accidental insertion of a “wrong” nucleotide into a growing DNA strand. Other mutations, equally accidental, include DNA deletions, insertions, duplications, inversions, and the like, any of which might escape repair. The causes of DNA damage can be chemical or physical and include spontaneous intracellular events (e.g., oxidative reactions) and environmental factors (e.g., radiation, exogenous chemicals).

These random events are in fact not rare, but frequent. Basing his calculations on studies of different kinds of DNA damage, Tomas Lindahl estimated that DNA-damaging events might be occurring at the rate of ten thousand per day! Lindahl realized that there must be some basic DNA repair mechanisms at work to protect cells against such a high rate of DNA damage. The discovery of the **base excision repair** mechanism earned Tomas Lindahl a share of the 2015 Nobel Prize in Chemistry.

Specific environmental factors that can damage DNA include UV light, X-rays, and other radiation, as well as chemicals (e.g., toxins, carcinogens, and even drugs). Both germline and somatic cells can be affected. While mutations can and do cause often debilitating diseases, it is instructive to keep the impact of mutations in perspective. Most mutations are in fact *silent*; they do not cause disease. And among mutations that could cause disease, much of the DNA damage is repaired. Cells correct more than 99.9% of mistaken base changes before they have a chance to become mutations. That is why we think of replication as a “faithful” process. Let’s look at some common types of DNA damage that are usually repaired:

- **Pyrimidine dimers:** the dimerization of adjacent pyrimidines (typically thymines, but occasionally cytosines) in a single DNA strand, caused by UV exposure
- **Depurination:** the *hydrolytic* removal of guanine or adenine from the #1 carbon of deoxyribose in a DNA strand
- **Deamination:** the hydrolytic removal of amino (-NH₂) groups from guanine (most common), cytosine, or adenine
- **Oxidative damage** of the deoxyribose sugar with any base (but most commonly with a purine base)
- Inappropriate **methylation** of any bases (but most commonly of purine bases)
- **DNA strand breakage** during replication or from radiation or chemical exposure

9.6 Molecular Consequences of Uncorrected DNA Damage

While bacteria also suffer DNA damage, we will focus here on eukaryotes, since they have evolved the most sophisticated repair mechanisms. Remember that unrepaired DNA damage **will** be passed on to daughter cells in mitosis and **might** be passed on to the next generation if the mutation occurs in a germline cell.

9.6.1 Depurination

Depurination is the spontaneous *hydrolytic* removal of guanine or adenine from the #1 carbon of deoxyribose in a DNA strand. Its frequency of five thousand depurinations per cell per day emphasizes the high rate of DNA damage that demands a fix! If not repaired, depurination results in a single base-pair deletion in the DNA of one chromosome after replication, leaving the DNA in the same region of the other chromosome unchanged. Figure 9.17 shows the effects.

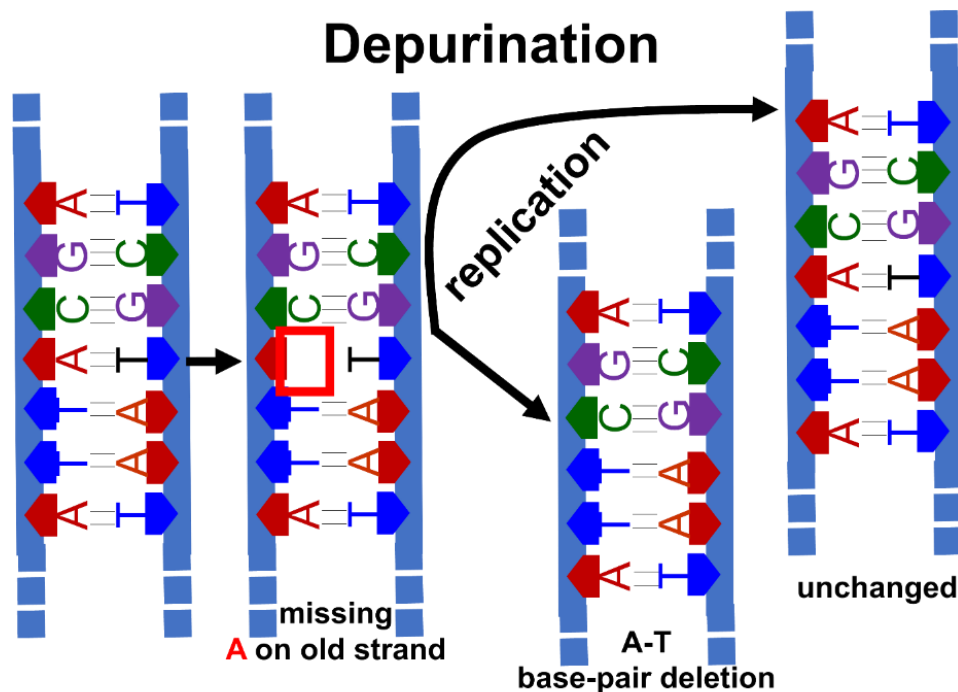


Fig. 9.17: Spontaneous depurination is the hydrolytic removal of a guanine or adenine from a nucleotide, resulting in a nucleotide deletion at that site in the DNA during replication.

The replisome ignores the depurinated nucleotide (an A in this example), jumping to the C in the template DNA. Unrepaired, one new double-stranded DNA will have a deletion, leaving the other new one with no mutation.

9.6.2 Pyrimidine Dimerization

Exposure of DNA to UV light can cause adjacent pyrimidines (commonly thymines; less often, cytosines) on a DNA strand to dimerize. Pyrimidine dimers form at a rate of a bit less than one hundred per cell per day! Uncorrected dimerization results in a two-base deletion in one chromosome, while the other remains unchanged (Figure 9.18, below).

Pyrimidine (Thymine) Dimerization

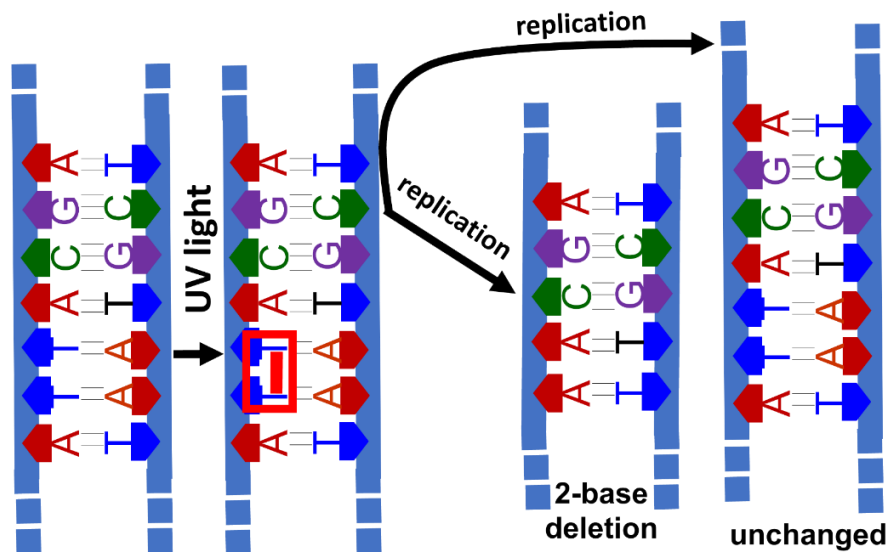


Fig. 9.18: Exposure of DNA to UV light can cause adjacent thymines to dimerize, resulting in deletion of two nucleotides at that site in the DNA during replication.

You can predict that the correction of this radiation-induced damage will either involve disrupting the dimers (in this example, thymine dimers) or removing the dimerized bases and replacing them with monomeric bases.

9.6.3 Deamination

Figure 9.19 shows the consequences of deamination to a DNA base sequence.

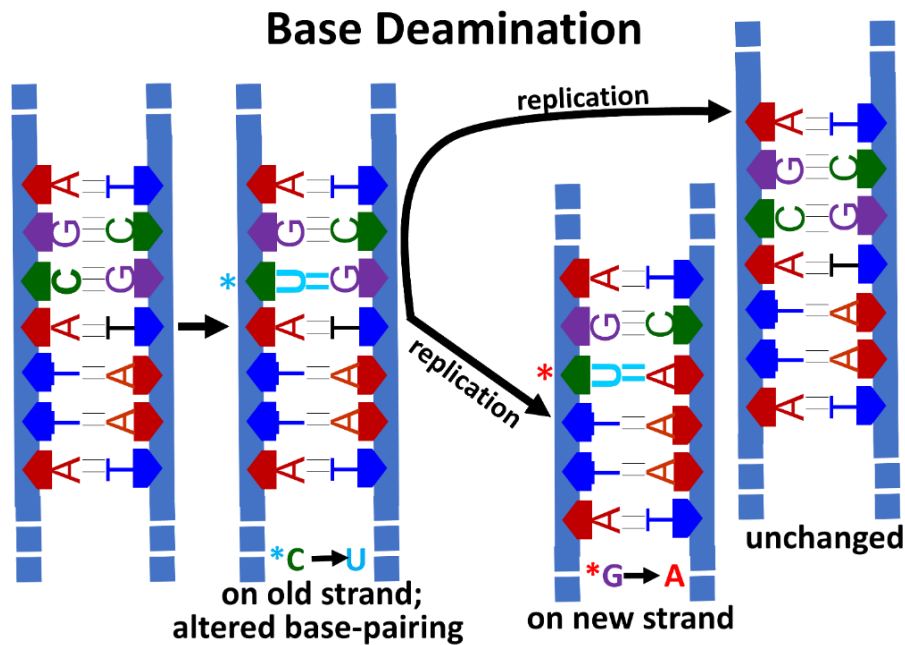


Fig. 9.19: Amino (-NH₂) group removal (deamination) from a base in one DNA strand results in a base substitution during replication. Use the asterisks to help follow the deamination and changes after replication.

Deamination is the hydrolytic removal of amino ($-NH_2$) groups from guanine (most common), cytosine, or adenine, at a rate of one hundred per cell per day. Deamination does not affect thymine (because it has no amino groups!). Uncorrected deamination results in a *base substitution* on one chromosome (or in this example, a T-A pair substitution for the original C-G) and no change on the other. Deamination of adenine or guanine results in unnatural bases (hypoxanthine and xanthine, respectively). These are easily recognized and corrected by DNA repair systems. Occasionally, some of the U-A base pairs remain unrepaired.

9.7 DNA Repair Mechanisms

Many enzymes and proteins are involved in DNA repair, including eleven out of fourteen DNA polymerases in humans! Three DNA polymerases are for replication). Some of these function in normal replication, mitosis and meiosis, but were co-opted for DNA repair activities. These molecular co-optations are so vital to normal cell function that some repair activities and molecular players are highly conserved in evolution. Among the DNA repair pathways that have been identified, we will look at *base excision repair*, *nucleotide excision repair*, *transcription-coupled repair*, *nonhomologous end-joining* and *homologous recombination*. Of these, the last is perhaps the most complex.

9.7.3 Base Excision Repair

On detection of an error (e.g., oxidization, an open-ring, deamination, or bases containing saturated C=C bonds), *DNA glycosylases* catalyze hydrolysis of the damaged base from its deoxyribose. For more on these enzymes, see ^{9.4}[DNA Glycosylases](#). Figure 9.20 shows *base excision repair*.

Base Excision Repair Following Deamination

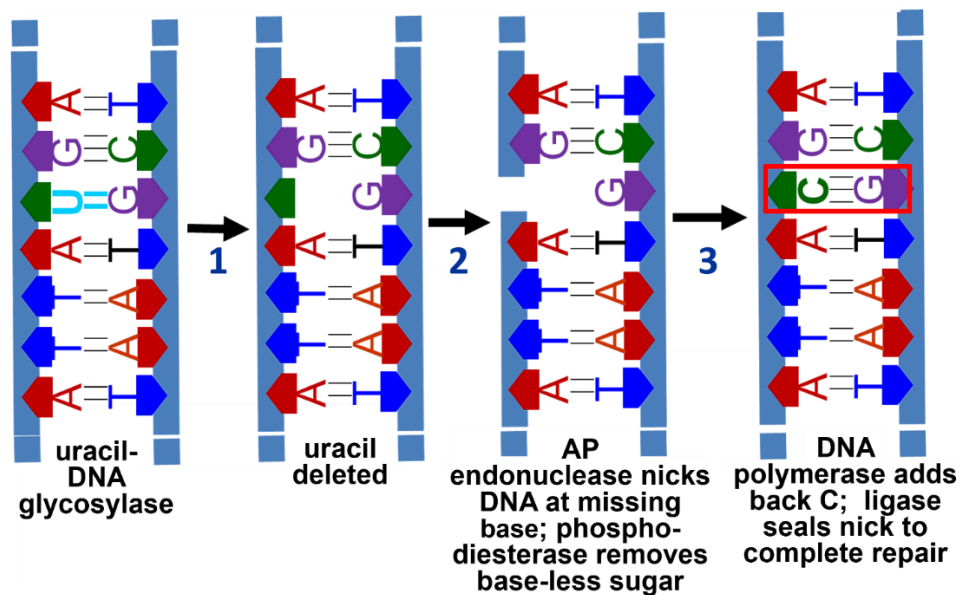


Fig. 9.20: In *Base Excision Repair*, (1) a glycosylase removes the sugar from the deaminated nucleotide; (2) an endonuclease nicks the strand at that nucleotide; (3) a DNA polymerase adds back a correct nucleotide and a DNA ligase seals the DNA strand to complete the repair.

9.7.4 Nucleotide Excision Repair

The discovery of *nucleotide excision repair* earned Aziz Sancar a share of the 2015 Nobel Prize in Chemistry. Figure 9.21 illustrates *nucleotide excision repair* for a pyrimidine dimer.

Nucleotide Excision Repair to Remove a Pyrimidine Dimer

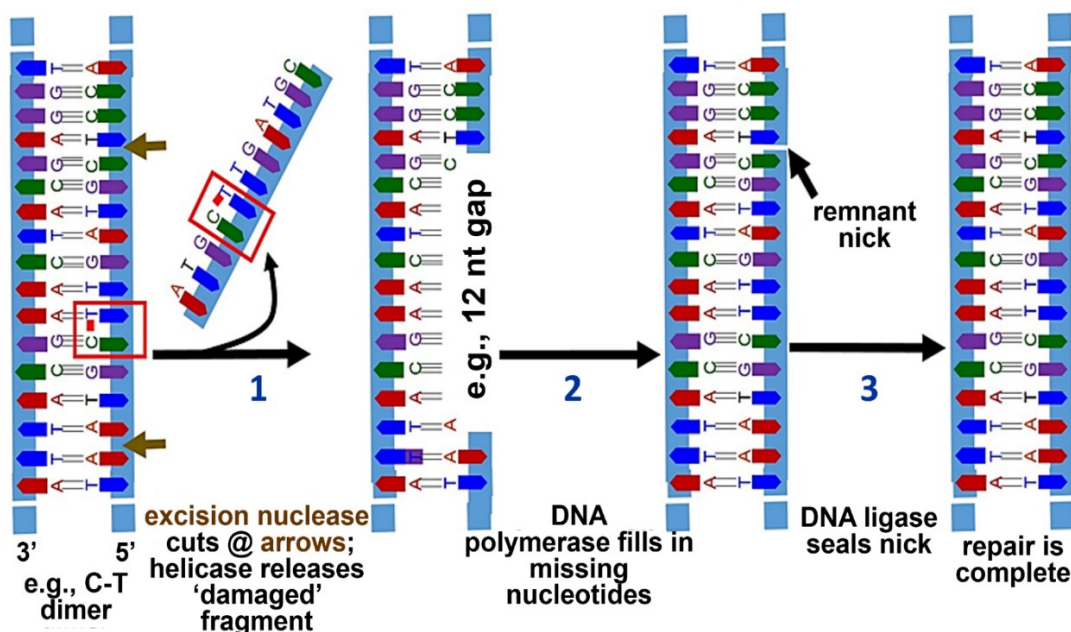


Fig. 9.21: In nucleotide excision repair, (1) a fragment containing the damaged DNA, e.g., a pyrimidine (C-T) dimer is excised; (2) a polymerase fills in the missing nucleotides in the gap, and (3) a DNA ligase seals the remaining nick, resulting in repaired DNA strands.

In this example, an *Excision Nuclease* recognizes a pyrimidine dimer, where it hydrolyzes phosphodiester linkages between nucleotides several bases away from either side of the dimer. A *DNA helicase* then unwinds and separates the DNA fragment containing the dimerized bases from the damaged DNA strand. Finally, *DNA polymerase* acts 5' – 3' to fill in the gap, and *DNA ligase* seals the remaining nick to complete the repair.

9.7.5 Mismatch Repair

DNA mismatch repair occurs when the proofreading DNA polymerase misses an incorrect base insertion into a new DNA strand. This repair mechanism relies on the fact that double-stranded DNA shows a specific pattern of methylation. The discovery of the *mismatch repair* mechanism earned Paul Modrich a share of the 2015 Nobel Prize in Chemistry.

The methylation patterns themselves are related to epigenetic patterns of gene activity and chromosome structure, which are expected to be inherited by daughter cells. When DNA replicates, the methyl groups on the template DNA strands remain, but the newly synthesized DNA is unmethylated. In fact, it takes some time for methylation enzymes to locate and to methylate the appropriate nucleotides in the new DNA. In the intervening time, several proteins and enzymes can detect inappropriate base pairing (the mismatches) and initiate mismatch repair.

The basic DNA mismatch repair process is illustrated in Figure 9.22.

DNA Mismatch Repair

When nucleotide mismatches occur in replication, *mismatch repair* efficiently fixes the problem, i.e., ~99% of the time!

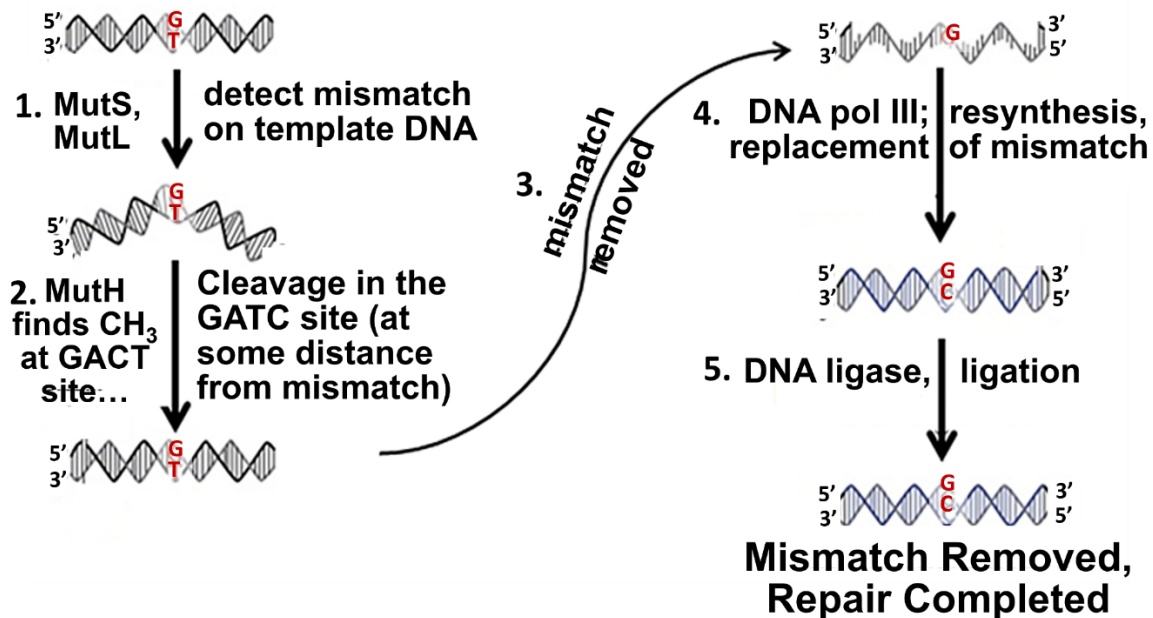


Fig. 9.22: Repair of a base mismatch in DNA begins when two *Mut* proteins (MutS and MutL) detect the mismatch (1); a third *Mut* protein (MutH) catalyzes cleavage at a methylated GATC site some distance from the mismatch (2), removing the mismatch (3); DNA polymerase III replaces the missing bases around the original mismatch (4); DNA ligase completes the repair (5).

Presumably, after mismatch repair is complete, the unmethylated new DNA is appropriately remethylated.

9.7.6 Transcription-Coupled Repair (in Eukaryotes)

If an RNA polymerase reading a template DNA encounters a nicked template or one with an unusual base substitution, it might stall transcription and seemingly “not know what to do next.” A normal transcript would not be made, and the cell might not survive. That’s no big deal in a tissue comprised of thousands if not millions of cells, right? But nevertheless, ***Transcription Coupled Repair*** exists!

In this repair pathway, if ***RNA polymerase*** encounters a DNA lesion (i.e., damaged DNA) while transcribing a template strand, it will indeed stall. This allows time for ***coupling proteins*** to reach the stalled polymerase and to enable repair machinery (e.g., by base or nucleotide excision) to make the repair. Once the repair is complete, the ***RNA polymerase*** “backs up” along the template strand (with the help of other factors) and resumes transcription of the corrected template.

9.7.5 Nonhomologous End-Joining

DNA replication errors can cause **double-stranded breaks**, as can environmental factors (ionizing radiation, oxidation). Repair by **nonhomologous end-joining** deletes damaged and adjacent DNA and rejoins the "cut" ends. When a double-stranded break is first recognized, nucleotides are **hydrolyzed** from the ends of both strands at the break-site, leaving "**blunt ends**." **Ku** among other proteins then bring DNA strands together and further hydrolyze single DNA strands, creating overlapping staggered (**complementary** or **sticky**) ends. The overlapping ends of these DNA strands form H-bonds. Finally, **DNA ligase** seals the H-bonded overlapping ends of DNA strands, leaving a repair with deleted bases. Repair by **nonhomologous end-joining** is illustrated in Figure 9.23.

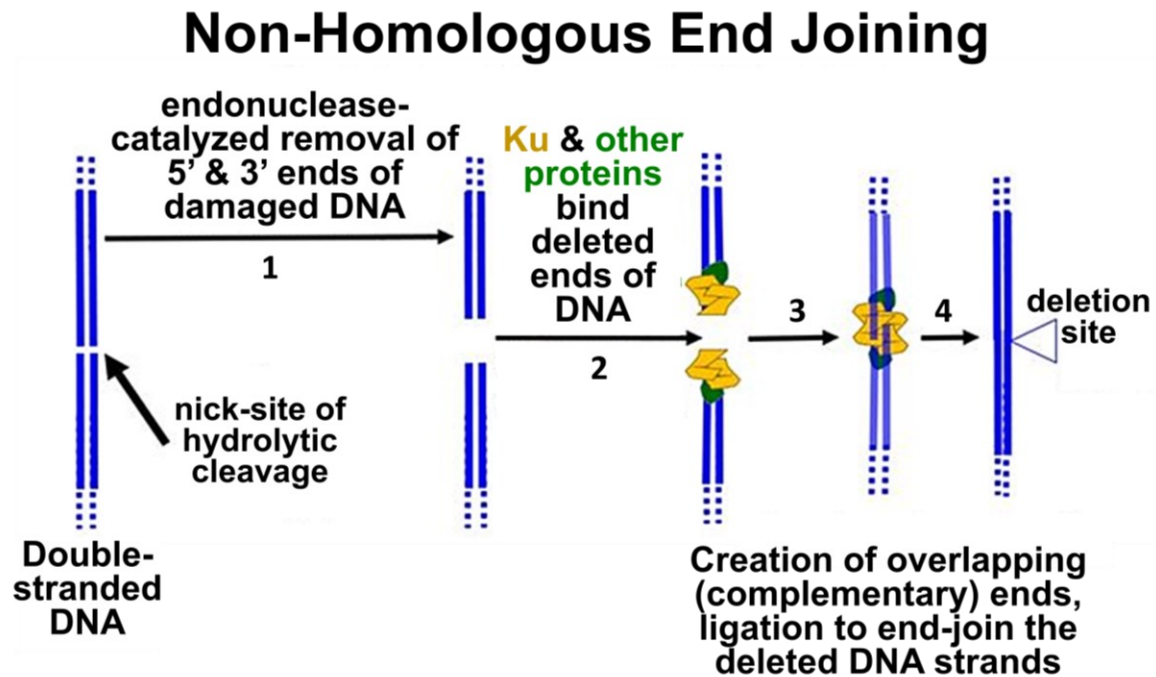


Fig. 9.23: During *Non-homologous End-Joining*, endonucleases find and trim back DNA at the break (1); specific proteins bind to the trimmed ends of the DNAs (2); catalysis of single-strand trimming forms complementary (*sticky*) ends that overlap the break site (3). After H-bonds form between the staggered sticky ends (4), a DNA ligase seals the ends of the overlapping DNA strands (4) completing an effective repair, albeit with a deletion.

In older people, there is evidence of more than two thousand "footprints" of this kind of repair per cell. How is this possible? This *quick-fix* repair often works with no ill effects because most of the eukaryotic genome does not encode genes or even regulatory DNA (the damage of which would be more serious).

9.7.6 Homologous Recombination

Homologous recombination is a complex but normal, frequent part of meiosis in eukaryotes. Recall that **homologous recombination** occurs in **synapsis** in the first cell division of **meiosis** (meiosis I). During synapsis, alignment of homologous chromosomes may lead to DNA breakage, an **exchange** of alleles, and ligation to reseal newly recombinant DNA molecules. Novel recombinations of variant **alleles** in the chromosomes of sperm and eggs ensure **genetic**

diversity in species. The key point is that DNA breakage is required to exchange alleles between **homologous chromosomes**. Consult the genetics chapter in an introductory biology textbook or the recombination chapter in a genetics text to be reminded of these events.

Cells use the same machinery to reseal DNA breaks during normal recombination and to repair DNA damaged by single- or double-stranded breakage. A single DNA strand that is nicked during replication can be repaired by recombination with the strands of homologous DNA that are being replicated on the other strand. A double-stranded break can be repaired using the same recombination machinery that operates on sister chromatids in meiosis. In both cases, the process accurately repairs damaged DNA **without any deletions**. These mechanisms are conserved in the cells of all species, further demonstrating an evolutionary imperative of accurate repair to the survival of species, no less important than the imperative to maintain the genetic diversity of species.

9.7.6.a Repair of a Single-Stranded Break

When a replisome reaches a break in one of the two strands of replicating DNA, the damage must be repaired, and the replication fork (RF) must be reestablished. Figure 9.24 illustrates the example of homologous recombination to repair a single stranded break at a replication fork.

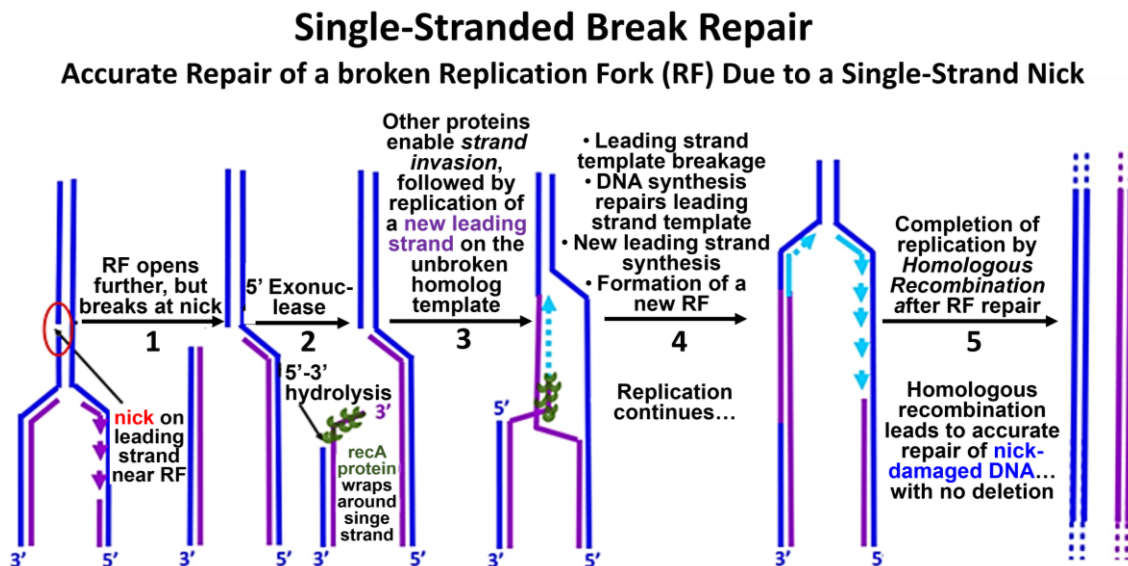


Fig. 9.24: Single-stranded break repair by *homologous recombination*: An RF reaches a break in the left template strand (1), template DNA behind the break is trimmed back (2), while the 3' end of the new DNA strand *invades* the new, (complementary) DNA strand replicating on the homologous parental template DNA and continues replicating (3). The missing DNA at the original break is re-synthesized (4). homologous recombination accurately repairs the break (5).

Such a break may have occurred prior to replication itself. Repair begins when the RF reaches the lesion. In the first step, a **5'-3' exonuclease** trims template DNA back along its newly synthesized complement. Next, **RecA** protein monomers (each with multiple DNA-binding sites) bind to the single-stranded DNA to form a **nucleoprotein filament**. With the help of additional proteins, the 3' end of the filament scans the other replicating strand for homologous sequences. When such sequences are found, the RecA-DNA filament binds to the homologous sequences, and the filament of new DNA "invades" the homologous (i.e.,

opposite) double-stranded DNA, separating its template from its newly replicated DNA. After this **strand invasion**, the replication of the leading strand continues from the 3' end of the invading strand. A new RF is established as the leading-strand template is broken and re-ligated to the original break. New lagging-strand replication then resumes at the newly rebuilt RF. The result is an accurate repair of the original damage, with no deletions or insertions of DNA.

RecA is a bacterial protein, an example of another one of those evolutionarily conserved proteins. Its homolog in the archaea is called **RadA**. In eukaryotes, the homolog is called **Rad51**, and it initiates **synapsis** during meiosis. Thus, it seems that a role for **RecA** and its conserved homologs in DNA repair predated its use in synapsis and crossing over in eukaryotes! For more about the functions of RecA protein and its homologs, see ^{9.5}[The Functions of RecA](#).

9.7.6.b Repair of a Double-Stranded Break

Homologous recombination can also repair a double-stranded DNA break with the aid of several enzymes and other proteins. Alternate repair pathways are summarized in Figure 9.25 (below). Here is a list of proteins involved in these homologous-recombination pathways:

- **MRX, MRN**: proteins that bind at double-stranded break and recruit other factors
- **Sae2**: an endonuclease active when phosphorylated to hydrolyze internal phosphodiester linkages
- **Sgs1**: a helicase that unwinds DNA that is under repair at a damaged RF
- **Exo1, Dna2**: single-strand exonucleases that hydrolyze terminal phosphodiester linkages
- **RPA, Rad51, DMC1**: proteins that bind to overhanging DNA to form a nucleoprotein filament and then initiate *strand invasion* at similar sequences

The activities of other enzymes involved in the repair are also identified in Figure 9.25 (below). But not shown in the illustration are products of the BRCA1 and BRCA2 genes which binding to Rad51 (the human RecA homolog). Expressed mainly in breast tissue, their wild-type (normal) protein products participate in homologous recombination repair of double-stranded DNA breaks.

When mutated and dysfunctional, BRCA1 and BRCA2 genes increase the likelihood of a woman getting breast cancer due to uncorrected DNA damage in breast cells. It doesn't help matters that the normal BRCA1 protein also plays a role in mismatch repair—and that the mutated protein can't!

To end this chapter, here is a bit of *weird science*! Consider the **tardigrade**, a tiny critter that can survive the vacuum in space, along with assorted forms of radiation, resisting DNA damage, and failing that, rapidly repairing DNA damage (^{9.6}[DNA Repair of the Tardigrade Genome](#)).

The Role of Homologous Recombination in Repair of a Double-Stranded Break

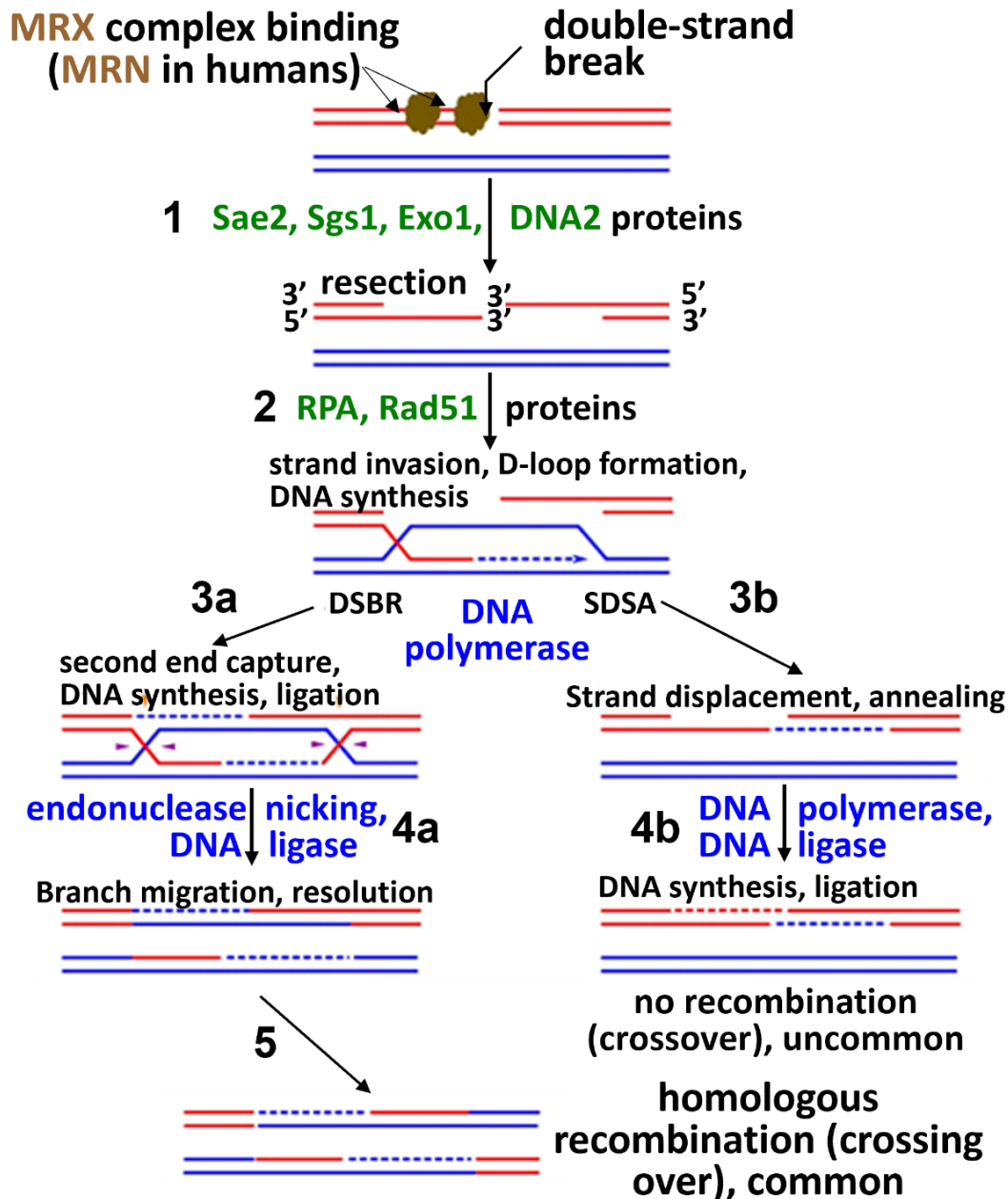


Fig. 9.25: In repairing a double-stranded DNA break, *MRX* or *MRN* proteins detect the break, after which several proteins (including *exonuclease*) enable trimming the strands at the break to make 'sticky ends' ((1). Other proteins, including *Rad51*, enable *strand invasion* while DNA polymerase replicates the invading strand (2). Alternate mechanisms called *second end capture* (3a) or *strand displacement* (3b) lead either to *branch migration* to resolve interlocking replicating DNA strands (4a), followed by *homologous recombination* to complete the repair (5), or to DNA synthesis and ligation to complete the repair without recombination (4b).

Some iText & VOP Key Words and Terms

base excision repair	high-speed blender	RecA protein
bidirectional replication	initiation	replication
clamp proteins	initiator proteins	replication fork
condensation reactions	lagging strand	replicons
deamination	leading strand	replisome
density-gradient centrifugation	methylation	satellite DNA
depurination	mutations	single-strand-binding proteins
direct repeats	nucleotide excision repair	supercoiling
discontinuous replication	Okazaki fragments	T2 phage
DNA ligase	origin of replication	tardigrade
DNA mismatch repair	phosphate backbone	telomerase
DNA polymerases I, II, & III	phosphodiester linkage	telomeres
DNA repair	primase	theta (θ) images
DNA sequence phylogeny	primer	topoisomerases
DNA strand breakage	processive replication	transcription-coupled repair
DNA topology	proofreading	transposase
elongation	pyrimidine dimers	VLP
env	pyrophosphate	VNTRs
helicase	RadA protein	

CHAPTER 9 WEB LINKS



Chapter 10

Transcription and RNA Processing

RNA transcription, RNA polymerases, initiation, elongation, termination, and processing

Reminder: For inactive links, google key words/terms for alternative resources.

SnRNP for Eukaryotic



Photo by G. Bergtrom

Home Movie Makers

10.1 Introduction

Transcription, the synthesis of RNA based on a DNA template, is the “central” step of the *Central Dogma* proposed by Crick in 1958. The basic steps of transcription are the same as for replication: **initiation**, **elongation**, and **termination**. The differences between transcription in prokaryotes and eukaryotes are in the details. Here are some:

- *E. coli* uses a single **RNA polymerase** enzyme to transcribe all kinds of RNAs, while eukaryotic cells use different RNA polymerases to catalyze the syntheses of **ribosomal RNA (rRNA)**, **transfer RNA (tRNA)**, and **messenger RNA (mRNA)**.
- In contrast to the eukaryotic genetic makeup, some bacterial genes are part of **operons** whose mRNAs encode multiple polypeptides.
- Most RNA transcripts in prokaryotes emerge from transcription ready to use.
- Eukaryotic transcripts synthesized as longer precursors undergo **processing** by **trimming**, **splicing**, or both!

- DNA in bacteria is virtually “naked” in the cytoplasm, while eukaryotic DNA is wrapped up in chromatin proteins in a nucleus.
- In bacterial cells, the association of ribosomes with mRNA and the translation of a polypeptide can begin even before the transcript is finished. This is because these cells have no nucleus. In eukaryotic cells, RNAs must exit the nucleus before they encounter *ribosomes* in the cytoplasm.

In this chapter, you will meet bacterial ***polycistronic*** mRNAs (transcripts of ***operons*** that encode more than one polypeptide). You'll read about eukaryotic ***split genes***, with ***introns*** and ***exons***). We will look at some details of transcription of the three major classes of RNA and then at how eukaryotes process precursor transcripts into mature, functional RNAs. Along the way, we will see one example of how protein structure has evolved to interact with DNA.

Learning Objectives

When you have mastered the information in this chapter, you should be able to do the following:

1. Discriminate between the three *steps of transcription* in prokaryotes and eukaryotes, and the *factors involved* in each.
2. State a hypothesis for why eukaryotes evolved complex *RNA processing* steps.
3. Speculate on why any cell in its “right mind” would have genes containing *introns* and *exons*, with the result that their transcripts would have to be processed by *splicing*.
4. Articulate the differences between *RNA and DNA structure*.
5. Explain the need for *sigma factors* in bacteria.
6. Speculate on *why* eukaryotes *do not have operons*.
7. List structural features of proteins that bind and *recognize specific DNA sequences* and speculate on why/how they came to differ.
8. Explain how proteins that *do not* bind specific DNA sequences *can* still bind to specific regions of the genome.
9. Suggest a reason for why bacteria don't polyadenylate most their mRNAs while eukaryotes do so.
10. Formulate a hypothesis for why bacteria do not “cap” their mRNAs.

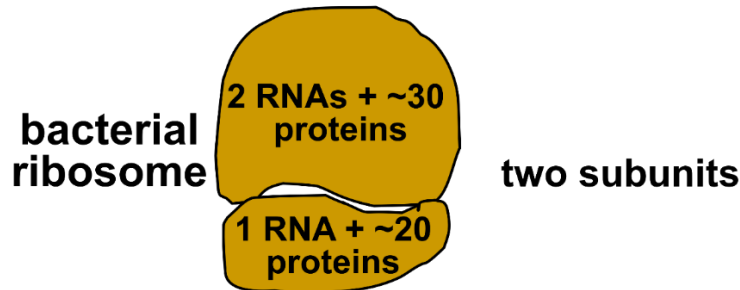
10.2 Overview of Transcription

Transcription, the synthesis of RNA, is the first step of the “central dogma” of DNA-to-protein information transfer. As we will see, some RNAs are translated into polypeptides, while others serve functional and even enzymatic roles in the cell. We begin with a look at the main kinds of RNA in cells.

10.2.1 The Major Types of Cellular RNA

All cells make three main kinds of RNA: ribosomal RNA (***rRNA***), transfer RNA (***tRNA***) and messenger RNA (***mRNA***). *rRNA* is a structural as well as enzymatic component of ribosomes, the protein-synthesizing machine in the cell; it's not surprising that rRNAs are by far the most abundant RNAs in a cell. In contrast, mRNAs are the least abundant. Three **rRNAs** and about fifty ribosomal proteins make up the two subunits of a bacterial ribosome, as illustrated in Figure 10.1 below.

rRNA (ribosomal RNA) ...part of ribosome *structure*



Eukaryotic cells have a larger ribosome with more proteins and an additional rRNA.

Fig. 10.1: Ribosomal RNAs (the 23S, 16S and 5S rRNAs) and as many as 50 proteins make up the bacterial ribosome shown here. This structure probably does not exist in isolation, but only when the subunits unite with an mRNA during translation. Eukaryotic ribosomal subunits are similar, but larger.

tRNAs decode base sequences of *mRNAs* into amino acids during protein synthesis (*translation*), thereby converting nucleic-acid-sequence information into the amino-acid sequences of polypeptides. The tRNAs that are attached to amino acids bind to ribosomes do so based on codon-anticodon recognition (Figure 10.2).

tRNA (transfer RNA) ...decodes genetic information

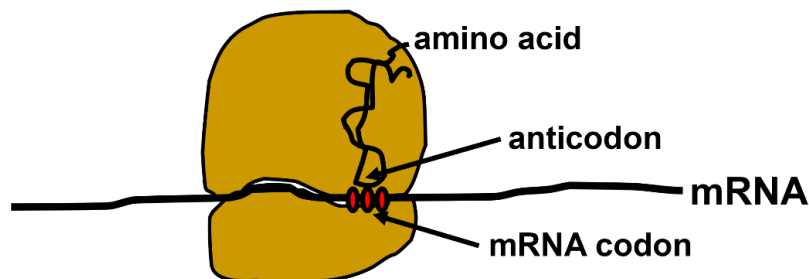


Fig. 10.2: Complete ribosomes form when the subunits associate with mRNA. Transfer RNAs (tRNAs) that decode the mRNA then 'enter' the ribosome. They are held in place by codon-anticodon complementarity, as well as by tRNA-ribosomal forces.



[186 Transcription Overview: Ribosomes and Ribosomal RNAs](#)



[187 Transcription Overview: Demonstrating the Major RNAs](#)



In 2009, ^{10.1}[Venkatraman Ramakrishnan](#), ^{10.2}[Thomas A. Steitz](#), and ^{10.3}[Ada Yonath](#) won a Nobel Prize in Chemistry for their studies of the structure and molecular biology of the ribosome.

The fact that genes reside inside the eukaryotic nucleus but that the synthesis of polypeptides (encoded by those genes) happens in the cytoplasm led to the proposal that there must be an mRNA. Sydney Brenner eventually confirmed the existence of mRNAs. Check out his classic experiment in Brenner S. (1961; *An unstable intermediate carrying information from genes to ribosomes for protein synthesis*. Nature 190:576-581).

Before we look at details of transcription, recall for future reference that multiple ribosomes can load an mRNA and move along it as *polyribosomes* (or *polysomes*), translating multiple copies of the same polypeptide (Figure 10.3).

mRNA (messenger RNA) ...encodes proteins

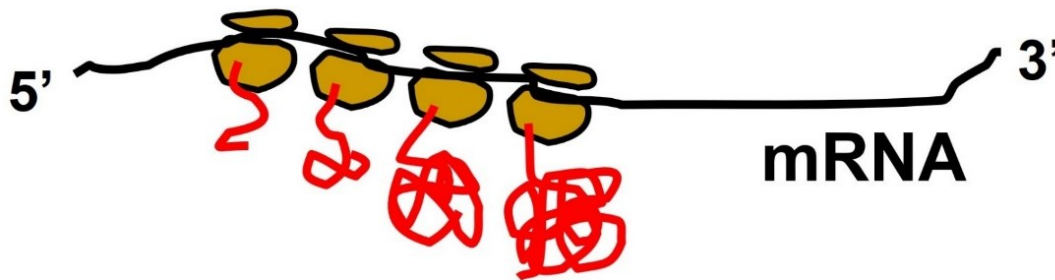


Fig. 10.3: *Polysomes* form when ribosomes assemble on an mRNA, one after another, initiating protein synthesis, each elongating and sequentially translating multiple polypeptides.

10.2.2 Key Steps of Transcription

As in replication, in transcription an *RNA polymerase* uses the template DNA strand of a gene to catalyze synthesis of a complementary, antiparallel RNA strand. Also, like replication, transcription is also error prone (though more so!). Here are some differences:

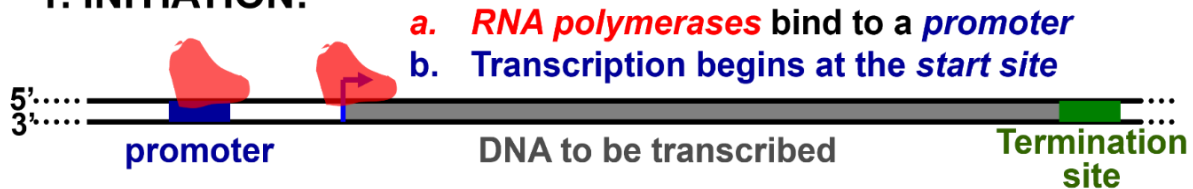
- RNA polymerases hydrolyze ribose nucleotide triphosphate (NTP) precursors, as they link the resulting nucleotide monophosphates (NMPs) to form RNA chains, while DNA polymerases use *deoxyribose nucleotide triphosphate* (dNTP) precursors.
- RNAs incorporate **uridines** (the *uracil* nucleotides) opposite a template *adenine* instead of **thymidines** (the *thymine* nucleotide). Thymidines end up opposite adenines in new DNA.
- In contrast to replication, RNA synthesis does not require a primer. With the help of *transcription initiation factors*, RNA polymerase locates the **transcription start site** of a gene and begins synthesis of a new RNA strand from scratch.

Several of the DNA sequences that characterize a gene are seen below in the summary of the basic steps of transcription in Figure 10.4. The *promoter* is the binding site for RNA

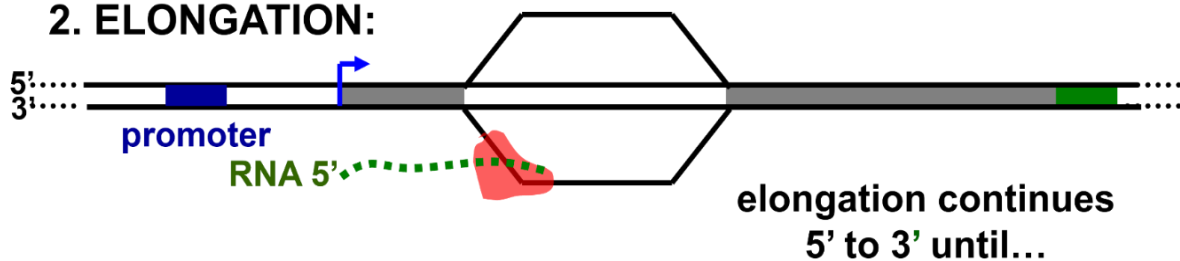
polymerase. It usually lies *upstream* of (5' to) the transcription start site (the bent arrow). Binding of the RNA polymerase positions the enzyme near the transcription start site, where it will start unwinding the double helix and begin synthesizing new RNA.

Transcription – the basic steps

1. INITIATION:



2. ELONGATION:



3. TERMINATION:

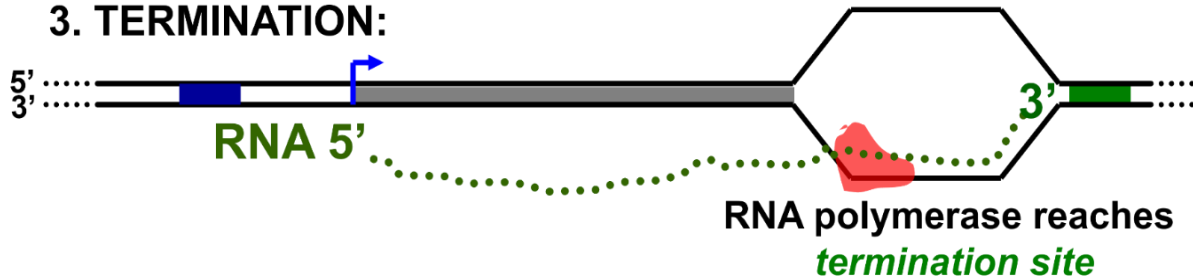


Fig. 10.4: The basic steps of transcription are (1) *Initiation*, the recognition of a promoter sequence near the transcription start site (the bent arrow) by *RNA polymerase*; (2) *Elongation*, the successive addition of nucleotides to a growing RNA strand; (3) *Termination*, the mechanism by which at the end of the gene, the newly made RNA is released.

In each of the three panels, the **transcription unit** is the DNA region to be transcribed, extending from the start site (the bent arrow to the right of the promoter) to a point just short of the termination site. Termination sites are typically *downstream* of (5' to) the transcription unit. By convention, *upstream* and *downstream* positions designate 5' and 3' regions of a given reference point on the DNA.



[188-2 Transcription Overview: The Basics of RNA Synthesis](#)



Some bacterial transcription units encode more than one mRNA. Bacterial *operons* are an example of this phenomenon. The resulting mRNAs can be translated into multiple polypeptides at the same time. In Figure 10.5 (below) RNA polymerase is transcribing an *operon* into a single mRNA molecule encoding three separate polypeptides.

Multiple genes of an operon are transcribed together and later translated as a single unit:

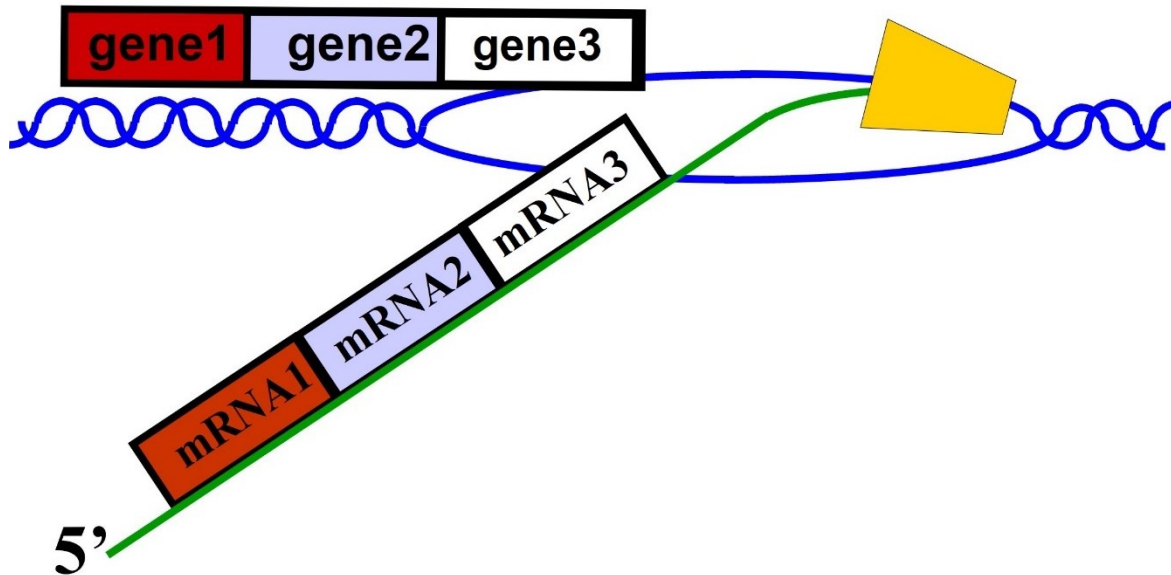


Fig. 10.5: An *operon* is a contiguous group of two or more genes in a bacterial genome that are transcribed as a single messenger RNA encoding all of the genes, all of which are then translated into polypeptides.

Transcription of all bacterial RNAs requires only one RNA polymerase. Different RNA polymerases catalyze rRNA, mRNA, and tRNA transcription in eukaryotes. We already noted that Roger Kornberg received the Nobel Prize in Medicine in 2006 for his discovery of the role of ***RNA polymerase II*** and other proteins involved in eukaryotic mRNA transcription.



[189 RNA Polymerases in Prokaryotes and Eukaryotes](#)



While mRNAs, rRNAs, and tRNAs are most of what cells transcribe. A growing number of other RNAs (e.g., *siRNAs*, *miRNAs*, and *lncRNAs*) are also transcribed. Some functions of these transcripts (including control of gene expression or other transcript use) are discussed in an upcoming chapter.

10.2.3 RNAs Are Extensively Processed After Transcription in Eukaryotes

Eukaryotic RNAs are processed (i.e., trimmed and chemically modified) from large precursor RNAs into mature, functional RNAs. These precursor RNAs (pre-RNAs, or *primary transcripts*) contain in their sequences the information necessary for their function in the cell.

Figure 10.6 provides an overview of the transcription and processing of the three major types of transcripts in eukaryotes.

Eukaryotic Transcript Processing

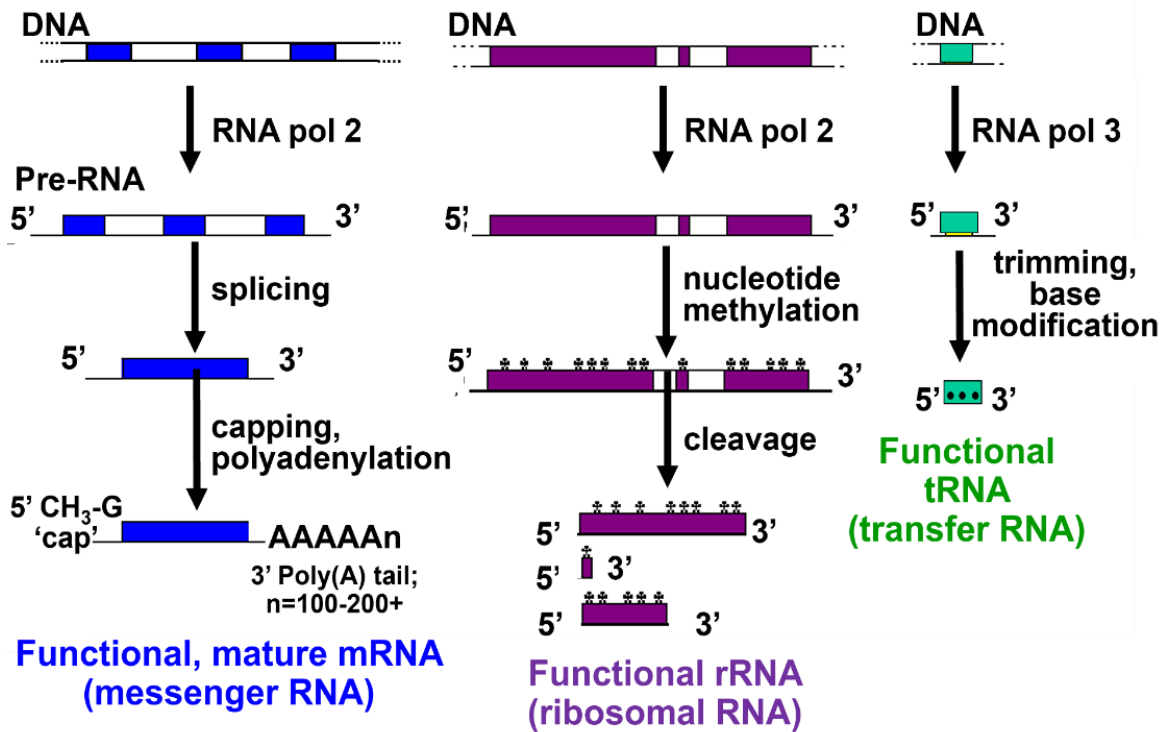


Fig. 10.6: Eukaryotic mRNAs, rRNAs, and tRNAs are transcribed by different RNA polymerases and are processed (hydrolyzed) by different mechanisms to yield functional RNAs. Pre mRNAs are spliced to remove intron, 'capped at their 5' ends and polyadenylated at their 3' ends (left). Pre-rRNAs are methylated and cleaved based on methylation patterns (middle). Pre-tRNAs processed by trimming their 5' and 3' ends and chemical modification of some bases (right).

To summarize the illustration:

- Many eukaryotic genes are "split," consisting of coding regions (**exons**) and non-coding (intervening) regions (**introns**). Transcription of split genes generates a primary mRNA (pre-mRNA) transcript. Pre-mRNA transcripts are **spliced** to remove the **introns** from the **exons**; exons are then ligated into a continuous mRNA. In some cases, the same pre-mRNA is spliced into alternate mRNAs that encode related but not identical polypeptides!
- Pre-rRNA is cleaved and/or trimmed (not spliced!) to make shorter mature rRNAs.
- Pre-tRNAs are trimmed, and some bases within the transcript are modified. Then three bases (not encoded by the tRNA gene) are enzymatically added to the 3' end.



[190 Posttranscriptional Processing; an Overview](#)

The details of transcription and processing differ substantially in prokaryotes and eukaryotes. Let's focus first on details of transcription itself and then RNA processing.

10.3 Details of Transcription

You can find a well-written summary of transcription in prokaryotes and eukaryotes on the National Institutes of Health (NIH) website at ^{10.4}[Transcription in Prokaryotes and Eukaryotes](#). At the link as well as here in this book, you will encounter proteins that bind DNA. Some proteins bind DNA to regulate transcription, *inducing* or *silencing* the transcription of a gene; we will discuss their role in the regulation of gene expression later. Other proteins interact with DNA simply to allow transcription; these include one or more that, along with RNA polymerase itself, must bind to the gene promoter to initiate transcription. We will look at bacterial transcription first.

10.2.4 Details of Transcription in Prokaryotes

In *E. coli*, a single RNA polymerase transcribes all kinds of RNA, associating with one of several sigma factor proteins (σ -factors) to initiate transcription. Different promoter sequences and their corresponding σ -factors play roles in the transcription of different genes (Figure. 10.7).

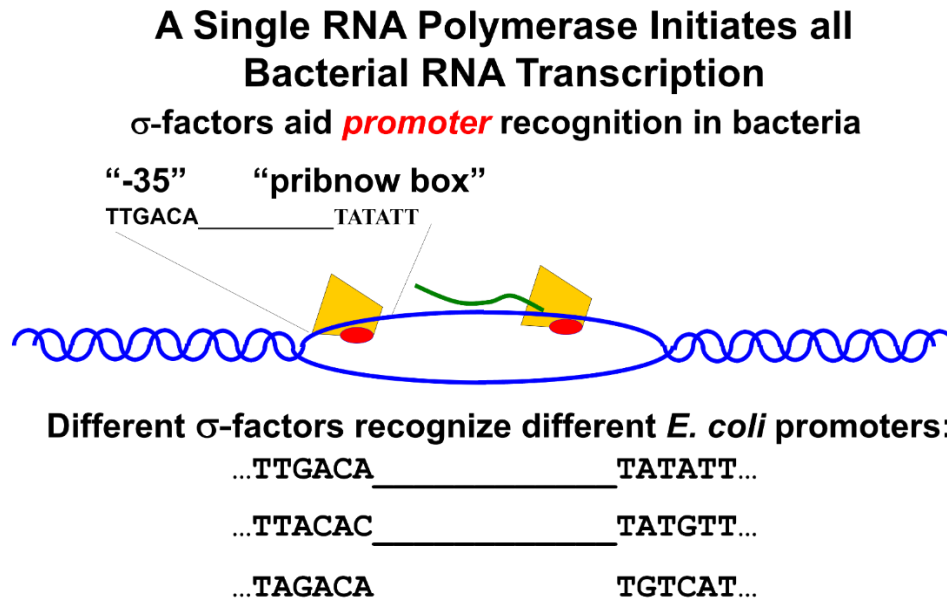


Fig. 10.7: All transcription begins when RNA polymerase binds to a two-part 'promoter' in mRNA (here shown as the '-35' and 'pribnow box' sequences). Bacterial RNA polymerase is associated with 'sigma factors' that allow it to recognize and bind to the promoter. Different 'sigma factors' recognize different promoter sequences.

In the absence of a σ -factor, *E. coli* RNA polymerase transcribes RNA, but it does so at a high rate and from random sequences in the chromosome. With a σ -factor bound to the RNA polymerase, the complex seemingly scans the DNA and recognizes and binds to the promoter sequence of a gene. The overall transcription rate is slower, but rather than random bits of the bacterial genome, only genes are transcribed! The **Pribnow box**, named for its discoverer, was the first RNA polymerase-binding sequence in a promoter to be characterized.

One way that bacteria regulate which genes are expressed is to selectively control the cellular concentrations of different σ -factors available to the *Pribnow box*. σ^{70} is the main

σ -factor of bacterial transcription initiation σ^{54} , a structurally unrelated factor regulates a variety of different genes, typically in response to stress (e.g., high temperature or antibiotic attack). Still other “alternative” σ -factors participate in activating specific genes or gene subsets (^{10.5}[Sigma Subunits and Bacterial Regulation](#)). Typically, the σ -factor falls off the RNA polymerase very soon after transcription is initiated, which then continues unwinding the double helix and elongating the transcript (Figure 10.8).

σ -Factors Function Only During Initiation

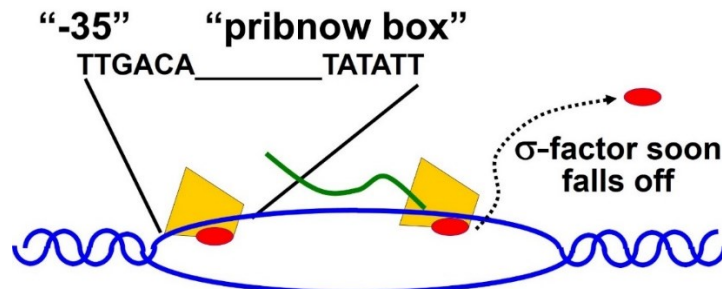


Fig. 10.8: After a bacterial RNA polymerase binds to the promoter (at the left), translation begins. No longer needed to initiate transcription, the 'sigma factor' dissociates from the polymerase (at the right), which then continues transcribing the RNA.

Elongation is the successive addition of nucleotides complementary to their DNA templates by forming phosphodiester linkages between nucleotides. The enzymatic condensation reactions of elongation are much like DNA polymerase-catalyzed elongation during replication. Transcription continues until RNA polymerase reaches the end of a gene. There are two ways that bacterial RNA polymerase “knows” when it has reached the end of a transcription unit.

One way involves the *rho* protein **termination factor**. As the RNA polymerase nears the 3' end of the nascent transcript, it transcribes a 72-base C-rich region, to which *rho* binds. *rho* is an ATP-dependent helicase that breaks the H-bonds between the RNA and the template DNA strand, preventing further transcription. Figure 10.9 (below) illustrates *rho*-dependent termination.

Rho-Dependent Transcription Termination

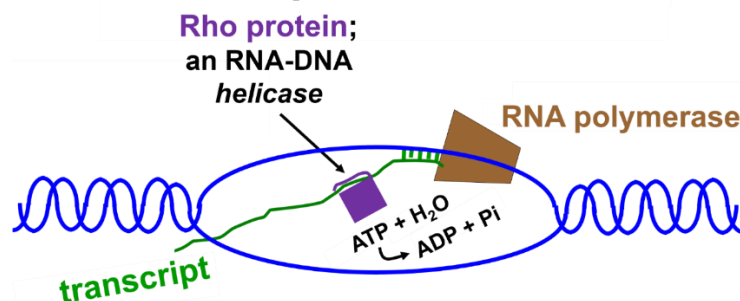


Fig. 10.9: *Rho*-dependent termination is one mechanism by which an RNA polymerase *knows* when it has finished transcribing. The *Rho* protein is an RNA-DNA helicase that recognizes a sequence near the end of the gene and transcript that *unwinds* the nascent RNA from the gene, ending transcription.

In the other mechanism of termination, a sequence near the 3' end of the transcript folds into a **hairpin loop** secondary structure, which serves as a termination signal, also causing dissociation of the RNA polymerase, template DNA, and the new RNA transcript. The role of the hairpin loop in ***rho-independent termination*** is shown in Figure 10.10.

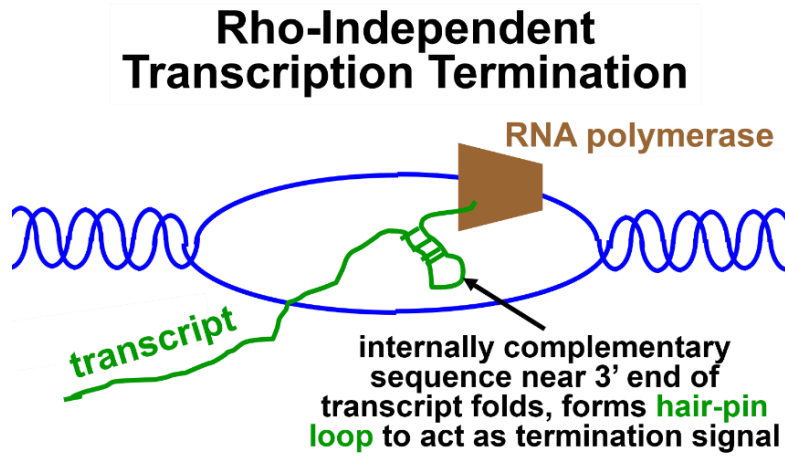


Fig. 10.10: *Rho-independent* termination is another mechanism for terminating bacterial transcription. A sequence near the end of the nascent transcript has internal complementary bases that fold into a *hairpin-loop* that acts as a *termination signal*.



[191 Details of Prokaryotic Transcription](#)

10.2.5 Details of Transcription in Eukaryotes

Unlike prokaryotes, eukaryotes use three different RNA polymerases to synthesize the three major RNAs (Figure 10.11).

Prokaryotic Transcription

Catalyzed by a single RNA polymerase

Eukaryotic Transcription

Catalyzed by 3 different RNA polymerases:

RNA pol I	RNA pol II	RNA pol III
28S, 18S, 5.8S rRNA	mRNA	4S, 5S tRNA rRNA
(>90%)	(<5%)	(~5%)

Fig. 10.11: The work of Eukaryotic transcription is more complex than that of prokaryotes. The latter transcribe all RNAs with a single RNA polymerase. Eukaryotes catalyze transcription of their rRNAs, mRNAs and tRNAs with different RNA polymerases.

With the help of initiation proteins, each RNA polymerase forms an *initiation complex* by combining with several **transcription factors** (TFs). Once the DNA at the start site of transcription unwinds, RNA polymerases catalyze the successive formation of phosphodiester linkages to elongate the transcript. Note that catalysis of the synthesis of most of the RNA in a eukaryotic cell (i.e., rRNAs) is by RNA polymerase I! These condensation reactions add ribose nucleotides to the free 3' end of a growing RNA molecule in reactions that are like those that elongate DNA strands. Unfortunately, the details of the eukaryotic transcription termination are not as well understood as they are in bacteria. Here we focus on initiation, followed by discussion of the processing of different eukaryotic RNAs into ready-to-use molecules.

10.3.2.a Eukaryotic mRNA Transcription

Transcription of eukaryotic mRNAs by *RNA polymerase II* begins with sequential assembly of a eukaryotic *initiation complex* at a gene promoter. The typical promoter for a protein-encoding gene contains a **-T-A-T-A- sequence motif**, or **TATA box**, and other short upstream sequences that recruit components of the initiation complex. The many steps of eukaryotic mRNA transcription initiation are illustrated in Figure 10.12 (below).

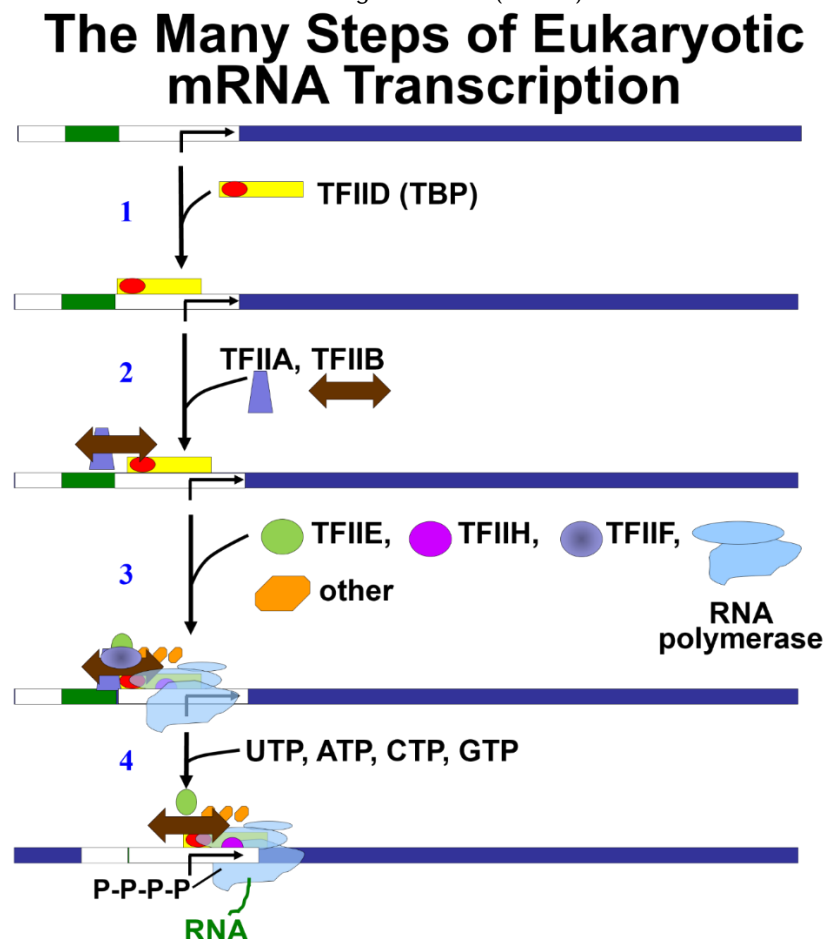


Fig. 10.12: Eukaryotic mRNA-transcription initiation starts with recognition of a promoter *TATA box* by initiation factor TFIID (TBP) (1). The association of a series of initiation factors follows (2, 3). RNA polymerase II binds last, and its phosphorylation starts mRNA transcription using the UTP, ATP, CTP, and GTP nucleotide precursors (4).

TATA-binding protein (*TBP*) first binds to the *TATA box*, along with *TFIID* (*transcription factor IID*). This intermediate in turn recruits *TFIIA* and *TFIIB*. Then *TFIIE*, *TFIIF*, and *TFIIH* are added to form the **core initiation complex**. Finally, several other initiation factors and RNA polymerase II bind to complete the mRNA transcription-**initiation complex**. Phosphorylation adds several phosphates to the amino terminus of the RNA polymerase, after which some of the transcription factors dissociate from the initiation complex. The remaining RNA-polymerase transcription-factor complex can now start making the mRNA.

The first task of the complete initiation complex is to unwind the template DNA strands at the start site of transcription. Unlike prokaryotic RNA polymerase, eukaryotic RNA Polymerase II does not have an inherent DNA helicase activity. For this, eukaryotic gene transcription relies on the multi-subunit TFIID protein, two of whose subunits have an ATP-dependent helicase activity. Once transcription is initiated, helicase activity is no longer required for continued elongation of the RNA strand. Consistent with the closer relationship of *archaea* to eukaryotes (rather than to prokaryotes), archaeal mRNA transcription initiation more closely resembles transcription in eukaryotes, albeit requiring fewer initiation factors during formation of an initiation complex.



[192-2 Eukaryotic mRNA Transcription](#)

A significant difference between prokaryotic and eukaryotic transcription is that RNA polymerases and other proteins involved at a eukaryotic gene promoter do not see naked DNA. Instead, they recognize specific DNA sequences behind a coat of chromatin proteins, and some of whose bases have been chemically modified.

On the other hand, all proteins that interact with DNA have a common a need to recognize the DNA sequences to which they must bind—within the double helix. In other words, they must see the bases in the interior of the helix, not its uniformly electronegative phosphodiester backbone surface. To this end, they must penetrate the DNA, usually through the **major groove** of the double helix. We will see that DNA regulatory proteins face the same problems in achieving specific shape-based interactions!



[193-2 Recognition of Transcription Factors at Promoters](#)

10.3.2.b Eukaryotic tRNA and 5S rRNA Transcription

It must have been something of a surprise to discover that the promoter sequence that binds RNA polymerase III to begin transcribing 5S rRNAs and tRNAs is **not** upstream of the transcribed sequence; instead, it is located within the transcribed sequence. To begin transcription of these RNAs, RNA polymerase III associates with structural relatives of the RNA polymerase II transcription factors (TFs) to form a core initiation complex at these “internal promoters.”

After additional, *gene-specific* TFs combine with this core complex, the finished initiation complex repositions RNA polymerase III from the internal promoter sequence to the upstream transcription start site to begin RNA synthesis. As a result, the final 5S rRNA and tRNA transcripts still contain retain the 'promoter' sequence! Note that, unlike initiation of mRNA transcription, RNA polymerase III-dependent transcription initiation does not require an independent DNA unwinding activity (i.e., a helicase). Figure 10.13 (below) illustrates the transcription of a 5S rRNA (or a tRNA) by RNA polymerase III.

RNA pol III Transcription of 5S rRNA or 4S tRNAs From Internal Promoters

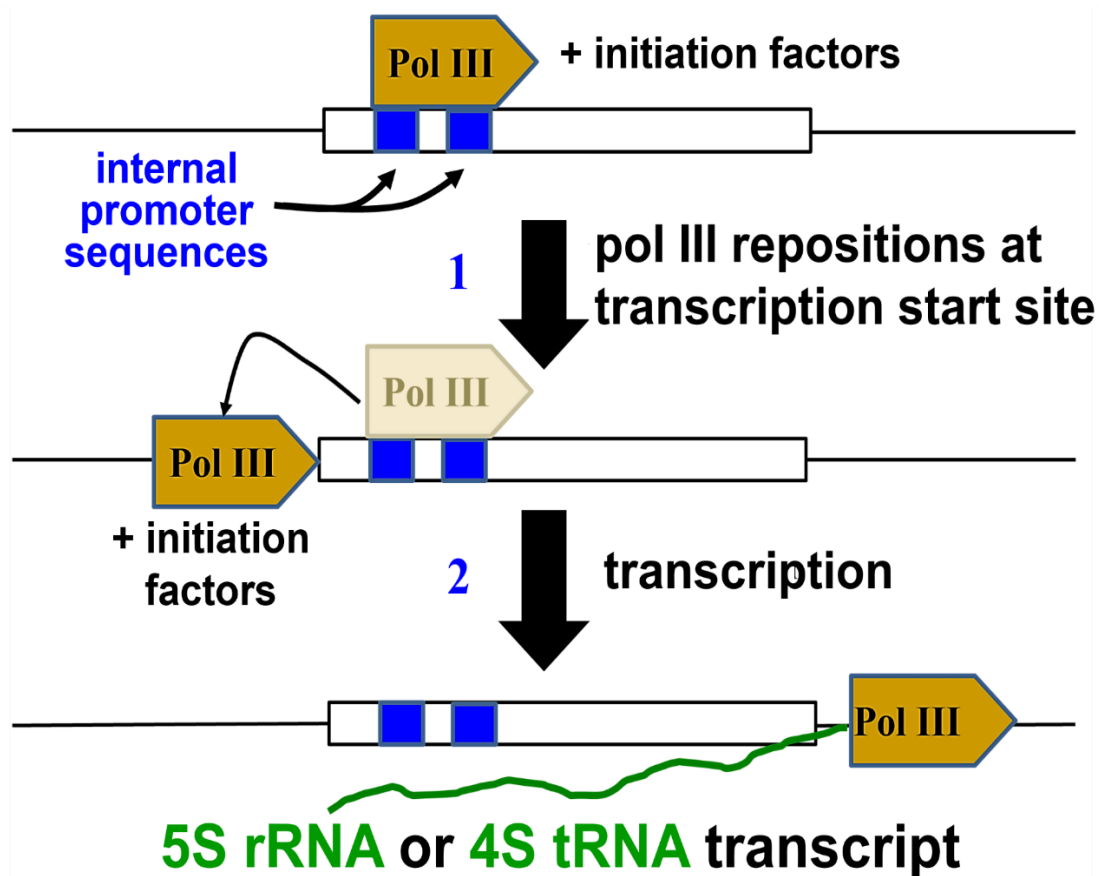


Fig. 10.13: RNA polymerase III initiates 5S rRNA and tRNA transcription from an *internal promoter* with the aid of transcription factors. After binding to the promoter, the polymerase repositions itself near the transcription start site (1). Transcription begins, producing 5S rRNAs and tRNAs that include their promoter sequences (2).

10.3.2.c Transcription of the Other Eukaryotic rRNAs

The production of the other rRNAs requires RNA polymerase I. A 45S precursor rRNA (*pre-rRNA*) is made and then processed into 28S, 18S, and 5.8S rRNAs (see section 10.5). RNA polymerase I and "core" TFs form a *core initiation complex* at the upstream promoters of the 45S genes. The addition of several gene-specific TFs completes assembly of the initiation complex, which then unwinds the promoter DNA, again without a separate helicase.

More information about the structure and function of eukaryotic transcription initiation complexes can be had at ^{10.6}[Three Transcription Initiation Complexes](#). As already noted, transcription termination is not as well understood in eukaryotes as in prokaryotes. Coupled termination and polyadenylation steps common to most prokaryotic mRNAs are discussed in more detail in the next section, and a useful summary can be found at the NIH-NCBI website, ^{10.7}[Eukaryotic Transcription Termination](#).

10.4 Details of mRNA Processing in Eukaryotic Nuclei

Eukaryotic mRNA primary transcripts undergo extensive processing, including *splicing*, *capping* and *polyadenylation*. The steps described here are considered in order of (sometimes overlapping!) occurrence. We begin with splicing—an mRNA phenomenon.

10.4.1 Spliceosomal Introns

The coding regions of bacterial genes are continuous. The discovery of eukaryotic split genes with introns and exons came as quite a surprise. It seemed incongruous for evolution to have stuck irrelevant DNA in the middle of coding DNA! For their discovery of split genes, Richard J. Roberts and Phillip A. Sharp shared the Nobel Prize for Physiology or Medicine in 1993. In fact, all but a few eukaryotic genes are split, and some have two or more *introns* separating bits of coding DNA, the *exons*. Figure 10.14 summarizes *splicing* to remove introns in pre-mRNAs and to splice exons (in the right order!) to make a mature, functional mRNA.

***RNA polymerase II* transcribes precursor mRNAs (pre-mRNAs) that are *spliced*:**

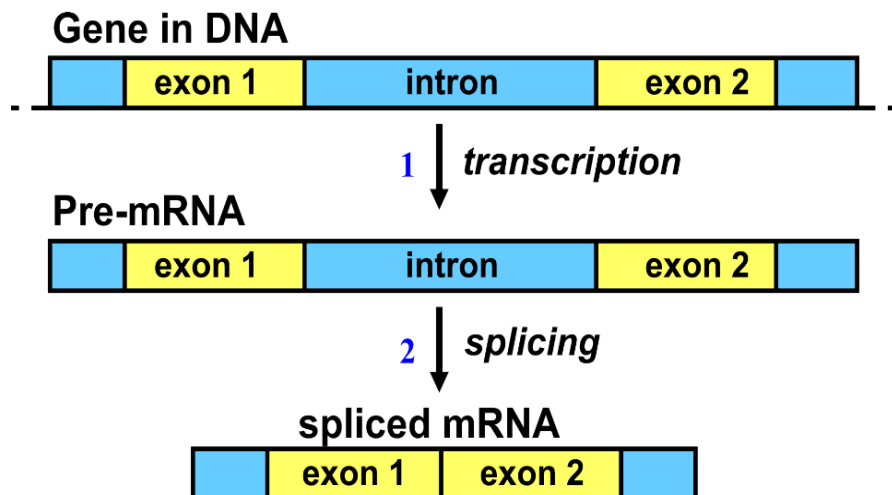


Fig. 10.14: Eukaryotic precursor mRNAs are encoded by split genes containing noncoding introns and coding exons. After transcription of a pre-mRNA (1), splicing removes introns and ligates exons into a contiguous polypeptide coding sequence (2).

snRNPs are particles composed of RNA and ribonuclear proteins. They bind to specific sites in an mRNA and then direct a sequential series of cuts and ligations (the *splicing*) necessary to process the mRNAs. The process was reminiscent of the way in which movies were edited (see the photo of an early splicing "machine" at the top of this chapter), hence the

term *splicing* to describe mRNA processing. Figure 10.15 below highlights the role of *small ribonuclear proteins* (*snRNPs*) in splicing.

snRNPs (*small nuclear RiboNuclear Proteins*) bind to pre-mRNA:

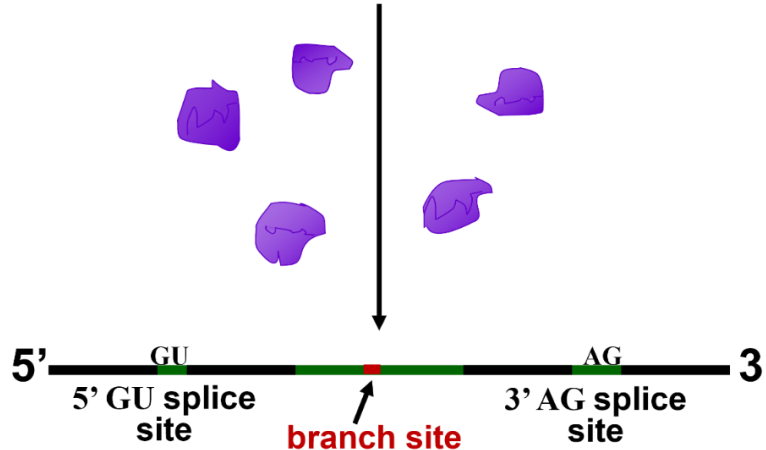
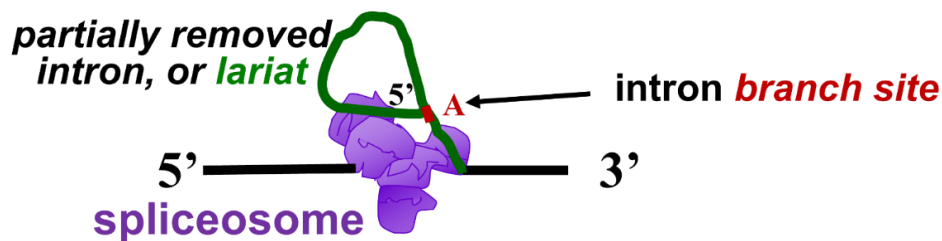


Fig. 10.15: Splicing a eukaryotic pre-mRNA involves association of the primary transcript (*pre-mRNA*) with five *snRNPs* (*small nuclear RiboNuclear Proteins*) that catalyze cleavage (hydrolysis) of the pre-mRNA at 5' and 3' splice sites (bottom of image), to be followed by ligation of the 3' to 5' exon ends.

When snRNPs (often pronounced “snurps”) bind to a pair of splice sites flanking a pre-mRNA intron, they form a *spliceosome*, which completes the splicing, including the removal of intermediate *lariat* structures of the intron. The last step is to ligate exons into a continuous mRNA, with all its codons intact and in the right order—an impressive trick for some pre-mRNAs with as many as fifty introns! Spliceosome action is summarized in Figure 10.16 and in the “*mRNA Splicing*” animation at one of the following links.

Forming a *Spliceosome*, sort of ‘Super-Enzyme’ to Orchestrate Splicing.



snRNP binds to pre-mRNA forms *spliceosome*; hydrolyses to remove intron start at 5' splice site.

Fig. 10.16: The binding of snRNPs to a pre-mRNA leads brings 5' and 3' splice sites close to each other, catalyzing the hydrolysis at the sites and forming a *lariat* structure of the intron remnants.



[194-2 The Discovery of Split Genes](#)



[195 mRNA Splicing](#)



10.4.2 Specific Nuclear-Body Proteins Facilitate snRNP Assembly and Function

Recall the organization of nuclei facilitated by nuclear bodies, discussed in an earlier chapter. Nuclear **speckles** are associated regions of high mRNA transcription, processing, and splicing. (Check out the animation at ^{10.8}[Nuclear Bodies and Transcription](#) for a 3D localization of markers of nuclear body activity.)

Cajal bodies (CBs) and **Gems** are nuclear bodies that are similar in size and have related functions in assembling spliceosomal snRNPs. Some splicing defects correlate with mutations in the **coil** protein associated with *Cajal bodies*; others correlate with mutations in **SMN** proteins normally associated with *Gems*. One hypothesis was that CBs and Gems interact in SnRNP and spliceosome assembly—but how?

Consider the results of an experiment in which fluorescence-tagged antibodies to **coilin** and the **SMN** protein were localized in undifferentiated and differentiated *neuroblastoma* cells (Figure 10.17).

Nuclear Body Localization

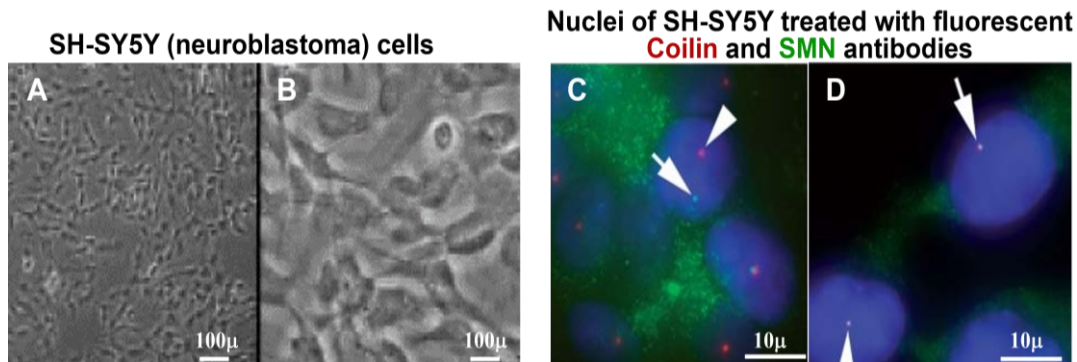


Fig. 10.17: Light (A, B) micrographs of undifferentiated (A) and differentiated (B) *neuroblastoma* cells show that Cajal bodies and Gems aggregate during differentiation. Fluorescent coilin and SMN antibodies localize separately in undifferentiated cells (C), but together after in differentiation (D). Since mutations in coilin and SMN proteins are associated with splicing defects, their co-localization after differentiation suggests that Cajal bodies and Gems have become involved in snRNP function.

Panels A and C above are undifferentiated cells in culture; panels B and D are cells that were stimulated to differentiate. In the fluorescence micrographs at the right, arrows point to fluorescent nuclear bodies. In panel C, the green, fluorescent antibodies to coilin (arrows) localize to CBs and the red fluorescent antibodies to SMN protein (arrowhead) binds to Gems—as expected. But in panel D, the two antibodies co-localize, suggesting that the CBs and Gems aggregate in the differentiated cells. This would explain the need for both coilin and SMN proteins to produce functional snRNPs. The CBs and gems may aggregate in differentiated cells because of an observed increase in expression of the SMN protein, producing more Gems and driving the association with CBs.

This and similar experiments demonstrate that different nuclear bodies do have specific functions, that they are not random structural artifacts sample preparation, and that they have in fact, evolved to organize nuclear activities in time and space in ways that are essential to the cell.

10.4.3 Group I and Group II *Self-Splicing* Introns

While eukaryotic *spliceosomal introns* are spliced using snRNPs as described above, *Group I* or *Group II* introns are removed by different mechanisms.

Group I introns interrupt mRNA and tRNA genes in bacteria and in the genomes of mitochondria and chloroplasts. They are also occasionally found in bacteriophage genes, but rarely in nuclear genes (and then only in lower eukaryotes). Group I introns are *self-splicing*! Thus, they are themselves *ribozymes*. They do not require a spliceosome with its snRNPs or other proteins for splicing activity. Instead, they fold into a secondary stem-loop structure that positions catalytic nucleotides at appropriate splice sites to excise their own introns and to religate the exons.

Group II introns are found in chloroplast and mitochondrial rRNA, mRNA, tRNA and some bacterial mRNAs. These can be quite long. They form complex stem-loop tertiary structures and self-splice, at least in a test tube! However, *Group II introns* encode proteins required for their own splicing in vivo. Like spliceosomal introns, they form a lariat structure at an A-residue branch site. All of this suggests that the mechanism of spliceosomal intron splicing evolved from that of *Group II* introns.

10.4.4 So Why Splicing?

Whether via spliceosomes or by self-splicing in organelles, the puzzle implied by any kind of splicing is why? Why do higher organisms have split genes in the first place? While the following discussion can apply to all splicing, it will reference mainly spliceosomal introns. Here are some answers to the question.

- Introns in the nuclear genes of some species are often longer (much longer!) than their exons. Because introns are *noncoding* (i.e., *non-informational*) making them large targets for mutation. In effect, most noncoding DNA—including *introns*—can buffer the ill effects of random mutations.
- You may recall that gene duplication on one chromosome and the loss of a copy from its homologue arise from unequal recombination (nonhomologous crossing over). It can occur when similar DNA sequences align during *synapsis* of meiosis. In unequal cross-over involving whole genes, one chromosome will lose the gene while its homologue will contain two copies of the gene. In an organism that inherits a chromosome with both gene copies, the duplicate can accumulate mutations if the other retains original function. The diverging gene then becomes part of a pool of selectable DNA, the grist of evolution. Descendants of organisms that inherit the duplicated genes have diversified the gene pool, again increasing the potential for evolution and species diversity.
- Unequal recombination can also occur between similar sequences (e.g., in introns) in the same or different genes, resulting in a sharing of exons between genes. After unequal recombination between the introns that flank an exon, one gene will acquire another exon while the other will lose it. Once again, if an organism retains a copy of the participating genes with original function, the organism can make the required protein and survive. Meanwhile, the gene with the extra exon may produce a similar protein, but one with a new structural domain and function.

- Just like a *complete duplicate gene*, one with a new exon that adds a new function to an old gene has been entered in the pool of selectable DNA. By creating proteins with different overall functions that nonetheless share at least one domain and one common function, the phenomenon of **exon shuffling** increases species diversity!

An example discussed earlier involves calcium-binding proteins that regulate many cellular processes. Structurally related *calcium* (Ca^{++}) *binding domains* are common to many otherwise structurally and functionally unrelated proteins. Consider exon shuffling in the unequal crossover (nonhomologous recombination) shown in Figure 10.18.

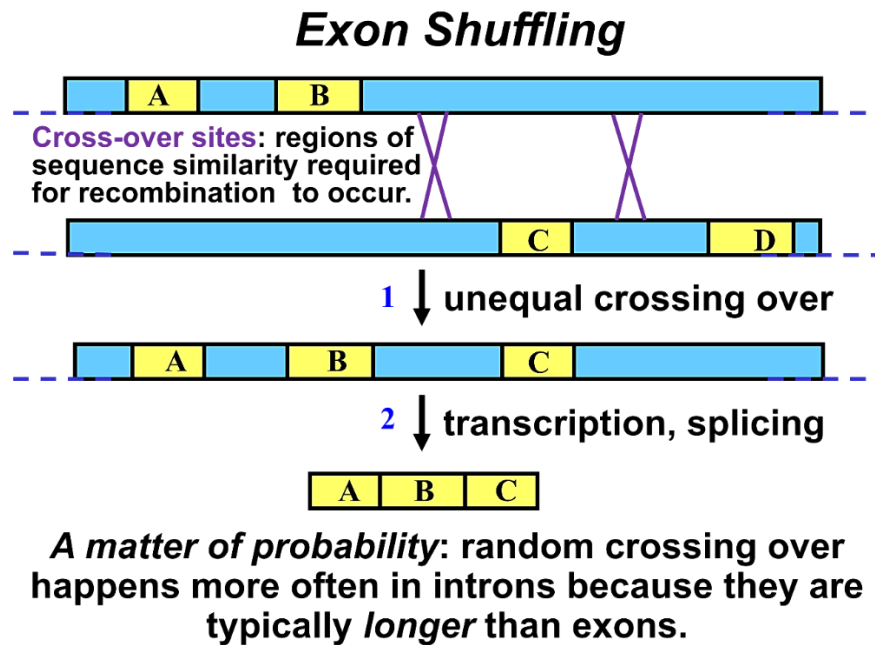


Fig. 10.18: *Exon shuffling*, the transfer or copying of an exon from one gene to another can occur by unequal recombination (cross-over) between different (nonhomologous) genes due to some sequence similarity within introns. The common sequences can align during synapsis in meiosis. If unequal crossing over occurs within introns or non-coding DNA (1) it creates a new gene, in this case one with three exons. Transcription and splicing of a pre-mRNA from this new gene produce a new protein.

In this example, regions of strong similarity exist in different (nonhomologous) introns in the same gene. These regions align during synapsis of meiosis. Unequal crossing over between the genes inserts exon C into one of the genes. The other gene loses the exon (not shown in the illustration).

In sum, introns are *buffers against deleterious mutations*, and they are equally valuable as potential targets for gene duplication and exon shuffling. This makes introns key players in creating *genetic diversity*, the hallmark of evolution.



[196-2 Origin of Introns](#)



[197 Intron Evolution: What was selected here?](#)



10.4.5 5-Prime Capping

Using GTP as a precursor, a *methyl guanosine CAP* added 5'-to-5' to an mRNA functions, in part, to help mRNAs leave the nucleus and associate with ribosomes. The CAP is added to an exposed 5' end, even as transcription and splicing are still in progress. A *capping enzyme* places a methylated guanosine residue at the 5' end of the mature mRNA. Figure 10.19 shows the 5' CAP structure (the check marks are 5'-3' linked nucleotides).

Capping: enzyme-catalyzed addition of CH₃-guanosine to 5' ends of transcripts:

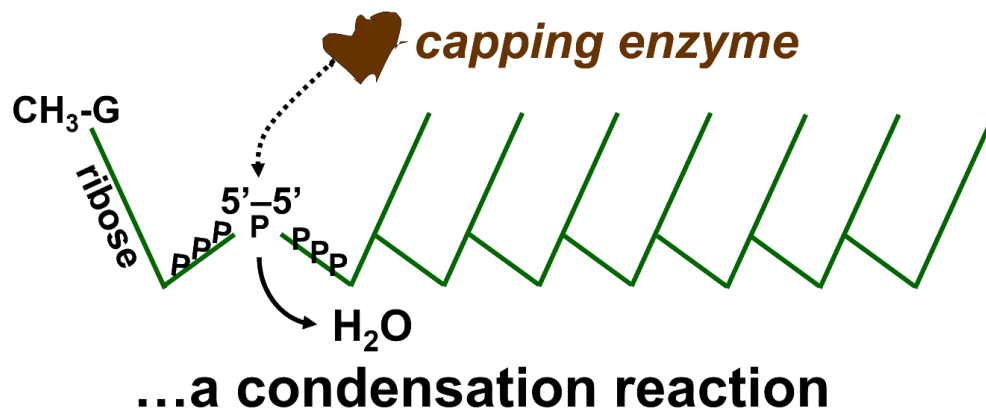


Fig. 10.19: Capping of mRNA results from catalysis of a 5'-to-5' condensation reaction linking the 5' end of an mRNA to a methylated guanine triphosphate (CH₃-GTP).

10.4.6 3-Prime Polyadenylation

A *poly(A) polymerase* catalyzes the addition of adenine monophosphates (AMPs) to the 3' end of most eukaryotic mRNAs, even before any splicing is complete. Polyadenylation requires ATP and can add several hundred AMPs to the 3' terminus of an mRNA. The enzyme binds to an -A-A-U-A-A- sequence near the 3' end of an mRNA and starts adding the AMPs. Polyadenylation after transcription termination is illustrated in Figure 10.20.

Eukaryotic mRNAs also get a 3' poly(A) tail...

poly(A) polymerase recognizes sequence near 3' end of mRNA...

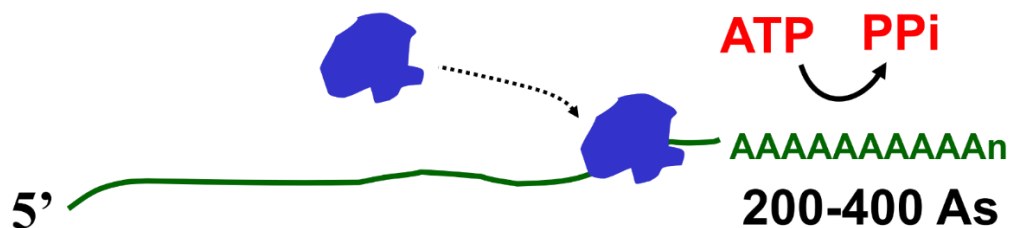


Fig. 10.20: mRNA *polyadenylation* is the addition of multiple adenine nucleotides to the 3' end of an mRNA, catalyzed by the enzyme *poly(A) polymerase*. Other than histone mRNAs, most eukaryotic mRNAs are polyadenylated.

The result of *polyadenylation* is a 3' ***poly(A) tail*** whose functions include assisting in the transit of mRNAs from the nucleus and regulating the half-life of mRNAs in the cytoplasm. The poly(A) tail shortens each time a ribosome finishes translating the mRNA.



[198 mRNA 5' Capping and 3' Polyadenylation](#)

10.5 Ribosomal RNA Processing in Eukaryotic Nuclei

In most eukaryotes, a large rRNA gene transcribes a 45S precursor transcript containing (from shortest to longest) 5.8S rRNA, 18S rRNA and 28S rRNA. The *S* comes from Theodor Svedberg, the biochemist who developed the *sedimentation velocity ultra-centrifugation* technique to separate molecules like RNA by size. The higher the *S* value, the larger the molecule and therefore the faster it moves through the viscous sugar gradient during centrifugation. RNA polymerase I transcribes 45S precursor rRNAs (pre-rRNAs) from multiple large *transcription units* in the genome. One such *unit* is shown in Figure 10.21.

45S rDNA *transcription units* transcribed as precursors are processed into 3 mature rRNAs:

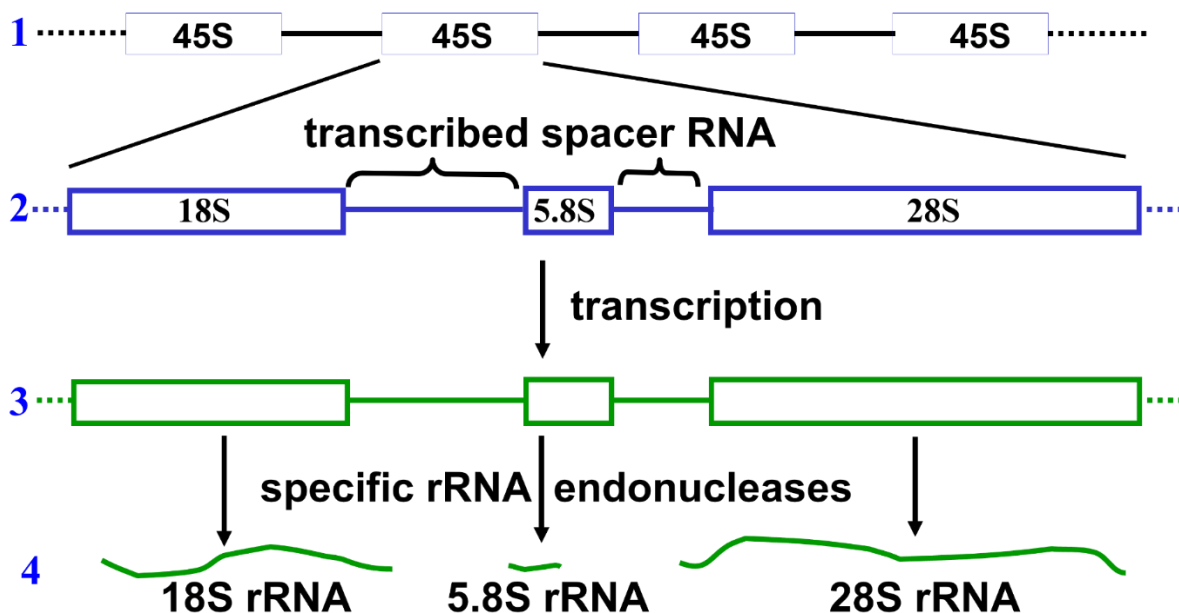


Fig. 10.21: The 45S eukaryotic rRNA genes (1) encode the 18S, 5.8S and 28S ribosomal RNAs (2). Pre-rRNA transcripts contains the rRNAs (3). Hydrolytic cleavage of the pre-rRNA (3) will separate the rRNAs and specific rRNA nucleases will hydrolyze spacer and flanking RNA, leaving behind mature rRNAs (4).

The 45S pre-rRNA is processed by hydrolytic cleavage. The many copies (two hundred to four hundred!) of the 45S gene in eukaryotic cells might be expected, since making proteins (and therefore ribosomes) will be an all-consuming cellular activity. In humans, 45S genes (45S rDNA) are distributed among five *acrocentric* chromosomes (those that have a centromere very near one end of the chromosome).

Inside the nucleus, the 45S rDNA in chromosomes is packed into the *nucleolus*. Because these genes are present in so many copies and are organized into a specific region of chromatin, it is possible to visualize 45S transcription in progress in electron micrographs such as the ones in Figure 10.22.

Light Micrographs of Amphibian Lampbrush Chromosomes

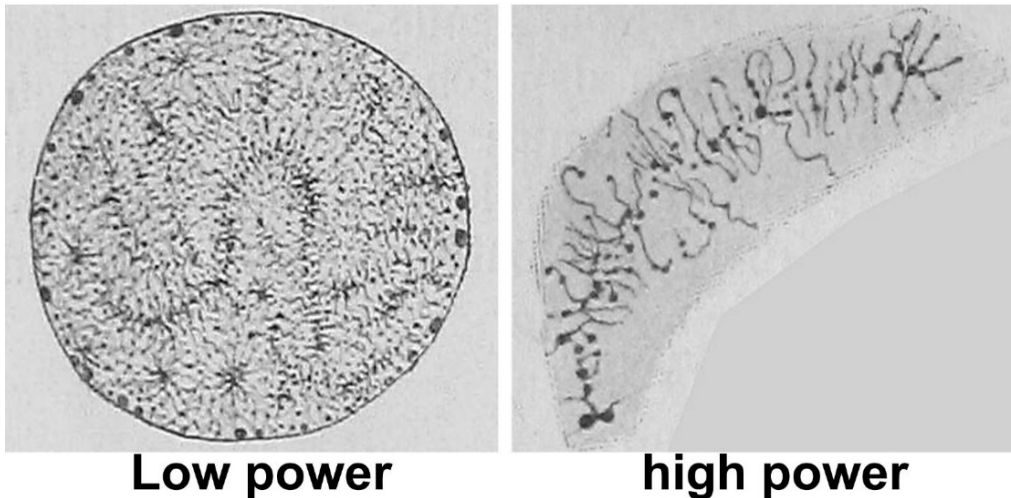


Fig. 10.22: As early as 1906, *chromatic fibers* were described in light micrographs of stained nucleoli of salamander oocytes. At high power (left) these *fibers* looked like a bottle brush used for cleaning bottles, or a lamp brush used to clean the chimneys of old-fashioned kerosene lamps. Today we know *chromatic fibers* as *lampbrush chromosomes*, and we also know that the bristles are nascent strands of rRNA arranged along transcribed 45S genes.

The term *lampbrush* comes from the shape of the 45S genes undergoing transcription; the RNAs extending from the DNA template look like an old-fashioned brush used to clean the chimney of a kerosene lamp. For a high-resolution transmission electron micrograph, see ^{10.9}[High Res Lampbrush Chromosomes](#).

Multiple gene copies encode 5S rRNAs. However, unlike the 45S rRNA genes, 5S rRNA genes may be spread among many chromosomes (seven in *Neurospora crassa*, the bread mold). Or in the case of humans, 5S RNA gene copies are found distributed along chromosome 1. The 5S rRNA genes are transcribed by RNA polymerase III with minimal posttranscriptional processing. As we noted, the promoters of the 5S genes are within the transcribed part of the genes, rather than upstream of their 5S transcription units.



[199-2 rRNA Transcription & Processing](#)



10.6 tRNA Processing in Eukaryotic Nuclei

RNA polymerase III also transcribes tRNA genes from internal promoters, but unlike the 5S rRNA genes, tRNA genes tend to cluster in the genome (Figure. 10.23, below).

Genomic map of the 4 kb region around the *MetA-MetB* operon. The map shows the locations of genes *Phe*, *Tyr*, *MetA*, *MetB*, *Asn*, *Ala*, *Leu*, and *Lys*. Arrows indicate the direction of transcription for each gene.

Also, unlike 5S rRNA genes, tRNA primary transcripts are processed by trimming, the enzymatic addition of a 3'-terminal-**C-C-A** triplet, and the modification of bases internal to the molecule. A yeast tRNA showing these modifications is illustrated in Figure 10.24.

Diagram illustrating the structure and modification of tRNA. The tRNA molecule is shown with its characteristic L-shaped tertiary structure. Key regions labeled include the D-loop, TΨC loop, and variable loop. Modified bases are indicated by blue boxes: D, m²G, m⁵C, m⁷G, m¹A, m²G, m⁵C, ψ, and C_m. The anticodon loop is at the bottom, containing the anticodon GmA A. The acceptor stem is at the top, with the amino acid attachment site (3' end) labeled. The sequence 5' to 3' is: pG-C-G-C-G-U-A-U-U-A-U. The 3' end is labeled 3' and has an OH group. The amino acid attachment site is labeled 'amino acid attachment site'. The acceptor stem is labeled 'acceptor stem'. The TΨC loop is labeled 'TΨC loop'. The D-loop is labeled 'D-loop'. The variable loop is labeled 'variable loop'. The anticodon loop is labeled 'anticodon loop'. The anticodon is labeled 'anticodon'.

The yeast tRNA folds into several *hairpin loops* based on internal H-bond formation between complementary bases in the molecule. An amino acid specific for the tRNA will be enzymatically added to the **A** residue of the 3' terminal **-C-C-A**.



10.7 Export of mRNA and Ribosomes from the Nucleus

The synthesis and processing of rRNAs are coincident with the assembly of the ribosomal subunits (Figure 10.25).

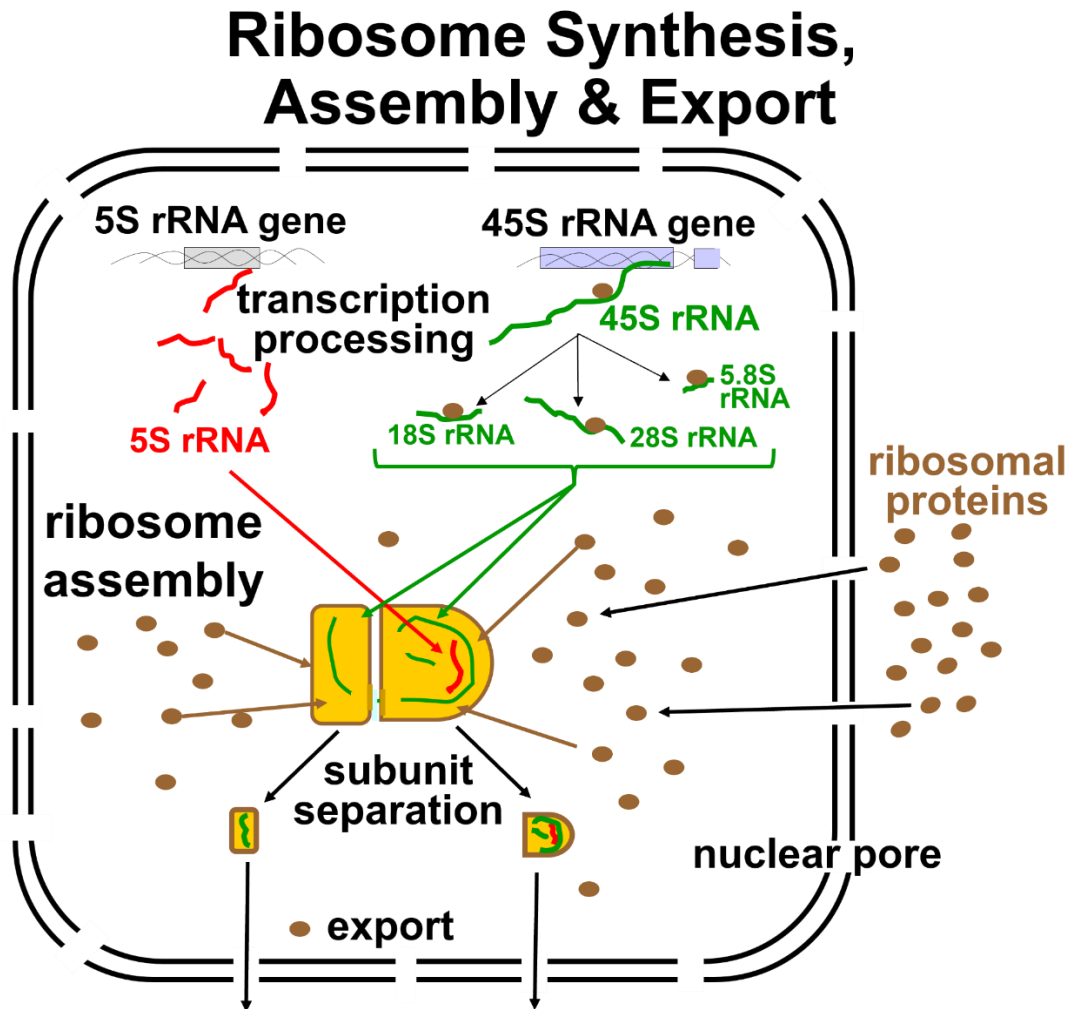


Fig. 10.25: Cells coordinate the export of ribosomal subunits from nucleus to cytoplasm with their assembly from mature rRNAs and ribosomal proteins imported from the cytoplasm.

The 45S pre-rRNAs initially bind to ribosomal proteins in the nucleolus (that big nuclear body!) to initiate assembly, and then serve as a scaffold for the continued addition of ribosomal proteins to both the small and the large ribosomal subunits. After the 5S rRNA is added to the nascent large ribosomal subunit, processing (cleavage) of 45S rRNA is completed, and the subunits are separated. The separated ribosomal subunits exit the nucleus to the cytoplasm, where they will associate with mRNAs to translate new proteins. To better understand what is going on, try summarizing what you see here in the correct order of steps. You can also see this process animated at this link: ^{10.10}[Ribosome Assembly & Nuclear Export](#).

Messenger RNAs are independently exported from the nucleus. Their 5' methyl guanosine caps and poly(A) tails collaborate to facilitate transport into the cytoplasm. We now understand that proteins in the nucleus participate in the export process: a *nuclear transport*

receptor binds along the mature (or maturing) mRNA; a *poly(A)-binding protein* binds along the poly(A) tail of the message; and another protein binds at or near the methyl guanosine CAP itself. These interactions enable transport of the mRNA through nuclear pores. After the mRNA is in the cytoplasm, the nuclear transport receptor recycles back into the nucleus while a *translation initiation factor* replaces the protein bound to the CAP. Figure 10.26 summarizes the nuclear transport process for mRNAs.

Specialized Proteins Mark Mature mRNAs for Nuclear Export Via Nuclear Pores

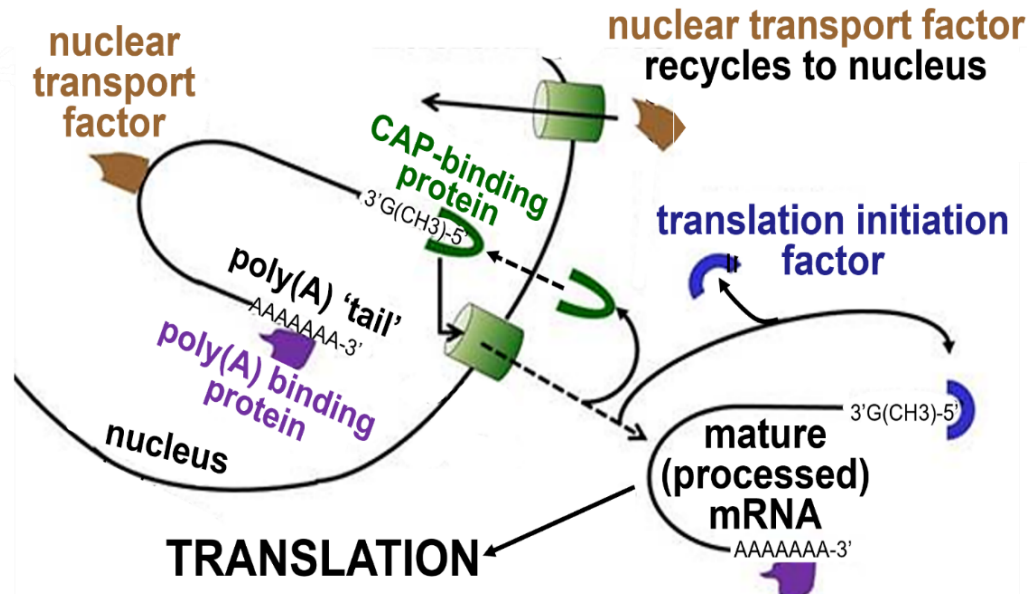


Fig. 10.26: The 5' guanosine CAP and 3' Poly(A) tail on mRNAs function during mRNA export from nucleus to cytoplasm. A nuclear transport factor along with specific 5' guanosine CAP and 3' poly(A) binding proteins associate with the mRNA during export. After export, the CAP binding protein and nuclear transport protein are recycled to the nucleus.

See a more detailed description of mRNA transport from the nucleus at ^{10.11}[mRNA Export from the Nucleus](#). The mature mRNA, now in the cytoplasm, is ready for translation, the process of protein synthesis mediated by ribosomes and a host of translation factors. The genetic code directs polypeptide synthesis during translation. We will discuss the details of translation shortly.

Some iText & VOP Key Words and Terms

45S pre-rRNA	introns	rRNA
45S pre-rRNA methylation	lariat	rRNA cleavage
5S rRNA	mature RNA transcript	rRNA endonucleases
5.8S rRNA	mRNA	σ -factor
16S rRNA	mRNA capping	snRNP
18S rRNA	mRNA polyadenylation	spacer RNA
23S rRNA	mRNA splicing	splice sites
28S rRNA	operons	spliceosome
45S rRNA	poly(A) polymerase	Svedberg unit
5' methyl guanosine capping	poly(A) tail	TATA-binding protein
adenine	polycistronic RNA	TBP
branch sites	Pribnow box	termination

crossing over	promoter	TFIIB, TFIIE, TFIIF, & TFIIH
cytosine	recombination	transcription
DNA-binding proteins	regulatory DNA sequence	transcription start site
<i>E. coli</i> RNA polymerase	regulatory factor	transcription unit
elongation	rho termination factor	translation
eukaryotic RNA polymerases	rho-independent termination	transposition
exon shuffling	ribonucleoproteins	transposons
exons	RNA polymerase I	tRNA
guanine	RNA polymerase II	tRNA processing
helix-turn-helix motif	RNA polymerase III	upstream vs downstream
initiation	RNA processing	uracil
internal promoters	RNA secondary structure	

CHAPTER 10 WEB LINKS



Chapter 11

The Genetic Code and Translation

The genetic code, tRNA (adapter) molecules, translation (protein synthesis)

Reminder: For inactive links, google key words/terms for alternative resources.

Captain Midnight, assisted by radio listeners with a Code-O-Graph (left), saved us from spies and sabotage in the run-up to WW II.



Then, the need for encryption & decryption got real, and...



...the allies got hold of the first of these German **Enigma Machines** that really helped us win the war.

11.1 Introduction

While evidence was accumulating that DNA was “the stuff of genes,” many still believed that proteins were genes—at least until the experiments of Hershey and Chase. It was in this somewhat ambivalent context that investigators wondered about how genes (whether proteins or nucleic acids) might encode metabolically functional proteins such as enzymes—which, of course, are also proteins. In 1944, the physicist-philosopher Erwin Schrödinger published *What Is Life? The Physical Aspect of the Living Cell*. Targeting the lay reader, Schrödinger took a stab at explaining life, heredity, and development in terms of physical laws. Here in his own words is what he had to say about the chemistry of heredity, which when deciphered would dictate *patterns* of development (abridging and highlighting courtesy of your author!):

THE HEREDITARY CODE-SCRIPT

Now, this pattern is known to be determined by the structure of that one cell, the fertilized egg. Moreover, we know that it is essentially determined by the structure of only a small part of that cell, its large nucleus. This **nucleus**...usually appears as a network of **chromatine**, distributed over the cell. But in the processes of cell division (mitosis and meiosis) it is seen to consist of a **set of particles**, usually fibre-shaped or rod-like, called the **chromosomes**, which number in man, 48 (2×24).... I ought to have spoken of two sets, in order to use the expression in the customary strict meaning of the biologist. For the two sets are almost entirely alike. As we shall see in a moment, one set comes from the mother (egg cell), one from the father (fertilizing spermatozoon). **It is these chromosomes**, or probably only an axial skeleton fibre of what we actually see under the microscope as the chromosome, **that contain in some kind of code-script the entire pattern of the individual's future development and of its functioning in the mature state**. Every complete set of chromosomes contains the full code; so there are, as a rule, two copies of the latter in the fertilized egg cell, which forms the earliest stage of the future individual. In calling the structure of the chromosome fibres a code-script we mean that the all-penetrating mind, once conceived by Laplace, to which every causal connection lay immediately open, could tell from their structure whether the egg would develop, under

suitable conditions, into a black cock or into a speckled hen, into a fly or a maize plant, a rhododendron, a beetle, a mouse or a woman. To which we may add, that the appearances of the egg cells are very often remarkably similar; and even when they are not, as in the case of the comparatively gigantic eggs of birds and reptiles, the difference has not been so much the relevant structures as in the nutritive material which in these cases is added for obvious reasons. **But the term code-script is, of course, too narrow. The chromosome structures...**are at the same time instrumental in bringing about the development they foreshadow. They **are law-code and executive power—or, to use another simile, they are architect's plan and builder's craft—in one.** (From Schrödinger E: *What is Life*. Cambridge, Cambridge University Press; 1944.)

Once again, keep in mind that these speculations presage the identification of DNA as the genetic chemical. The terms *code*, *broken*, and *deciphered* themselves came from the recent World War II history. Winning the war relied on *breaking* enemy codes (see the ***Enigma Machine*** at the top of this chapter), as well as on hiding strategic battle information from the enemy (recall or look up the history of the Navajo *Code Talkers*). In this chapter we will look at the experiments that broke the ***genetic code*** in messenger RNA, and as a result in the genes..., in DNA). We'll describe the elegant experiments that deciphered at first, only one of the amino-acid *codewords*, then a few more of these three-base ***codons***, and finally, all sixty-four possible codons that could be 'written' using four different nucleotides. Of these, it turns out that sixty-one codons code for amino acids and three are ***stop codons***. Experiments like those that broke the genetic code also led to our understanding of the mechanism of protein synthesis. Earlier studies had indicated that genes and proteins are colinear—that is, that the length of a gene was directly proportional to the polypeptide it encoded. If so, it follows that the lengths of mRNAs are also colinear with (i.e., proportional to the lengths of) their translation products. Finally, ***colinearity*** also suggested the obvious hypothesis that translation proceeded in three steps (***initiation***, ***elongation***, and ***termination***), just like transcription. We now know that initiation is also a multi-step process for assembling an ***initiation complex*** at a ***start codon*** near the 5' end of the mRNA. The machine itself is made up of ribosomes, mRNA, ***initiation factors*** and requires a source of chemical energy for assembly. Since mature mRNAs are longer than needed to specify a polypeptide (even after splicing!), one function of *initiation factors* is to correctly position the ribosome and associated proteins near a *start codon*. The start codon specifies the first amino acid in a new polypeptide. Once the *initiation complex* forms, elongation begins. Cycles of condensation reactions on the ribosome connect amino acids by peptide linkages, growing the chain from its amino end to its carboxyl end. Translation finishes when the ribosome moving along the mRNA encounters a stop codon. We will look at how we came to understand these discrete steps of translation.

Learning Objectives

When you have mastered the information in this chapter, you should be able to do the following:

1. Compare and contrast the *mechanisms* and *energetics* of initiation, elongation, and termination in translation vs transcription.
2. Speculate on why *the genetic code is universal* (or nearly so).
3. Justify early thinking about a *four-base genetic code*.
4. Justify early thinking about an *overlapping genetic code* (for example, one in which the last base of a codon could be the first base of the next codon in an mRNA).

- ## 11.2 Overview of the Genetic Code

11.2.1 The Universal, Degenerate Genetic Code

Standard Genetic Code – RNA Codon Table

		nonpolar	polar	basic	acidic	(stop codon)			
1st base	2nd base								3rd base
	U		C		A		G		
U	UUU	(Phe/F)	UCU	(Ser/S) Serine	UAU	(Tyr/Y) Tyrosine	UGU	(Cys/C) Cysteine	U
	UUC	Phenylalanine	UCC		UAC		UGC		C
	UUA		UCA		UAA	Stop (Ochre)	UGA	Stop (Opal)	A
	UUG		UCG		UAG	Stop (Amber)	UGG	(Trp/W) Tryptophan	G
C	CUU	(Leu/L) Leucine	CCU	(Pro/P) Proline	CAU	(His/H) Histidine	CGU	(Arg/R) Arginine	U
	CUC		CCC		CAC	CGC	C		
	CUA		CCA		CAA	(Gln/Q) Glutamine	CGA		A
	CUG		CCG		CAG	CGG	G		
A	AUU	(Ile/I) Isoleucine	ACU	(Thr/T) Threonine	AAU	(Asn/N) Asparagine	AGU	(Ser/S) Serine	U
	AUC		ACC		AAC	AGC	C		
	AUA		ACA		AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine	A
	AUG ^[A]	(Met/M) Methionine	ACG		AAG	AGG		G	
G	GUU	(Val/V) Valine	GCU	(Ala/A) Alanine	GAU	(Asp/D) Aspartic acid	GGU	(Gly/G) Glycine	U
	GUC		GCC		GAC	GGC	C		
	GUA		GCA		GAA	(Glu/E) Glutamic acid	GGA		A
	GUG		GCG		GAG	GGG	G		

227

A look at this table reveals that there is a single codon for each of two amino acids, methionine and tryptophan, but two or more for each of the other eighteen amino acids (and in one case six!). For this reason, we say that the genetic code is **degenerate**. The table highlights the **AUG** codon. At some point, evolution dictated that methionine will be the first amino acid of polypeptides. It is in fact the **start codon** for all polypeptides *as well as* for placing methionine in a polypeptide where it is needed for its function. And even though all polypeptides start with an N-terminal methionine, most lose it after translation. Finally, while the last amino acid in a polypeptide can be any amino acid consistent with the function of the protein being made, the table also highlights three **stop codons** in the *standard genetic code* that “tell” ribosomes the location of the last amino acid of a polypeptide. We will see in considerable detail that ribosomes are the living, biological equivalent an Enigma machine, translating messages encrypted in DNA and RNA into polypeptides. And the key to this remarkable mRNA translation machine is the tRNA *decoding device*. To function as the key, each amino acid attaches to a specific tRNA whose short sequence contains a three-base *anticodon* that is complementary to an mRNA codon. Enzymatic reactions catalyze the *dehydration synthesis* reactions that link amino acids by *peptide bonds* in the order specified by mRNA codons.



[201 The Genetic Code Dictionary](#)



11.2.2 Comments on the Nature and Evolution of Genetic Information

The nearly universal genetic code found in viruses and all organism, from bacteria to humans) implies that the code was “fixed” early in evolution. It is probable that portions of the code were in place even before life began. Once in place however, the genetic code was highly constrained against evolutionary change. The degeneracy of the genetic code enabled (and contributed to) this constraint by permitting base changes in most codons without an effect on the amino acids that they encoded. We can compare our gene and other DNA sequences to those of different organisms precisely because the genetic code is universal and resistant to change. This is what allows us to compare genomes and establish evolutionary relationships between larger groups of organisms (genus, family, order), between species, and even between individuals of a species. In addition to constraints imposed by a universal genetic code, the genomes of some organisms show *codon bias*, a preference to use some codons over the use of others in their genes. For example, codon bias is seen in organisms that favor codons rich in A and T or in organisms that prefer codons richer in G and C. Interestingly, codon bias in genes often accompanies a corresponding genome-wide nucleotide bias. Thus, an organism with an A-T codon bias may also have an A-T biased genome. Likewise, a G-C rich genome can exhibit a bias toward the use of G and C in their codons. If you check back, you can recognize genomic nucleotide bias in Chargaff's base ratios!

Finally, we often think of genetic information in a genome as genes that specify protein sequences, in other words, *coding DNA*. On the other hand, obvious examples of *noncoding* but critically informational DNA include the genes for ribosomal and transfer RNAs. The relative amounts of *informational* DNA (genes for polypeptides, rRNAs, and tRNAs) and *noninformational* DNA range across species, although they are higher in prokaryotes than

eukaryotes. In fact, ~88% of the *E. coli* genome encodes polypeptides, compared to the humans, in which that figure is less than ~1.5%!

Some of *less* obvious informative DNA in higher organisms is transcribed (e.g., introns). But other *informative* DNA is never transcribed. They may not be called genes, the latter include regulatory DNA sequences, sequences that support chromosome structure and other DNA that contribute to development and phenotype.

As for truly noninformative (useless) DNA in a eukaryotic genome, that amount is steadily shrinking as genome sequencing reveals novel DNA sequences, new potential genes, and thus, new RNAs with new functions (topics covered elsewhere in this text).

11.3 Gene and Protein Colinearity and Triplet Codons

Serious efforts to understand how proteins are encoded began after Watson and Crick used the experimental evidence of Maurice Wilkins and Rosalind Franklin (among others) to determine the structure of DNA. Most hypotheses about the genetic code assumed colinearity, namely that DNA (i.e., genes) and polypeptides were *colinear*.

11.3.1 Colinearity of Genes and Proteins

For genes and proteins, colinearity just means that the length of a DNA sequence in a gene is proportional to the length of the polypeptide encoded by the gene. The gene-mapping experiments in *E. coli* (which we have already discussed) certainly supported this hypothesis. Figure 11.2 illustrates the concept of the colinearity of genes and proteins.

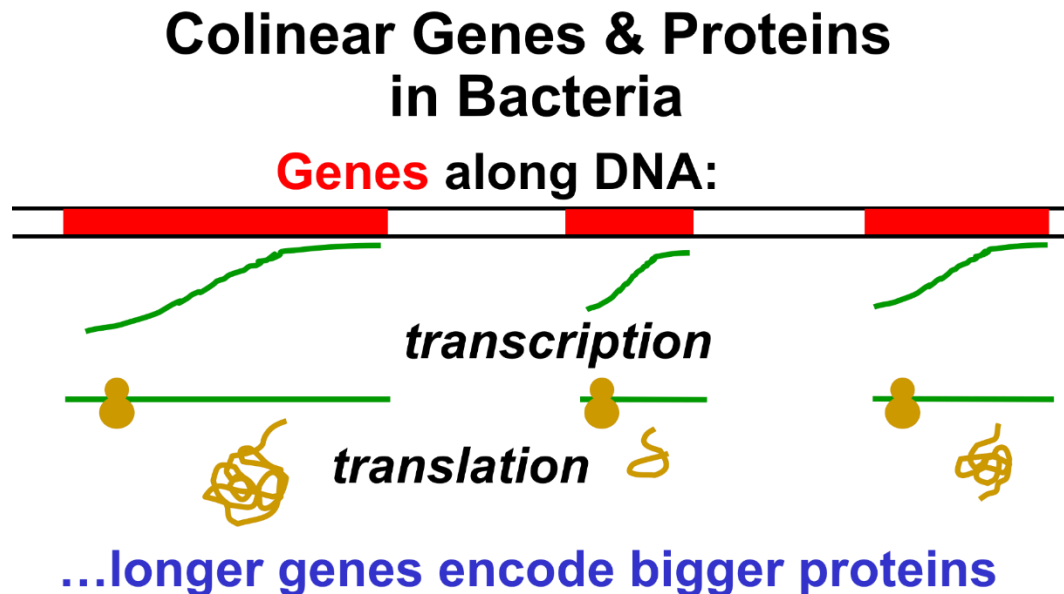


Fig. 11.2: Colinearity of genes and proteins (polypeptides) in bacteria.

If the genetic code is colinear with the polypeptides it encodes, then a one-base *codon* obviously would not work because such a code would only account for four amino acids. A two-

base genetic code also doesn't work because it could only account for sixteen (4^2) of the twenty amino acids found in proteins. However, three-nucleotide codons could code for a maximum of sixty-four (4^3) amino acids, more than enough to encode the twenty amino acids. And of course, a four-base code also works; it satisfies the expectation that genes and proteins are colinear, with the "advantage" that there would be 256 (4^4) possible codons to choose from.

11.3.2 How Is a Linear Genetic Code "Read" to Account for All genes in the genome of an Organism?

George Gamow (a Soviet-American physicist working at the George Washington University) was the first to propose **triplet codons** to encode the twenty amino acids. This was the simplest hypothesis to account for the colinearity of genes and proteins, and for encoding twenty amino acids. Once the concept of colinearity was accepted, a remaining concern was: Is there enough DNA in an organism's genome to fit all the codons needed to make all of its proteins? Assuming that genomes did not have a lot of extra DNA lying around, scientists still wondered how genetic information might be compressed into short DNA sequences consistent with colinearity and assumptions about the number of genes required by an organism. One idea assumed that there were twenty *meaningful* three-base codons (one for each amino acid) and forty-four *meaningless* codons, and that the *meaningful* codons in a gene (i.e., an mRNA) would be read and translated in an overlapping manner. A code where codons overlap by one base is shown in Figure 11.3 (below). You can figure out how compressed a gene could get with codons that overlapped by two bases.

An Overlapping Genetic Code

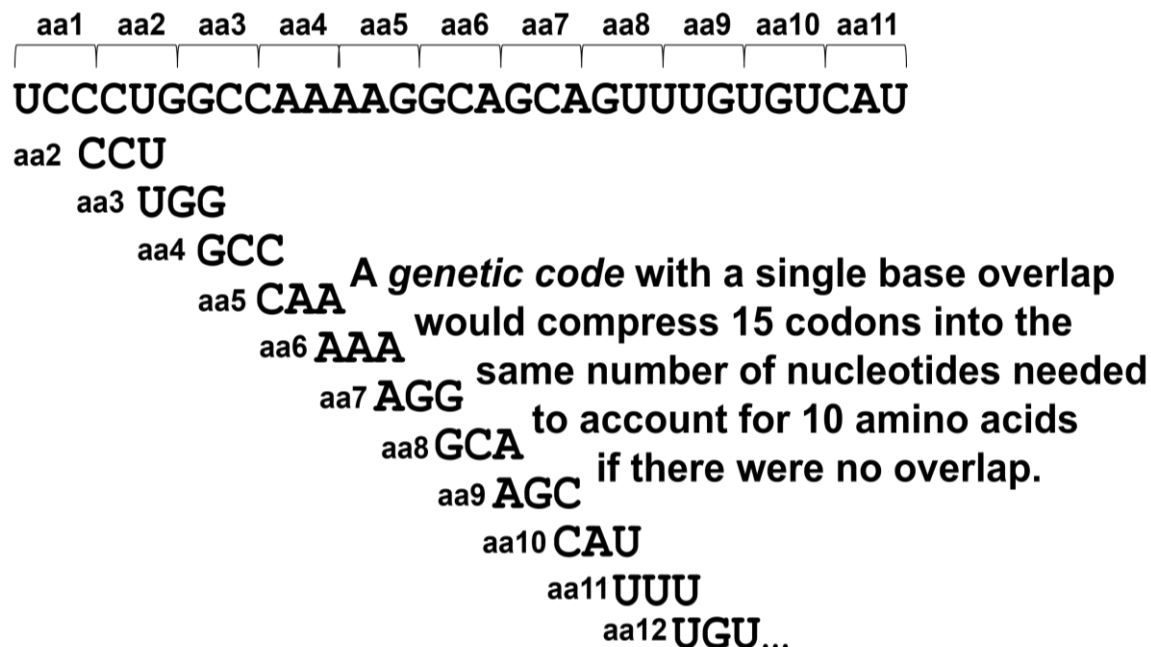


Fig. 11.3: A single base overlapping genetic code would fit more genetic information in less DNA!

As attractive as an overlapping codon hypothesis was in achieving genomic economies, it sank of its own weight almost as soon as it was floated! If you look carefully at

the example above, you can see that each succeeding amino acid would have to start with a specific base. A look back at the table of sixty-four triplet codons quickly shows that only one of sixteen amino acids, those that begin with a **C**, can follow the first one in the illustration. Based on the amino-acid sequences already accumulating in the literature, it was clear that virtually any amino acid could follow another in a polypeptide. Therefore, overlapping genetic codes are untenable. The genetic code must be nonoverlapping!

Sydney Brenner and Francis Crick performed elegant experiments directly demonstrating the nonoverlapping genetic code. They showed that bacteria with a single base deletion (and, likewise, a double base deletion) in the coding region of a gene failed to make the expected protein. On the other hand, a bacterium containing a mutant version of a gene in which three bases were deleted *was* able to make the protein. But the protein it made was less active than the protein made by bacteria with genes that had no deletions.

The next issue was whether there were only twenty *meaningful* codons and forty-four *meaningless* ones. If only twenty codons encoded amino acids, how would the translation machinery know the correct twenty to translate? What would prevent the translational machinery from "reading the wrong" triplets (i.e., reading an mRNA *out of phase*)?

For example, if the translation machine began reading an mRNA from the second or third base of a codon, wouldn't it likely encounter a meaningless three-base sequence in short order? An alternate hypothesis speculated that the code was *punctuated*. That is, perhaps there were the chemical equivalents of commas between the meaningful triplets. The commas would of course, be additional nucleotides. In this punctuated code, the translation machinery would recognize the "commas" and would not translate *any* meaningless three-base triplets, avoiding out-of-phase translation attempts. Of course, a code with *nucleotide commas* would increase the amount of DNA needed to specify a polypeptide by a third! Finally, Crick proposed the *Comma-less Genetic Code*. He divided the sixty-four triplets into twenty *meaningful* codons (encoding the amino acids) and forty-four *meaningless* ones that did not encode amino acids. His code was such that when the twenty *meaningful* codons are placed in any order, any of the triplets read in overlap would be among the forty-four *meaningless* codons. In fact, he could arrange several different sets of twenty and forty-four triplet codons with this property! Crick had cleverly shown how to read the triplets in correct sequence without nucleotide commas.



[202 Speculations about a Triplet Code](#)



Of course, we know now that while the genetic code is indeed comma-less, *it is not comma-less* in the sense that Crick had envisioned. What's more, thanks to experiments to be described next, we know that ribosomes read the correct codons in the right order because they know exactly where to start reading the mRNA!

11.4 Breaking the Genetic Code

Marshall W. Nirenberg and J. Heinrich Matthaei decoded the first triplet. They fractionated *E. coli* as shown below in Figure 11.4, and then identified which fractions were required for *cell-free protein synthesis*, i.e., **in vitro translation**.

Nirenberg et al. break first few codewords by finding the message!

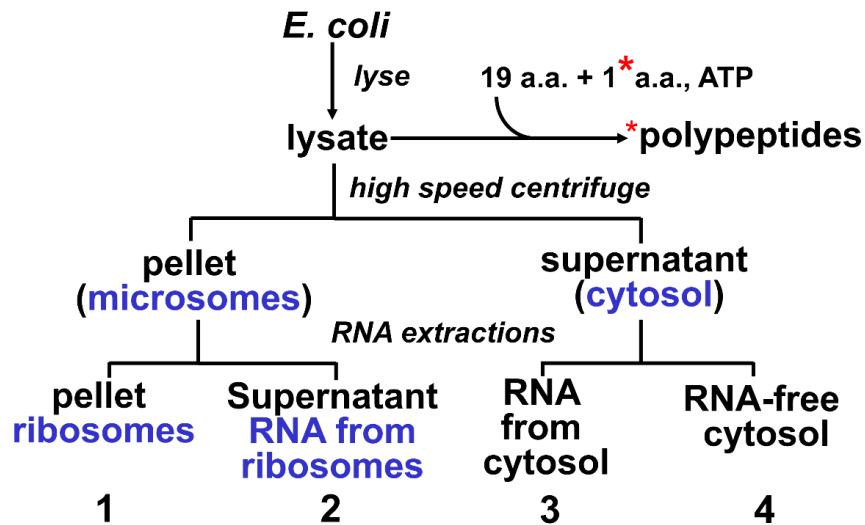


Fig. 11.4: Nirenberg and Matthaei fractionated bacterial cells and recombined different combinations of fractions (e.g., 1+2, 1+3, 1+4, 2+3 etc.) proving the requirement for a “messenger RNA” fraction (#2) in *reconstituted* cell components to synthesize proteins.

Check out the original work in the classic paper by Nirenberg MW and Matthaei JH [(1961) *The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribo-nucleotides. Proc. Natl. Acad. Sci. USA 47:1588-1602*]. Various combinations of the isolated cell fractions were added back together, along with amino acids (one of which was radioactive) and ATP (as an energy source). After the mixture underwent a short incubation, Nirenberg and Matthaei looked for the presence of high molecular weight radioactive proteins as evidence of *in vitro translation*. They found that all four final fractions (1–4 in Figure 11.4) must be added together to make radioactive proteins in the test tube. One of the essential cell fractions consisted of RNA that had been gently extracted from ribosomes (fraction 2 in the illustration). Reasoning that this RNA might be mRNA, they substituted a synthetic poly(U) preparation for this fraction in their cell-free protein-synthesizing mix, expecting poly(U) to encode a simple repeating amino acid. Nirenberg and Matthaei set up twenty reaction tubes, with a different amino acid in each... and made only polyphenylalanine! (Figure 11.5).

Nirenberg substitutes poly (U) for fraction 2

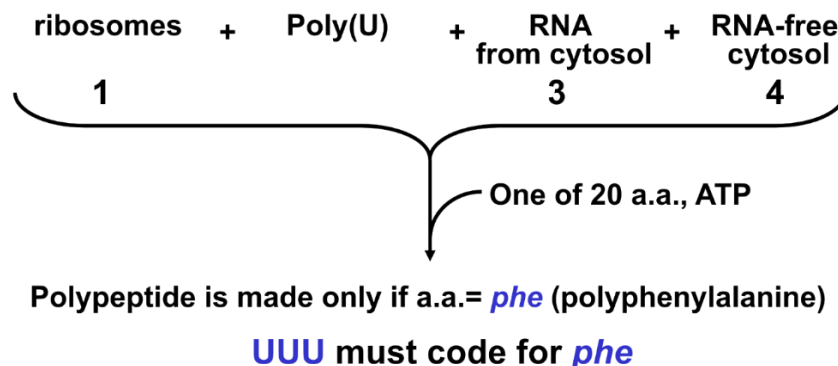


Fig. 11.5: Nirenberg and Matthaei’s reconstitution of fractionated bacterial cells decodes the first codon.

Thus, the triplet codon UUU means *phenylalanine*. When other polynucleotides were synthesized by H. G. Khorana, poly(A) and poly(C) were shown in quick succession to make polylysine and polyproline using this experimental protocol. Thus, AAA and CCC must encode *lysine* and *proline*, respectively. With much ingenuity and despite a bit more difficulty, additional codons were deciphered using poly di- and trinucleotides in cell-free systems.



203 Deciphering the First Codon

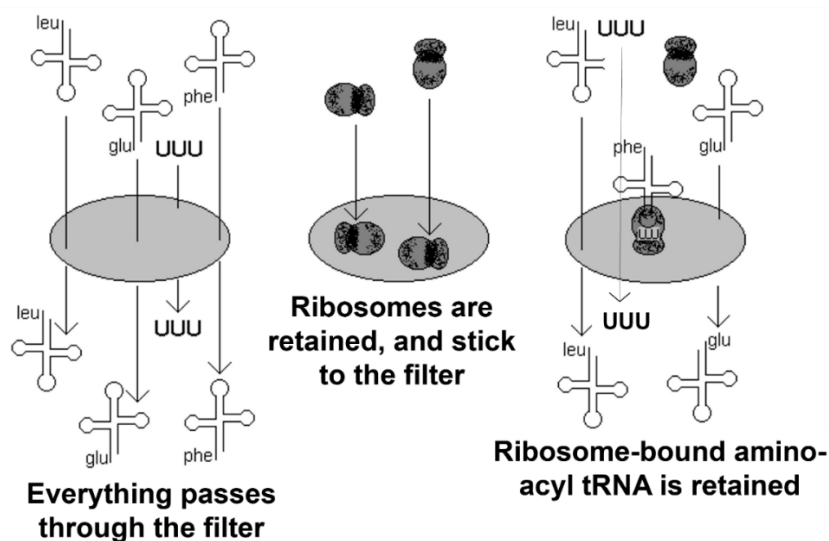
Nirenberg and Khorana shared the 1968 Nobel Prize in Physiology or Medicine for with R. W. Holley their contributions to our understanding of protein synthesis (see Holley's contribution below).

The process of deciphering the rest of the genetic code relied on Crick's realization that, chemically, amino acids have no attraction for either DNA or RNA (or triplets thereof). Instead, he predicted the existence of an **adaptor molecule** that would contain nucleic acid and amino acid information *on the same molecule*. Today we recognize this molecule as **tRNA**, the genetic *decoding device*.

Nirenberg and Philip Leder designed the experiment that pretty much *broke* the rest of the genetic code. They added individual amino acids to separate test tubes containing tRNAs, in effect causing the synthesis of specific aminoacyl-tRNAs. They then mixed their amino acid-bound tRNAs with isolated ribosomes and synthetic triplet codons. They had previously shown that synthetic three-nucleotide fragments would bind to ribosomes, so they hypothesized that triplet-bound ribosomes would in turn, bind appropriate amino acid-bound tRNAs. Figure 11.6 summarizes their experiment.

Breaking the Rest of the Code

Aminoacyl-tRNAs help crack the genetic code:



These experiments matched each codon to an amino acid!

Fig. 11.6: Nirenberg and Leder's experiment led to breaking the entire genetic code.

In this experiment, various combinations of tRNA, ribosomes, and aminoacyl-tRNAs were placed over a filter. Nirenberg and Leder knew that aminoacyl tRNAs alone passed through the filter and that ribosomes did not. They predicted that triplets would associate with the ribosomes, that this complex would bind the tRNA with the amino acid encoded by the bound triplet, and that the filters would retain this three-part complex, allowing identification of the amino acid retained on the filter. They would then be able to match each amino acid to a triplet codon(s). Ultimately, sixty-one of the codons were matched to specific amino acids: Since 61 codons specify 20 amino acids, the genetic code was *degenerate*!



204 Deciphering All 64 Triplet Codons

After the code was largely deciphered, Robert Holley sequenced and characterized a yeast tRNA—which is what earned him his share of the Nobel Prize (above). This first successful sequencing of a nucleic acid was possible because the RNA was short and contained several modified bases that facilitated the sequencing chemistry. Holley found the amino acid alanine at one end of his tRNA, and he found one of the anticodons for an alanine codon roughly in the middle of the tRNA sequence. From regions of internal sequence complementarity, Holley predicted that this (and other) tRNAs would fold and assume a *stem-loop* (or *cloverleaf*) structure with a central **anticodon loop**. Figure 11.7 shows the stem-loop structure of a phenylalanine tRNA and a subsequent computer-generated the molecular structure with a now-familiar “L”-shape). Note the **amino acid attachment site** at the 3’ end at the top of the molecule and the **anticodon loop** at the bottom.

The tRNA Decoding Device: Structure of a Phenylalanine tRNA

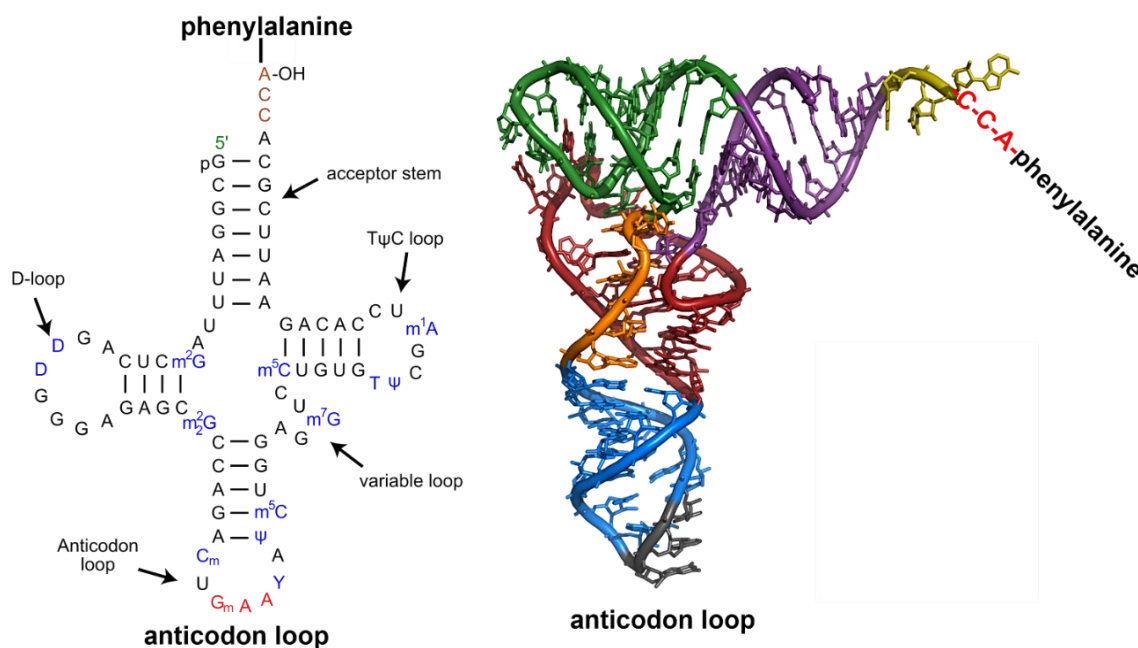


Fig. 11.7: Cloverleaf (left) and computer-generated 3D (right) structures of a phenylalanyl tRNA.



205-2 tRNA Structure and Base Modifications

After a brief overview of translation, we'll break this process down into its three steps and see how aminoacyl-tRNAs function in the initiation and elongation steps of translation, and we'll look at the special role of an *initiator* tRNA.

11.5 Translation

Like any polymerization in a cell, translation occurs in three steps: **initiation** brings a ribosome, mRNA, and an *initiator* tRNA together to form an initiation complex. **Elongation** is the successive addition of amino acids to a growing polypeptide. **Termination** is signaled by sequences (one of the stop codons) in the mRNA and by protein **termination factors** that interrupt elongation and release a finished polypeptide.

The events of translation occur at A, P, and E sites on the ribosome (Figure 11.8).

A, P and E 'sites': Based Largely on rRNA Conformation

A = amino acid entry (binding) site

P = 'peptidyl-tRNA binding site

E = exit site

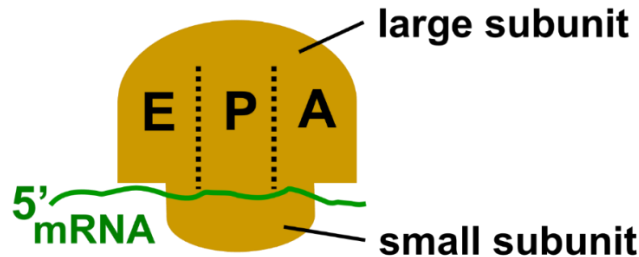


Fig. 11.8: A, P, and E sites, mostly on the large ribosomal subunit, are involved in mRNA translation.

11.5.1 Even Before Initiation—Making Aminoacyl-tRNAs

Translation is perhaps the most energy-intensive job a cell must do, starting with attaching amino acids to their tRNAs. The amino-acylation reaction is the same for all amino acids. A specific **aminoacyl-tRNA synthase** attaches a tRNA to (i.e., **charges**) an appropriate amino acid. This process of charging tRNAs has its own three steps and requires ATP (Figure 11.9).

Amino Acid Activation in Three Steps

1. Amino acid binds to a specific **aminoacyl-tRNA synthase** along with **ATP**. After ATP hydrolysis, AMP is linked to the amino acid:



2. The amino acid is transferred to the enzyme:



3. A tRNA with the correct anticodon binds the enzyme, the amino acid is transferred to the tRNA, and the enzyme is released:

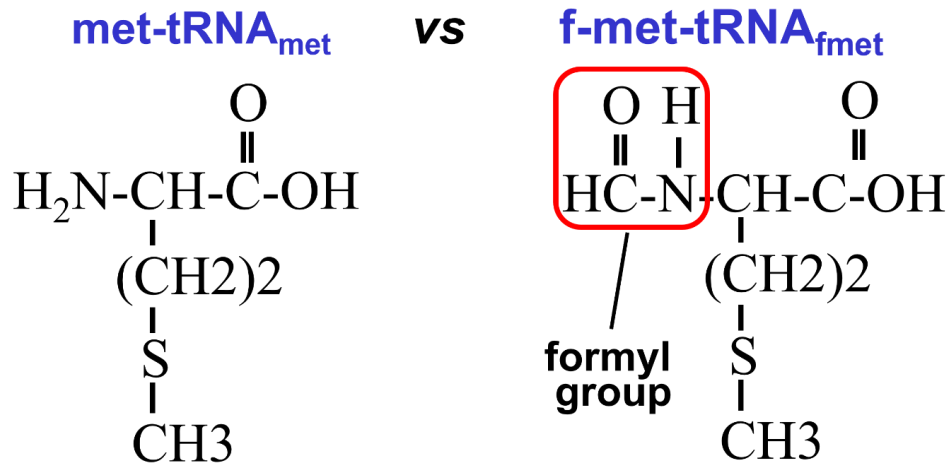


Fig. 11.9: The three steps of amino acid *activation* (tRNA acylation) require ATP hydrolysis.

In the first step, ATP and an appropriate amino acid bind to the aminoacyl-tRNA synthase. The ATP is hydrolyzed, releasing a pyrophosphate (PPi), and leaving an enzyme-AMP-amino acid complex, which transfers the amino acid to the enzyme, releasing the AMP.

Finally, the tRNA binds to the enzyme; the amino acid is transferred to the tRNA; and the intact enzyme is regenerated and released. The charged tRNA is now ready for use in translation. Several studies had already established that polypeptides are synthesized from their amino (N-) terminal end to their carboxyl (C-) terminal end. When it became possible to determine the amino-acid sequences of polypeptides, it turned out that around 40% of *E. coli* proteins had an N-terminal methionine. This suggested that *all* proteins begin with a methionine, but that the methionine was subsequently removed in a **posttranslational processing** step. It also turned out that, even though there is only one methionine codon, two different tRNAs for methionine could be isolated. One of the tRNAs was bound to a methionine modified by *formylation*, called **formyl-methionine-tRNA_{fmet}** (*fmet-tRNA_f* for short). The other was **methionine-tRNA_{met}** (*met-tRNA_{met}* for short), whose methionine was not formylated. *tRNA_{met}* and *tRNA_f* each have an anticodon to AUG, the only codon for methionine. But they have different base sequences encoded by different tRNA genes. *tRNA_{met}* is used to insert an encoded methionine anywhere in a polypeptide. *tRNA_f*, the *initiator* tRNA, is only used to start new polypeptides with formyl-methionine. In prokaryotes, the amino group of the methionine on *met-tRNA_f* is *formylated* by a **formylating enzyme** to make the *fmet-tRNA_f*. This enzyme does not recognize the methionine on *met-tRNA_{met}*. Methionine and formylated methionine structures are compared in Figure 11.10.

One codon for methionine, but Marcker & Sanger find 2 tRNAs, attached to either:



**Both tRNAs have *different* base sequences...,
but the *same* anticodon!**

Fig. 11.10: Discovery of *met-tRNA* and *formyl-met-tRNA_f*, despite only one codon for methionine.

In *E. coli*, a **formylase** enzyme removes the formyl group from all N-terminal formyl methionines at some point after translation has begun. As noted, methionines themselves (and sometimes more N-terminal amino acids) are also removed from ~60% of *E. coli* polypeptides. While eukaryotes inherited both a *tRNA_{met}*, and a *tRNA_f* (using only *met-tRNA_f* during initiation), methionine on the latter is never formylated in the first place! What's more, methionine is absent from the N-terminus of nearly all mature eukaryotic polypeptides.

Apparently, early in evolution, the need for an initiator tRNA must have ensured a correct starting point for translation on an mRNA and therefore the growth of a polypeptide from one end to the other (i.e., from its N- to C-termini).

At that time, formylation of the N-terminal methionine may have served to block accidental addition of amino acids to the N-terminus of a polypeptide. Today, formylation is a kind of *molecular appendix* in bacteria. In eukaryotes at least, evolution has selected other features to replace actual formylation as a protector of the N-terminus of polypeptides.

11.5.2 Translation *Initiation*

Now that we have charged the tRNAs, we can look more closely at the three steps of translation. Understanding translation initiation began with cell fractionation of *E. coli*, the purification of molecular components required for cell-free (in vitro) protein synthesis, and finally, *reconstitution* experiments. Cellular RNA was purified and the 30S ribosomal subunit was separated from a ribosomal extract. These were then added to initiation-factor proteins purified from the bacterial cells. Reconstitution experiments revealed that when added to each other in the correct order, the separated fractions, along with the purified initiation factors, formed a stable 30S ribosomal subunit–mRNA complex. We now know that a short **Shine-Dalgarno** sequence in the **5' untranslated region (5' UTR)** of the mRNA forms H-bonds with its complementary sequence in the **16S rRNA** in 30S ribosomal subunit. The Shine-Dalgarno sequence is just upstream of the initiator AUG codon. The first of 3 *initiation* steps is a stable binding of the small (30S) ribosomal subunit to the mRNA, requiring two initiation factors **IF1** and **IF3**, which also bind to the 30S ribosomal subunit, shown in Figure 11.11.

Initiation:

- **IF3 keeps 50S subunit from binding**
- **IF1 enables 30S subunit to bind mRNA via H-bonding to 16S rRNA**

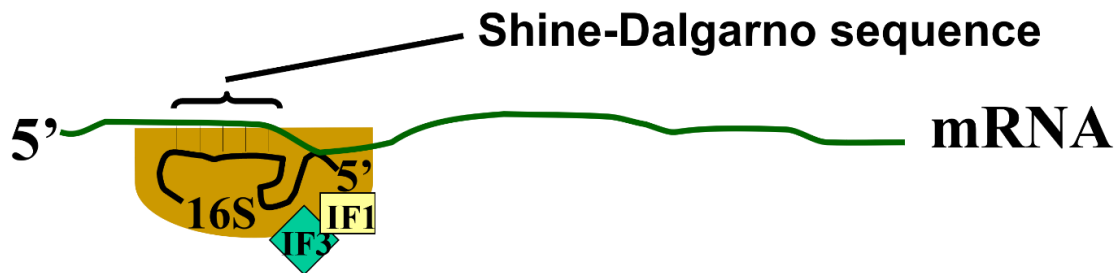


Fig. 11.11: *E. coli* translation *Initiation* was experimentally divided into 3 steps. In step 1 shown here, the small (30S) ribosomal subunit binds to mRNA. *IF1* (*initiation factor 1*) enables complementary H-bonding between a region of the 16S rRNA to the Shine-Dalgarno sequence near the 5' end of the mRNA. *IF3* (*initiation factor 3*) prevents premature attachment of the large (50S) ribosomal subunit.



[206 Translation Initiation: mRNA Associates with 30S Ribosomal Subunit](#)



Next, with the help of GTP and another initiation factor (**IF2**), the initiator *fmet-tRNA^f* recognizes and binds to the AUG start codon found in all mRNAs. Some call the resulting

structure (seen in Figure 11.12) the **Initiation Complex**, which includes the 30S ribosomal subunit, IFs 1, 2, and 3, and the *fmet-tRNA_{fmet}*.

Initiation: cytosol fractionation & reconstitution reveal:

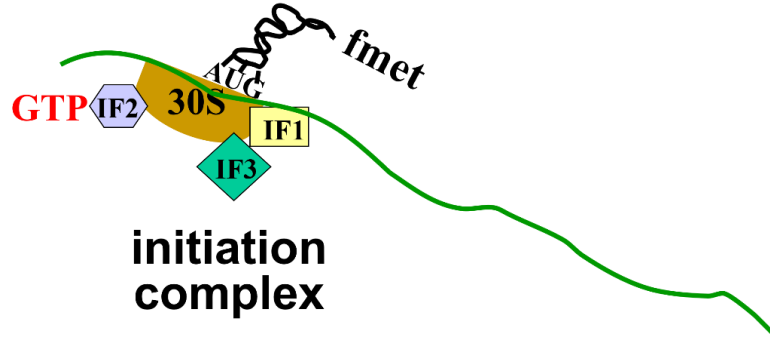


Fig. 11.12: In the second bacterial translation initiation step, a *GTP-bound IF2* enables *fmet-tRNA_{fmet}* (the first aminoacyl tRNA) to bind to the mRNA-bound 30S ribosomal subunit, making an *initiation complex*.



[207 Initiation Complex Formation](#)

In the last initiation step, the large ribosomal subunit binds to this complex; IFs 1, 2, and 3 disassociate from the ribosome; and the initiator *fmet-tRNA_{fmet}* is now in the *P* site. Some prefer to call the structure formed at this point the initiation complex (Figure 11.13).

Initiation: Completion of Ribosome Assembly

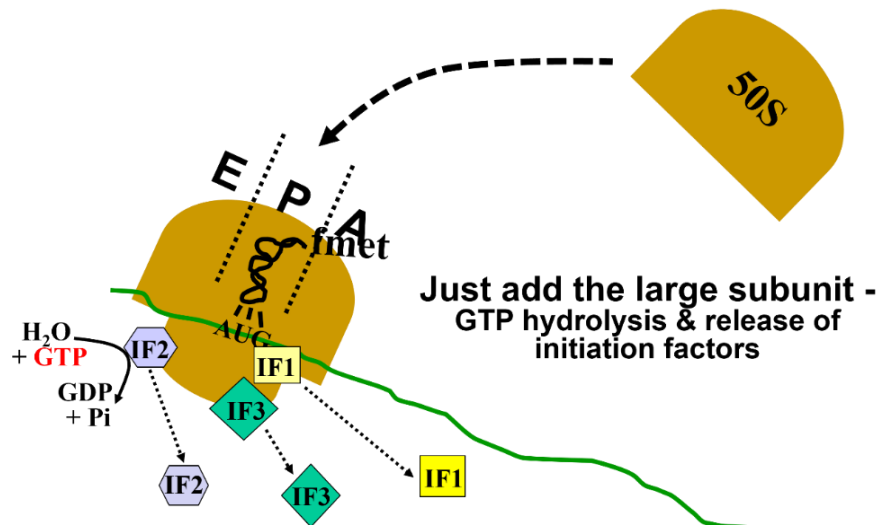


Fig. 11.13: In step three of *translation initiation*, GTP on IF2 is hydrolyzed, the three initiation factors dissociate from the 30S subunit, and the 50S subunit binds to the *initiation complex*.



[208 Adding the Large Ribosomal Subunit](#)

Initiation can happen multiple times on a single mRNA, forming the polyribosome, or polysome (already described in chapter 1). Each of the complexes formed above will engage in polypeptide elongation, described next.

11.5.3 Translation Elongation

Elongation is a sequence of protein factor-mediated condensation reactions and ribosome movements along an mRNA. As you will see, polypeptide elongation proceeds in three discrete stages and requires a considerable input of free energy.

11.5.3.a Translation Elongation 1

Elongation starts with the entry of the second aminoacyl tRNA (*aa2-tRNA_{aa2}*) into the A site of the ribosome drawn in by the codon-anticodon interaction and GTP hydrolysis. To get there, a GTP-bound *Elongation Factor Tu* (**EFTu-GTP**) binds to an incoming aminoacyl-tRNA, in this case *aa2-tRNA_{aa2}* (arrow 1). The resulting ternary complex enters the ribosome A site, activating its GTPase function (arrow 2), hydrolyzing the GTP and releasing the reaction products EFTu-GDP and Pi, from the ribosome (arrow 3). To keep elongation moving along, *elongation factor Ts* (**EFTs**) catalyzes a GTP-GDP exchange on EFTu, regenerating EFTu-GTP (arrow 4). Figure 11.14 illustrates these activities of bacterial translation elongation.

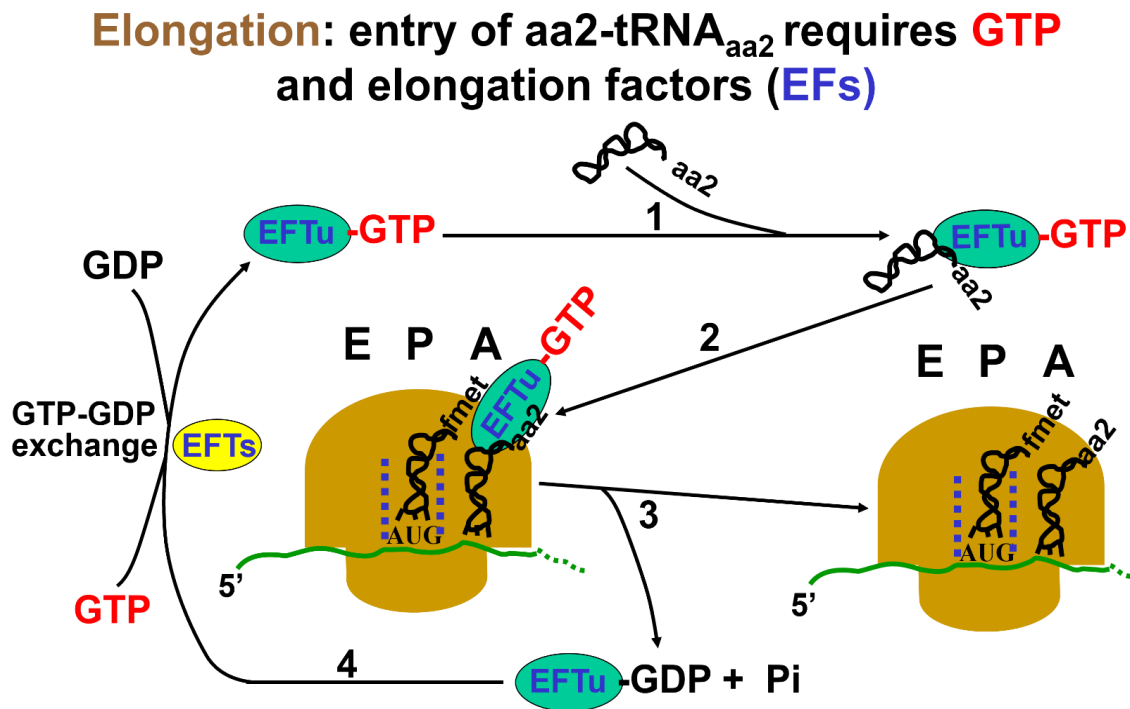


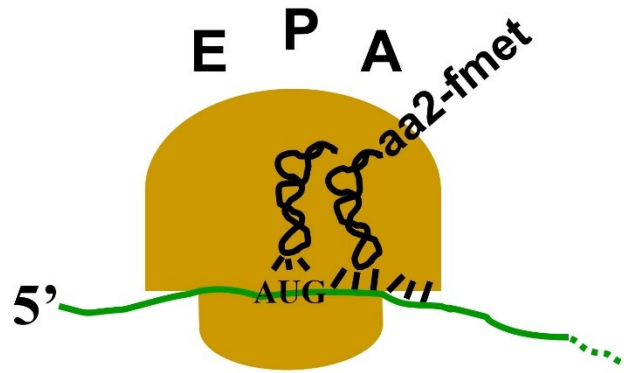
Fig. 11.14: Entry of a second aminoacyl-tRNA into the ribosomal A site starts when the aminoacyl-tRNA binds EFTu-GTP, a GTP-bound elongation factor (1). The *aa2-tRNA_{aa2}*-EFTu-GTP complex enters the ribosome A site (2). The now active GTPase function of EFTu-GTP hydrolyzes the GTP and the resulting EFTu-GDP and Pi come off the ribosome (3). Elongation factor Ts (EFTs) swaps GDP for GTP on EFTu to regenerating EFTu-GTP (4). These steps will be repeated in each elongation cycle of translation.



11.5.3.b Translation Elongation 2

Peptidyl transferase, a **ribozyme** component of the ribosome itself, links the incoming amino acid to a growing chain in a condensation reaction. Figure 11.15 shows the results.

Elongation: *peptide bond formation*



Peptidyl transferase: ribozyme activity of rRNA; spans the A & P sites; catalyzes peptide bond formation (a condensation reaction) between *fmet* and the second amino acid in the A site.

Fig. 11.15: In step 2 of translation elongation, a *peptidyl transferase ribozyme* activity of the 23S rRNA (on the 50S ribosomal subunit) catalyzes hydrolysis of *fmet* from the *fmet*-tRNA, and formation of a peptide bond (linkage) between the *fmet* and the second amino acid (*aa2*), still linked to its *tRNA_{aa2}*.

In this reaction, the *fmet* is transferred from the initiator *tRNA_f* to *aa2*-*tRNA_{aa2}* in the A site, forming a peptide linkage with *aa2*, leaving the *tRNA_f* in the P site.



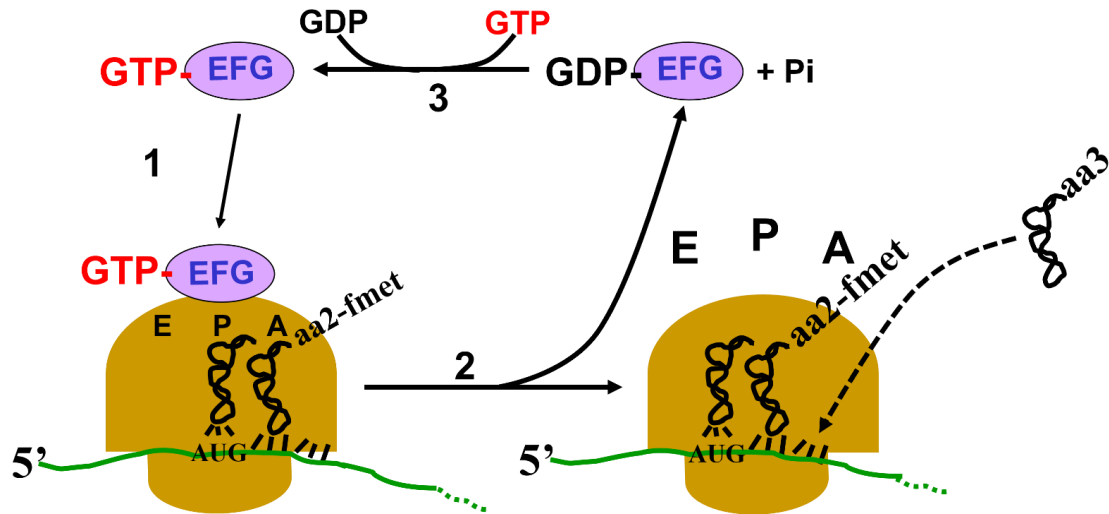
[210 Elongation: A Ribozyme Catalyzes Peptide Linkage Formation](#)

11.5.3.c Translation Elongation 3

For translation to continue, the ribosome must move along the mRNA to expose the next codon in its A site. This begins when GTP binds to *elongation factor G* (**EFG**). After formation of the dipeptidyl-tRNA on the ribosome, **GTP-EFG** binds to the 50S ribosomal subunit where it acts as a GTP-dependent **translocase**, hydrolyzing the GTP to power the movement (**translocation**) of the ribosome along an mRNA to expose a new codon on the A site and shifting the dipeptidyl-tRNA to the P site. The third tRNA (*tRNA_{aa3}*) now binds in the A site; *tRNA_{aa2}*, now attached to a dipeptide, is in the P site. The *tRNA_{fmet}* (*tRNA_f*) that started in the A site, is now in the E site of the ribosome. Note the free energy cost of elongation: at 3 NTPs per cycle, translation is the most expensive polymer synthesis reaction in cells!

The *tRNA_f*, no longer attached to an amino acid, will exit the E site as the next (third) *aa*-tRNA enters the empty A site, based on a specific codon-anticodon interaction (assisted by elongation factors and powered by GTP hydrolysis) to begin another cycle of elongation. Figure 11.16 (below) illustrates the movement of the ribosome along the mRNA.

Elongation: Translocation Along mRNA by *Elongation Factor G (EFG)*, a Transient *translocase*



Note that one full cycle of elongation has cost 3 **NTPs**, one each

- to make aa-tRNA,
- to bring each aa-tRNA to A site
- to move ribosome on mRNA!

Fig. 11.16: Elongation continues as the ribosome moves along the mRNA. Binding of *elongation factor G* (**GTP-EFG**) to the 50S ribosomal subunit activates its *translocase* that then catalyzes GTP hydrolysis to power ribosome movement along the mRNA. The first (fmet) tRNA in the E site is ready to exit the ribosome, the dipeptidyl-tRNA is in the peptidyl (P) site, and the next codon is exposed in the A site. Entry of the next aminoacyl-tRNA will begin the next round of elongation, repeating the events in Fig. 11.14.



[211 Elongation: Translocase Moves Ribosome along mRNA](#)



[212 Adding the Third Amino Acid](#)



[213 Big Translation Energy Costs](#)



[214 The Fates of fMet and Met: Cases of Posttranslational Processing](#)



As polypeptides elongate, they pass through a groove in the large ribosomal subunit. As they emerge, the previously noted *E. coli* *formylase* catalyzes removal of the formyl group from the now-exposed N-terminal initiation fmet of all growing polypeptides. While about 40% of *E. coli* polypeptides begin with a methionine, specific proteases catalyze the hydrolytic removal of the methionine, and sometimes more N-terminal amino acids) from the other 60%. All of these N-terminal modifications are examples of posttranslational processing.

11.5.4 Translation Termination

A ribosome's translation of an mRNA ends when translocation exposes one of the three stop codons in the A site of the ribosome. These *stop codons*, creatively called *ochre*, *amber*, and *opal* (UAA, UAG, and UGA, respectively) are situated some distance from the 3' end of an mRNA. The region between a stop codon and the end of the mRNA is called the **3' untranslated region** of the messenger RNA (**3' UTR**).

Since there is no aminoacyl-tRNA with an anticodon to any of the stop codons, the ribosome actually stalls, and the translation slow-down is just long enough for a protein **termination factor** to enter the A site. This interaction causes the release of the new polypeptide and the disassembly of the ribosomal subunits from the mRNA. The process requires energy from yet another GTP hydrolysis. After dissociation, ribosomal subunits can be reassembled with an mRNA for another round of protein synthesis. Figure 11.17 illustrates translation termination.

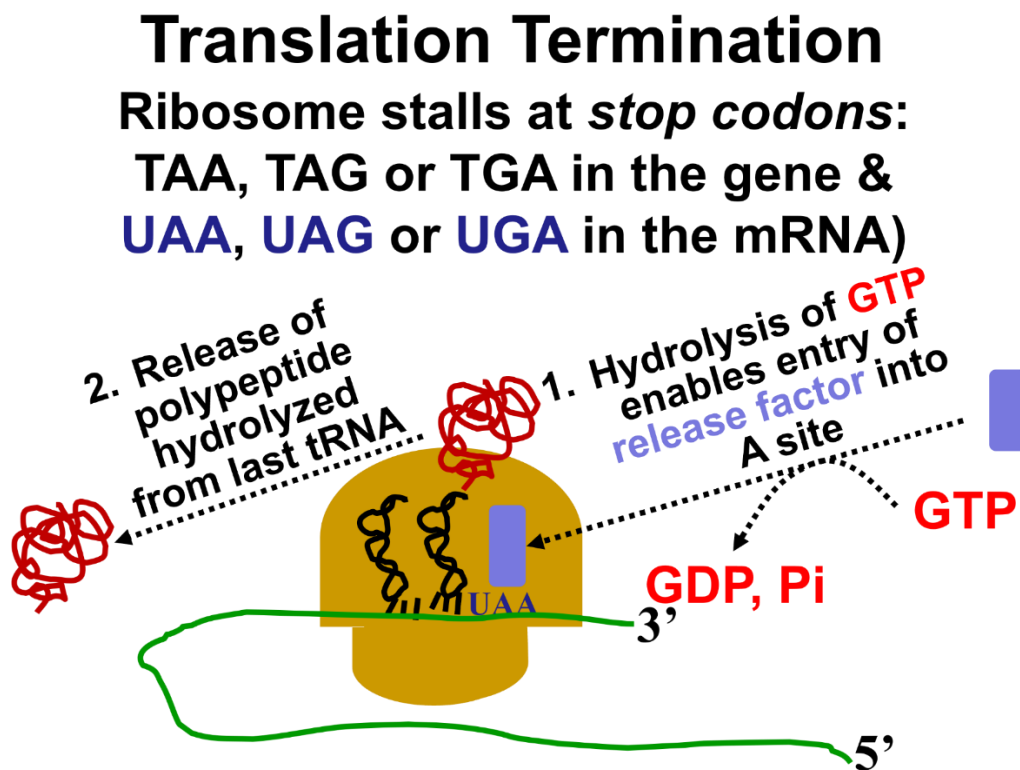


Fig. 11.17: Bacterial translation ends when the ribosome reaches a *termination codon*. Since there are no aminoacyl-tRNAs with anticodons to bind to them in the A site, a *termination factor* enters the A site instead, signaling the separation of mRNA, the new polypeptide, and the ribosomal subunits.



[215-2 Translation Termination](#)

We've seen some examples of posttranslational processing (formyl group removal in *E. coli*, N-terminal methionine removal from most polypeptides, etc.). Most proteins, especially in eukaryotes, undergo additional steps of posttranslational processing before becoming biologically active. We will see examples in upcoming chapters.

11.6 How Can the Genetic Code Be Degenerate and Accurate at the Same Time?

The A-T and G-C complementarity in DNA is known as standard Watson-Crick base pairing. Thinking about how a redundant, degenerate genetic code could work, Francis Crick realized that most often, the first and second bases of codons for the same amino acid were fixed, while the third base was variable (except of course, for methionine and tryptophan codons!). In his *Wobble Hypothesis*, he proposed that the third base in the tRNA anticodon could recognize more than one base in a corresponding mRNA codon because it was able to shift (i.e., wobble) while the other two base pairings were fixed. According to Crick's hypothesis, the wobble would be due to a real freedom of movement, or flexibility of the third base in the anticodon, allowing the formation of thermodynamically stable non-Watson-Crick base pairs.

Some iText & VOP Key Words and Terms

64 codons	free vs bound ribosomes	ribosome
adapter molecules	genetic code	small ribosomal subunit
amino terminus	initiation	start codon
aminoacyl tRNA	initiation complex	stop codons
aminoacyl-tRNA synthase	initiation factors	termination
Amino acid attachment site	initiator tRNA	termination factor
anticodon	large ribosomal subunit	translocation
AUG	meaningful codons	triplets
bacterial bound ribosomes	mRNA, tRNA	tRNA vs tRNA _{aa2}
carboxyl-terminus	ochre, amber, opal	UAA, UAG, UGA
colinearity	peptide linkage	universal genetic code
comma-less genetic code	peptidyl transferase	UUU
degenerate genetic code	polypeptide	Wobble Hypothesis
elongation	polysome	

CHAPTER 11 WEB LINKS



11.1

Chapter 12

Regulation of Transcription and Epigenetic Inheritance

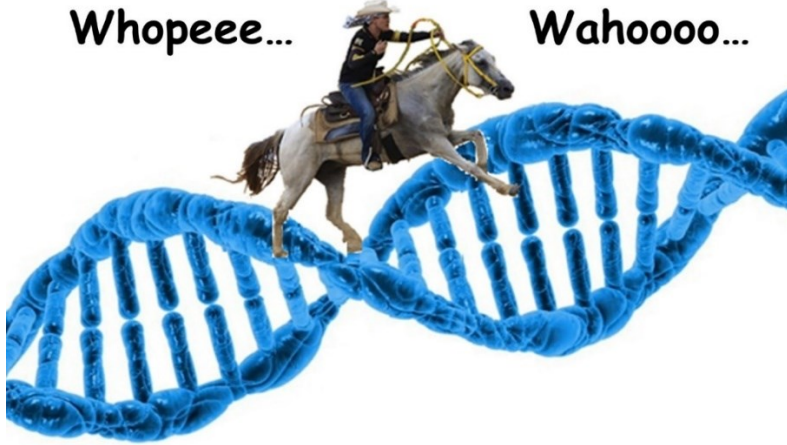
Gene repression and induction (prokaryotes); multiple transcription factors (eukaryotes); regulatory elements in DNA; memories of gene regulation (epigenetics)

Reminder: For inactive *links*, google key words/terms for alternative resources.

Epi G Rides the Genome

Whopee...

Wahoooo...



12.1 Introduction

Cells regulate their metabolism in several ways. We have already seen that allosterically regulated enzymes monitor the cellular levels of metabolites. Recall that the glycolytic intermediates rise and fall in cells based on cellular energy needs, binding to or dissociating from **allosteric sites**. In response to interactions with **allosteric effectors**, allosteric enzymes cause an increase or decrease in catalytic activity.

Cells can also control absolute levels of enzymes and other proteins by turning genes on and off, typically by controlling transcription. **Transcription regulation** usually starts with extracellular environmental signaling. The signals are chemicals in the air, in the water, or (in the case of multicellular organisms) in the blood, lymph, or other extracellular fluids. Bacterial and protist genes often respond to environmental **toxins** or fluctuating **nutrient levels**. Familiar **signal molecules** in higher organisms include **hormones**, which are released at the appropriate time in a sequential **developmental program** of gene expression or in response to nutrient levels in body fluids.

Some signal molecules enter cells and bind to specific **intracellular** (and in a few cases, **intranuclear**) **receptors** to convey their instructions. Others bind to **cell surface receptors**, which **transduce** their "information" into intracellular molecular signals. When signals indicate the need for gene regulation, responding cells ultimately produce **transcription factors**. These in turn recognize and bind to specific **regulatory DNA sequences** associated with the genes that they control.

DNA sequences that bind transcription factors are relatively short. They can lie **proximal** (close) to the transcription start site of a gene, and/or (in the case of eukaryotes) **distal** to (far from) it. We will see that some of these regulatory DNA sequences are **enhancers**,

turning on or increasing gene transcription. Others are ***silencers***, down-regulating or suppressing transcription of a gene. Finally, eukaryotic DNA regulatory sequences are hidden behind a thicket of chromatin proteins. As gene expression patterns change during development, or in response to signaling molecules (like hormones), chromatin is reorganized, cells differentiate, and new tissues and organs form. To this end, new gene expression patterns and chromatin organization in cells must be remembered by their progeny.

In this chapter, we look at the path from cell recognition of a signal molecule to the interaction of regulatory proteins with DNA in prokaryotic and with DNA and chromatin in eukaryotic cells. We also consider how eukaryotic cells *remember* and pass on to their progeny instructions that alter chromatin configuration and patterns of gene expression—topics in the field of **epigenetics**.

Learning Objectives

When you have mastered the information in this chapter, you should be able to do the following:

1. Compare and contrast *transcription factors* and so-called *cis-acting elements*.
2. Discuss the role of *DNA bending* in the regulation of gene expression.
3. Explain the benefits of organizing bacterial genes into *operons* and *why some* bacterial genes are not part of *operons*.
4. Compare and contrast regulation of the *lac* and *trp* operons in *E. coli*.
5. Define and describe *regulatory genes* and *structural genes* in *E. coli*.
6. Discuss why a fourth gene was suspected in *lac-operon* regulation.
7. Distinguish between gene *repression* and *derepression* and between *positive* and *negative* gene regulation, using real or hypothetical examples. For example, explain how it is possible to have repression by positive regulation.
8. Draw and label all functional regions of prokaryotic and eukaryotic genes.
9. Define and articulate the differences (if any!) between *gene expression*, gene regulation, and *transcription regulation*.
10. Describe the *transcription-initiation complex* of a regulated gene in eukaryotes.
11. Define a gene.
12. Compare and contrast different mechanisms of gene regulation in eukaryotic cells.
13. Distinguish between the roles of enhancers and other *cis-acting elements* in transcription regulation.
14. Compare and contrast the *genome* and the *epigenome*.

12.2 Gene Regulation in Prokaryotes: The Lactose (*lac*) Operon

Many prokaryotic genes are organized into ***operons***: linked genes transcribed into a single mRNA that encodes two or more proteins. Operons usually encode proteins with related functions. Regulating the activity of an operon (rather than multiple single genes encoding single proteins) allows better coordination of the synthesis of several proteins at once. In *E. coli*, the regulated ***lac operon*** encodes three enzymes involved in the metabolism of ***lactose*** (an alternative nutrient to glucose). Regulation of an operon (or of a single gene for that matter) may occur by ***repression*** or by ***induction***. When a small metabolite in a cell binds to a regulatory ***repressor*** or ***inducer*** protein, the protein undergoes an allosteric change that allows it to bind to or to unbind from a regulatory DNA sequence.

We will see examples of such regulation in the *lac* and *trp* (tryptophan) operons. Regulation of the *lac*-operon genes is an example of **gene repression** as well as **gene induction**. We'll see that *trp*-operon regulation is by **gene repression**. In both operons, changes in levels of intracellular metabolites reflect the metabolic status of the cell, eliciting appropriate changes in gene transcription. Let's start with a look at *lac*-operon regulation.



216 Overview of Prokaryotic Gene Regulation

The mRNA transcribed from the *lac operon* is simultaneously translated into three enzymes, as shown in Figure 12.1.

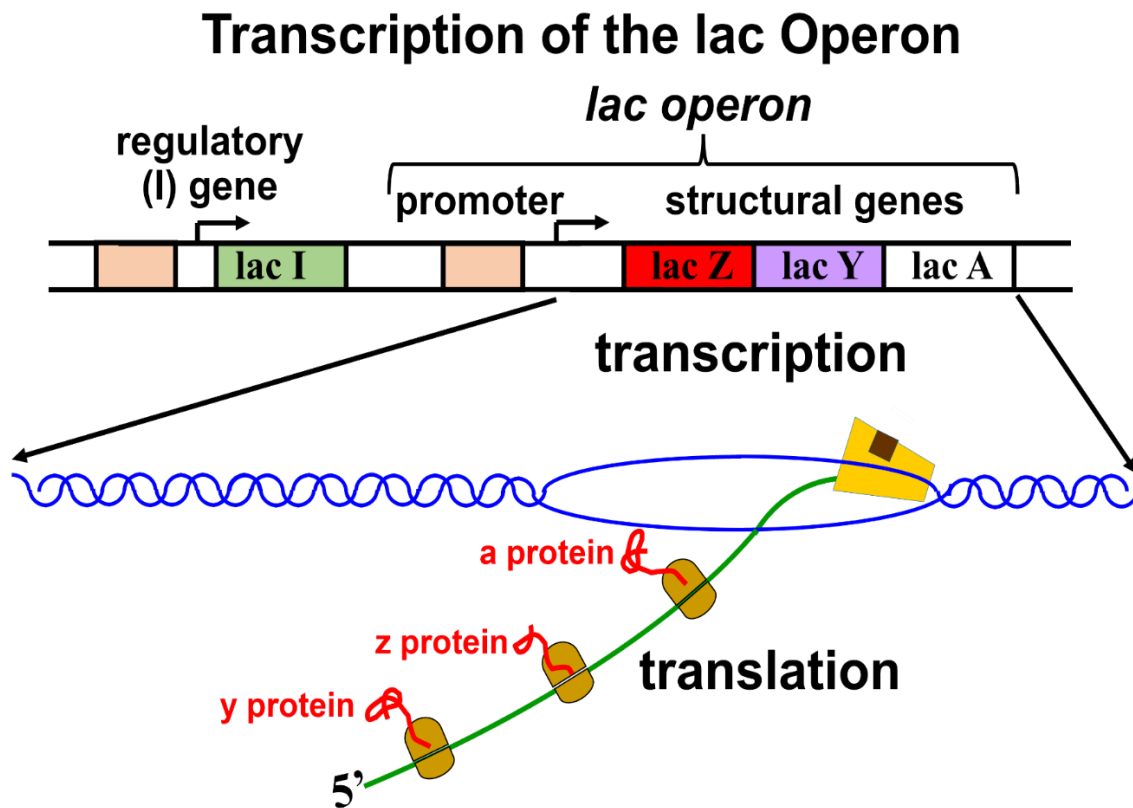


Fig. 12.1: *E. coli lac* operon transcription produces a single mRNA that encodes Z, Y and A polypeptides involved in lactose metabolism. A regulatory (*lac I*) gene is to the left of the operon.

12.2.1 Working Out Regulation of the *lac* Operon in *E. coli*

In the animal digestive tract (including ours), genes of the *E. coli lac operon* regulate the use of *lactose* as an alternative nutrient to glucose. Think "cheese" instead of "chocolate"! The operon consists of *lacZ*, *lacY*, and *lacA* genes, which are all **structural genes** encoding proteins that participate in cell structure and metabolic function. As already noted, the *lac operon* is transcribed into an mRNA encoding the Z, Y, and A proteins. Let's zoom in on a single *lac* operon below in Figure 12.2, to see how a gene on the left (the *lac I* gene) controls the transcription of all the genes in the operon.

lac Operon Gene Organization and Proximity to the *lac I* Regulatory Gene

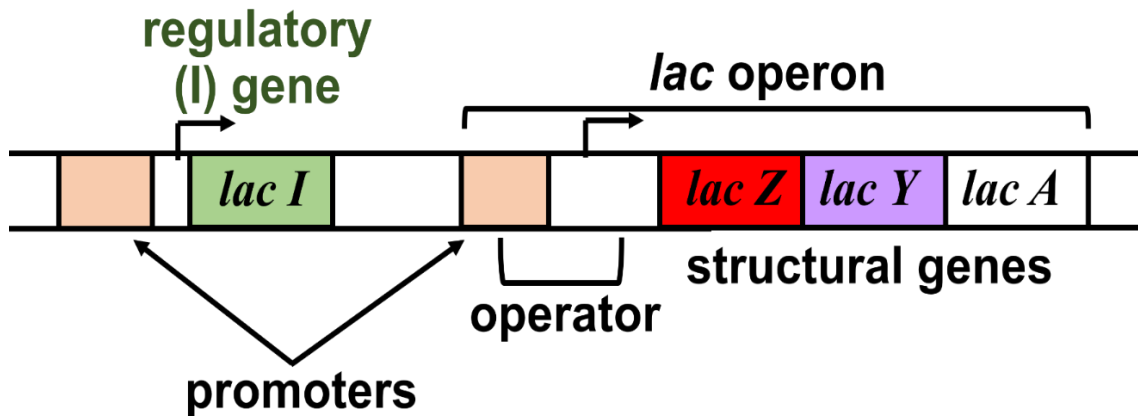


Fig. 12.2: The *lacZ*, *lacY*, and *lacA* genes of the *lac* operon are all controlled by the single promoter to the left of the *lacZ* gene. A regulatory *lac I* gene at the left of the operon with its own promoter.

The *lacZ* gene encodes ***β-galactosidase***, the enzyme that breaks lactose, a disaccharide, into *galactose* and glucose. The *lacY* gene encodes lactose ***permease***, a membrane protein that facilitates lactose entry into the cells. The role of the *lacA* gene (encoding a ***transacetylase***) in lactose energy metabolism is not well understood. The ***lacI* gene** to the left of the *lac Z* gene is a **regulatory gene** (distinct from *structural genes*). Regulatory genes encode proteins that interact with the regulatory DNA sequences associated with a gene to control transcription. As we describe next, the **operator** sequence separating the *lacI* and *lacZ* genes is a transcription-regulatory DNA sequence.

The *E. coli lac* operon is usually silent (repressed) because these cells prefer glucose as an energy and carbon source. In the presence of sufficient glucose, a **repressor protein** (the *lacI*-gene product) is bound to the **operator**, blocking transcription of the *lac* operon. Even if lactose is available, cells will not be able to use it as an alternative energy and carbon source when glucose levels are adequate. However, when glucose levels drop, the *lac* operon is active, and the three enzyme products are translated. We will see how *lac*-operon transcription by both **derepression** and direct **induction** can lead to maximal transcription of the *lac* genes only when necessary (i.e., in the presence of lactose and absence of glucose). Let's look at some of the classic experiments that led to our understanding of *E. coli*-gene regulation in general, and of the *lac* operon in particular.

In the late 1950s and early 1960s, François Jacob and Jacques Monod were studying the use of different sugars as carbon sources by *E. coli*. They knew that *wild-type E. coli* would **not** make the β -galactosidase, the β -galactoside permease, or the β -galactoside transacetylase proteins when grown on glucose. Of course, they also knew that the cells would switch to lactose for growth and reproduction if they were deprived of glucose, producing the three enzymes of lactose metabolism. Therefore, they then searched for, found, and isolated different *E. coli* mutants that could not grow on lactose, even when there was no glucose in the growth medium. Here are some of the mutants they studied:

1. One mutant failed to make active β -galactosidase but did make permease.
2. One mutant failed to make active permease but made normal amounts of β -galactosidase.

3. Another mutant failed to make transacetylase but could still metabolize lactose in the absence of glucose. Hence the uncertainty of its role in lactose metabolism.
4. Curiously, one mutant strain failed to make any of the three enzymes!

Since double mutants are very rare and triple mutants even rarer, François Jacob and Jacques Monod inferred that the activation of all three genes in the presence of lactose were controlled together in some way. In fact, it was this discovery that defined the operon as a set of genes transcribed as a single mRNA, whose expression might therefore be effectively coordinated by a simple chemical “on-off” switch. They later characterized the repressor protein produced by the *lacI* gene as this switch. F. Jacob, J. Monod, and A. Lwoff shared the Nobel Prize in Physiology or Medicine in 1965 for their work on bacterial gene regulation. We now know that there are several layers of *lac*-operon regulation. *Negative* and *positive* regulation of the *operon* depends on two regulatory proteins to control the rate of lactose metabolism.

12.2.2 Negative Regulation of the *lac* Operon by Lactose

Repression of *lac*-operon activity involves a repressor protein that must be removed for gene expression to occur. The repressor protein product of the *I* gene is always made and present in *E. coli* cells; expression of the *I* gene is **constitutive**, meaning it is unregulated. In the absence of lactose in the growth medium, the repressor protein binds tightly to the operator DNA. While **RNA polymerase** is bound to the promoter and ready to transcribe the operon, the presence of the repressor protein—which is bound to the operator sequence between the RNA polymerase and the Z-gene transcription start site—physically blocks its forward movement. Under these conditions, little or no transcript is made. Figure 12.3 shows the *lac* repressor protein in its repressed state.

The *lac* Operon is Usually Repressed

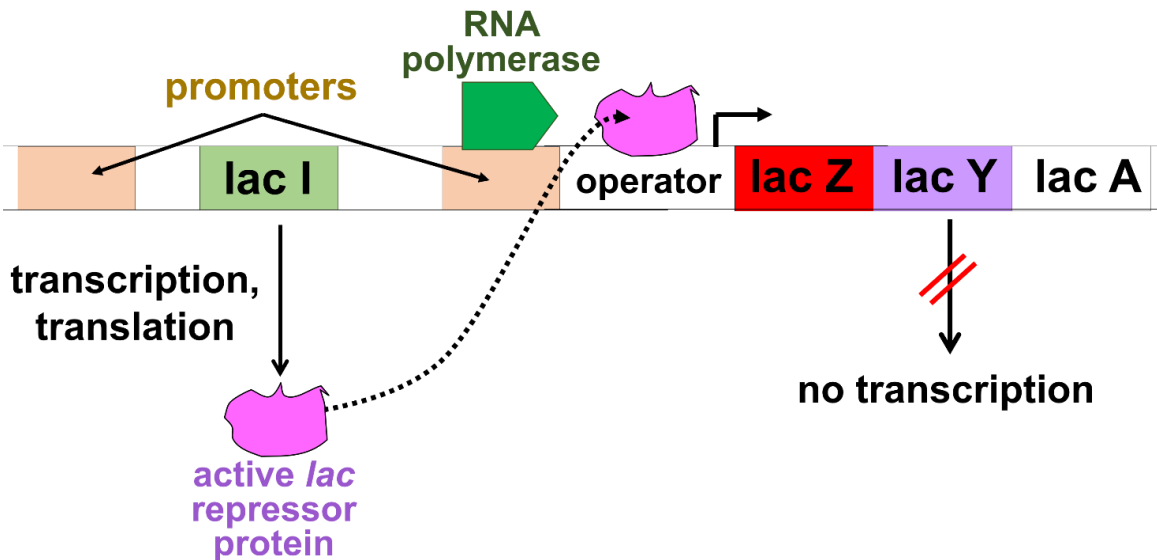


Fig. 12.3: *Negative regulation of the lac operon:* 'Negative regulation' of the *lac* operon: The *lac* operon in this illustration cannot be transcribed because a '*lac* repressor' protein is bound to the operator sequence between the promoter and transcription start site of the operon, blocking RNA polymerase movement.

If cells are grown in the presence of lactose, the lactose entering the cells is converted to **allolactose**. Allolactose in turn binds to the repressor sitting on the operator DNA to form a two-part complex, as shown in Figure 12.4.

Excess Allolactose Binds *lac* Repressor

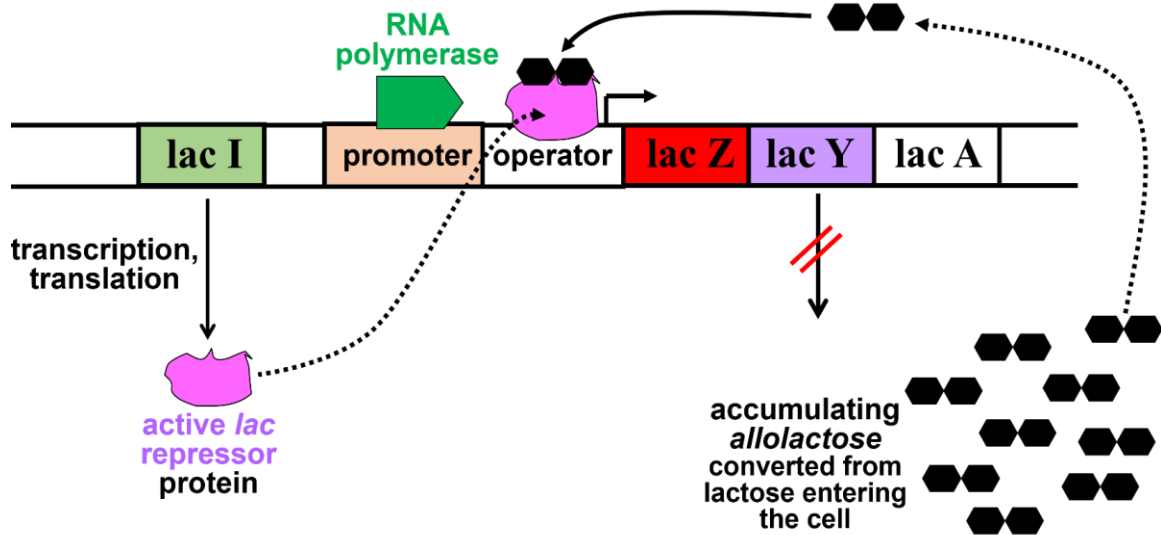


Fig. 12.4: Negative regulation of the *lac* operon: If present, lactose enters cells and is converted to *allolactose*. Allolactose is a *lac* operon *inducer*. if it accumulates in the cell, it will bind the *lac* repressor protein and reduce its affinity for the *lac* operon operator sequence.

The allosterically altered repressor dissociates from the operator, and RNA polymerase can resume transcribing the *lac*-operon genes (Figure 12.5).

lac Operon Derepressed, Some Transcription

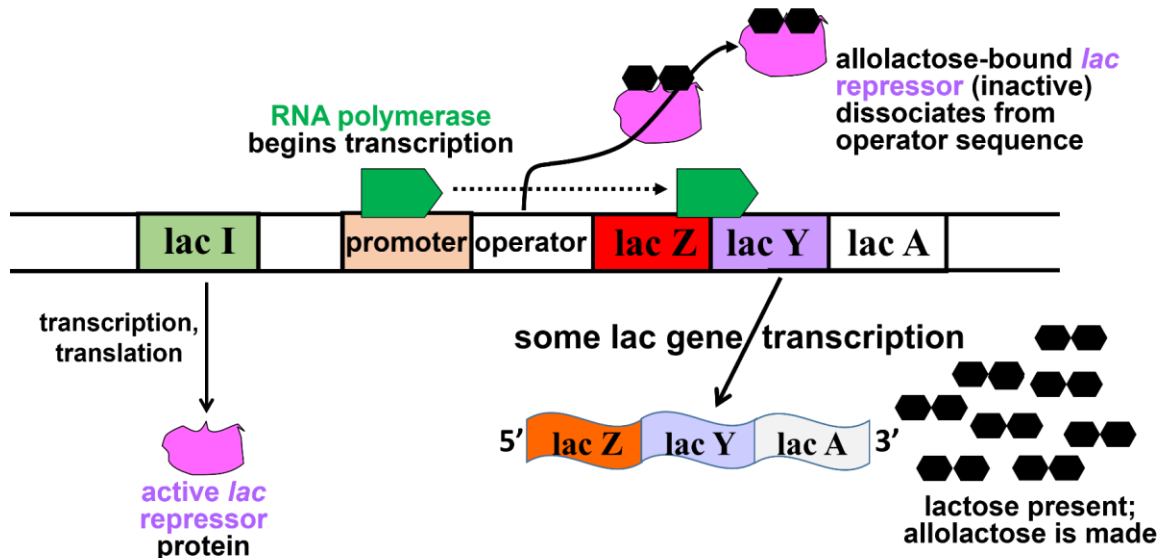


Fig. 12.5: Negative regulation of the *lac* operon: Allolactose 'induces' *lac* gene transcription by binding to the *lac* repressor. The resulting allosteric change lowers the affinity of the repressor for the operator sequence. The repressor then dissociates from the operator sequence, allowing RNA polymerase to move forward and transcribe a low level of the *lac* genes.

12.2.3 Positive Regulation of the *lac* Operon: Induction by Catabolite Activation

The second control mechanism regulating *lac*-operon expression is mediated by CAP (cAMP-bound **catabolite activator protein** or cAMP receptor protein). When glucose is available, cellular levels of cAMP (cyclic AMP) are kept low in the cells, and CAP is in an inactive conformation. On the other hand, if glucose levels are low, levels of cAMP levels go up and bind to the CAP; CAP is then active. If lactose levels are also low, the cAMP-bound CAP will have no effect. But, if lactose is present and glucose levels are low, then allolactose will bind to the *lac* repressor, causing it to change conformation and dissociate from the operator region. Under these conditions, the cAMP-bound CAP can bind to the operator in lieu of the repressor protein. When this happens, rather than blocking RNA polymerase, the activated cAMP-bound CAP induces even more efficient *lac*-operon transcription. The result is synthesis of higher levels of *lac* enzymes, which in turn, facilitates efficient cellular use of lactose as an alternative energy source to glucose. Figure 12.6 illustrates this maximal *lac*-operon activation under high lactose and low glucose conditions.

Maximal Induction of *lac* Gene Transcription When Operon is De-repressed *and* Induced

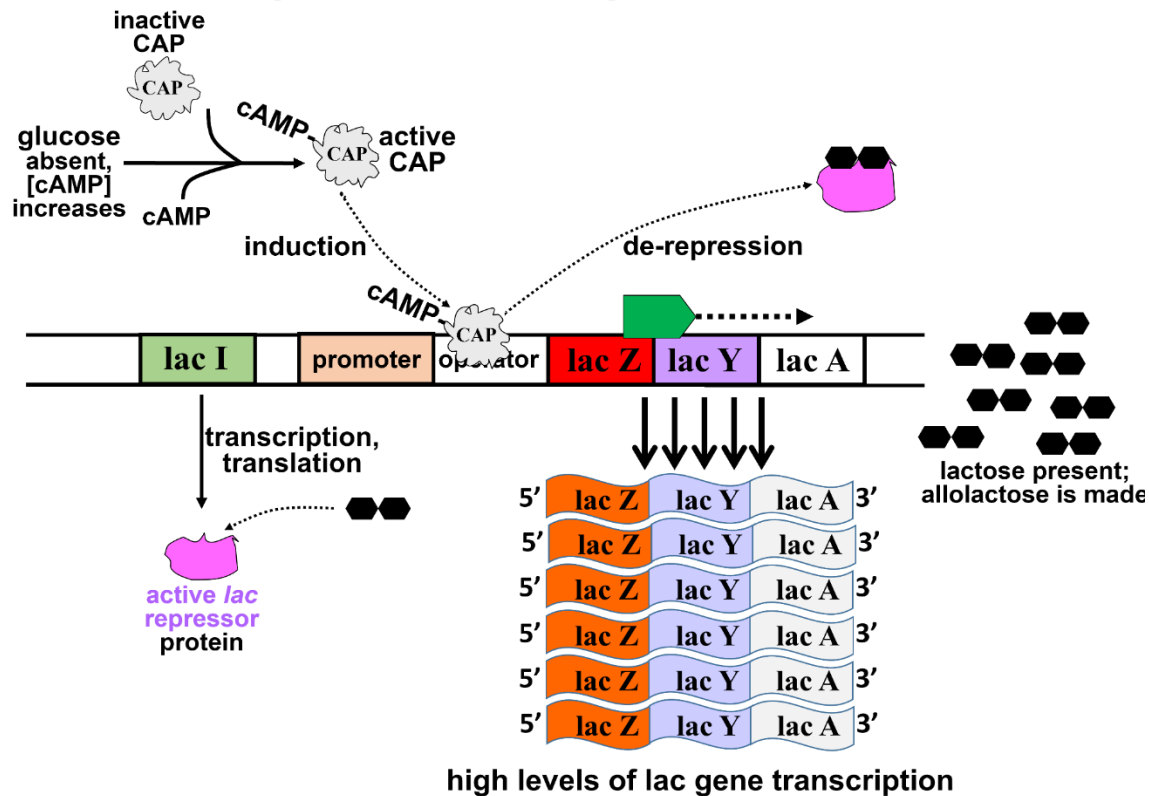


Fig. 12.6: Positive regulation of the *lac* operon: If lactose is present in the absence of glucose, the operon is *derepressed* AND *induced*. Increased cAMP binds the *catabolite activator protein* (CAP), which binds to the *lac* operator (instead of the *lac* repressor protein), inducing maximal *lac*-gene transcription. By positive regulation



cAMP-bound-CAP is a transcription *inducer*. It forces DNA in the promoter-operator region to bend (Figure 12.7).

cAMP-cap Protein Causes DNA to *Bend*!

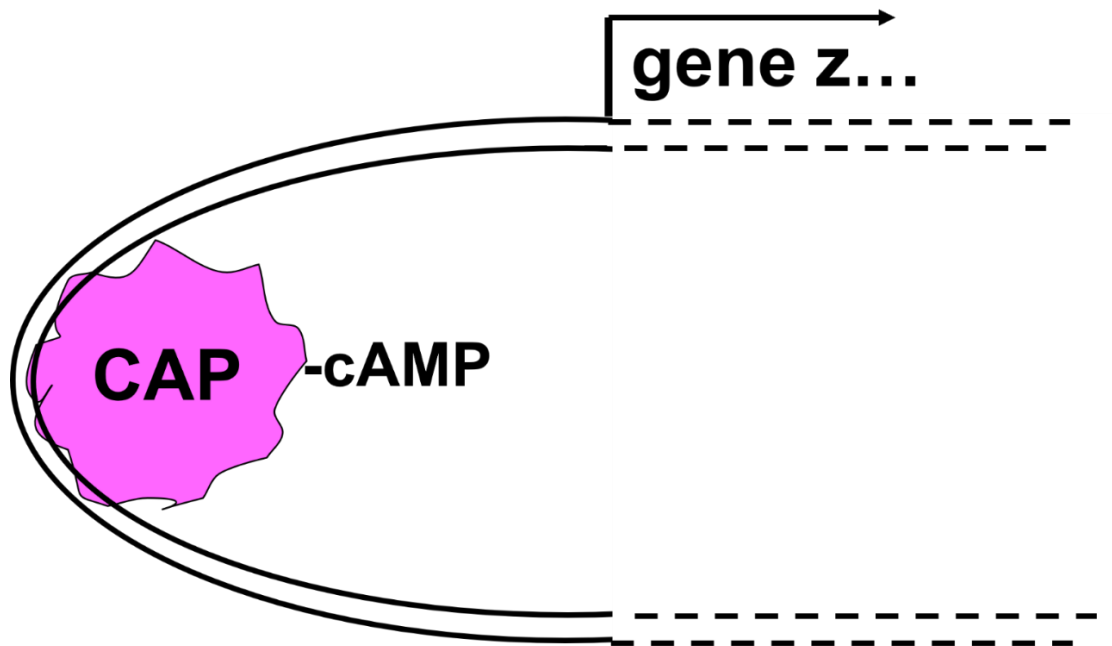


Fig. 12.7: *Positive regulation of the lac operon:* the cAMP-CAP-bound operator bends the double helix. This loosens the H-bonds between the bases, making transcription-factor access to template strands easier. Bending the double helix to facilitate transcription is also a regulatory feature in eukaryotes.

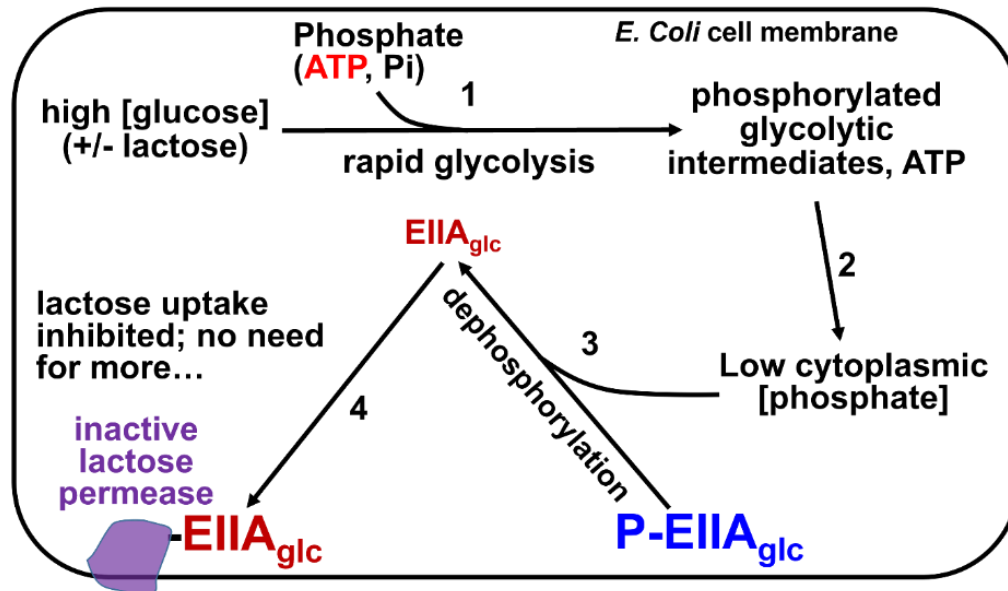
Binding of cAMP-CAP to the double helix loosens the H-bonds between the strands, making it easier for RNA polymerase to find and to bind the promoter and for transcription to begin.

In recent years, additional layers of gene regulation in the *lac*-operon have been uncovered. In one case, the ability of *lac permease* to transport lactose across the cell membrane is regulated. In another, additional operator sequences have been discovered to interact with a multimeric repressor to control *lac*-gene expression. We'll consider *inducer exclusion* first.

12.2.4 *lac*-Operon Regulation by Inducer Exclusion

When glucose levels are high (even in the presence of lactose), phosphate is consumed to phosphorylate glycolytic intermediates. This keeps cytoplasmic phosphate levels low. Under these conditions, an unphosphorylated elongation factor (EIIA^{Glc}) binds to the *lactose permease* in the cell membrane, preventing the enzyme from bringing lactose into the cell. The roles of phosphorylated and unphosphorylated EIIA^{Glc} in regulating the *lac* operon are shown below in Figure 12.8.

Inducer Exclusion (lactose/allolactose exclusion) responds to high glucose levels to prevent lactose uptake



1. High [glucose] leads to faster glycolysis;
2. Rapid glycolysis reduces cellular [phosphate]
3. Low cytoplasmic [phosphate] causes dephosphorylation of **P-EIIA_{glc}** to **EIIA_{glc}**
4. The **EIIA_{glc}** binds to and inhibits the **lactose permease** enzyme

Fig. 12.8: *lac*-operon regulation by *inducer exclusion*: High glucose levels accelerate glycolysis, depleting cellular phosphate, leading to protein dephosphorylation of proteins. One such proteins is P-EIIA^{Glc}, a translation elongation factor). Without the phosphate, EIIA^{Glc} inhibits *lactose permease* (the *lacA* protein). Lactose can't enter cells, the *lac* operon is needlessly transcribed.

High glucose levels block lactose entry into the cells, effectively preventing allolactose formation and derepression of the *lac* operon. Inducer exclusion is thus a logical way for the cells to handle an abundance of glucose, whether or not lactose is present. On the other hand, if glucose levels are low in the growth medium, phosphate concentrations in the cells rise sufficiently for a specific kinase to phosphorylate the EIIA^{Glc}. Then, phosphorylated EIIA^{Glc} undergoes an allosteric change and dissociates from the lactose permease, making it active, so that more lactose can enter the cell. In other words, the inducer is not "excluded" under these conditions!

The kinase that phosphorylates EIIA^{Glc} is part of a phosphoenolpyruvate- (PEP-) dependent *phosphotransferase system* (PTS) cascade. When extracellular glucose levels are low, the cell activates the PTS system to bring whatever glucose is around into the cell. The last enzyme in the PTS *phosphorylation cascade* is the kinase that phosphorylates EIIA^{Glc}. Now activated, phosphorylated EIIA^{Glc} dissociates from the lactose permease. The now-active permease can bring available lactose from the medium into the cell.

12.2.5 Structure of the *lac*-Repressor Protein and Additional Operator Sequences

The *lac* repressor encoded by the I gene is a tetramer of identical subunits (Figure 12.9).

lac Repressor Structure

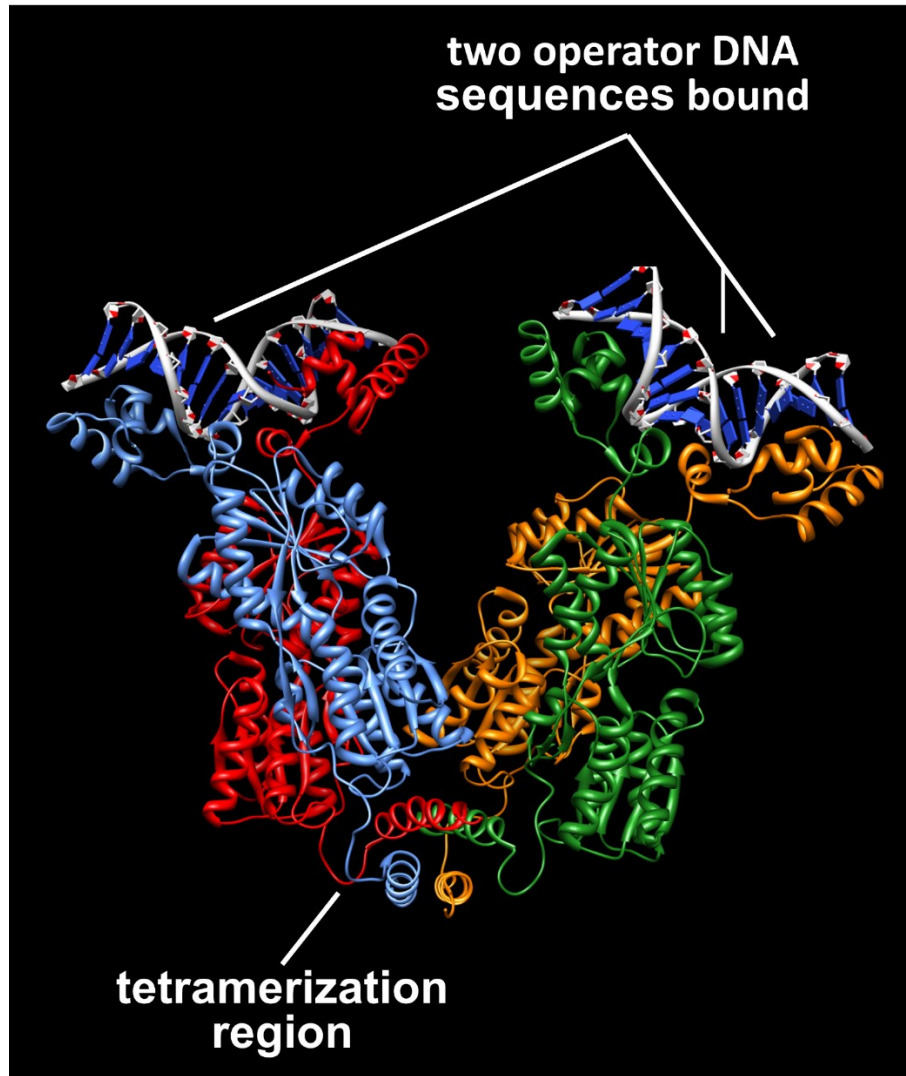


Fig. 12.9: Computer-generated structure of the tetrameric *lac* repressor, which is bound to DNA via *helix-turn-helix* motifs.

Each subunit contains a *helix-turn-helix motif* capable of binding to DNA. However, the operator DNA sequence downstream of the promoter in the operon consists of a pair of *inverted repeats* spaced apart in such a way that they can only interact with two of the repressor subunits, leaving the function of the other two subunits unknown...that is, until recently! Two more operator regions were recently characterized in the *lac* operon. *Operator 2* (**O2**) is within the *lac Z* gene itself, and **O3** is near the end but still within the *lac I* gene. Apart from their unusual location within actual genes, these operators, which interact with the remaining two repressor subunits, went undetected at first because mutations in the O2 or O3 regions individually do not contribute substantially to the effect of lactose in derepressing the *lac* operon. Only mutating both regions at the same time results in a substantial reduction in the binding of the repressor to the operon.

12.3 Gene Regulation in Prokaryotes: the Tryptophan (*trp*) Operon

Let's look at two of the ways that *E. coli* controls cellular levels of its metabolites such as amino acids. Recall feedback inhibition, in which the synthesis of a cellular metabolite is inhibited when an excess of the metabolite binds to and allosterically inhibits the activity of an enzyme in the pathway to its own synthesis (Figure 5.4). Synthesis of the amino acid **tryptophan (*trp*)** is allosterically regulated in this way. The 5 polypeptides encoded by the *trp* operon comprise the three enzymes in the *trp* biosynthetic pathway as shown in Figure 12.10: **holoenzyme 1** is a dimer of proteins encoded by genes 4 and 5 of the operon; **enzyme 2** is a polypeptide encoded by *gene 3*; **holoenzyme 3** is a dimer of proteins encoded by genes 1 and 2. When cell growth slows down, the cells initially accumulate excess *trp*, some of which will bind to and allosterically inhibit the dimeric *holoenzyme 1*, the first enzyme in the pathway. The result of the inhibition of intermediate **B** production is of course, a slowdown in *trp* synthesis.

Bacterial cells use a second feedback mechanism to slow down *trp* synthesis. During rapid growth, the *trp* operon is active and will continually synthesize the amino acid. The "on" state of the *trp* operon is illustrated in Figure 12.10.

Active Transcription of the Tryptophan (*trp*) Operon

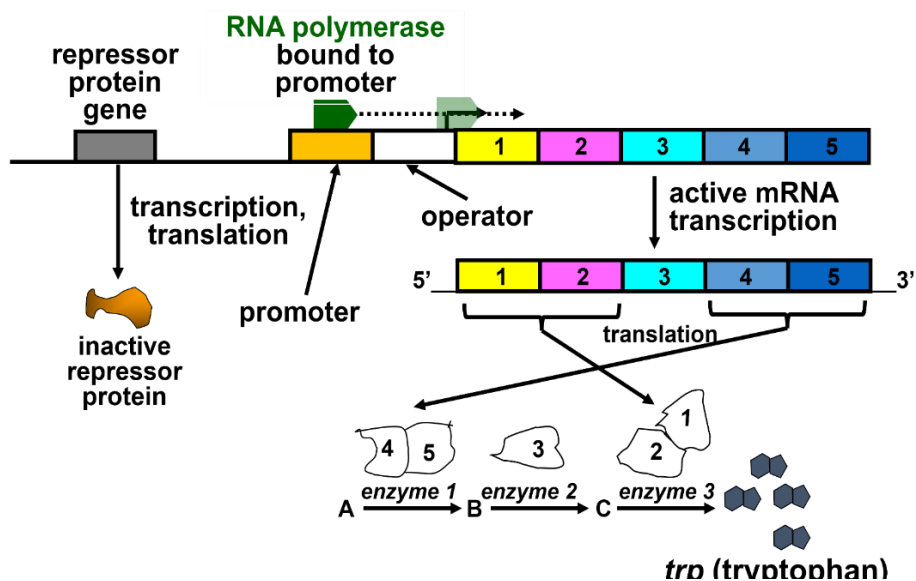
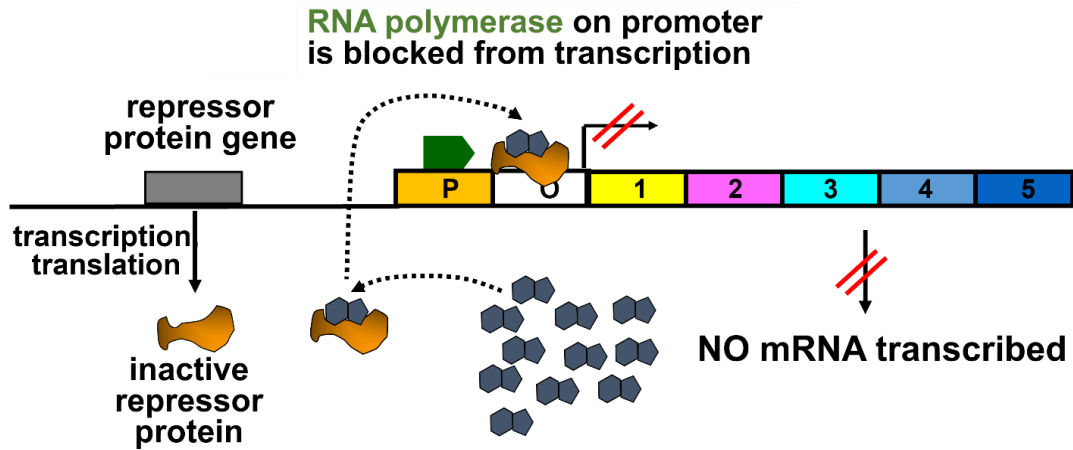


Fig. 12.10: Transcription of the five-gene tryptophan (*trp*) operon: the five resulting proteins are enzymes in the pathway for *trp* synthesis. A *trp*-repressor gene to the left of the operator is always transcribed.

On the other hand, if tryptophan consumption slows down, tryptophan accumulates in the cytoplasm. Excess tryptophan will then bind to the *trp* repressor, a protein encoded by the repressor protein gene. The *trp*-bound repressor then binds to the *trp* operator, blocking RNA polymerase from transcribing the operon. The repression of the *trp* operon by *trp* is shown in Fig. 12.11 (below). In this scenario, tryptophan is a **co-repressor**. The function of a co-repressor is to bind to a repressor protein and change its conformation so that it can bind to the operator.

Silencing the trp Operon with a Co-Repressor



When *trp* accumulates to excess it becomes a co-repressor...

Fig. 12.11: When there is enough tryptophan (*trp*) in cells, it acts as a corepressor, binding to and changing the conformation of the *trp*-repressor, which then recognizes and binds the *trp*-operon operator (O), blocking RNA polymerase from moving past the promoter (P), thus repressing operon transcription.



[219 Repression of the Tryptophan \(*trp*\) Operon](#)



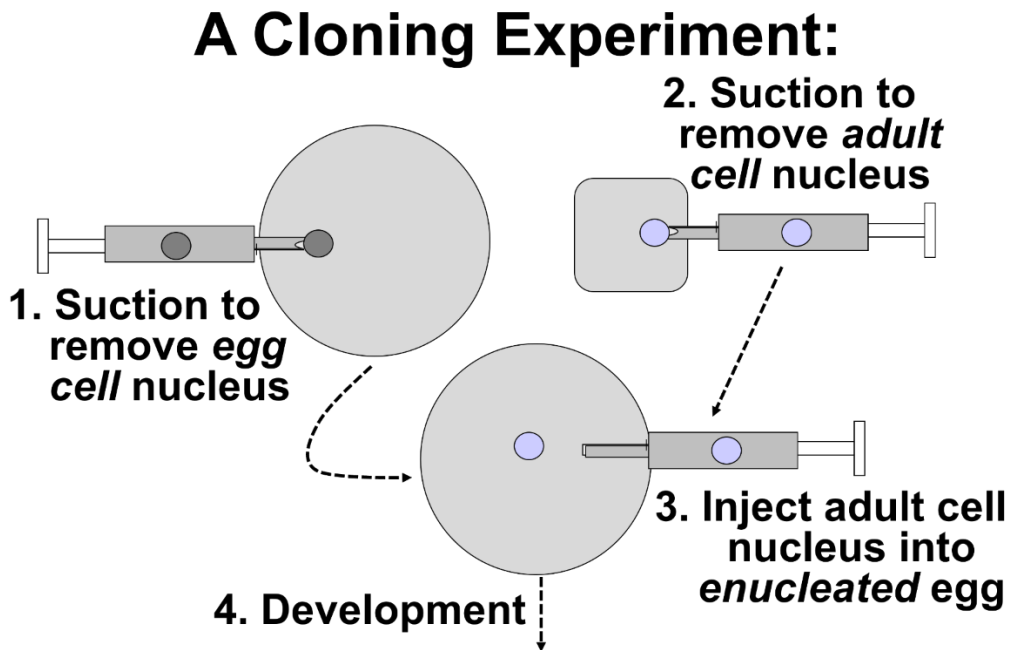
12.4 The Problem with Unregulated (*Housekeeping*) Genes in All Cells

Before we turn our attention to the regulation of gene expression in eukaryotes, consider for a moment the expression of **constitutive** (i.e., unregulated, or **housekeeping**) genes that are always active. The requirement that some genes always be "on" raises questions about cellular priorities of gene expression. *Constitutive* gene products include sets of many polypeptides that form large macromolecular complexes in cells; they also include enzyme sets that participate in vital biochemical pathways. How do cells maintain such polypeptides in stoichiometrically reasonable amounts? Or can their levels rise or fall transiently without much effect? Recent studies suggest that transcription of *housekeeping* genes is, in fact, not coordinated!

We also saw that the efficiency of glycolysis relies in part on allosteric regulatory mechanisms that evolved to control the activities of glycolytic enzymes rather than their transcription. While this takes care of some element of metabolic control, a problem remains. Recall that protein synthesis is energy-intensive: each peptide linkage costs three NTPs (not to mention the waste of an additional NTP per phosphodiester linkage made in the transcription of an mRNA!). The overproduction of proteins would seem to be a waste of energy and thus an evolutionary "mistake." While we may not know just how expensive it is to express housekeeping genes, whatever the energy expenses are, they are the cost of evolving complex structures and biochemical pathways vital to their everyday function and survival. This is a good example of a major precept of evolutionary theory: evolution is an accumulation of improvements brought about by (one can almost say *jerry-rigged* by) natural selection; it is not the result of perfect design in advance. Now back to our focus on regulated gene expression in eukaryotes.

12.5 Gene Regulation in Eukaryotes

Let's recall an experiment described earlier and illustrated in Figure 12.12.



Results: e.g., normal adult clones of e.g., cloning in frogs, sheep, cows...

Conclusion: nuclei of all cells of an organism contains the same genome.

Fig. 12.12: Replacement of an egg's nucleus with that from an adult cell enabled the cloning of embryos and eventually, of complete organisms. This proved that adult cells contain all the genes required to program and to complete the development of the organism.

Results of this experiment provided the evidence that even very different cells of an organism contain the same genes. In fact, in any multicellular eukaryotic organism, every cell contains the same genes in their DNA. Therefore, the different cell types in an organism differ *not* in which genes they contain, but which sets of genes they express! Looked at another way, cells differentiate when they turn on new genes and turn off old ones. Thus, gene regulation produces different sets of gene products during differentiation, leading to cells that look and function differently in the organism.



[220 An Experiment: All of an Organism's Cells Have the Same Genome](#)



In eukaryotes many steps lie between the transcription of an mRNA and the accumulation of a polypeptide end-product—many more steps than occur in prokaryotes. Eleven of these eukaryotic steps are shown below in Figure 12.13, which lays out the pathway from gene to protein.

Eleven Steps to Protein Synthesis Where Gene Expression Might Be Controlled

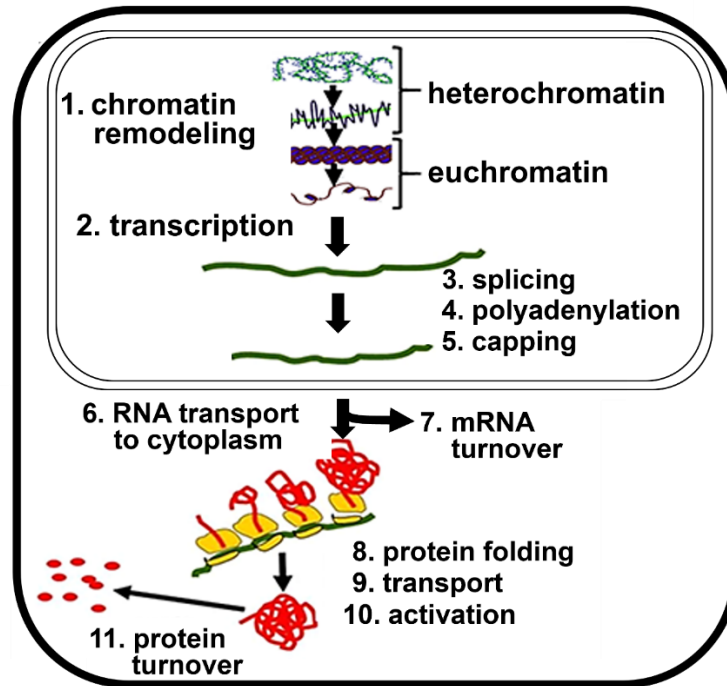


Fig. 12.13: A potential to regulate cellular metabolism by controlling mRNA and/or protein levels or activity.

Theoretically, eukaryotic cells can turn on, turn off, speed up, or slow down any step(s) in this pathway, changing the steady-state concentration of a polypeptide in the cells. But the expression of a gene is usually controlled at only one or a few steps. As in bacteria, control of transcription initiation is a common mechanism for regulating gene expression (in principle if not in detail).



[221-2 Many Options for Regulating Eukaryotic Genes](#)



12.5.1 Complexities of Eukaryotic Gene Regulation

Eukaryotic gene regulation is more complex than in prokaryotes. This is in part because their genomes are larger and because they encode more genes. For example, the *E. coli* and human genomes house about five and twenty-five thousand genes, respectively. Furthermore, unlike prokaryotes, eukaryotes produce more than twenty-five thousand proteins by *alternative splicing* to recombine mRNA coding regions and—in a few cases—by initiating transcription from *alternative start sites* in the same gene. And of course, the activity of many more genes must be coordinated without the benefit of multigene operons! Eukaryotic gene regulation is also more complicated because all nuclear DNA is wrapped in protein in chromatin and chromosomes. Last but not least, all organisms control gene activity with **transcription factors** that must penetrate chromatin to bind to specific DNA sequences (**cis-regulatory elements**). In eukaryotes, these elements can be **proximal to** (near) the promoter of a gene or **distal to** (quite far from) the gene they regulate.

Figure 12.14 *maps* a eukaryotic gene to show its typical components (e.g., its promoter, introns, and exons) and its associated *cis-acting* regulatory elements.

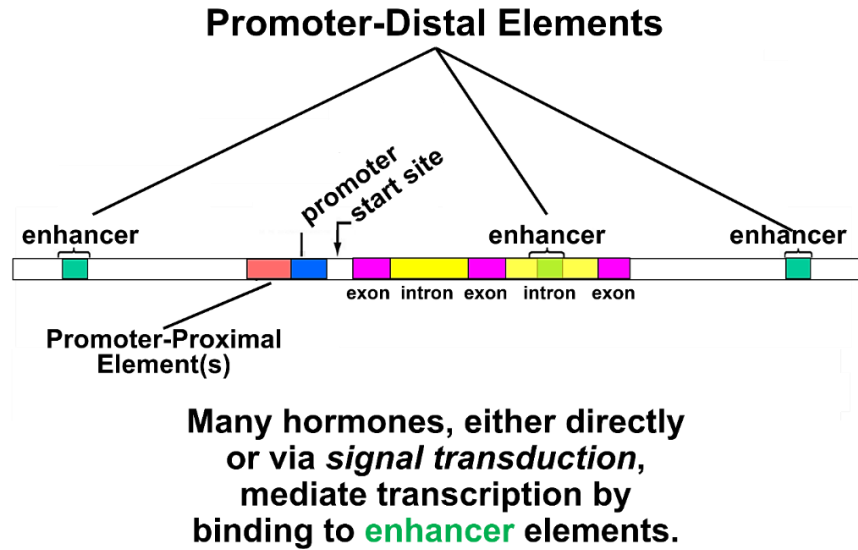


Fig. 12.14: Eukaryotic regulatory DNA sequences (*CIS* elements) are *proximal* (near a gene promoter) or *distal* (far from the promoter). Here, a *proximal* element lies just left of the gene promoter, while *distal* elements (in this case, 'enhancers') lie upstream (to the left) of the promoter, downstream (to the right) of the last exon of the gene, and in the middle of an intron. Regulatory proteins (sometimes called 'TRANS' factors) bind to proximal and distal *CIS* elements to control transcription of regulated genes.

Enhancers are typical *distal* *cis* elements that recognize and bind transcription factors to *increase* the rate of transcription of a gene. Oddly enough, these short DNA elements can be in the 5' or 3' untranslated region of the gene or even within introns, often thousands of base pairs away from promoters and transcription start-sites of the genes they control. *Upstream* regulatory regions of eukaryotic genes (to the left of a gene promoter as shown in Figure 12.14) often have distal binding sites for more than a few transcription factors, some with positive (*enhancing*) and others with negative (*silencing*) effects. Of course, which of these DNA regions are active in controlling a gene depends on which transcription factor(s) are present in the nucleus. Sets of positive regulators can work together to coordinate and to maximize gene expression when needed and sets of negative regulators may bind negative regulatory elements to silence a gene.



[222-2 Transcription Factors Bind DNA Near and Far](#)



We saw in eukaryotes that mRNA transcription initiation involves many transcription factors (*TFs*) and RNA polymerase II acting at a gene promoter to form a ***transcription-preinitiation complex***. *TFIID* (the ***TATA-binding protein***) is one of the first factors to bind DNA in the promoter region, causing it to bend, much like the CAP-protein in bacteria. Once bound, *TFIID* recruits other transcription factors to the promoter. As in bacteria, bending DNA loosens the H-bonds between bases, facilitating unwinding of the double helix near the gene. As the eukaryotic DNA bends, regulatory proteins on distal enhancer sequences are brought close to proteins bound to regulatory elements that are more proximal to the promoter, as shown in Figure 12.15 (below).

Assembly of a Eukaryotic Transcription Complex

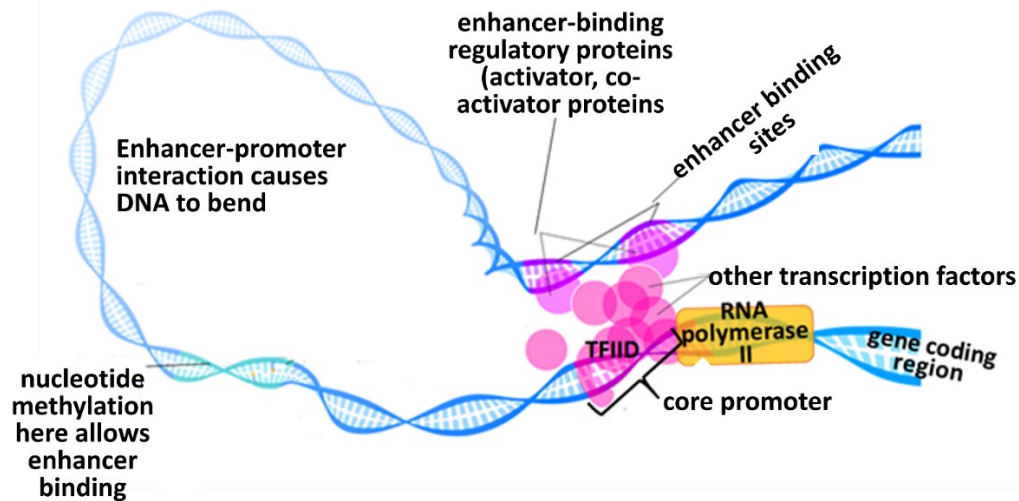


Fig. 12.15: Assembly of a eukaryotic transcription complex on a regulated gene: distal regulatory protein/DNA interactions can cause DNA bending, recruiting transcription-initiation factors and proximal regulatory factors to the transcription complex to control gene expression.

Nucleotide *methylation sites* may facilitate regulatory protein-enhancer binding. When such regulatory proteins, here called *activators* of transcription, bind to their enhancers, they acquire an affinity for protein *cofactors*, which allow recognition and binding to other proteins in a transcription-initiation complex. This results in bending the DNA, which then makes it easier for RNA polymerase II to initiate transcription from the appropriate strand of DNA.



[223-2 Assembling a Eukaryotic Transcription-Initiation Complex](#)

It is worth reminding ourselves that allosteric (shape) changes in proteins allow DNA-protein interactions (in fact, any interaction between macromolecules). The *lac* repressor we saw earlier is a transcription factor with *helix-turn-helix* DNA-binding motifs. This motif and two others (*zinc finger* and *leucine zipper*) that characterize DNA-binding proteins are illustrated in Figure 12.16.

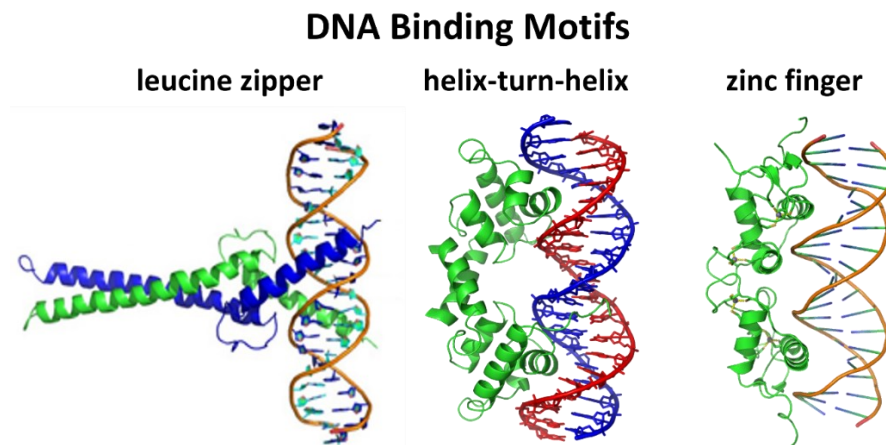


Fig. 12.16: Three DNA-binding motifs commonly found in trans-acting (i.e., regulatory) protein factors.

The DNA-binding motifs shown (zipper, helices, zinc fingers) bind regulatory DNA sequences that are “visible” to the transcription factor in the major groove of the double helix.



[224-2 Transcription Factor Domains: Motifs Bind Specific DNA Sequences](#)

We will look next at some common ways in which eukaryotic cells are signaled to turn genes on or off in order to increase or decrease their rates of transcription. As we describe different gene regulatory strategies, remember that eukaryotic cells regulate gene expression in response to changes in extracellular as well as intracellular environments. These can be unscheduled, unpredictable changes in blood or extracellular fluid composition (e.g., ions and small metabolites), or they can be changes for the long term, dictated by genetic programs of differentiation and development. Some changes in gene expression even obey *circadian* (daily) rhythms, like the ticking of a clock.

In eukaryotes, changes in gene expression, expected or not, are usually mediated by the timely release of chemical signals from specialized cells (e.g., hormones, cytokines, and growth factors). We will focus on some better-understood models of the gene regulation that is caused by these chemical signals.

12.5.2 Regulation of Gene Expression by Hormones That Enter Cells and Those That Don't

Gene-regulatory (cis) elements in DNA and the transcription factors that bind to them have coevolved—but not only that! Eukaryotic organisms have evolved complete pathways that respond to environmental or programmed developmental cues and that lead to an appropriate cellular response. In eukaryotes, chemicals released by some cells signal other cells to respond, effectively coordinating the activity of the whole organism. Hormones released by cells in endocrine glands are well-understood signal molecules. Hormones affect *target cells* elsewhere in the body.



[225 Chemicals That Control Gene Expression](#)

12.5.2.a How Steroid Hormones Regulate Transcription

Steroid hormones cross the cell membranes to have their effects. Common steroid hormones include testosterone, estrogens, progesterone, glucocorticoids, and mineral corticoids. Once inside the target cell, such hormones bind to a cytoplasmic or nuclear ***steroid-hormone receptor*** protein to form a ***steroid-hormone-receptor complex***.

The receptor may be in the cytoplasm or in the nucleus, but in the end the hormone-receptor complex must bind to DNA-regulatory elements of a gene, either to enhance or to silence transcription. Therefore, a steroid hormone must cross the plasma membrane and may also need to cross the nuclear envelope. You can follow the binding of a steroid hormone to a cytoplasmic receptor in Figure 12.17. Here the hormone (depicted as a red triangle in the

illustration) enters the cell. An allosteric change in the receptor releases a protein subunit called Hsp90 (the black rectangle). The remaining hormone-bound receptor enters the nucleus.

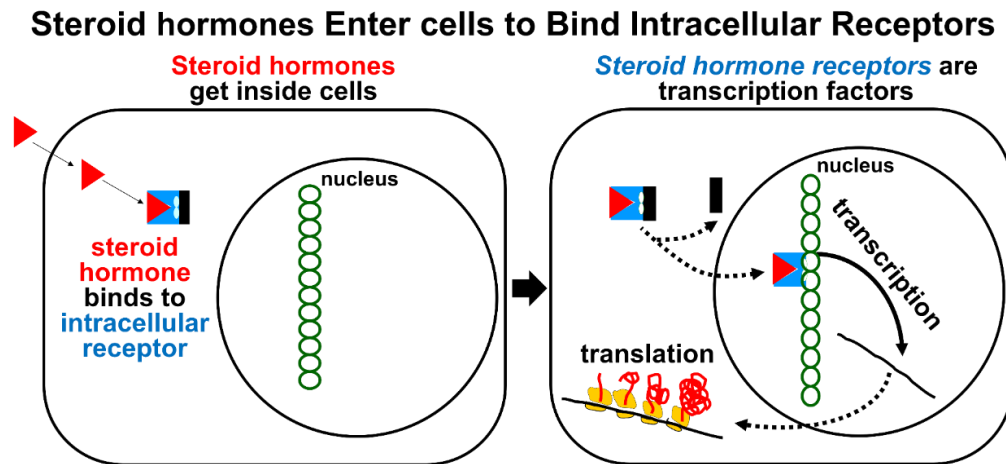


Fig. 12.17: Once in a target cell, some *steroid hormones* bind a *receptor protein* in the cytoplasm, changing its shape. The resulting hormone-receptor complex translocates to the nucleus where it acts as a transcription factor, binding to regulatory DNA sequences to activate, repress or modulate gene transcription. Some steroid hormones find their receptors already in the nucleus!

The fascinating thing about Hsp90 is that it was first discovered in cells subjected to heat stress. When the temperature gets high enough, cells shut down most of transcription and instead transcribe Hsp90 and/or other special *heat-shock* genes. The resulting ***heat-shock proteins*** seem to protect the cells against metabolic damage until temperatures return to normal. Of course, most cells never experience such high temperatures, so the evolutionary significance of this protective mechanism is unclear. But as we now know, heat-shock proteins have critical cellular functions, in this case blocking the DNA-binding site of a hormone receptor until a specific steroid hormone binds to it.

Back to hormone action! No longer associated with the Hsp90 protein, the receptor bound to its hormone *cofactor* now binds to a *cis-acting* transcription-control element in the DNA, turning the transcription of a gene on or off. The hormone receptors for some steroid hormones are already in the nucleus of the cell, so the hormone must cross not only the plasma membrane but also the nuclear envelope in order to access the receptor. As for steroid-hormone functions, we've seen their roles in activating genes for gluconeogenesis. Steroid hormones also control sexual development and reproductive cycling in females, sexual development and sperm maturation in males, salt and mineral homeostasis in the blood, metamorphosis in arthropods, and more, all by regulating gene expression.



[226 Steroid Hormones Regulate Gene Transcription](#)

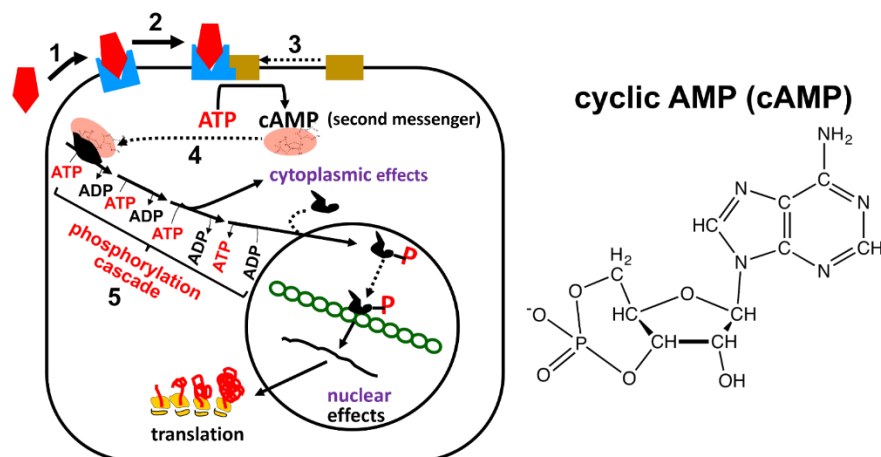


12.5.2.b How Protein Hormones Regulate Transcription

Compared to steroids, proteins are large and soluble and have highly charged surfaces. Even protein-derived hormones like adrenalin, though small, are charged and hydrophilic. Large and/or hydrophilic signal molecules cannot get across the phospholipid barrier of the plasma

membrane. To have any effect at all, they must bind to receptors on the surface of cells. Typically, these receptors are membrane glycoproteins. The information carried by protein hormones must be conveyed into the cell indirectly by a process called **signal transduction**. One of two well-known pathways of signal transduction is shown in the left panel in Figure 12.18.

Second messenger-Mediated Signal Transduction by e.g., Hormones that Can't Get Into Cells:



1. A **Hormone** binds membrane **receptor**
2. A **hormone receptor** changes shape, binds **adenylate cyclase** enzyme
3. **cAMP** synthesized from **ATP**
4. **cAMP** activates a kinase...
5. leading to a **cascade of phosphorylation** with **cytoplasmic and/or nuclear** effects.

Fig. 12.18: Some signal molecules can't enter cells. They get their message into the cell by *signal transduction*. Here, a large or polar signal molecule (e.g., a hormone) binds a membrane receptor (1). The binding changes the shape of the receptor (2), which can then bind to a cell membrane protein (3), which becomes an activated enzyme. This enzyme catalyzes the formation of cAMP (the structure seen at the right). cAMP binds and activates the first in a series of protein kinases (4), starting a cascade of protein phosphorylations, a *phosphorylation cascade* (5) that leads to a cellular response, e.g., the transcription regulation shown here, or changes in cytoplasmic activities.

In this pathway, the binding of the hormone to a cell-surface receptor results in formation of **cAMP**, a **second messenger** (Figure 12.18, right panel). cAMP forms when the hormone-receptor in the membrane binds to and activates a membrane-bound **adenylate cyclase** enzyme. Once formed, cAMP binds to and activates a protein kinase to initiate a protein **phosphorylation cascade** in the cytoplasm.

The *phosphorylation cascade* is a series of consecutive protein phosphorylations, the last of which is often an activated transcription factor which enters the nucleus and bind to a *cis-regulatory DNA sequence* to turn a gene on or off or modulate its expression. As suggested in the illustration, that last phosphorylation can also lead to one of many different cytoplasmic responses.

Whether by regulating genes or their enzyme products, signal transduction regulate biochemical pathways to control cellular metabolism. cAMP was the first *second-messenger* metabolite to be discovered. When hormones bind to a membrane receptor that is itself a

protein kinase, hormone-receptor binding causes an allosteric change, activating the linked *receptor kinase*. This starts a phosphorylation cascade without a second messenger. The cascade ends with the activation of a transcription factor, as illustrated in Figure 12.19.

Signal Transduction by e.g., chemical messengers that can't get inside cells - Receptor is Enzyme

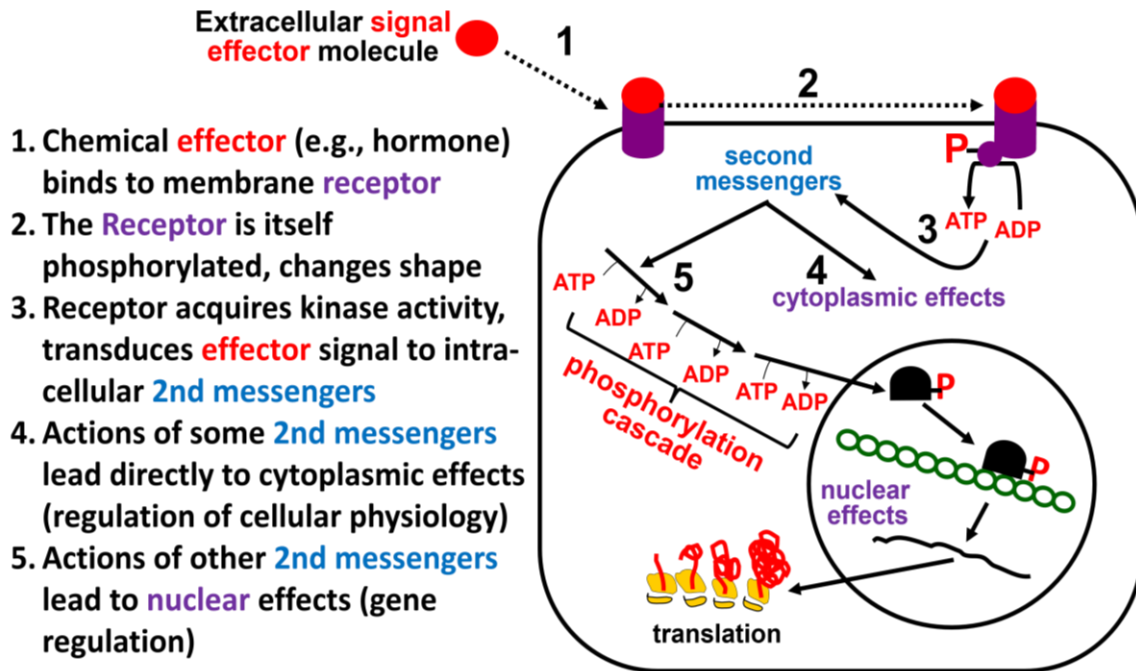


Fig. 12.19: *Enzyme-linked receptors* transduce hormonal signals directly, activating receptor kinases on the cytoplasmic cell surface, initiating a phosphorylation cascade, which lead to any of several responses.

We'll take a closer look at signal transduction in another chapter.



[227 Signal Transduction Can Lead to Gene Regulation](#)

12.6 Regulating Eukaryotic Gene Expression Means Contending with Chromatin

Consider again the different levels of chromatin structure, as seen in Figure 12.20. (below). Transcription factors bind specific DNA sequences by detecting them through the grooves (mainly the major groove) in the double helix. The drawing reminds us, however, that unlike the nearly naked DNA of bacteria, eukaryotic (nuclear) DNA is coated with proteins that, in aggregate, are greater in mass than the DNA that they cover. The protein-DNA complex of the genome is, of course, chromatin. Here's another reminder: DNA coated with histone proteins forms the 9 nm diameter *beads-on-a-string* structure in which the beads are the *nucleosomes*. The association of specific nonhistone proteins with the nucleosome "necklace" causes it to fold over on itself to form the *30 nm solenoid*.

Organization of DNA in Chromatin

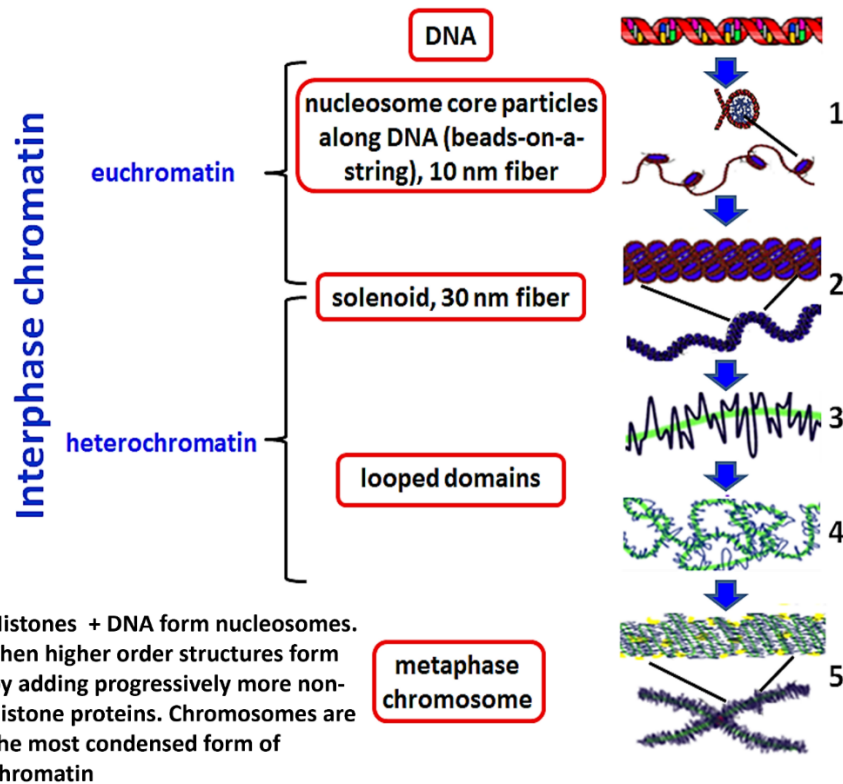


Fig. 12.20: Chromatin organization: different levels of chromatin structure result from differential association of DNA with histone and non-histone chromosomal proteins.

As we saw earlier, it is possible to selectively extract chromatin. Take a second look at the results of typical extractions of chromatin from isolated nuclei (Figure 12.21).

Chromatin Extraction

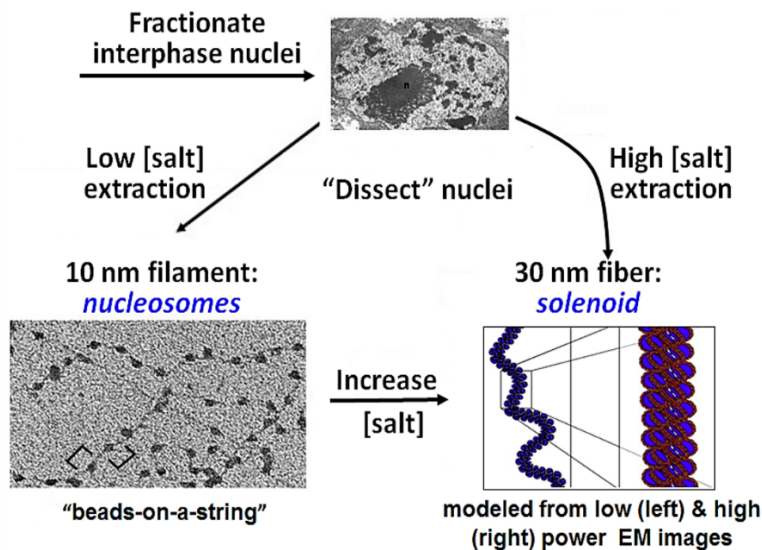


Fig. 12.21: High-salt chromatin extraction from nuclei, or high-salt treatment of 10 nm filaments, yields 30 nm *solenoid* structures, essentially coils of 10 nm filaments.

Further accretion of nonhistone proteins leads to more folding and the formation of the **euchromatin** and **heterochromatin** seen in nondividing cell nuclei. In dividing cells, the chromatin further *condenses* to form the **chromosomes** that separate during either *mitosis* or *meiosis*.

Recall that a biochemical analysis of the 10 nm filament extract revealed that the DNA wraps around octamers of histone-protein - that is, the nucleosomes (the beads) in this beads-on-a-string structure. Histone proteins are highly conserved in eukaryotic evolution; they are not found in prokaryotes. They are also very basic (having many *lysine* and *arginine* residues) and therefore very positively charged. This explains why they can arrange themselves uniformly along DNA, binding to the negatively charged *phosphodiester backbone* of DNA in the double helix.

Since the DNA in euchromatin is less tightly packed than it is in heterochromatin, perhaps active genes are to be found in euchromatin and *not* in heterochromatin. Experiments in which total nuclear chromatin extracts were isolated and treated with the enzyme deoxyribonuclease (DNase) revealed that the DNA in active genes was degraded more rapidly than nontranscribed DNA. More detail on these experiments can be found at the following two links.



[228 Question: Is Euchromatic DNA Transcribed?](#)



[229 Experiment and Answer: Euchromatin is Transcribed](#)



The results of such experiments are consistent with the suggestion that active genes are more accessible to DNase because they are in less-coiled or less-condensed chromatin. DNA in more-condensed chromatin is surrounded by more proteins; thus, it is less accessible to and protected from DNase attack. When packed up in chromosomes during mitosis or meiosis, all genes are largely inactive.

Regulating gene transcription must occur in non-dividing cells or during the interphase of cells, times when it is easier to silence or activate genes by **chromatin remodeling** (i.e., changing the shape of chromatin). Changing chromatin conformation involves enzyme-catalyzed modification of chromatin proteins and DNA. For example, chromatin can be modified by histone acetylation, and de-acetylation, methylation, and phosphorylation reactions, catalyzed by *histone acetyltransferases* (*HAT* enzymes), de-acetylases, methyl transferases and kinases, respectively.

Acetylation of lysines near the amino end of histones H2B and H4 tends to unwind nucleosomes and open the underlying DNA for transcription. De-acetylation promotes condensation of the chromatin in the affected regions of DNA. Likewise, the methylation of lysines or arginines (basic amino acids that characterize histones!) in H3 and H4 typically opens DNA for transcription, while demethylation has the opposite effect. These chemical modifications affect recruitment of other proteins that alter chromatin conformation and

ultimately activate or block transcription. This reversible acetylation and its effect on chromatin are illustrated in Figure 12.22 (below).

Chemical Modifications Reorganize Chromatin to Regulate Gene Expression

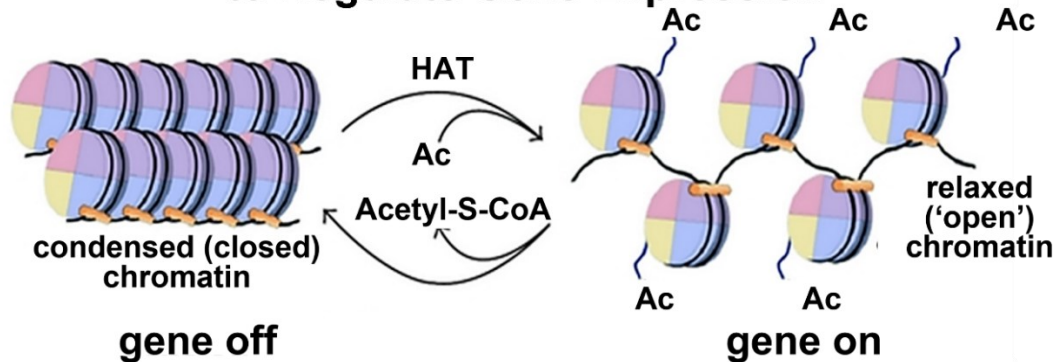


Fig. 12.22: The chemical modification of histones e.g., acetylation catalyzed by *HAT* enzymes (*Histone Acetyl-Transferases*) shown here, can change chromatin conformation, *opening* or *closing* it to transcription.

Epigenetic alterations can also account for changes in broader features of chromatin structure. For example, *topologically associated domains* (TADs) are megabase regions of DNA that associate with each other. The boundary DNA between TADs is methylated and a *CTCF* protein bound to those regions is thought to *insulate* the TADs from one another, minimizing interactions between TADs. CTCF is ubiquitous in the nucleus (on chromatin at least) and thought generally to mediate chromatin interactions at a distance. Disruption of this *insulation* is seen in cancer cells and can result from mutation (e.g., of CTCF) or epigenetic alteration. In fact, *hypermethylation* of the boundary DNA between TADs is associated with many cancers. For more details, check out ^{12.1}[Epigenetic Cancer Chromatin Signatures](#) (P. Pinoli et al., 2020).

12.7 Regulating all Genes on a Chromosome at Once

Recall that male cells have only one X chromosome and that one of the X chromosomes in human female adult somatic cells is inactivated, visible in the nucleus as a *Barr body*. Inactivation of one of the X chromosomes in *Drosophila* females. But both males and females require X chromosome gene expression during embryogenesis. At that time, given the difference in X chromosome *gene dosage* between males and females, how do male embryos get by with only on X chromosome? Experiments looking at gene dosage via the expression of X chromosome genes in male *and* female flies revealed similar levels of gene products. It turns out that a protein called **CLAMP** (Chromatin-Linked Adaptor for Male-specific lethal Protein) was shown to bind to **GAGA** nucleotide repeats lying between the genes for histones 3 and 4. There are about 100 repeats of the five-gene histone locus on X chromosomes, and so, about 100 GAGA repeats. All those CLAMP proteins bound to those GAGA repeats can interact with a nuclear body called **HLB** (Histone Locus Body). The HLB was previously implicated in higher levels of X chromosome gene expression. A recent study suggests that once bound to the GAGA sequences, CLAMP recruits **MSL** (Male-Specific-Lethal) proteins to form an **MSL complex**. This complex (the associated HLB/CLAMP/ MSL) has been shown to boost X chromosome gene transcription in male fruit flies. The resulting MSL protein complexes that form then globally increase X chromosome gene expression, compensating for the lower X

gene dosage in males. Read the original research at L. E. Reider et al. (2018) *Genes & Development* 31:1-15). And finally, there is emerging evidence that the HLB action may also be involved in inactivation of an entire female X chromosome later in embryogenesis in females!

12.8 Mechanoreceptors: Capturing Nonchemical Signals

A discussion of hormone and other signal molecule-based signal transduction is incomplete without acknowledging mechanoreceptors on cell surfaces that respond to mechanical stimuli, such as touch, pain, temperature, pressure, stretching, and even changes in posture. The latter are mediated by *proprioceptors* (*proprio*, "one's own self"), or self-awareness receptors. In joints, they respond to changes in posture to let us know where our arms and legs are at any given moment! **Piezoreceptors** (*piezo*, meaning "squeeze" or "press") respond to pressure or stretch to open a Ca^{++} channel, as illustrated in Figure 12.23 below.

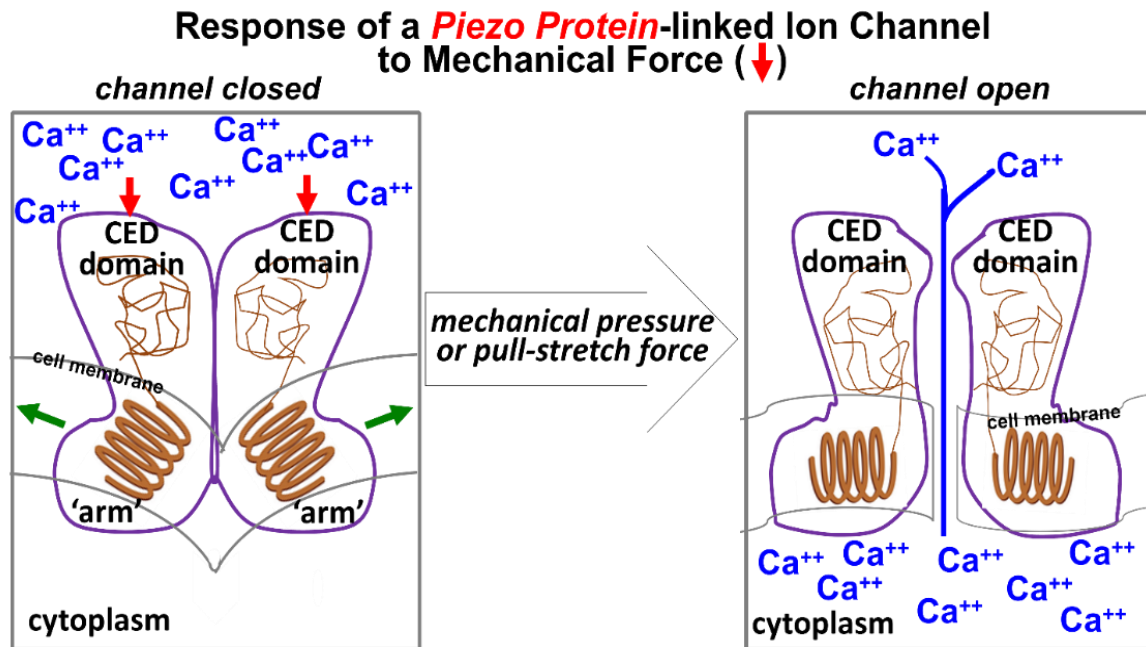


Fig. 12.23: Extracellular pressure on a *CED domain* (red arrows) and activation of spring-like arms stretch the cell membrane (green arrows) to opens a piezo-linked ion channel.

Piezoreceptors in cells lining the urinary bladder signal when the bladder is full, with the expected response. In blood vessel endothelial cells these alert the cells to changes in blood pressure and mediate appropriate responses. In sensory cells in the skin, piezoreceptors enable nerve-ending responses to touch or pain. The domain structure of one *piezoreceptor* protein was shown to include *springy* domains that could respond to mechanical forces. For more details check out ^{12.2}[About Piezoreceptors](#) (Lin et al., 2019). David Julius and Ardem Patapoutian shared the 2021 *Nobel Prize in Physiology or Medicine* for their mechanoreceptor research. Patapoutian discovered the *Piezo1* and *Piezo2* cation channels that mediate piezoreception, and the TRPM8 cation channel for cold perception. Julius identified the *TRPV1* ion channel for heat (e.g., hot pepper!) perception and independently discovered TRPM8.

12.9 Epigenetics

Aristotle thought that an embryo emerged from an amorphous mass, a “less fully concocted seed with a nutritive soul and all bodily parts.” The much-later development of the microscope led to more detailed (if inaccurate) descriptions of embryonic development. In 1677, no less a luminary than Anton van Leeuwenhoek (who first saw and described pond-water protozoa and other animalcules), looked at a human sperm with his microscope and thought he saw a miniature human inside! The tiny human, or *homunculus*, became the epitome of *preformation* theory.

William Harvey, also in the 17th century, described morphological changes in the developing embryos of chickens (and other animals). Harvey coined the term *epigenesis* to counter the notion that tiny adult structures in eggs or sperm simply grew bigger during embryonic gestation. Other experiments had led embryologists to conclude that the physical and chemical environment of an embryo strongly affected development. Thus temperature, pH, and in the case of chicken eggs, the position of incubation can affect embryonic development.

In a series of elegant experiments Hans Speeman reported in 1924 that cells associated with differentiation of one region of an embryo could be transplanted to a different part of the same embryo, or to another embryo entirely, where it would induce new tissue development. He won the 1935 Nobel Prize in Physiology and Medicine for his discovery of embryonic *organizers* that induced morphogenesis. Other embryologists showed that cells killed by freezing or boiling still induced morphogenesis after being placed on an embryo. This clearly implicated a role for actual chemicals in embryogenesis. The fact that differences in physical or chemical environment could affect embryonic development led many to conclude that environment played the dominant role and that genes played only a minor one in an organism's ultimate phenotype. Unlike most of his fellow embryologists, Conrad Waddington believed in a more equitable role of genes and environment in determining phenotype. Adapting the term epigenesis, he coined the term *epigenetics* to describe the impact of environment on embryonic development (1942, *The Epigenotype. Endeavour. 1*: 18–20).

At the time, the concept of epigenetics led to the *nature* vs. *nurture* controversy. On the nature side, inheritance was mainly genetic, while the nurture side gave a dominant role to environmental chemistry. We now know that environmental differences can and do cause individuals with the same genes (genotype) to vary in appearance (phenotype). The modern version of the *nature* vs. *nurture* argument has more to do with complex traits. For example, how much do genetics vs environment influence intelligence, psychology, and behavior. There is much to-do and little evidence to resolve these questions, and likely too many factors affecting these traits to separate them experimentally. Today, epigenetic (*epi* meaning ‘over’ or above) studies look at protein interactions in eukaryotes affecting gene expression, in other words at interactions superimposed on genes. These interactions change the structure NOT of genes or DNA, but of the proteins (and other molecules) that affect how and when genes are expressed. As we've seen, control of transcription involves transcription factors that recognize and bind to regulatory DNA sequences such as enhancers or silencers. These protein-DNA interactions often require selective structural changes in chromatin conformation around genes. These changes can be profound and stable, and they are not easily undone.

12.9.1 Epigenetic Inheritance in Somatic Cells

Examples of somatic cell epigenetics include the inheritance of chromatin protein alterations that accompany changes in gene expression that occur in development. Given the right signal, say a hormone at the right time, a few cells respond with chromatin rearrangements and new patterns of gene expression that can define a cell as *differentiated*. Hundreds, even thousands of changes to chromatin and gene expression accompany progress from fertilized egg to fully mature organism. Every one of these changes in a cell is passed on to future cell generations of cells by mitosis, accounting for the correct formation different tissues and organs in the organism. In this way, the many different *epigenomes* representing our differentiated cells are *heritable*. Somatic cell epigenetics is therefore the study of when and how undifferentiated cells (embryonic and later, adult stem cells) acquire their differentiated characteristics and how they can then pass on this information to progeny cells. As we'll see shortly, epigenetic inheritance is not limited to somatic cells, but can span generations!

To help us understand this new aspect of inheritance and evolution, consider Jean-Baptiste Lamarck's belief that the ancestors of giraffes had short necks but evolved longer and longer necks because longer necks enabled them to reach food higher up in the trees. That new character would be inherited by the next giraffe generation. According to Lamarck, evolution was *purposeful*, with the goal of improvement. Later, Charles Darwin published his ideas about evolution by *natural selection*, where nature selects from preexisting traits in individuals (the raw material of evolution). The individual that just randomly happens to have a useful trait then has a survival (reproductive) edge in an altered environment. Evolution is thus *not purposeful*. Later still, with the rediscovery of Mendel's genetic experiments, it became clear that it's an organism's genes that are inherited, are passed down the generations, and are the basis of an organism's traits. By the start of the twentieth century, Lamarck's notion of purposefully acquired characters was discarded.

With this brief summary, let's look at epigenetic inheritance across generations of sexually reproductive species.

12.9.2 Epigenetic Inheritance in the Germline

Epigenetic inheritance implies an epigenetic blueprint in addition to our DNA blueprint. This means that in addition to the passing on of genes from male and female parents, epigenomic characteristics (which genes are expressed and when) are also passed on from one generation to the next. Waddington suspected as much early on, calling the phenomenon *genetic assimilation*, and once again he created controversy! Does genetic assimilation make Lamarck right after all? Prominent developmental biologists accused Waddington of promoting Lamarck's notions of *purposeful evolution* (the idea that what is best or necessary for the survival of an organism leads it to evolve a new gene for the purpose). Waddington and others denied the accusation, trying to explain how epigenetic information might be heritable, without leading to *purposeful* evolution. So, is there, in fact, an *epigenetic code*? Let's look at data from the small Swedish town of Överkalix that led to renewed interest in epigenetic phenomena. Consider the meticulous harvest, birth, illness, death and other demographic and health records collected and analyzed by L. O. Bygren and colleagues at Sweden's Karolinska Institute. A sample of Bygren's data is illustrated below in Figure 12.24.

Grandparent	Food Supply	Grandson relative risk of death from cardiac disease & diabetes	Granddaughter relative risk of death from cardiac disease & diabetes
Grandfather	poor	-35%	no change
Grandfather	abundant	+67%	no change
Grandmother	poor	no change	-49%
Grandmother	abundant	no change	+113%

It's as if the environment was indeed causing an *acquired change* in the grandparent that is passed not to one, but through two generations..., and in a sex-specific way!

Fig. 12.24: Summary of health, birth, death, and demographic records for the town of Överkalix.

It looked to the good doctor as if environment were influencing germline inheritance, as if the environment were indeed causing an *acquired change* in the grandparent that is passed not to one, but through two generations—and in a sex-specific way!



[230 Epigenetic Inheritance: First Inkling](#)

This phenomenon was subsequently demonstrated experimentally. For example, rat pups born to rats exposed to a toxin while pregnant, suffered a variety of illnesses. This might be expected if the toxic effects on the mother were visited on the developing pups (e.g., through the placenta). However, when the diseased male rat pups matured and mated with females, the pups in the new litter grew up suffering the same maladies as the male parent. This was the case *even though the pregnant females were not exposed to the toxins*. Because the original female was already pregnant when she was exposed, the germline cells (eggs and sperm) of her litter had not suffered mutations in utero. This could only mean that the epigenetic patterns of gene expression caused by the toxin in pup germline cells (those destined to become sperm and eggs) in utero were retained during growth to sexual maturity, to be passed on to their progeny even while gestating in a normal, unexposed female.



[231-2 Experimental Demonstration of Germ-Line Epigenetic Inheritance](#)

For some interesting experimental findings on how diet influences epigenetic change in *Drosophila*, look at ^{12.3}[Diet and a Heritable Fly Epigenome](#). For an amusing (but accurate) take on epigenetics, check out the YouTube at ^{12.4}[Epigenetics Explained!](#). So, given the reality of

epigenetic inheritance, let's consider a most intriguing question: Is epigenetic inheritance (like Mendelian genetic inheritance) the result of natural selection?

Several studies have correlated adverse conditions during pregnancy in animals (including humans) with high incidences of health, behavioral and other anomalies in the adults. This was the case in Holland for children and grandchildren of mothers pregnant during the famine of the last winter of WWII, a season that became known as the ***Dutch Hunger Winter of 1944-1945***. As adults, these descendants had higher incidences heart disease, obesity, and diabetes. Such correlations were attributed to epigenetic changes in embryonic DNA *in utero* (read more at ^{12.5}[Dutch Grandkids After the Hunger Winter](#)).

Here is the question that should occur to us: Why should so many survivors of life in an undernourished womb have suffered the same epigenetic changes at the same time? A study of methylation patterns in specific DNA regions in survivors of *Dutch Hunger Winter* pregnancies and a control group that missed the famine suggests an intriguing answer. The study found unique DNA methylation patterns in the surviving adults. To explain the survival of so many newborns that then suffered adult health anomalies, it was suggested that epigenetic modifications occur at random in embryos..., and then that some of these (i.e., the ones that *marked* the suffering adults) were selected during the pregnancy because they actually conferred a survival advantage to embryos (see a more complete discussion at ^{12.6}[Natural Selection in utero](#)).

Did the many cases of epigenetic inheritance by large cohorts of individuals result from survival advantages of ***epigenetic natural selection***? In other words, could epigenetic inheritance be subject to natural selection in the same way as genetic inheritance..., with more complex results?

These days, the term epigenetics describes how heritable change in chromatin modification (chromatin remodeling) affects gene expression. We now know that epigenetic configurations of chromatin that are most stable include modifications such as *acetylation*, *phosphorylation*, *methylation* of histones and *methylation and phosphorylation of DNA*. Such changes can convert the 30nm fiber to the 10 nm 'beads-on-a-string' nucleosome necklace... and *vice versa*. Chromatin remodeling can also lead to altered patterns of gene expression, whether during normal development or when caused by environmental factors such as abundance or limits on nutrition, toxins/poisons, or other life-style choices. The active study of DNA methylation patterns even has its own name, *methyloomics*! Check out ^{12.7}[Epigenetics Nomenclature](#) for more epigenetic nomenclature.

Let's close this chapter with some socially, culturally, and personally relevant observations and questions.

If environmental causes of epigenetic change persist in a unique or isolated location or demographic, could the natural selection of epigenetic traits result in geographic or demographic differences in epigenetic characters just as it has done for genetic characters? Can we trace an epigenetic ancestry in the same way as our genetic one? That seems to be the case from recent comparisons of methylation patterns of DNA between Mexican and Puerto Ricans of Hispanic origin. DNA regions known to be sensitive to diesel emissions or intra-

uterine exposure to tobacco (and even social stressors) were differentially methylated in the two groups. Read more about this at ^{12.8}[Heritable Effects of Environmental Exposure](#)

Now think about the epigenetic consequences of smog that Los Angelinos endured and of lead toxicity to folks drinking lead-laden water in Flint Michigan (especially in children) the epigenetic effects of which remain to be seen. As you can see, epigenetics remains a hot topic, with much more yet to be concerned about! If environmental causes of epigenetic change persist in a unique or isolated location or demographic, could the natural selection of epigenetic traits result in geographic or demographic differences in epigenetic characters, just as it has done for genetic characters?

Can we trace an epigenetic ancestry in the same way as our genetic one? That seems to be the case in a recent comparison of DNA methylation patterns between Mexicans and Puerto Ricans of Hispanic origin. The study revealed that DNA regions known to be sensitive to diesel emissions or intra-uterine exposure to tobacco (and even social stressors) were differentially methylated in the two groups. See details of this study at ^{12.9}[Epigenetics of Culture and Ethnicity](#).

Finally, thinking about epigenetic selection of clearly harmful traits, can we start to see them as analogous to the genetic selection of, say, β -globin gene mutations for sickle cell anemia that confer a survival advantage in one context (the threat of malarial parasites in mosquitoes), but that are harmful in another?

Here are some bottom lines that you cannot avoid!

Can you be sure that your smoking habit *will not* affect the health of your children or grandchildren? What about eating habits? Drinking? *It is not just a little* scary to know that I have a gullible germline epigenome that can be influenced by *my* behavior, good and bad, and that my children (and maybe grandchildren) will inherit my epigenetic legacy long before they get my house and money.

And even that may not be the limit; epigenetic memory in *Caenorhabditis elegans* can stretch to as many as 24 generations! (^{12.10}[Epigenetic Memory in *C. elegans*](#)) Also, read about epigenetic inheritance resulting from Dad's cocaine use ^{12.11}[Sins of the Fathers](#).

Some iText & VOP Key Words and Terms

10 nm fiber	fully differentiated cells	miRNA (micro RNA)
3' nontranscribed DNA	galactose	negative regulation
30 nm solenoid fiber	galactoside transacetylase	PTS
5' nontranscribed DNA	gene activation	regulatory genes
adaptive immune system	gene derepression	second messenger
adult stem cells	gene expression	nucleosomes
allolactose	gene induction	O ₁ and O ₂ lac operators
antisense RNA	gene regulation	operator
basic vs nonbasic proteins	gene repression	operon regulation
beads-on-a-string	HAT enzymes	PEP-dependent P-transferase
β -galactosidase	helix-turn-helix motif	phage DNA
cAMP	heterochromatin	phosphodiester backbone

cAMP-receptor protein	histone acetylation	phosphorylation cascade
CAP protein	histone kinases	pluripotent cells
CAT box	histone-methyl transferases	positive regulation
catabolite-activator protein	histone methylation	promoter
chromatin remodeling	histone phosphorylation	proximal regulatory element
cis-acting elements	housekeeping genes	signal transduction
condensed chromatin	inducer exclusion	steroid-hormone receptors
developmental program	interphase	steroid hormones
differential gene expression	introns	structural genes
distal regulatory element	lac operon	TATA box
DNA bending	lacI gene	tetrameric-lac repressor
DNase	lactose	totipotent cells
embryonic stem cells	lactose permease	transcription factors
enhancers	lactose repressor	transcription regulation
environmental signals	lacZ, lacY, and lacA genes	transcription start site
epigenome	leucine-zipper motif	trp operon
euchromatin	levels of chromatin structure	trp repressor
exons	major groove	zinc-finger motif
extended chromatin	minor groove	

CHAPTER 12 WEB LINKS



12.13



12.14



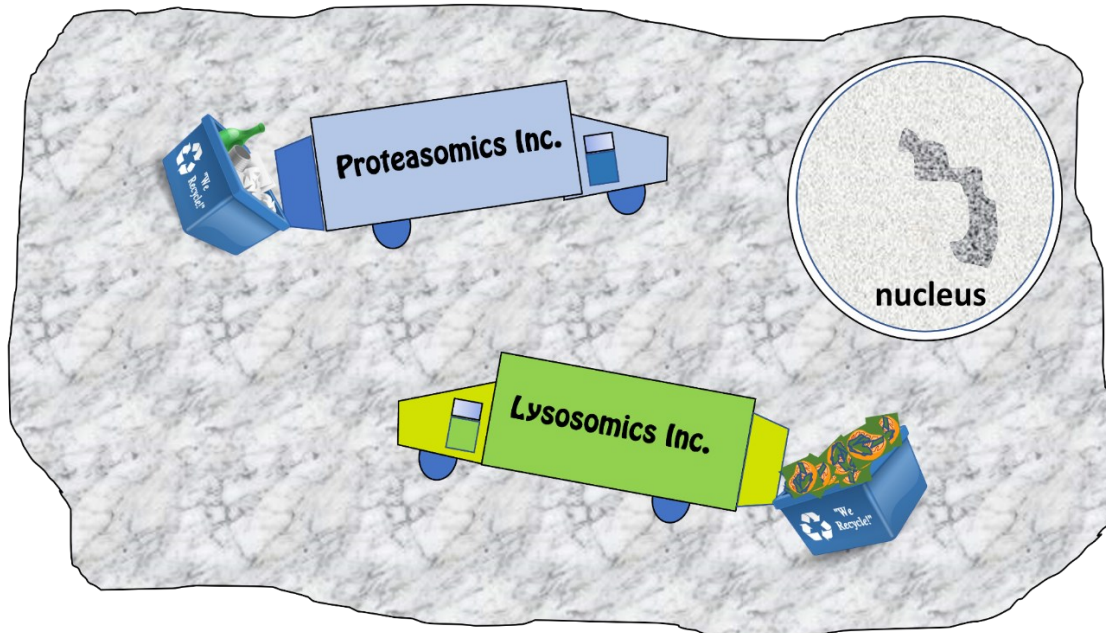
Chapter 13:

Posttranscriptional Regulation of Gene Expression

Regulating gene expression with short and long noncoding RNAs, CRISPR-Cas, and protein phosphorylation; controlling gene expression by regulating translation

Reminder: For inactive links, google key words/terms for alternative resources.

SOMICS – The Art of Cellular Waste Management



Recyclers of Almost Everything!

13.1 Introduction

The metabolic potential of cells is flexible, depending on various mechanisms that ultimately determine the levels and activities of the proteins that dictate a cell's metabolic state. We have seen some of these regulatory mechanisms:

- Extracellular chemical signals or developmental chemical prompts that regulate transcription.
- *Allosteric regulation* by cellular metabolites and chemical modification, (e.g., by *phosphorylation* or *dephosphorylation*) can control the activity of enzymes or other proteins.

In this chapter, we look at different kinds of ***posttranscriptional regulation***, events that take place sometime between mRNA transcription and the maturation of the biological protein activity of a protein. These control mechanisms are most diverse in eukaryotes. Like other pathways for regulating gene expression, posttranscriptional regulation begins with extracellular chemical signaling. Responses include changes in the rate of polypeptide translation and changes in macromolecular ***turnover rate***, such as changes in the ***half-life*** of specific RNAs and proteins in cells.

Regardless of mechanism, each up- or down-regulation of (i.e., increase or decrease in) gene expression contributes to changes in the **steady state** of a particular RNA or protein required for proper cell function. In considering *posttranscriptional regulation*, we will see how cells use specific proteins and different noncoding RNA transcripts to target unwanted proteins or RNAs for degradation.

Learning Objectives

When you have mastered the information in this chapter, you should be able to do the following:

1. Explain what it is about *C. elegans* that makes it a model organism for studying the development and regulation of gene expression.
2. Compare and contrast the origins and functions of *miRNA* and *siRNA*.
3. Search for examples of the *miRNAs*, *siRNAs*, lncRNAs, and circRNAs that regulate the expression of specific genes, and explain their mechanisms.
4. Explain how a *riboswitch* functions to control bacterial gene expression.
5. Explain the evolutionary origins and roles of the bacterial *CRISPR-Cas immune* system components.
6. Explain how *eif2* activity is modulated to coordinate a red cell's heme and globin levels.
7. Describe how eukaryotic cells degrade unwanted proteins and speculate on how bacteria might do so.
8. Answer the question, "How did *junk DNA* arise?"
9. Explain ways in which so-called "junk DNA" has value.

13.2 Posttranscriptional Control of Gene Expression

Not too long ago, we thought that very little of the eukaryotic genome was ever transcribed. We also thought that the only noncoding RNAs were tRNAs and rRNAs. Now we know that other RNAs play roles in gene regulation and in the degradation of *spent* cellular DNA or unwanted foreign DNA. These are discussed in detail next.

13.2.1 Riboswitches

The *riboswitch* is a bacterial transcription mechanism for regulating gene expression. While this mechanism is not specifically posttranscriptional, it is included here because the action occurs after transcription initiation and aborts the completion of an mRNA. When the mRNA for an enzyme in the guanine synthesis pathway is transcribed, it folds into stem-and-loop structures. Enzyme synthesis will continue for as long as the cell needs to make guanine. But if guanine accumulates in the cell, excess guanine will bind stem-loop elements near the 5' end of the mRNA, causing the RNA polymerase and the partially completed mRNA to dissociate from the DNA, prematurely ending transcription.

The ability to form folded stem-loop structures at the 5' ends of bacterial mRNAs seems to have allowed the evolution of translation-regulation strategies. Guanine interaction with the stem-loop structure of an emerging 5' mRNA can abort its own transcription; similarly, small metabolite/mRNA and even protein/mRNA interactions can also regulate (in this case prevent) translation. The basis of guanine riboswitch-regulation of the expression of a guanine-synthesis pathway enzyme is shown below in Figure 13.1.

A Bacterial Guanine Microswitch

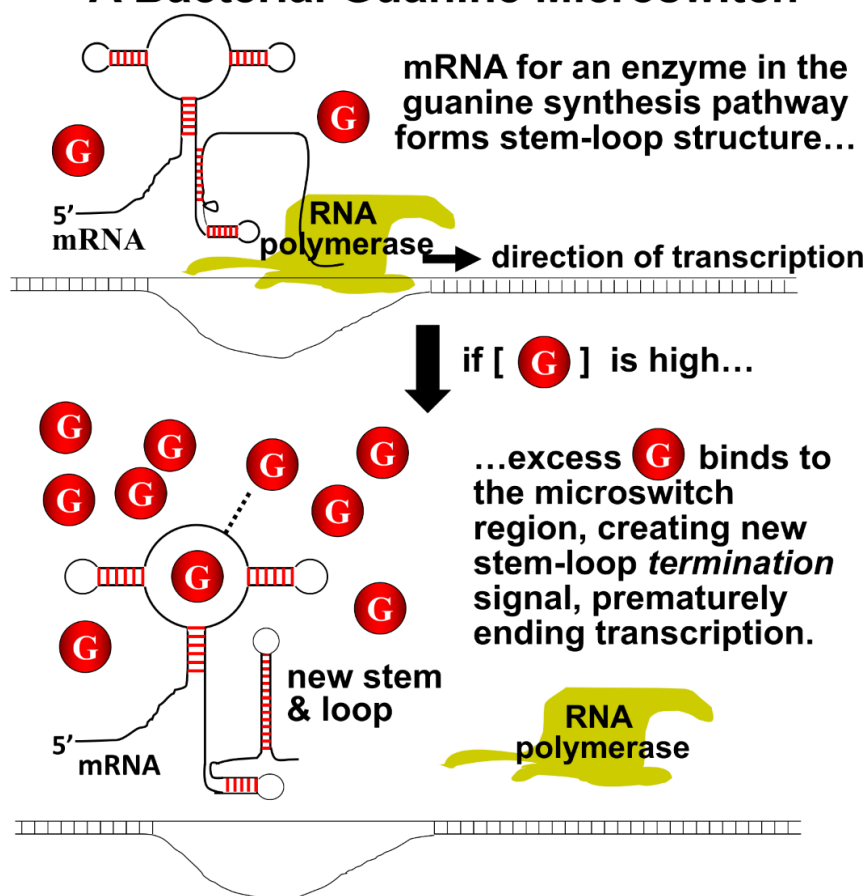


Fig. 13.1: Bacterial guanine riboswitch: The “switch” is the action of an mRNA encoding an enzyme in the guanine-synthesis pathway. Excess guanine binds and distorts the stem-loop structure of the mRNA, causing termination of further gene transcription.

As we will see shortly, 5' mRNA folded structures also play a role in eukaryotic translation regulation.



[232 Riboswitches Interrupt Bacterial Transcription](#)



[233 Small Metabolites Also Regulate Bacterial mRNA Translation](#)



13.2.2 CRISPR/Cas: A Prokaryotic Adaptive Immune System

In higher organisms, the *immune system* is *adaptive*. It remembers prior exposure to a pathogen and can thus mount a response to a second exposure to the same pathogen. The discovery of an “*adaptive immune system*” in many prokaryotes was therefore something of a surprise. The CRISPR story begins in 1987 in Japan, with the discovery by Y. Ishino (et al.) of a cluster of regularly interspaced short palindromic DNA sequences in *E. coli*. Clustered DNA repeats

were known of course but they were always consecutive and uninterrupted by other sequences (think rRNA genes). Ishino's cluster was at the time of unknown origin and function.



[233a Discovery of the CRISPR Phenomenon](#)



In 1993 similar clustered repeats were found in several strains of *Mycobacterium tuberculosis* (the cause of tuberculosis), but the interspaced sequences interrupting the palindromic repeats differed between the strains. By 2001 these clustered, interrupted repeats, by then found in many prokaryotes (both bacterial and archaeobacterial), were named **CRISPR** (**C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats). The discovery of an array of **Cas**-gene repeats was linked to some of the CRISPR repeats in 2002. This was followed in 2005 by the revelation that some of the interspaced sequences were derived from nonhost origins, specifically phage DNA and plasmids, and in the same year, the protein encoded by the **Cas** gene was shown to have helicase and nuclease activity. This led to a suggestion that CRISPR and **Cas** were part of a bacterial *adaptive immune system*, a bacterial response to phage infection or other invasive DNA and (for the lucky bacterial survivor), a protective mechanism against a second infection! In 2012 Jennifer Doudna and Emmanuelle Charpentier realized that the CRISPR/Cas9 system in streptococcal bacteria might be able to rapidly and accurately edit any DNA sequence, including genomic DNA..., and then proceeded to prove it! Let's look more closely at CRISPR and Cas DNAs to see how they work as an adaptive immune system and how they can be used to edit DNA.

CRISPR RNAs are derived from phage transcripts that have interacted with CRISPR-associated (**Cas**) proteins. They make up the **CRISPR/Cas** system, which seems to have evolved to fight off viral infections by targeting phage DNA for destruction. When viral DNA gets into a cell during a phage infection, it can generate a CRISPR/Cas-gene array in the bacterial genome, with **spacer** DNA sequences separating repeats of the CRISPR genes. These remnants of a phage infection are the *memory* of this *prokaryotic immune system*. When a phage attempts to re-infect a previously exposed cell, **spacer RNAs** and **Cas** genes are transcribed. After **Cas** mRNA translation, the **Cas** protein and **spacer RNAs** will engage and target the incoming phage DNA for destruction to prevent infection. CRISPR/Cas systems (there are more than one!) *remember* prior phage attacks and transmit that memory to progeny cells. The CRISPR/Cas9 system in *Streptococcus pyogenes* is one of the simplest of the CRISPR/Cas immune defense systems (Figure 13.2).

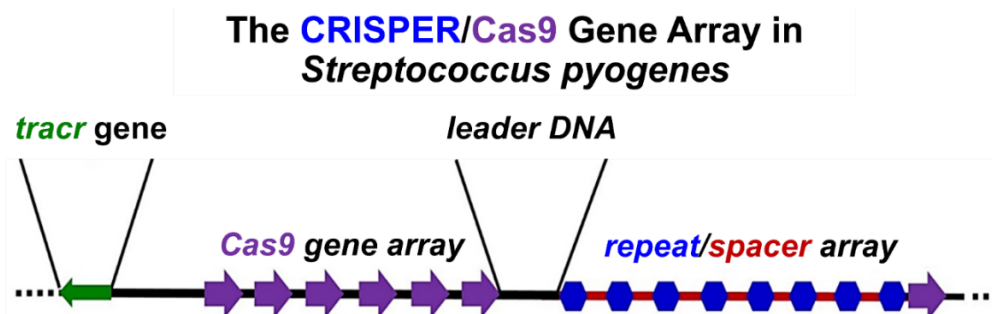


Fig. 13.2: The array of CRISPR/Cas9 genes in *S. pyogenes*, including a *tracr* gene, multiple *Cas9* genes, CRISPR DNA repeats separated by phage-derived spacer DNA, and a last *Cas9* gene. The phage-derived spacer DNAs are relics from the survival of an earlier phage infection.

The CRISPR/Cas gene array consists of the following components:

- **Cas**: Genes native to host cells
- **CRISPR**: Twenty-four to forty-eight base-pair repeats native to host cells
- **Spacer DNA**: DNA between CRISPR repeats—typically, phage DNA from prior phage infection or plasmid transformation
- **Leader DNA**: DNA containing a promoter for the transcription of CRISPR/spacer RNA
- **tracr gene**: Gene encoding a transcription-activator (tracr) RNA (but not included in all systems)

Let's look at CRISPR/Cas in action.

13.2.2.a The CRISPR/Cas “Immune” Response

Consider the mechanism of action of this prokaryotic immune system. The action begins when infectious phage DNA gets into the cell (Figure 13.3).

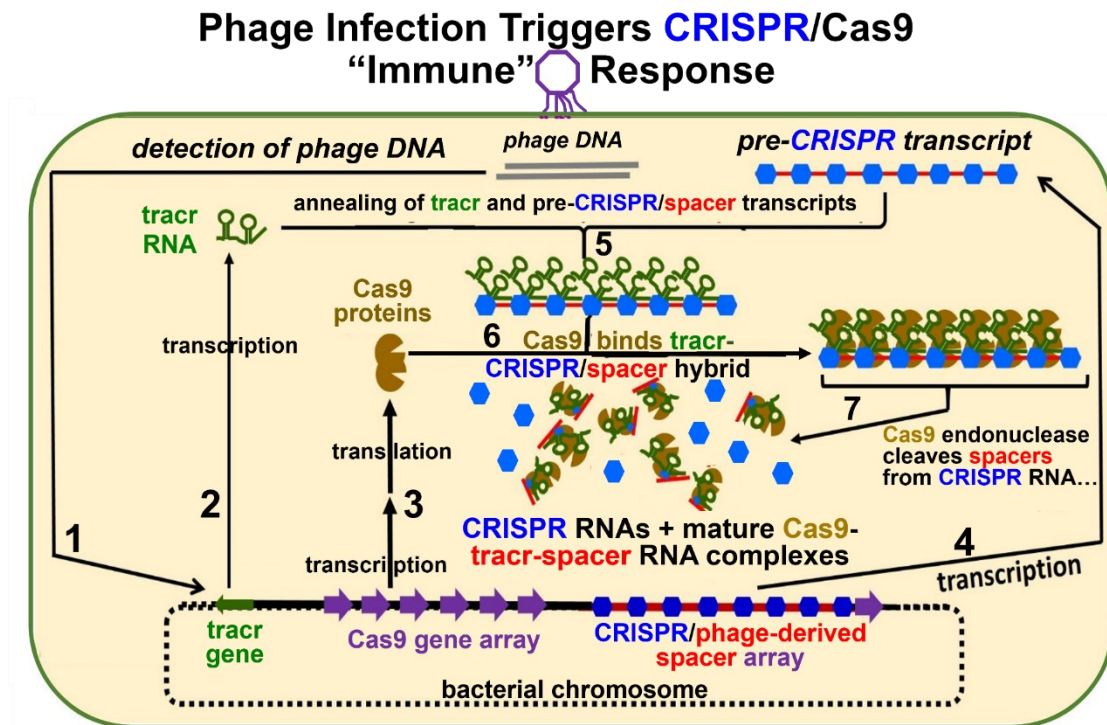


Fig. 13.3: Activation of the CRISPR/Cas9 array

Let's summarize what has happened here:

- Incoming phage DNA was detected after phage infection (1).
- The *tracr* (2) and *Cas9* (3) genes are transcribed, along with the CRISPR/phage-derived spacer region (4). *Cas9* mRNAs are then translated to make the *Cas9* protein. Remember, the spacer DNAs in the CRISPR region are the legacy of a prior phage infection.
- CRISPR/phage-derived spacer transcripts forms H-bonds with a complementary region of the *tracr* RNA (5) and the hybridized RNAs associate with *Cas9* proteins (6).
- *Cas9* endonucleases hydrolyze spacer RNA from CRISPR RNA sequences (7). The spacer RNAs remain associated with the complex, while the actual, imperfectly palindromic CRISPR sequences (shown in blue in the illustration) fall off.

In the next steps we'll look at how *phage-derived spacer DNA* from a prior infection protect progeny bacteria from a new phage infection. When transcribed, these spacer RNAs are now called **guide RNAs (gRNAs)**. They will *guide* mature Cas9/tracrRNA/spacer RNA complexes to the new incoming phage DNA. Figure 13.4 illustrates these events and subsequent events.

Upon Infection, Guide RNA (gRNA) in Mature Cas9-tracrRNA-Spacer RNA Complexes Inactivate Phage DNA

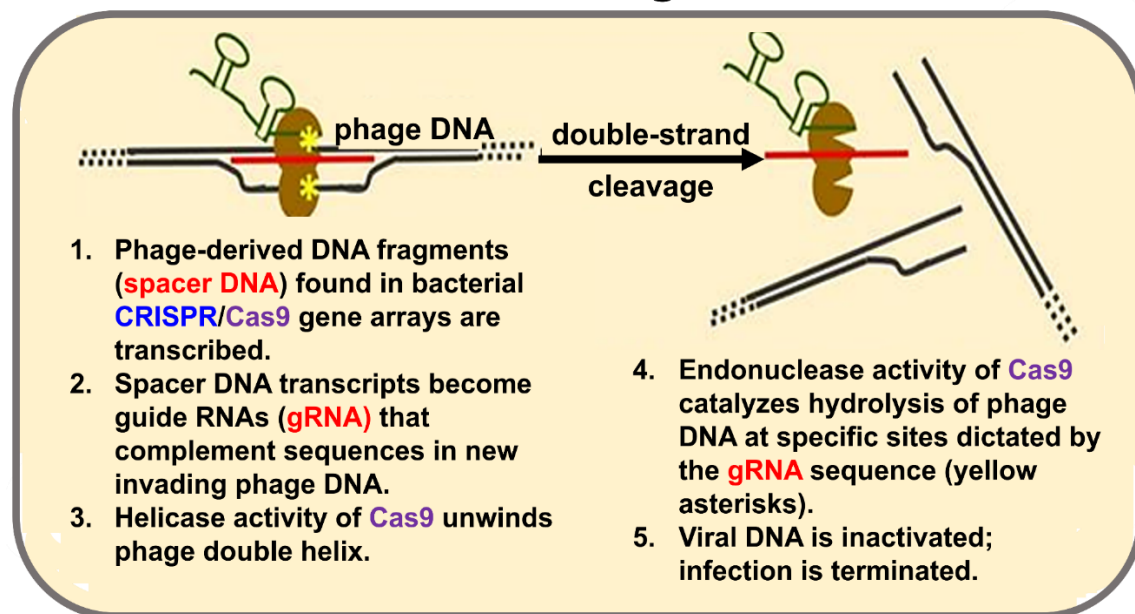


Fig. 13.4: Phage spacers in a CRISPR/Cas9 array (derived from an earlier phage infection) protect against phage re-infection.

Once again, let's summarize:

- a) *Spacer RNA* (i.e., *gRNA*) in the complex targets *incoming phage DNA*.
- b) *Cas helicase* unwinds the *incoming phage DNA* at complementary regions.
- c) *gRNA* H-bonds to the *incoming phage DNA*.
- d) *Cas endonucleases* create a double-stranded break (hydrolytic cleavage) *at specific sites* in the *incoming phage DNA*. Because precise cleavage is guided by RNA molecules, CRISPR/Cas endonucleases are classified as *type V restriction enzymes*.
- e) The *incoming phage DNA* is destroyed, and a new phage infection is aborted.

Check out ^{13.1}[More About CRISPR](#) to learn more about how bacteria acquire spacer DNAs (and therefore how this primitive adaptive immune system “remembers”) in the first place.

13.2.2.b Using CRISPR/Cas to Edit/Engineer Genes

Early studies demonstrated the reproducible cleavage of incoming phage DNA at specific nucleotides. Several labs quickly realized that it might be possible to adapt the system to cut DNA at any specific nucleotide in a target DNA!

It has turned out that the system works both in vivo and in vitro, allowing virtually unlimited potential for editing genes and RNAs in a test tube—or in *any* cell. Here is the basic process:

- a) Engineer *gDNA* with a Cas-specific DNA sequence that targets a desired target in genomic DNA.
- b) Fuse the *gDNA* to *tracr DNA* to make a *single guide DNA (sgDNA)*, so that it can be made as a single guide transcript (*sgRNA*).
- c) Engineer a *CRISPR/Cas9* gene array that substitutes this *sgDNA* for its original *spacer* DNAs.
- d) Place the engineered array into a plasmid next to *regulated promoters*.
- e) *Transform* cells by "*electroporation*"—which works for almost any cell type!
- f) *Activate the promoter* to transcribe the CRISPR/Cas9 genes....

For demonstrating the power of CRISPR-Cas precision gene editing, Jennifer Doudna and Emmanuelle Charpentier earned the 2020 Nobel Prize for Chemistry. The applications are indeed powerful... and controversial!

13.2.2.c CRISPR—The Power and the Controversy

The application of gene editing with CRISPR/Cas systems has already facilitated studies of gene function in vitro, in cells and in whole organisms. For a description of CRISPR/Cas applications already on the market see ^{13.2}[CRISPR Apps from NEB](#). The efficiency of specific gene editing using CRISPR/Cas systems holds great promise for understanding basic gene structure and function, for determining the genetic basis of disease, and for accelerating the search for gene therapies.

Here are a few examples of how CRISPR/Cas editing is being applied.

- One can now engineer an sgRNA with mutations that target specific sites almost anywhere in chromosomal DNA. The sgRNA is cloned into the CRISPR/Cas9 array on a plasmid. After transformation of appropriate cells, the engineered CRISPR/Cas9 forms a complex with target-DNA sequences. After both strands of the target DNA have been nicked, DNA repair can insert the mutated guide sequences into the target DNA. The result is loss or acquisition of DNA sequences at *specific, exact sites*, or *precision gene editing*. The ability to do this in living cells has excited scientists in both the basic and clinical research communities.
- Before transforming cells, one can engineer the *CRISPR/Cas9* gene array onto the plasmid to eliminate both *endonuclease* activities from the Cas protein. Upon transcription of the array in transformed cells, the *CRISPR/Cas9-sgRNA* still finds an *sgRNA*-targeted gene. However, lacking Cas protein endonuclease activities, the complex that forms just sits there *blocking transcription*. This technique is sometimes referred to as *CRISPRi* (*CRISPR interference*), by analogy to *RNA interference* or *RNAi* (see section 13.2.3.a below). Applied to organisms (and not just in vitro or to cells), it mimics the much more difficult *knockout*-mutation experiments that have been used in studies of the behavior of cells or organisms that have been made unable to express a specific protein.

Several newer, highly efficient CRISPR/Cas systems are exciting prospects for rapid, targeted gene therapies to fight disease and their potential to alter entire populations using **gene drive** technologies (see ^{13.3}[Gene Drive](#)) to insert modified genes into germline cells of a target organism. Gene drive could eliminate pesticide resistance in unwanted insects and herbicide resistance in undesirable plants. It could even erase whole mosquito or invasive species populations. For a recent *proof-of-concept* research, see Kryou et al. (2018) *A CRISPR-Cas9 Gene Drive Targeting doublesex Causes Complete Population Suppression In Caged Anopheles gambiae Mosquitoes*. Nat. Biotech. 1062-1066. But the use of CRISPR-based *gene drive* to control populations is also controversial (see J. Adler, (2016) *A World Without Mosquitoes*. Smithsonian, 47, 36-42).

- It is possible to edit out, i.e., delete an entire chromosome from cells. This bit of “global” genetic engineering depends on our ability to identify multiple unique sequences on a single chromosome and then to target these sites for CRISPR/Cas treatment. When the system is activated, the chromosome is cut at those sites, fragmenting it beyond the capacity of DNA-repair mechanisms to fix the situation. Imagine the possibilities for therapeutically removing from a human genome, **aneuploidies**—for example, extra chromosomes such as those found in **trisomy 21** (the cause of *Down syndrome*) or those correlated with some cancers. See ^{13.4}[CRISPR/Cas9 Deletes Whole Chromosome](#) to learn more.
- For yet another twist on CRISPR power, check ^{13.5}[CRISPR-Based Gene ID in Fixed Cell Chromatin](#) to read about a technique to identify and to locate specific DNA sequences in cells without disturbing chromatin structure.

As with any powerful technique (or powerful person for that matter!), behind the power as you might guess, lurks a dark side. The “dark side” of the CRISPR/Cas (and the related newer gene-editing technologies) was quickly recognized by its practitioners, most notably Jennifer Doudna herself. If for no other reason than its efficiency and simplicity, applications of CRISPR for precision gene editing has raised ethical issues. Clearly the potential exists for abuse or even for use with no beneficial purpose at all, and the availability of CRISPR kits only intensifies the concerns. We can be hopeful that, as in all discussions of biological ethics, scientists are very much engaged in the conversation. Listen to Doudna talk about CRISPR and its potential to do good in the world and its potential to be abused by the foolish and unscrupulous (^{13.6}[Dr. Doudna Talks About CRISPR](#)).

We will no doubt continue to edit genes with CRISPR-based technologies, but if you still have qualms, maybe RNA editing will be a useful alternative. Check out the link at ^{13.7}[Why Edit RNA?](#) for an overview of the possibilities! As a final point, “*mice and men*” (and women and babies too) have antibodies to Cas9 proteins, suggesting prior exposure to microbial CRISPR/Cas9 antigens. This observation may limit clinical applications of the technology! See ^{13.8}[Uncertain Future of CRISPR/Cas9 Technology](#).

13.2.3 The Small RNAs: miRNA, siRNA and piRNA in Eukaryotes

Micro RNAs (miRNAs) and small interfering RNAs (siRNAs) are found in *Caenorhabditis elegans*, a small nematode (roundworm) that quickly became a model for studies of cell and

molecular biology and development. You can see this roundworm in the fluorescence micrograph in Figure 13.5.

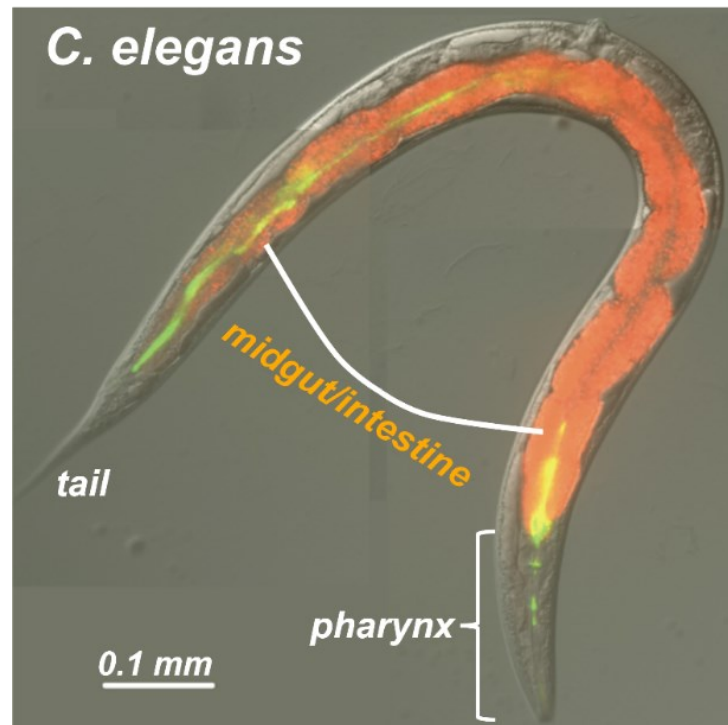


Fig. 13.5: Anatomy of the roundworm *Caenorhabditis elegans*.

What are the attractions of *C. elegans*? Its ~21,700 genes are comparable to the ~25,000 genes in a human genome, its gene products produce an adult worm consisting of just 1,031 cells, these few cells are organized into all of the major organs found in higher organisms, and finally, it's possible to trace the embryonic origins of each adult cell!

13.2.3.a Small Interfering RNA (siRNA)

siRNA was first found in plants and in *C. elegans*. But they are common in many higher organisms. siRNAs were so named because they *interfere* with the function of other RNAs foreign to the cell or organism. Their action was dubbed *RNA interference (RNAi)*. For their discovery of siRNAs, A. Z. Fire and C. C. Mello shared the 2006 Nobel Prize in Physiology or Medicine. Figure 13.6 (below) illustrates the action of siRNA targeting foreign DNA.

When cells recognize foreign double stranded RNAs (e.g., some viral RNA genomes) as alien, a *nuclease* called **dicer** hydrolyzes them. The resulting short, double-stranded hydrolysis products (the **siRNAs**) combine with **RNAi Induced Silencing Complex** or **RISC** proteins. The *antisense* siRNA strand in the resulting **siRNA-RISC** complex then binds to complementary regions of foreign RNAs, targeting them for degradation. Cellular use of RISC to control gene expression in this way may have derived from the use of RISC proteins by miRNAs as part of a cellular defense mechanism, to be discussed next.

Custom-designed siRNAs have been used to disable the expression of specific genes to study their function *in vivo* and *in vitro*. Both siRNAs and miRNAs are being investigated as

possible therapeutic tools to interfere with RNAs whose expression leads to cancer or other diseases.

Small Interfering RNA (siRNA) Degrades Foreign RNA

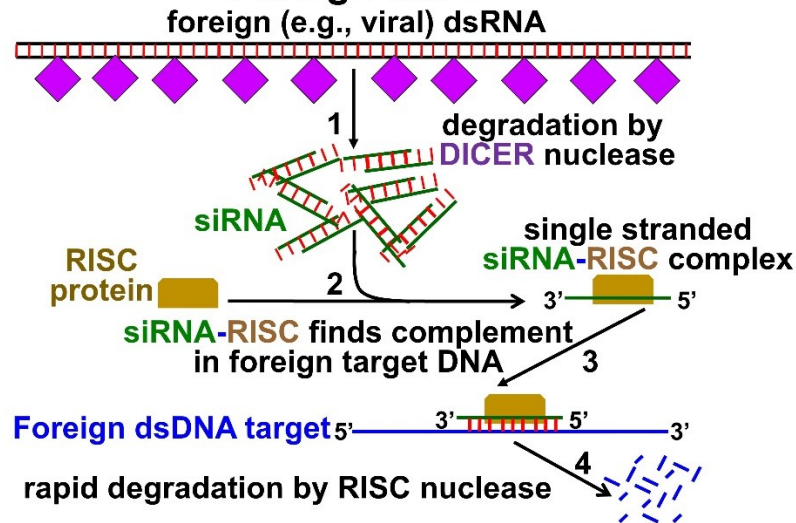


Fig. 13.6: *siRNAs* (small interfering RNAs) form after a *dicer* nuclease recognizes and digests foreign RNA into (*siRNAs*) short digest fragments (1) that bind to *RISC* protein to form a single-stranded *siRNA-RISC complex* (2). The complex then finds and H-bonds to complementary sequences in remaining foreign RNA (3), targeting them for destruction by the *RISC* nuclease (4).



[234 siRNA Posttranscriptional Regulation](#)



[235 Did siRNA Co-opt RISC as a Strategy to Trash, Corrupt, or Wear Out RNA?](#)

Check out the unexpected results of an experiment in which RNAi was used to block embryonic expression of the *orthodenticle* (*odt*) gene that is normally required for the growth of horns in a dung beetle. See the results at ^{13.9}[Genetically Modified Beetles Grow a Functional 3rd Eye](#). As expected, this *knock-out* mutation blocked horn growth. What was unexpected however was the development of an eye in the middle of the beetle's head (the "third eye" in the micrograph). The third eye not only looks like an eye but also functions as one. This was demonstrated by preventing normal eye development in *odt*-knockout mutants. The third eye was responsive to light! Keep in mind that this was a *beetle* with a third eye, not *Drosophila*! Sometimes, especially with new technologies, we can learn as much by studying an unexpected critter as from a model research organism!

13.2.3.b Micro RNAs (miRNA)

miRNAs are widely distributed in eukaryotes, where they target unwanted *endogenous* cellular RNAs for degradation. After transcription, pre-miRNAs fold into hairpin loop structures that bind to protein complexes made up of *DGCR8*, *RNase III* (among others). They are then

processed into mature miRNAs. Like SiRNAs, mature miRNAs combine with *RISC* proteins. The RISC-protein/miRNA complex targets damaged or obsolete mRNAs. An estimated 250 miRNAs in humans may be sufficient to H-bond to diverse target RNAs; only targets with strong complementarity to a RISC-protein/miRNA complex will be degraded. Figure 13.7 shows the pathway from the transcription of a pre-miRNA, its processing, and ultimately to targeting unwanted mRNA for degradation.

Micro RNAs (miRNAs) Degrade Old or Damaged DNAs

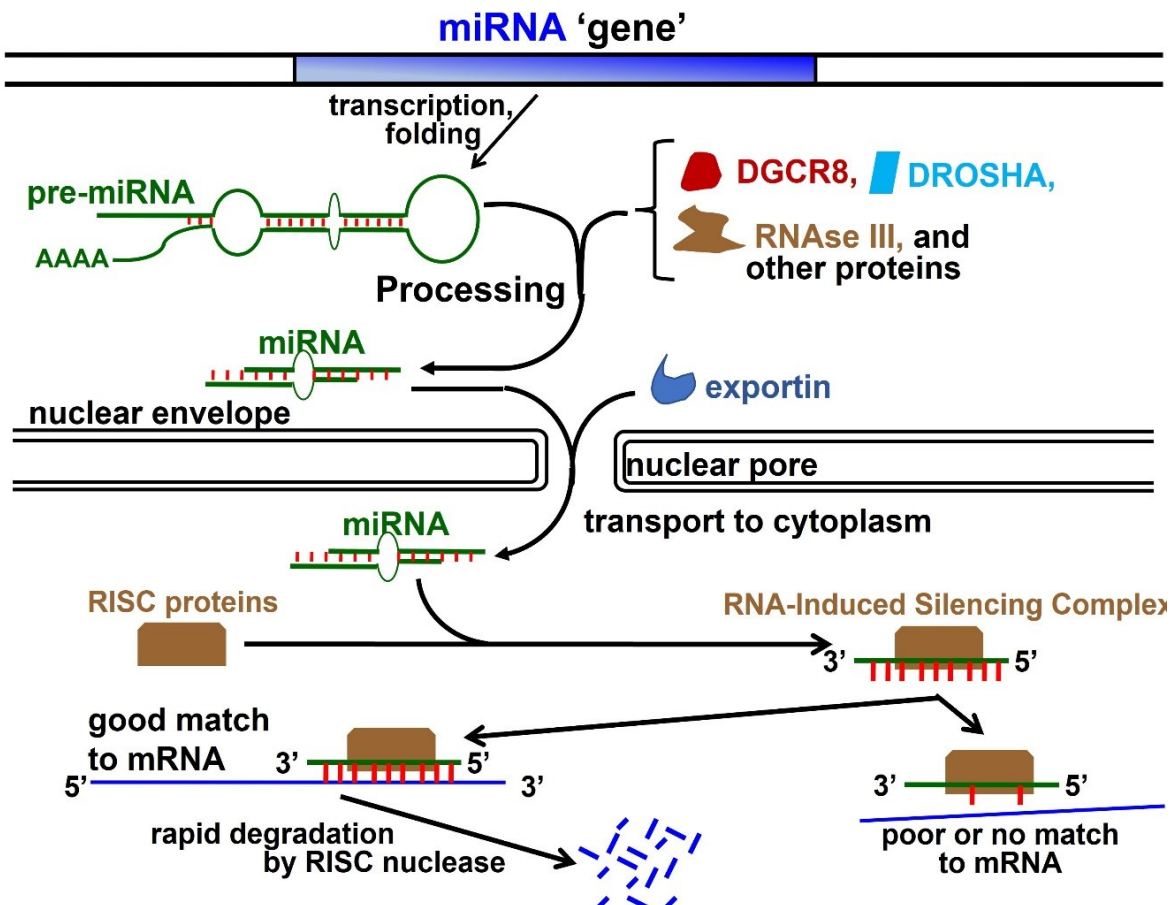


Fig. 13.7: Formation and mechanism of action of miRNA (micro-RNA). miRNAs degrade unwanted (e.g., damaged or old) RNA. RISC protein genes are expressed and combine with mature miRNAs to form a *RISC-Induced Silencing Complex* that targets unwanted mRNAs for destruction.



[236 miRNA Posttranscriptional Regulation](#)

A recent study found that mutant embryonic stem cells (ESCs) unable to make the *DGCR8* also fail to differentiate normal neurons, suggesting a role for miRNA in *embryonic development*. An experiment was designed to ask if the mutant *DGCR8* protein disrupted development in some novel way, or just by inhibiting pre-mi-RNA processing. An *miRNA* look-alike (*mi-R-302*) was engineered that did not bind *DGCR8* and thus, was not processed. When the *mi-R-302*

was introduced into the mutant *ESCs*, these cells went on to differentiate into neuronal cells, even though the look-alike molecule was not processed. confirming the involvement of miRNA in development.

13.2.3.c Piwi-Interacting RNAs

By some estimates, *PIWI-Interacting RNA* (*piRNA*) is the largest separate class of non-coding small RNA. It was first shown to be involved in destroying transposon transcripts to abort transposition. The basics of piRNA synthesis and action in *D. melanogaster* are illustrated in Fig. 13.8 below.

piRNA (Piwi) Counteracts Transposition by Two Mechanisms:

- Destroying the Transposon
- Blocking its Transcription

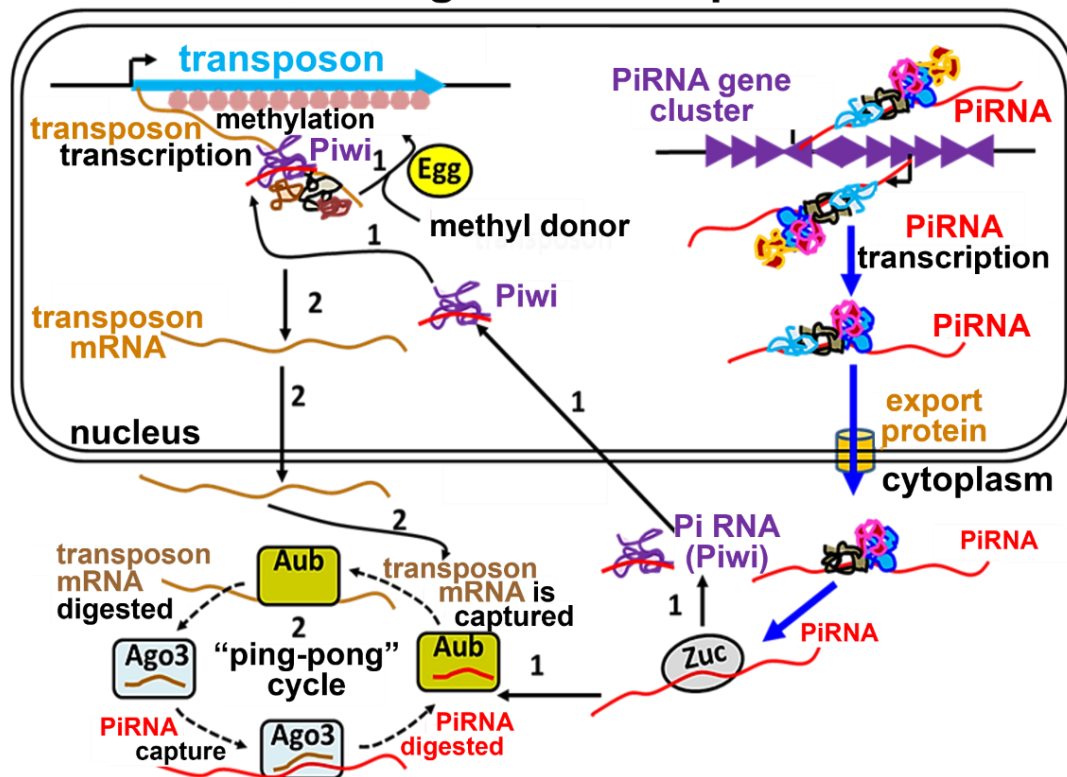


Fig. 13.8: PiRNA is transcribed from clustered Pi genes. PiRNAs associate with nuclear proteins, to be exported, proteins and all, to the cytoplasm (starting upper right, (thick blue arrows). In the cytoplasm, PiRNA engages with a Zuc protein. Thereafter, PiRNA follows one of two pathways (black arrows labeled 1 or 2) to block transposition in *D. melanogaster* testis, illustrated here

The synthesis of piRNAs (follow the thick blue arrows) begins when the nascent primary piRNAs transcribed from a piRNA gene cluster associate with proteins. Those proteins remaining on the transcript after export to the cytoplasm are exchanged for a zucchini (*Zuc*) nuclease. **Primary piRNA** is then hydrolyzed to generate the mature piRNA. piRNA blocks transposition by two pathways (black arrows 1 and 2). In pathway 1, Zuc nuclease cleaves

primary piRNA to make mature piRNA, which is picked up by the *Piwi protein*. The latter joins other proteins on the nascent transposon mRNA, activating *Eggless (Egg)*, a histone methyltransferase. This enzyme catalyzes methylation of a histone to remodel chromatin to heterochromatin, thereby silencing further transposon mRNA transcription. In pathway 2, nascent transposon mRNA associates with proteins that target the mRNA for cytoplasmic destruction. In the cytoplasm, the transposon mRNA is captured by the *Argonaute Aub* (aubergine) protein by hybridization to complementary sequences in mature piRNA, a portion of which has already been transferred from *Zuc* to the *AUB* protein. The transposon mRNA is then digested, preventing actual transposition. Now, a remaining transposon mRNA fragment is transferred to **Ago3** (Argonaute-3), where it can hybridize to complementary sequences in intact primary piRNA from *Zuc* and then digest it to mature piRNA. *Aub* and *Ago-3* are part of a *ping-pong* cycle. The *ping-pong* cycle is an alternate source of piRNA for the *Aub* protein, while *Zuc*-mediated maturation of piRNA is the fork leading to the two pathways that block transposition. Like siRNAs and miRNAs, piRNAs also use a RISC pathway to degrade unwanted RNAs, notably in germline cells. Learn more about piRNA at ^{13.10}[PIWI-Interacting RNA](#) and ^{13.11}[PIWI Proteins and PIWI-Interacting RNA](#).

13.2.4 Long Noncoding RNAs

Long noncoding RNAs (lncRNAs) are a yet another class of noncoding eukaryotic RNAs. The latter include transcripts of antisense, intronic, intergenic DNAs, as well as of *pseudogenes* and *retrotransposons* (pseudogenes are recognizable but mutant genes whose transcripts are nonfunctional; retrotransposons are one kind of transposon or mobile DNA element). While some *lncRNAs* might turn out to be incidental transcripts that the cell simply destroys, others have a role in regulating gene expression. A recently discovered lncRNA is *XistAR*, which—along with its gene product, *Xist*—is required to form *Barr bodies*. Recall that *Barr bodies* form in human females when one of the X chromosomes in somatic cells is inactivated. For a review of lncRNAs, see Lee, J. T. (2012). *Epigenetic Regulation by Long Noncoding RNAs*; Science 338, 1435-1439).

A recent article (at ^{13.12}[lncRNAs and smORFs](#)) summarizes the discovery that some long noncoding RNAs contain short open reading frames (*smORFs*) that are actually translated into short peptides of at as many as thirty amino acids! Who knows? The human genome may indeed contain more than twenty-one thousand to twenty-five thousand protein-coding genes after all!

13.2.5 Circular RNAs (circRNA)

Though discovered more than twenty years ago, circular RNAs (circRNAs) are made in different eukaryotic cell types. Look at ^{13.13}[Circular RNAs-circRNA](#) to learn more about this peculiar result of *alternative* splicing. At first circRNAs were hard to isolate. When they were isolated, they contained “scrambled” exonic sequences and were therefore thought to be nonfunctional errors of mRNA splicing. In fact, circRNAs are fairly stable. Their levels can rise and fall in patterns suggesting that they are functional molecules. Levels of one circRNA, called *circRims1*, rise specifically during neural development. In mice, other circRNAs

accumulate during synapse formation, likely influencing how these neurons will ultimately develop and function.

Thus, circRNAs don't appear to be "molecular mistakes." In fact, errors occurring during their synthesis may be correlated with disease! Speculations about circRNA functions include roles in gene regulation, particularly regulation of the genes or mRNAs from which they themselves are derived.

13.3 "Junk DNA" in Perspective

Not long ago we thought that less than 5% of a eukaryotic genome was transcribed (i.e., as mRNA, rRNA, and tRNA), and that much of the nontranscribed genome served either a structural function or no function at all! That "functionless" genetic material, labeled "*junk DNA*," included nondescript intergenic sequences, mutant pseudogenes, "dead" transposons, long stretches of intronic DNA, and more. Thus, junk DNA was defined as DNA we could do without. Junk DNAs were thought to be accidental riders in our genomes, hitchhikers picked up on the evolutionary road. But while miRNA genes are a small proportion of a eukaryotic genome, their discovery—and that of more abundant *lnc RNAs*—suggests a far greater amount of functional DNA in the genome.

Might there be, in fact, no such thing as "junk DNA"?

The debate continues. Read all about it at ^{13,14}[Junk DNA—Not so useless After All](#) and ^{13,15}[Only 8.2% of Human DNA is Functional](#). Perhaps we should rethink what it means for DNA to be "junk," without "genetic purpose," since more than 90% of our DNA has no known genetic purpose, and its maintenance surely comes at an energy cost.

At the same time, all our DNA is grist for future selection, a source of the diversity required for long-term survival. The same natural selection that picks up "hitchhiker" DNA sequences, as we have seen, can at some point put them to work!

13.4 The RNA Methylome

You can think call this an *RNA epitranscriptome* if you like! Recall that methyl groups direct the cleavage of ribosomal RNAs from eukaryotic 45S pre-RNA transcripts. tRNAs (among other transcripts) are also posttranscriptionally modified. In the absence of a known function, these modifications were thought to be nonfunctional. But are they?

13.5 Eukaryotic Translation Regulation

Messenger RNAs (mRNAs) are made to be translated, and so they are! Nonetheless, translation is regulated, largely by controlling translation initiation. Let's be reminded about the basics of eukaryotic *initiation*. In many respects, the overall process is similar to prokaryotic translation initiation, described elsewhere. The 40S ribosomal subunit itself can bind to and scan an mRNA, seeking the start site of an ORF (open reading frame) encoding a polypeptide. When a GTP-bound *eukaryotic initiation factor 2* (**GTP-eIF2**) binds with met-

tRNA_f, it forms a *ternary complex (TC)*. The TC can associate with the scanning 40S subunit. When a TC-associated scanning subunit encounters the start site of the ORF, scanning stalls. Additional eukaryotic initiation factors (eIFs) help form the *initiation complex*, positioning the initiator tRNA anticodon over the start site AUG in the mRNA. The initiation complex then recruits the large (60S) ribosomal subunit. Since most of the regulation of translation involves controlling the activity of initiation factors and the structure of the promoter region of genes, watch for these interactions in the following descriptions. Let's begin with a close look at Look at translation initiation in eukaryotes (Figure 13.9). Recall that the binding of the 60S ribosomal subunit to the initiation complex (lower left) causes the release of all the eIFs and the hydrolysis of the GTP on eIF2; GDP remains bound to eIF2 (upper right).

Initiating Eukaryotic Polypeptide Synthesis

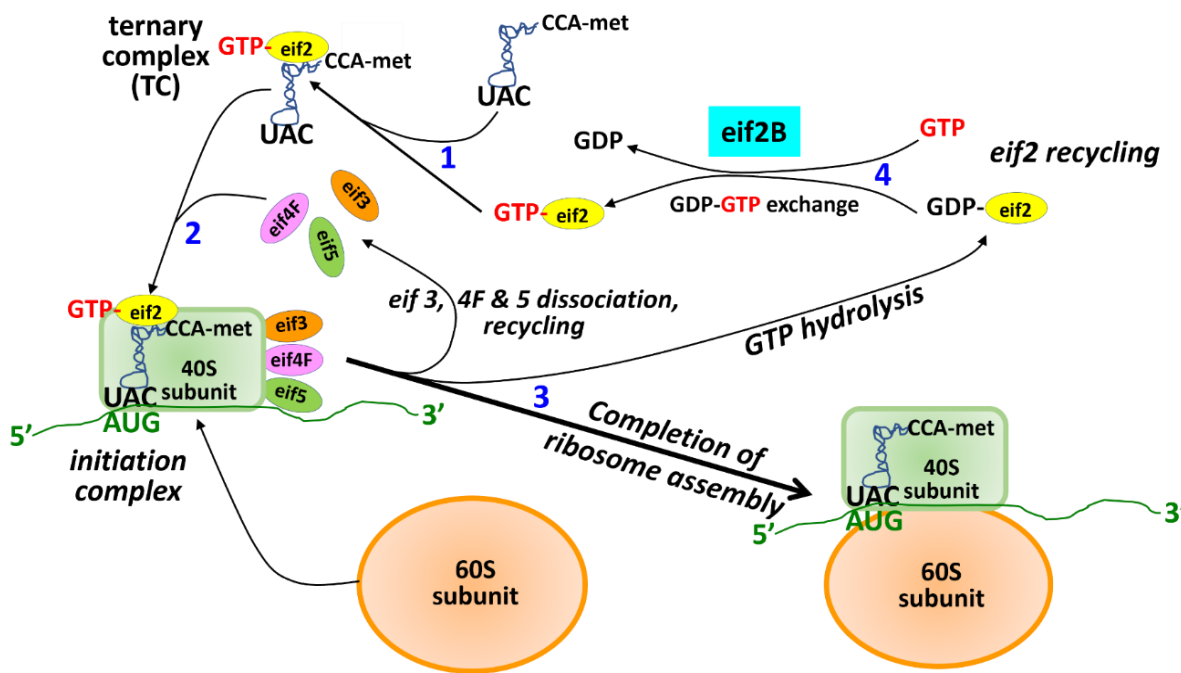
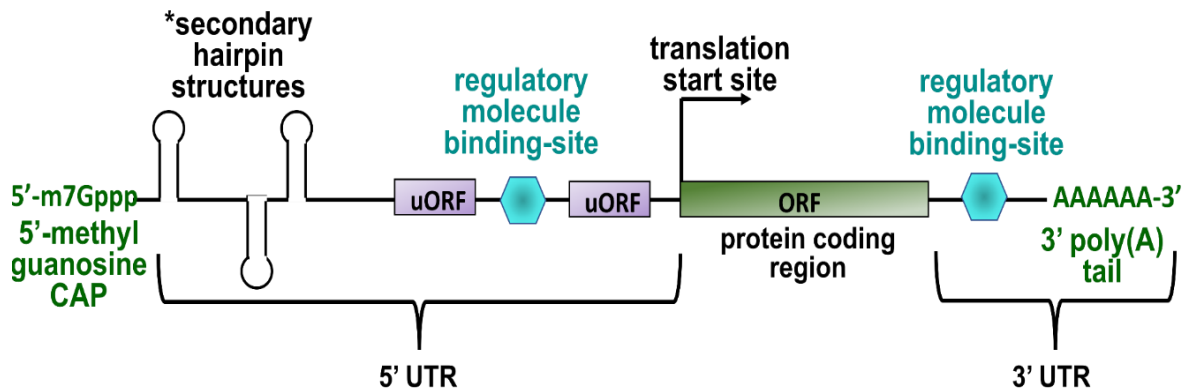


Fig. 13.9: Review of the steps in Eukaryotic translation initiation.

For protein syntheses to continue, new GTP must replace GDP on eIF2. Another initiation factor, **eIF2B**, facilitates this GTP/GDP swap, recycling GTP-eIF2 for use in initiation (i.e., the association of eIF2, the initiator tRNA_f and formyl methionine to form the TC). The regulation of translation is superimposed on these basic processes. But where? What steps in initiation are controlled?

We know that the 5' methyl-guanosine caps and poly(A) tails on mRNAs are required for efficient translation, because mRNAs engineered to lack one and/or the other are poorly translated. But there is little evidence that cells control capping, polyadenylation, or the structures themselves. Instead, translation is regulated largely by targeting interactions with structural features and sequences in the 5' region of mRNAs. Figure 13.10 below shows key structural motifs in a gene that can contribute to translational control.

mRNA Regions/Sequences With Roles in Translation Control



***Some hairpin structures enable CAP-independent translation initiation; others inhibit initiation.**

Fig. 13.10: Regions and specific sequences known to be involved in regulating translation.

Regulation of translation may be global, affecting the synthesis of many polypeptides; or it may be specific, affecting a single polypeptide. The global regulation involves changes in the activity of eIFs that would typically affect all cellular protein synthesis. Specific regulation involves sequences or regions on one or a few mRNAs in turn can recognize and bind specific regulatory proteins and/or other molecules. Such specific interactions would control translation of only those mRNAs, without affecting overall protein biosynthesis.

We will consider three examples of the translational control of gene expression.

13.5.1 Specific Translation Control by mRNA-Binding Proteins

Translation initiation complexes typically scan the 5' UTR of an mRNA. When the complex finds the translation start site, it can bind the large subunit and begin translating the polypeptide. Ferritin is a cytoplasmic iron-binding protein, necessary because iron is of course, insoluble and must be bound to proteins, both in cells and in the blood (the latter is a story for another time and place!). Once in the cell, iron binds to the *ferritin*. As you may guess, cellular iron metabolism depends on the amount of ferritin in the cell, which is regulated at the level of translation. Here we consider the control of *ferritin* synthesis, shown in Figure 13.11, (below).

Ferritin is in fact a two-subunit protein. The 5' UTRs of the mRNAs for both polypeptide chains contain iron-responsive elements (IREs). These IREs are stem-loop structures that recognize and specifically bind *iron-regulatory proteins (IRP1, IRP2)*. The initiation complex scans the 5' UTR of an mRNA. When it finds the normal translation start site, it can bind the large subunit and begin translating the polypeptide. In iron-deficient cells, scanning by the initiation complex is thought to be physically blocked by steric hindrance.

Specific Inhibition of Ferritin Polypeptide Synthesis By Iron Regulatory Proteins (IRPs)

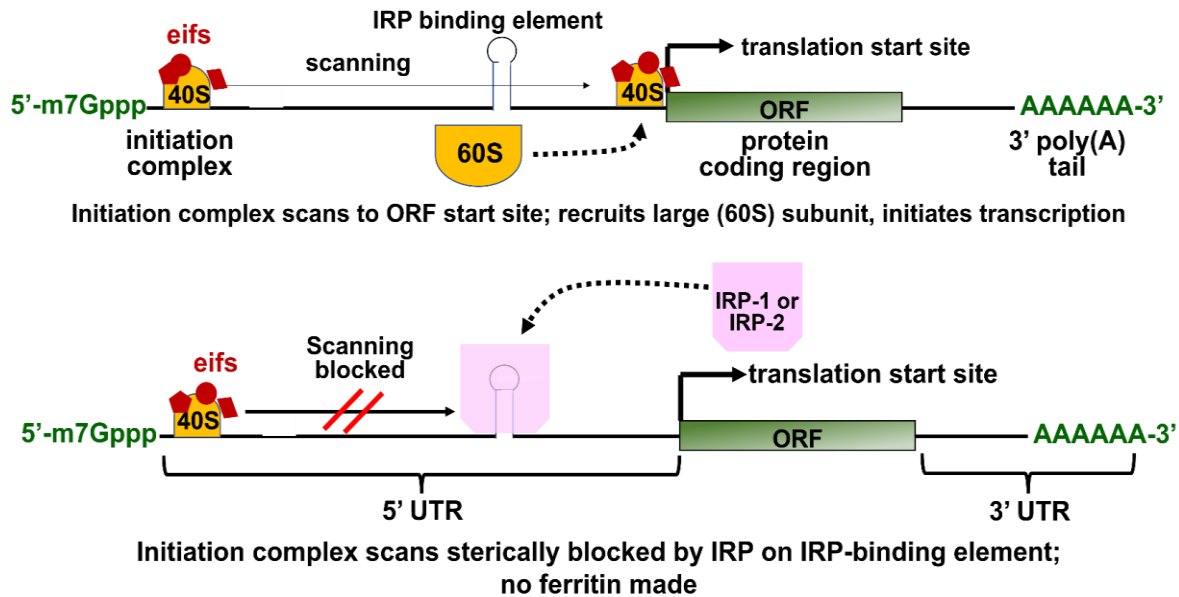


Fig. 13.11: Iron-regulatory proteins disrupt promoter scanning to inhibit ferritin-polypeptide synthesis.

13.5.2 Coordinating Heme and Globin Synthesis

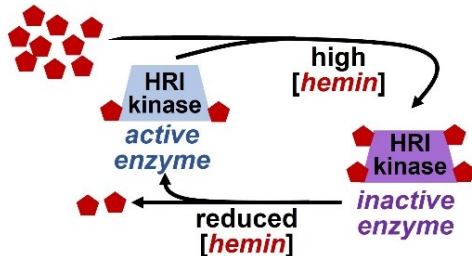
Reticulocytes (the precursors to **erythrocytes**, the red blood cells in mammals) synthesize **globin** proteins. They also synthesize **heme**, an iron-bound porphyrin-ring molecule. Each globin must bind to a single heme to make a hemoglobin protein subunit. Clearly, it would not do for a reticulocyte to make too much globin protein and not enough heme, or *vice versa*.

In fact, **hemin** (a precursor to heme) regulates the translation initiation of both α and β globin mRNAs. Recall that, in order to sustain globin mRNA translation, the GDP on **GDP-eIF2** generated after each cycle of translation elongation must be exchanged for fresh GTP. This is normally facilitated by initiation factor **eIF2B**. **GDP-eIF2** can be phosphorylated or unphosphorylated (active). Making sure that globin is not under- or overproduced relative to heme biosynthesis involves controlling levels of active vs inactive **eIF2B** by **hemin**. Hemin accumulates when there is not enough globin polypeptide to combine with heme in the cell.

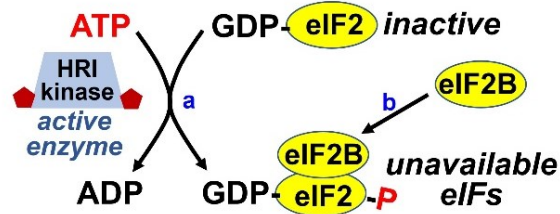
Figure 13.12 below illustrates globin mRNA translation regulation by hemin. At low hemin levels, expect reticulocytes to slow down globin translation. Low Hemin levels activates an **HRI kinase**, causing phosphorylation of GDP-eIF2. **eIF2B**, instead of catalyzing the GTP/GDP swap, binds to GDP-eIF2-P. Thus, neither GDP-eIF2 nor eIF2B are available for the swap. When cellular hemin levels catch up to, or exceed globin levels, HRI kinase is itself partially dephosphorylated and becomes inactive. The result is that GDP-eIF2 can't be phosphorylated, existing phosphorylated GDP-eIF2-P is dephosphorylated, and the GDP-eIF2 and eIF2B dissociate, making both available for the GTP/GDP exchange and a resumption of globin synthesis.

Coordinating Heme and Globin Synthesis By Regulating Globin mRNA Translation Initiation

Hemin regulates HRI kinase activity. If **[hemin]** is high, HRI kinase is phosphorylated. At low **[hemin]** HRI kinase is dephosphorylated:

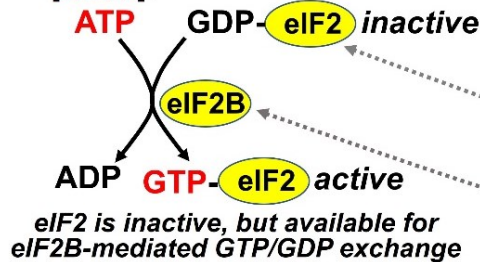


1. At low **[hemin]**, (a) an **active HRI kinase** phosphorylates GDP-eIF2. Then eIF2B binds the GDP-eIF2-P (b), making both eIFB and eIF2 unavailable for translation:



Globin synthesis slows to meet low **hemin** levels.

3. Active eIF2B enables the GTP/GDP swap on eIF2 to make the active **GTP-eIF2**; globin translation speeds up to meet the high **[hemin]**:



2. At high **[hemin]**, **inactive HRI kinase** can no longer phosphorylate eIF2, existing eIFB/eIF2-P complexes dissociate, and the eIF2-P dephosphorylated by a phosphorylase:

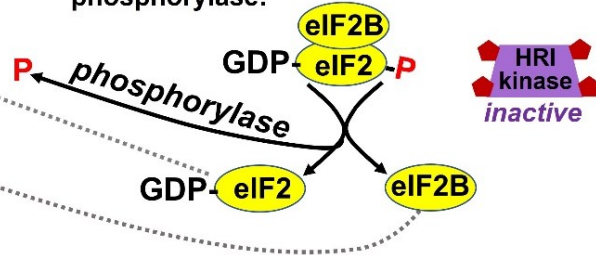


Fig. 13.12: Hemin (a heme precursor) and globin polypeptide syntheses are coordinated (panels 1-3). At high **[hemin]**, globin translation is maximized; at low **[hemin]**, globin translation slows down.



[237 Translation Regulation of Globin Polypeptide Synthesis](#)

13.5.3 Translational Regulation of Yeast GCN4

Like the coordination of heme and globin production, regulation of the yeast **GCN4** protein is based on controlling the ability of the cells to swap GTP for GDP on eIF2. But this regulation is quite a bit more complex even though yeast is a more primitive eukaryote.

It turns out that GCN4 is a **global transcription factor**, controlling the transcription of as many as thirty genes in pathways for the synthesis of nineteen out of the twenty amino acids! The discovery that amino acid starvation caused yeast cells to increase production of amino acids led to the discovery of the **General Amino Acid Control (GAAC)** mechanism involving GCN4. **GCN** is short for **General Control Nondepressible**, referring to its global, positive regulatory effects. It turns out that the GCN4 protein is also involved in stress-gene expression, glycogen homeostasis, purine biosynthesis... the list just goes on. In fact, the GCN4 protein is involved in the action of up to 10% of all yeast genes! Here we focus on the GAAC mechanism.

Yeast cells with ample amino acids don't need to make more. GCN4 levels are low in these cells. But if the cells are starved of amino acids, GCN4 levels rise as much as tenfold in two hours, causing increased general amino acid synthesis. Amino acid starvation initially signals an increase in the activity of *GCN2*, another *protein kinase* that catalyzes GDP-eIF2 phosphorylation (Figure 13.13), which we just saw, cannot exchange GTP for GDP on eIF2.

Effect of Amino Acid Starvation on Translation

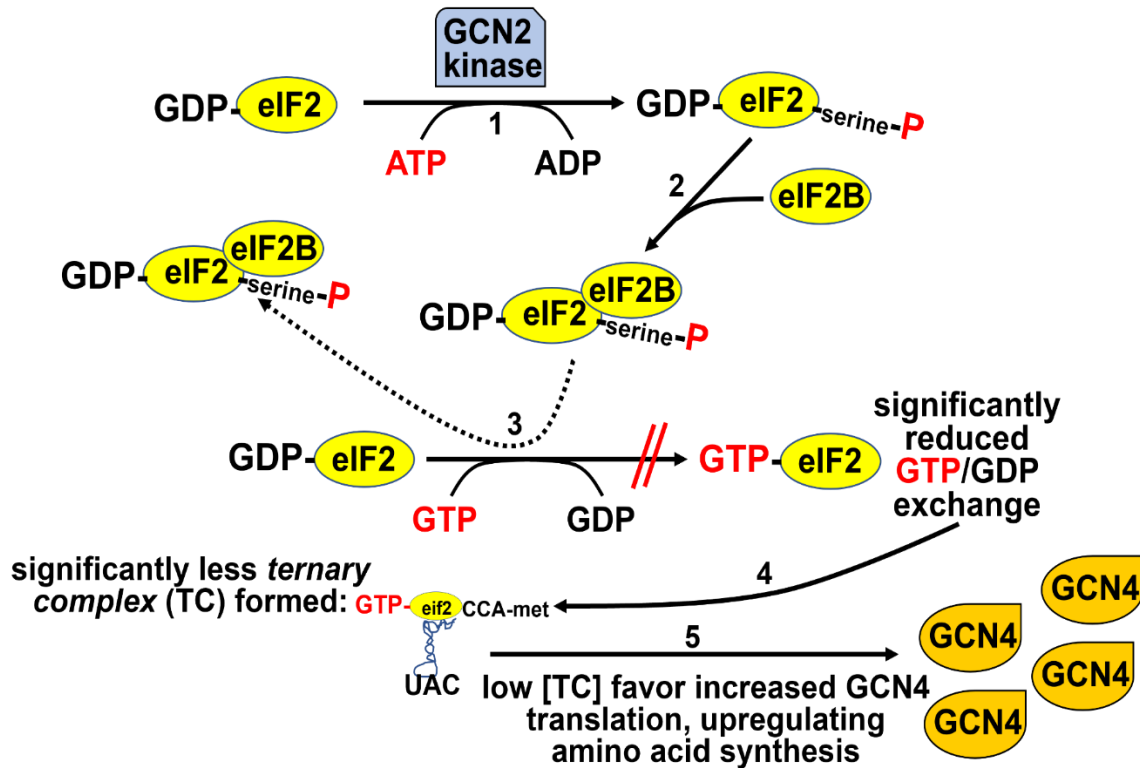


Fig. 13.13: Amino acid starvation increases a *GCN2 kinase* activity that phosphorylates GDP-eIF2 (1). eIF2B then binds GDP-eIF2-P (2), blocking GTP/GDP exchange (3). The resulting drop in ternary complexes (4) favors GCN4 translation, leading to maximal amino acid synthesis (5).

There is a paradox here. You would expect a slowdown in GTP-eIF2 regeneration to inhibit overall protein synthesis, and it does. However, the reduced levels of GTP-eIF2 somehow also stimulate translation of the GCN4 mRNA, leading to more transcription of the amino acid synthesis genes. In other words, amino acid starvation leads yeast cells to use available substrates to make their own amino acids, so that protein synthesis can continue... at the same time as initiation complex formation is disabled! Let's accept that paradox for now and look at how amino acid starvation leads to increased translation of the GCN4 protein and the upregulation of amino acid biosynthesis pathways. To begin with, we are going to need to understand the structure of GCN4 mRNA (Figure 13.14). Note the four short upstream open reading frames (*uORFs*) in the 5' untranslated region (5' UTR) of GCN4 mRNA.

GCN4 mRNA with Multiple Short 5' Open Reading Frames

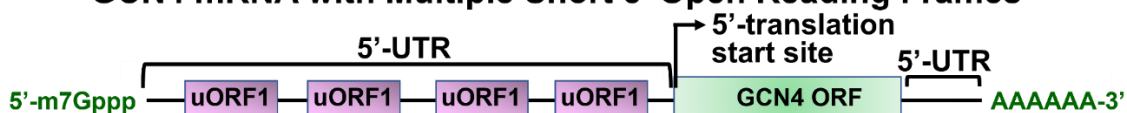


Fig. 13.14: Short open reading frames (*uORFs*) in the GCN4 mRNA 5' UTR act to regulate translation.

Recall that eIF2 binds with GTP and the initiator met-tRNA to form a ternary complex (TC). As we also noted earlier, TC-associated 40S ribosomal subunits scan mRNAs to find translation start sites of ORFs, allowing initiation complexes to form. When a 60S ribosomal subunit binds, polypeptide translation starts. But while uORFs in the GCN4 mRNA encode only a few amino acids before encountering a stop codon, they can also be recognized during scanning. When TCs and 40S subunits are plentiful, they engage uORFs in preference to the GCN4-coding region ORF (Figure 13.15).

Adequate Amino Acids in Medium Lead to Low Levels of GCN4 Translation

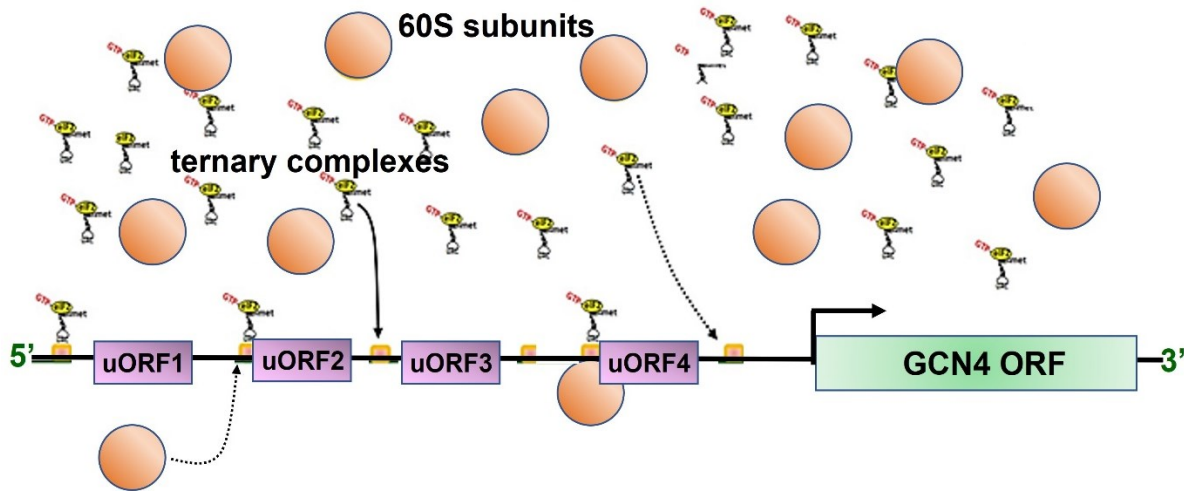


Fig. 13.15: At normal amino acid levels, abundant **TC-40S** (ternary) complexes scanning for ORFs cluster at **uORF** start sites, slowing scanning to the GCN4 ORF and thus inhibit GCN4 translation.

Under these conditions, active eIF2B allows the GTP/GDP swap on GDP-eIF2, leading to efficient GTP-eIF2 recycling and high TC levels. TCs were bound to the small (40S) subunits during scanning and/or at the start sites of uORFs. At this point, they form initiation complexes that bind 60S ribosomal subunits and begin uORF translation.

The effect of high TC levels, leads to high levels of initiation complexes that in turn drive scanning for the GCN4 ORF, slowing down scanning the upstream ORFs (uORFs). thereby inhibiting initiation complex formation at the actual GCN4 ORF. The result is increased transcription of GCN4 mRNA and GCN4 protein synthesis.

What happens in amino acid-starved cultures of yeast cells, when GTP-eIF2 cannot be efficiently regenerated and GTP/initiator met-tRNA ternary complexes are in short supply? To review, amino acid starvation signals an increase in GCN2 kinase activity, resulting in phosphorylation and inactivation of eIF2B. Inactive *phospho*-eIF2 will not facilitate the GTP/GDP swap at GDP-eIF2, inhibiting overall protein synthesis. The resulting reduction in GTP-eIF2 also lowers the levels of TC and TC-associated 40S subunits. In Figure 13.16 (below), see how this phenomenon upregulates GCN4 translation, even as the translation of other mRNAs has declined.

Amino Acid Starvation Leads to High Levels of GCN4 Translation

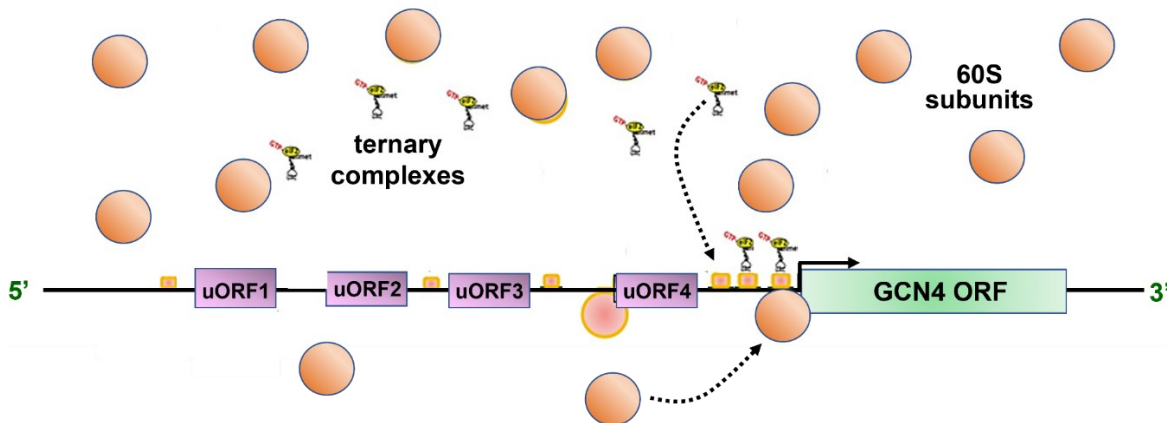


Fig. 13.16: During *amino acid starvation* there are fewer **TC-40S** complexes scanning for ORFs and they more readily bind to the GCN4 ORF, increasing GCN4 synthesis. At high levels, GCN4 regulates the expression of many genes, including those encoding enzymes needed for amino acid synthesis.

Under these conditions, the formation of initiation complexes at uORFs is less likely, thereby making the scanning for open reading frames by the remaining TC-40S complexes more likely to reach the GCN4 ORF, resulting in an increase in GCN4 translation! The increased production of the gene regulatory GCN4 protein then turns on virtually all of the genes for enzymes of amino acid biosynthesis. The cells thus make their own amino acids when the medium has few to contribute! For a good review of translation-level regulation in general and of GCN4, see ^{13,16}[GCN4 Translation Regulation](#).

13.6 Protein Turnover in Eukaryotic Cells: Regulating Protein Half-Life

We have already seen that organelles have a finite life span or **half-life**, defined as the time for half of them to disappear in the absence of a regeneration (new synthesis) of the organelles. Recall that lysosomes help to destroy worn-out mitochondria, including their molecular components. Recall also the small RNAs (especially *miRNA*) that destroy old or damaged cellular RNAs.

As we already know, the steady-state level of any cellular structure or molecule exists when the rate of its manufacture or synthesis is balanced by the rate of its turnover. Of course, steady-state levels of things can change. For example, the level of gene expression (the amount of a final RNA or protein gene product in a cell) can change if rates of transcription, processing, or turnover change. We should also expect the same for the steady-state levels of cellular proteins. Here we consider the factors that govern the half-life of cellular proteins, a property of all cellular molecules and structures. The half-lives of different proteins seem to be inherent in their structure. Thus, some amino acid side chains are more exposed at the surface of the protein and are thus more susceptible to change or damage over time than others. Proteins with fewer “vulnerable” amino acids should have a longer half-life than those with more of them. Proteins damaged by errors of translation, folding, or processing gone awry or just worn out from use or “old age” will be targeted for destruction. The mechanism for detecting and destroying unwanted, old, damaged, or misbegotten proteins involves a 76-amino acid polypeptide called **ubiquitin** that binds to target proteins and delivers them to a

large complex of polypeptides called the **proteasome**. The pathway to protein destruction is shown below in Figure 13.17. Use the illustration to help you follow the numbered steps in the outline that follows the illustration.

Ubiquitin Targets and Delivers Old and Damaged proteins to Proteasomes for Destruction

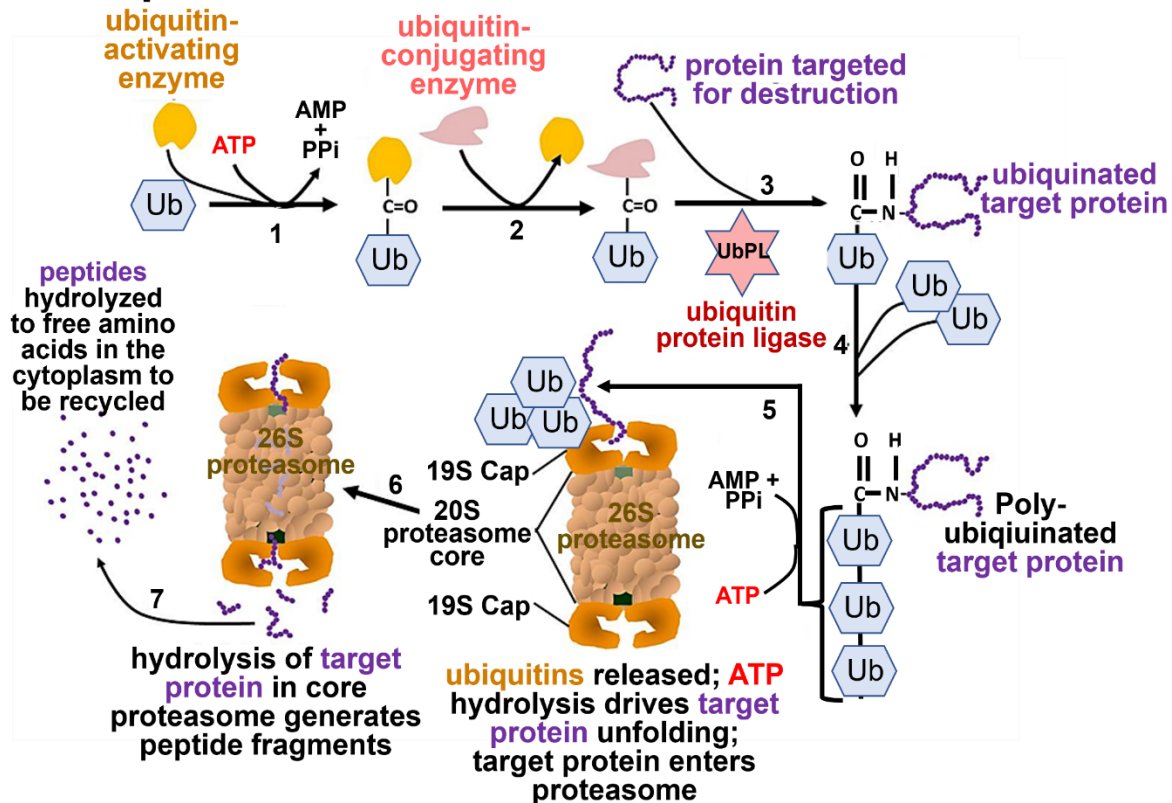


Fig. 13.17: Proteasomes are the main recycler of old and damaged protein in a cell. In this mechanism, an activating enzyme binds ubiquitin (1), as the ubiquitin is activated, a conjugation protein replaces the activation enzyme on ubiquitin (2). A protein targeted for destruction is covalently linked to ubiquitin (3). A ubiquitin ligase recruits more ubiquitins to the complex (4). Next, ATP is hydrolyzed as the complex binds to a proteasome Cap. The target protein dissociates from the complex and enters the proteasome (6) to be hydrolyzed and the amino acids recycled (7).

1. The first step is to activate a ubiquitin. This starts when ATP hydrolysis fuels the binding of ubiquitin to a **ubiquitin-activating enzyme**.
2. A **ubiquitin-conjugating enzyme** then replaces the ubiquitin-activation enzyme, forming a complex with ubiquitin.
3. Proteins destined for destruction replace ubiquitin-conjugating enzymes on the ubiquitin.
4. Several more ubiquitins then bind to this complex.
5. The **poly-ubiquitinated protein** delivers its protein to one of the 19S 'CAP' structures of a proteasome,

6. After binding to the proteasome-*cap* structure of a proteasome, the poly-ubiquitinated target proteins dissociate and ubiquitins are released and recycled as the target protein unfolds (powered by ATP hydrolysis) and enters a 20S core proteasome.
7. The target protein is digested to short peptide fragments by proteolytic enzymes in the interior of the proteasome core. The fragments are released from the CAP-complex at the other end of the proteasome and digested down to free amino acids in the cytoplasm.

There is a mind-boggling variety of proteins in a cell—including as many as six hundred different ubiquitin proteins, encoded by as many genes! Presumably, each ubiquitin handles a subclass of proteins based on common features of their structure. The structurally complex 26S proteasome is smaller than a eukaryotic small ribosomal subunit (40S). Nevertheless, it is one of the largest cytoplasmic particles, even without the benefit of any RNA in its structure! To see different animated versions of *ubiquitin* and *proteasome* in action, look at ^{13,17}[Proteasome in Action-1](#) and ^{13,18}[Proteasome in Action-2](#).

For their discovery of the ubiquitin-proteasome system and its proteolytic actions, A. Ciechanover, A. Hershko, and I. Rose shared the 2004 Nobel Prize in Chemistry.

Some iText & VOP Key Words and Terms

19S proteasome- <i>cap</i> complex	GDP-eIF2	riboswitch
20S proteasome complex	Gene Drive	RISC
Amino acid starvation	gene editing	RISC endonuclease
Ago-3	global transcription factor	RISC proteins
Argonaute proteins	globin	RNAi (RNA interference)
Aub	gRNA	RNA turnover rates
Barr bodies	GTP/GDP swap	RNA-induced silencing complex
<i>Caenorhabditis elegans</i>	GTP-eIF2 recycling	sgRNA (single-guide RNA)
Cas	Half-life	siRNA (small interfering RNA)
Cas helicase activity	HCR kinase	small RNAs
Cas9 endonuclease	heme	smORF
circRNA (circular RNA)	hemin	spacer RNA
CRISPR	Initiation complex scanning	steady state
CRISPR interference	Iron-regulatory protein	<i>Streptococcus pyogenes</i>
CRISPR/Cas	IRP	tracr DNA
CRISPR/Cas9 gene array	junk DNA	tracr gene
CRISPRi	lncRNA	tracr RNA
dicer	long noncoding RNA	translation elongation
Egg (Eggless) methyltransferase	“ping-pong” cycle	ubiquitin
eIF2 phosphorylation	miRNA (micro RNA)	ubiquitination
eIF2B	mRNA scanning	uORF
EIIA ^{Glc}	Piwi protein	XistAR
ferritin	piRNA	yeast GCN4
GAAC	proteasome	Zuc nuclease
GCN2	Protein turnover rates	

CHAPTER 13 WEB LINKS

13.1



13.2



13.3



13.4



13.5



13.6



13.7



13.8



13.9



13.10



13.11



13.12



13.13



13.14



13.15



13.16



13.17



13.18



Chapter 14

Repetitive DNA: A Eukaryotic Genomic Phenomenon

Minisatellite DNA, microsatellite DNA, ribosomal RNA genes; transposons – selfish junk DNA or architects of the genome?

Reminder: For inactive *links*, google key words/terms for alternative resources.

Too Much of a Good Thing???



Or Not???

14.1 Introduction

Because of their small size, bacterial genomes have few **repetitive DNA** sequences. In contrast, repetitive DNA sequences can make up as much as half of a eukaryotic genome. Much of this repeated DNA consists of identical or nearly identical sequences of varying length repeated many times (up to a million or more times!). Examples include **satellite DNA** (*minisatellite* and *microsatellite* DNAs) and **transposons**, or **transposable elements**. Here we'll consider experiments that first revealed the existence and proportion of repeated DNA in genomes. Next, we describe Barbara McClintock's even earlier (and pretty amazing!) discovery of transposable elements. After we describe the different classes of transposons and different mechanisms of **transposition**, we tackle the question of why they and other repetitive DNAs even exist. Elsewhere we introduced the notion of **junk DNA** as DNA sequences that are not transcribed and serve no known purpose. Are repeated DNA sequences really **junk DNA**? Are transposable elements just **junk**? We now understand and continue to learn that transposons and other repetitive DNAs can have specific functions, from regulating gene expression to reshaping genomes to increasing genetic diversity in evolution. Some are even transcribed! Far from being "junk," much of this redundant DNA exists in genomes because of evolutionary selection.

Learning Objectives

When you have mastered the information in this chapter, you should be able to do the following:

1. Compare and contrast *renaturation kinetic data*.
2. Explain *CoT curves* and *DNA complexity*.
3. List physical and chemical properties of *main-band* and *satellite DNAs*.
4. Outline an experiment to determine if a given sequence of DNA is repetitive or not.
5. Summarize how Barbara McClintock revealed the genetics of maize *mosaicism*.

6. Outline the experiments suggesting that the *Ds* gene moves from one locus to another in the maize genome.
7. Compare and contrast *cut-and-paste* and *replicative transposition*.
8. Compare the behaviors of *autonomous* and *nonautonomous* transposons.
9. List the differences between *Mu phage* infection and transposition.
10. Describe the common structural features of transposons.
11. Compare the mechanisms of *LINE* and *SINE* transposition.
12. Speculate on how a species with a high *transposon load* manages to avoid potentially lethal consequences of transposition.
13. Speculate on which came first in evolution: DNA transposons, RNA transposons, or retroviruses, and explain your reasoning.

14.2 The Complexity of Genomic DNA

By the 1960s, when Roy Britten and Eric Davidson were studying eukaryotic gene regulation, they knew that there was more than enough DNA to account for the genes needed to encode an organism. It was also likely that DNA was more structurally complex than originally thought. They knew that ***cesium chloride (CsCl) density gradient centrifugation*** separated molecules based on differences in density and that fragmented DNA would separate into main and minor bands of different density in the centrifuge tube. The minor band was dubbed ***satellite DNA***, recalling the *Sputnik* satellite that had recently been launched by Russia (or moons as satellites of planets!). DNA bands with different densities could not exist if the proportions of A, G, T, and C in DNA (already known to be species-specific) were the same throughout a genome. Instead, there must be regions of DNA that are richer in A-T than G-C pairs, and vice versa. Analysis of the satellite bands that moved farther on the gradient (i.e., that were denser) than the main band were indeed richer in G-C content. Those that lay above the main band were more A-T-rich.

Consider early estimates of how many genes it might take to make a human, mouse, chicken, or petunia: about a hundred thousand! Such estimates may have been based on notions of how many proteins a eukaryotic life requires. We know now that it takes fewer! Nevertheless, the genome of a typical eukaryote contains one hundred to one thousand times more DNA than necessary to account for even the inflated one hundred thousand gene estimate. So, how *do* we explain all the seemingly *extra* DNA? Britten and Davidson performed a series of elegant experiments to measure DNA ***renaturation kinetics*** that revealed some physical characteristics of genes and what seemed to account for the excess DNA. Let's look at these experiments in some detail.

14.2.1 The Renaturation Kinetic Protocol

The first step in a renaturation kinetic experiment is to shear DNA isolates to an average size of 10 Kbp (kilobase pairs) by pushing high molecular weight DNA through a hypodermic needle at constant pressure. The resulting double-stranded fragments (***dsDNA fragments***) are next heated to 100°C to ***denature*** (separate) the two strands. The solutions are then cooled to 60°C to allow the single-stranded DNA (***ssDNA***) fragments to slowly re-form complementary double strands. At different times after incubation at 60°C, the partially renatured DNA was sampled, and ssDNA and dsDNA were separated and quantified.

The experiment is summarized below in Figure 14.1.

Here is an experiment that reveals the *complexity* of a eukaryotic genome.

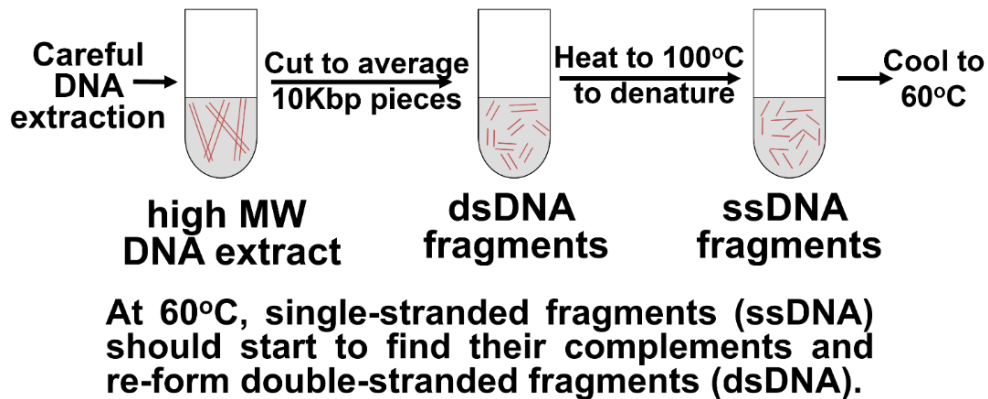


Fig. 14.1: Renaturation kinetics: double-stranded DNA (dsDNA) is mechanically cut to ~10 Kbp fragments, heated to denature the DNA, and then cooled to allow the ss DNAs find and renature with their complements.

14.2.2 Renaturation Kinetic Data

Britten and Davidson then plotted the percentage of DNA that had renatured over time. Figure 14.2 is the plot of data from a renaturation kinetics experiment using *rat* DNA, showing the rate of dsDNA formed at different times (out to many days!).

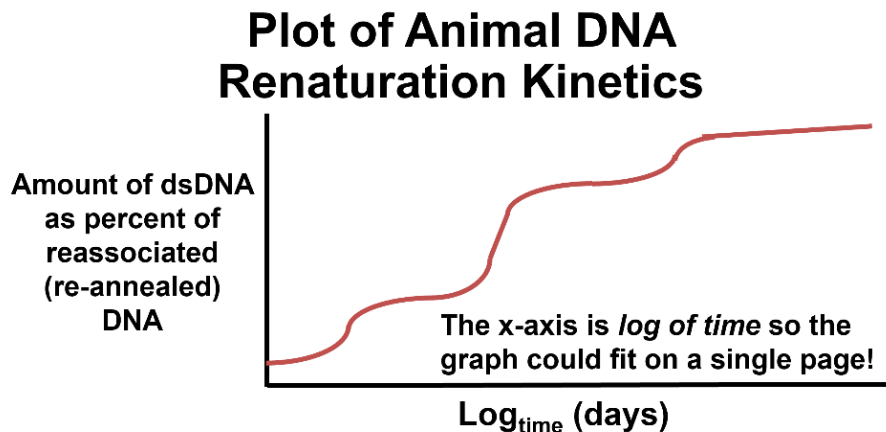


Fig. 14.2: Plot of rat dsDNA formed over time during renaturation of denatured DNA.

In this example, the DNA fragments could be placed in three main groups with different overall rates of renaturation. Britten and Davidson hypothesized that the dsDNA that had formed most rapidly was composed of sequences that must be more *highly repetitive* than the rest of the DNA. The rat genome also had a lesser amount moderately repeated (*middle-repetitive*) dsDNA fragments (which took longer to anneal than the highly repetitive fraction), and even less of a very-slowly reannealing *unique* sequence DNA fraction (which took the longest time to re-anneal). The latter sequences were so rare in fact, that it could take days for them to re-form double strands, and they were classified as nonrepetitive, unique- (or nearly unique-) sequence DNA, as illustrated below in Figure 14.3.

Interpretation of a Renaturation Kinetic Plot

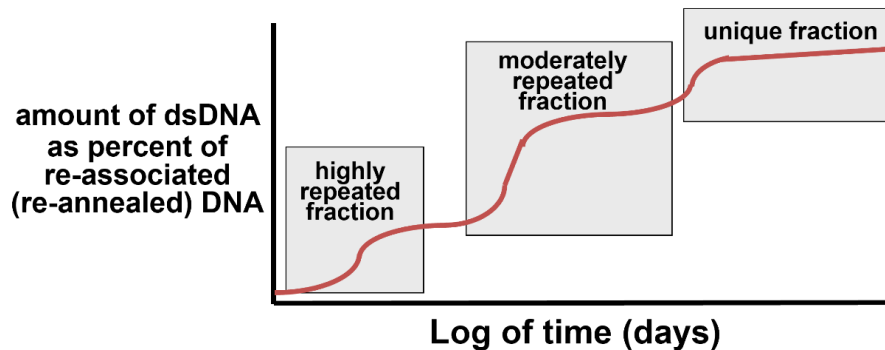


Fig. 14.3: The three “phases” of the curve in Fig. 14.2 have here been highlighted to identify the three fractions of repeated and almost unique DNA sequences in the rat genome.

It became clear that the rat genome, and in fact most eukaryotic genomes, consists of different classes of DNA that differ in their redundancy. From the graph, a surprisingly large fraction of the genome was repetitive, to a greater or lesser extent.



[238 Discovery of Repetitive DNA](#)



When renaturation kinetics were determined for *E. coli* DNA, only one “redundancy class” of DNA was seen (Figure 14.4).

Bacterial DNA Renaturation Kinetics Plot

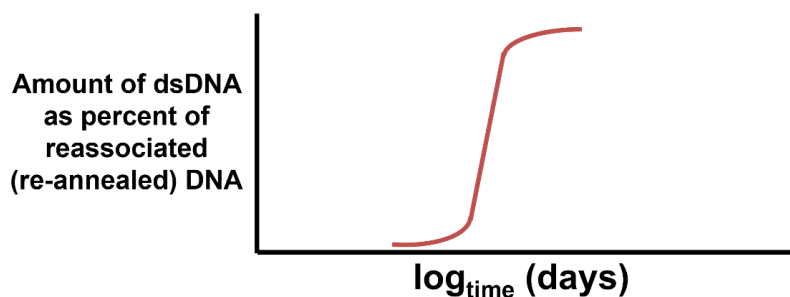


Fig. 14.4: Plot of *E. coli* dsDNA formed over time during renaturation of denatured DNA.

Based on *E. coli* gene-mapping studies and the small size of the *E. coli* “chromosome,” the reasonable assumption was that there is little room for “extra” DNA in a bacterial genome and that the single class of DNA on this plot must be unique-sequence DNA.

14.2.3 Genomic Complexity

Britten and Davidson defined the relative amounts of repeated and unique (or single-copy) DNA sequences in an organism's genome as its **genomic complexity**. Thus, prokaryotic genomes have a lower genomic complexity than eukaryotes. Using the same data as is in the

previous two graphs, Britten and Davidson demonstrated the difference between eukaryotic and prokaryotic *genome complexity* by a simple expedient. Instead of plotting the fraction of dsDNA formed vs time of renaturation, they plotted the percentage of reassociated DNA against the **concentration of the renatured DNA multiplied by the time the DNA took to reanneal** (the *CoT* value). When *CoT* values from rat and *E. coli* renaturation data are plotted on the same graph, you get the *CoT curves* in the graph in Figure 14.5.

DNA Complexity: the percent of a genome that is unique, moderately repeated, or highly repeated.

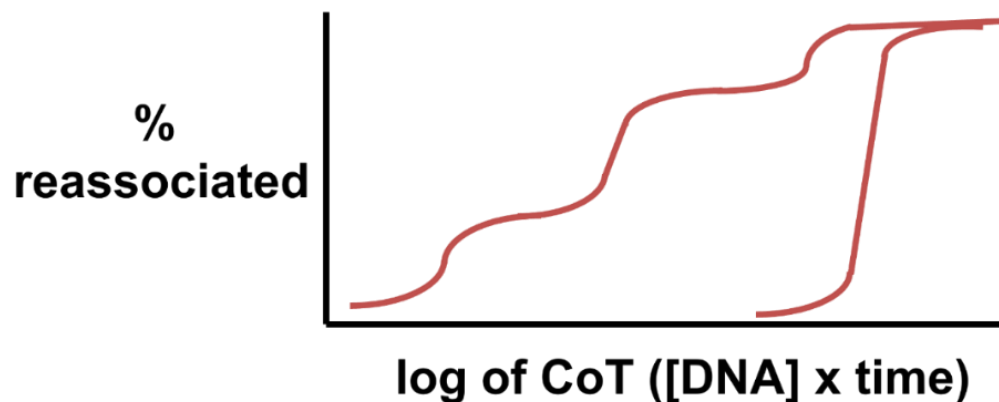


Fig. 14.5: DNA complexity is revealed by plotting rat and *E. coli* DNA renaturation kinetics as the percentage of reassociated dsDNA over CoT (Concentration of reassociated dsDNA \times time).

This deceptively simple extra calculation (from the same data!) allows direct comparison of the *complexities* of different genomes. These *CoT curves* tell us that $\sim 100\%$ of the bacterial genome consists of unique sequences (curve at the far right), compared to the rat genome, which has two DNA redundancy classes and (at the right of the curve), only a small fraction of unique- sequence DNA. Prokaryotic genomes are indeed largely composed of unique (nonrepetitive) sequence DNA that must include single-copy genes (or operons) that encode proteins, ribosomal RNAs, and transfer RNAs.



[239 CoT Curves and DNA Complexity Explained!](#)



14.2.4 Functional Differences between CoT Classes of DNA

The next questions, of course, were what kinds of sequences are repeated and what kinds of sequences are "unique" in eukaryotic DNA? Eukaryotic satellite DNAs, transposons, and ribosomal RNA genes were early suspects.

To start answering these questions, satellite DNA was isolated from the CsCl gradients, made radioactive, and then heated to separate the DNA strands. In a separate experiment, renaturing rat DNA was sampled at different times of renaturation. The isolated *CoT fractions*

were once again denatured and mixed with heat-denatured radioactive satellite DNA probe. The mixture was then cooled a second time to allow renaturation. The experimental protocol is illustrated in Figure 14.6.

Low CoT Value DNA Contains Satellite Sequences

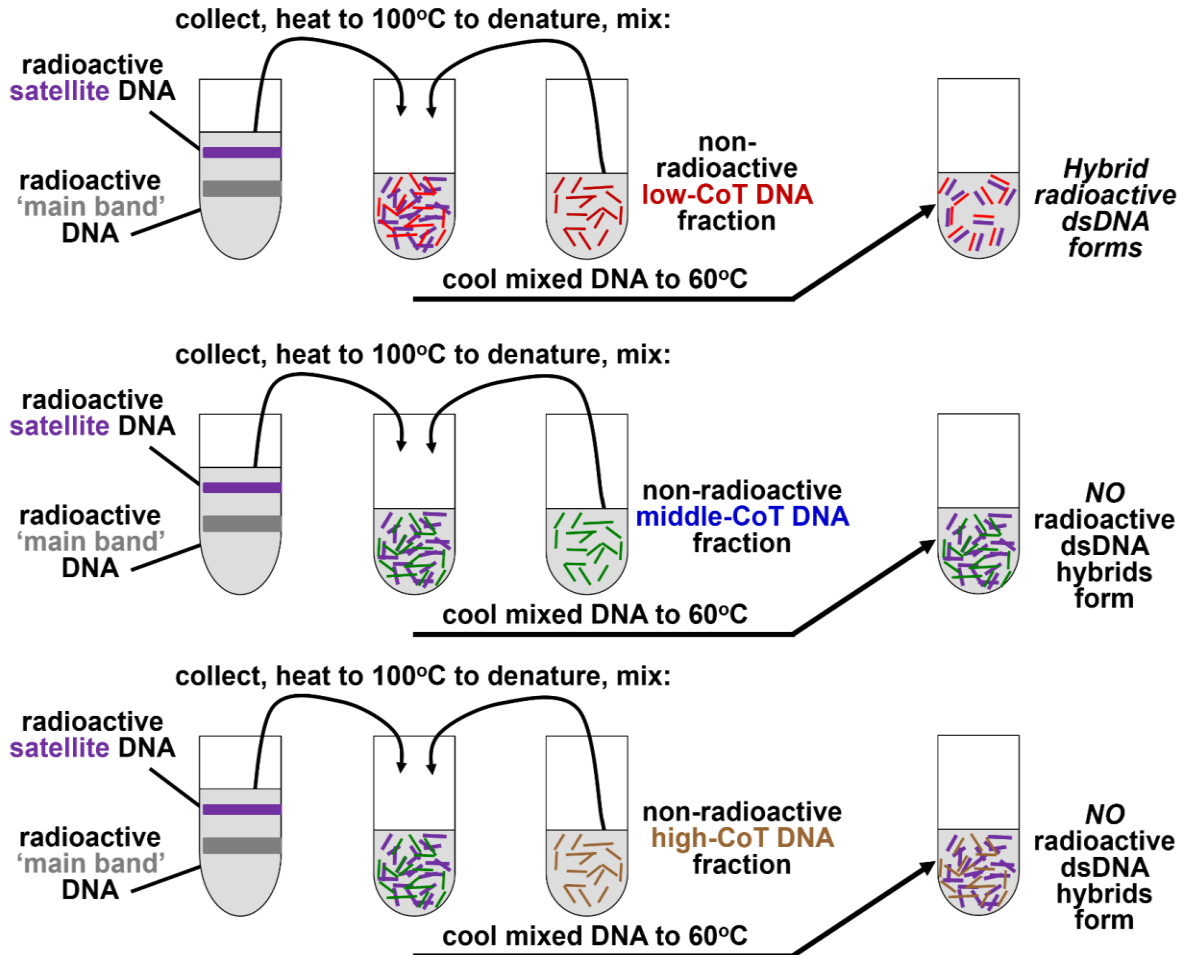


Fig. 14.6: Isolated CoT fractions of eukaryotic DNA were separately mixed with radioactive satellite DNA isolated from a CsCl density gradient, then reheated to denature all of the DNA, and finally, cooled to allow satellite DNA find complementary fragments in the CoT fractions.

The results of this experiment showed that radioactive satellite DNA only annealed to DNA from the *low-CoT* fraction of DNA. Satellite DNA is thus *highly repeated* in the eukaryotic genome.

In similar experiments, isolated radioactive rRNAs formed radioactive RNA-DNA hybrids when mixed and cooled with the denatured *middle CoT* of eukaryotic DNA. Thus, rRNA genes were moderately repetitive. With the advent of recombinant DNA technologies, the redundancy of other kinds of DNA was explored by probing renatured DNA fractions from renaturation kinetics experiments using cloned genes encoding rRNAs, mRNAs, transposons, and other sequences. Table 14.1 (below) summarizes the results of such experiments.

Table 14.1

Classes, Properties, and Some Functions of Repetitive DNAs in Eukaryotes

Class	Type	% of genome, e.g. mammals	Sub-type	Organization */or Properties	Typical Unit Length (bp)	Copy #, e.g., mammals	Location in genome	Function (if Known) with examples
Highly Repetitive	Satellite DNAs	10-15%	Micro-satellite DNA	Tandem repeats (*VNTRs)	2-8 bp	1000+	Centromeres, Hetero-chromatin, Dispersed	Spindle fiber attachment, gene regulation
			Mini-satellite DNA	Tandem repeats (*VNTRs)	10-60 bp	1000+	Dispersed	Gene Regulation
			Telomeres (sub-category of mini-satellite DNA)	Tandem Repeats	4-6 bp	~2500	Chromosome ends	Prevent chromosome shortening during replication
Moderately Repetitive	Transposons	20-45% (>80% in Maize!)	DNA Transposon	Move via DNA intermediates ("cut-&-Paste")	Up to 7K bp	10,000+	Dispersed	P-element, Mariner, Ac, Ds
			LINEs Retro-transposon SINEs	Move via RNA intermediates	Up to 7Kbp	~500,000	Dispersed	L1
				Move via RNA intermediates	80-400 bp	Up to 10 ⁶	Dispersed	Alu
	rRNA genes	<1%	45S rRNA genes	Tandem 45S rDNA repeats	13.7 Kbp	Up to- 400	Nucleolus	Translation
			5S rRNA genes	Tandem 5SrDNA repeats	120 bp	200-300	Dispersed	Translation
Unique Sequence	Transcribed genes	50-60%	Protein-coding genes (introns + exons)	~25000 in humans	variable	One or a few (e.g., in gene families)	Dispersed	Just about everything else!

* VNTR: Variable Number of Tandem Repeats

The table compares properties (lengths, copy number, functions, percentage of the genome, location in the genome, etc.) of repetitive-sequence DNAs. The fact most of a eukaryotic genome is made up of repeated DNA and that transposons can be as much as 80% of a genome—came as a surprise! We'll focus next on the different kinds of *transposable elements*.



[240 Identifying Different Kinds of DNA Each CoT Fraction](#)



[241 Some Repetitive DNA Functions](#)

14.3 The “Jumping Genes” of Maize

Barbara McClintock's report that bits of DNA could jump around and integrate into new loci in DNA was so dramatic and arcane that many thought the phenomenon was either a one-off or not real! Only with the subsequent discovery of transposons in bacteria (and in other eukaryotes) were McClintock's jumping genes finally recognized for what they were.

To begin the tale, let's look at the maize reproduction in Figure 14.7 (below).

Maize Reproduction

Maize is Monoecious
separate male and female plants.

Pollen tubes form; sperm reaches ovules (eggs) at base of silks.

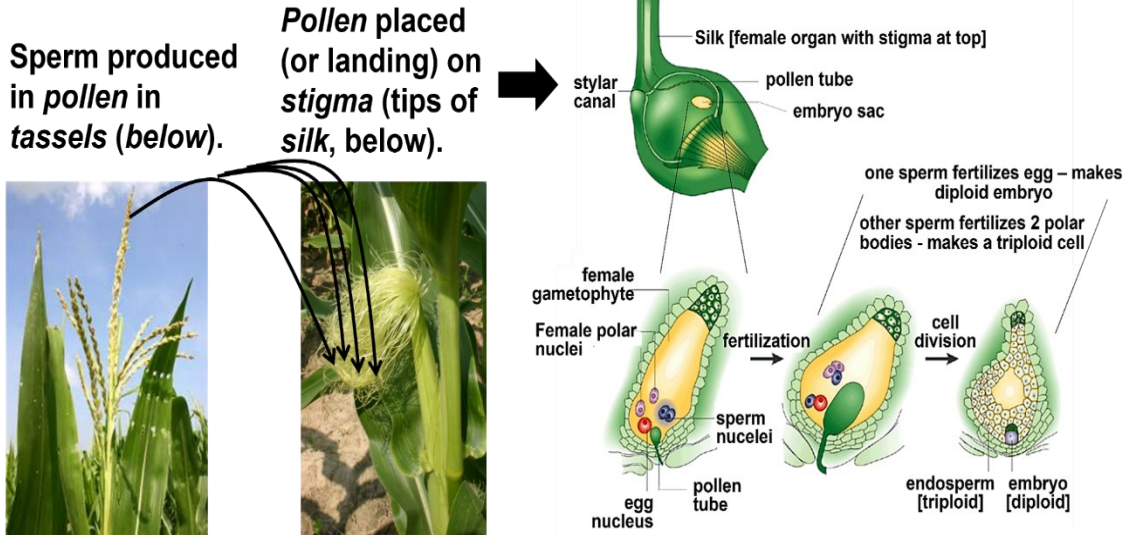


Fig. 14.7: Life cycle of maize: sperm from male plant pollen land on *stigma* atop corn *silk* to tunnel through a *pollen tube* of their own making to the base of the silk. There, one sperm fertilizes the egg; another fertilizes polar bodies left over from oogenesis, which become triploid endosperm cells.

The different colors of corn seeds (kernels) result from **anthocyanin** pigments, which are expressed differentially by cells of the *aleurone* tissue, derived from the triploid endosperm. McClintock was studying the inheritance of color variation, which ranged from colorless (white or yellow, due to an absence of *anthocyanins*) to brown, purple, spotted, or streaked. The **mosaic** of kernel colors is vivid in the corn cobs in the photo in Figure 14.8.

Variegated Corn



Fig. 14.8: Examples of mosaic corn cobs with different color kernels on the same cob

Clearly, kernel color is inherited, and the inheritance of colorless and purple seed color does indeed follow Mendelian rules. But the genetics of mosaicism does not. Mosaic color patterns after genetic crosses were not consistent, implying that the mutations responsible for kernel colors were not due to mutations in germ cells. Rather, genes controlling anthocyanin synthesis must have been undergoing mutations in the somatic cells that would become (or already were) the ones in which the pigments were produced.



[242 What Interested McClintock about Maize](#)

14.3.1 Discovering the Genes of Mosaicism: The Unstable Ds Gene

As we describe McClintock's experiments, keep in mind that her research and intuitions about gene regulation and *epigenetic* inheritance came long before molecular technologies made it possible to prove or to name these phenomena. McClintock was looking for a genetic explanation for seed-color variation in the 1940s and early 1950s. DNA structure had only recently been published, and gene cloning and DNA sequencing were decades into the future! Her only available technologies were based on an understanding of Mendelian allelic assortment in traditional breeding studies. Since seed color is expressed in cells derived from endosperm, McClintock knew that the inheritance of the kernel-color *phenotype* must be studied against a *triploid* genetic background. She was also aware of speculations that the variegated-color phenotype might result when a specific *unstable mutation*, which usually produced colorless kernels, "reverted" in some cells but not in others, thereby creating a spotted or streaked phenotype. Just what made for an "unstable mutation" was, of course unknown. McClintock identified three genes involved in seed-kernel coloration and ultimately solved this puzzle.

Two of the genes studied by McClintock controlled the presence vs absence of kernel color. These are the **C** and **Bz** genes:

1. **C'** is the dominant *inhibitor allele*, so-called because if even one copy were present, the kernels would be *colorless* (yellow), regardless of the rest of the genetic background.
2. **Bz** and **bz** are dominant and recessive alleles of the Bz gene, respectively. In the absence of a dominant C' allele, the presence of a Bz allele would lead to purple kernels. If the bz allele were present without *both C' and Bz* alleles, the kernels would be dark brown.

The third gene—the one required to get the variegated kernel color—was the **Ds** (*Dissociator*) gene. McClintock knew that without a viable **Ds** gene, kernels were either colored or colorless, depending on the possible genotypes dictated by the C and Bz alleles. In other words, the **Ds** gene must suffer the "*unstable mutations*" that led to variegated kernel color.

The mutations occurred at random among aleurone layer cells, so the mutations must have been occurring in a region of *chromosomal instability* (prone to damage or breakage) in some cells but not in others. Let's look at what McClintock did to figure out what was going on in corn-kernel color genetics. Having already demonstrated crossing-over in maize (another

remarkable achievement!), McClintock *mapped* the C', Bz, and Ds genes to *chromosome 9*. She then selectively mated corn with genotypes shown in the protocol in Figure 14.9.

Expectations for Cross Between a Female Triple Recessive (for C, B and Ds genes with a male Homozygous for the Same Genes

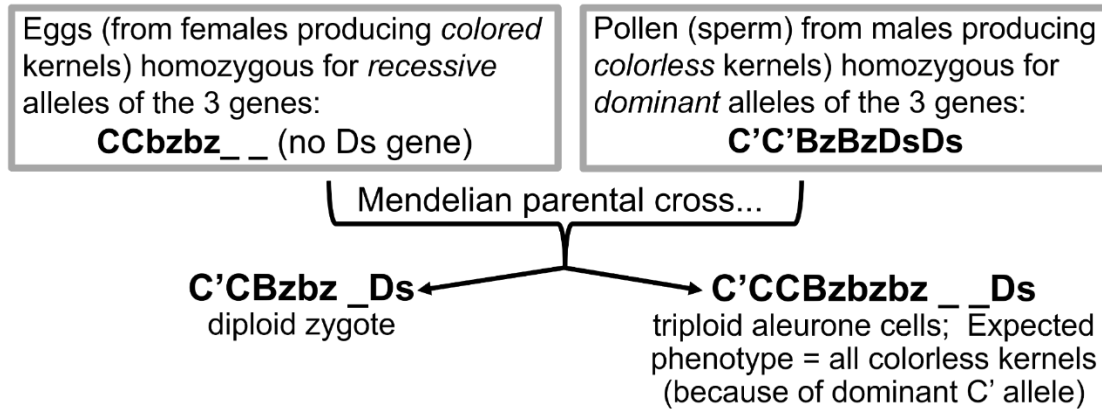


Fig. 14.9: An experimental cross between CCbzbz triploid (triple recessive) female with no Ds genes and C'C'BzBzDsDs (triple dominant) male with Ds genes, showing predicted genotypes and kernel color phenotypes, based on Mendelian assumptions.

Remember that triploid cell genotypes are being considered in this illustration! You can refer to the phenotypic effects inherent in three-gene allelic backgrounds as we follow McClintock's cross. Her cross of a homozygous recessive with a homozygous dominant plant should ring a bell! Let's look more closely at this cross. Figure 14.10 shows the expected triploid genotypes from the cross. Aleurone cells resulting from this cross should all be colorless (yellow) because of the presence of the dominant C' allele.

Expected of Triploid Aleurone Cell Genotypes From the Cross in Figure 14.9

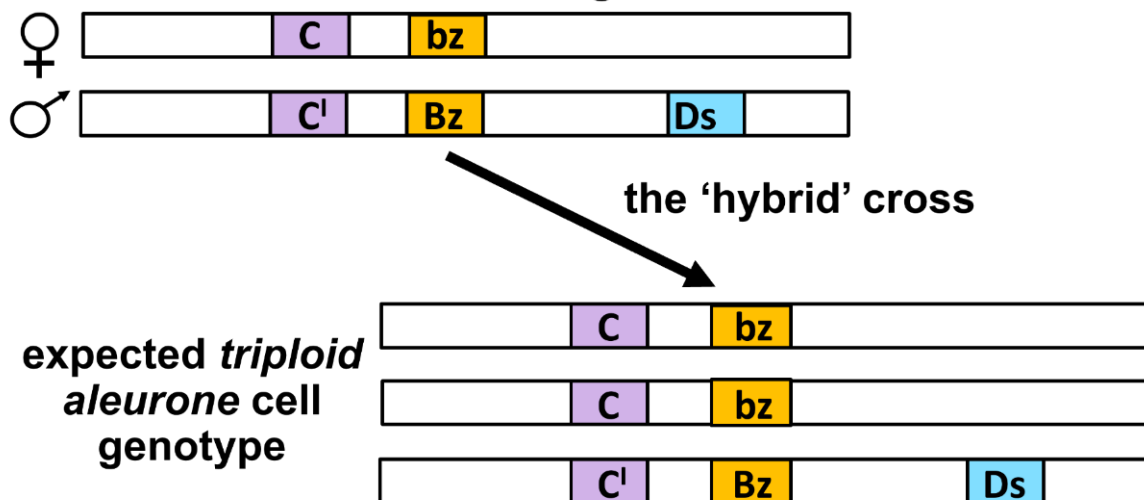


Fig. 14.10: Triploid genotypes expected for the cross in Fig. 14.9: The starting homozygous recessive female and homozygous male genotypes (upper) are mated. The expected triploid genotypes for this hybrid cross are shown (below).

There were indeed many *colorless* kernels on the *hybrid* cob, as expected, but there were also many *mosaic* kernels with dark *spots* or *streaks* against a colorless background.

McClintock's interpretation of events is illustrated in Figure 14.11 (below).

McClintock's Interpretation of the Hybrid Cross

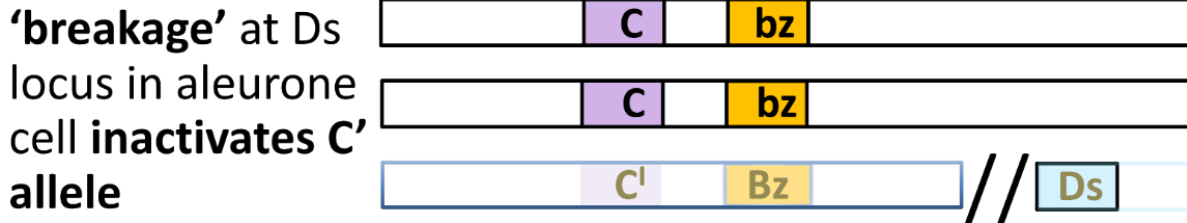


Fig. 14.11: McClintock's interpretation of results of the *triploid cross* in Fig. 14.9

According to McClintock, if some aleurone-layer cells in some kernels were to suffer chromosomal breakage at the ***Ds*** (*Dissociator*) locus (indicated by the double slash, //), the C' allele would be inactivated. Without a functional C' allele, the **operative** genotype in the affected cells is **CCbzbz**. These cells will revert to making the brown pigment as directed by the bz allele. When these cells divide, they form clusters of brown cells that are surrounded by cells with an unbroken chromosome (and thus an active C' allele), creating a mosaic—that is, the appearance of pigment spots or streaks in some kernels, against the otherwise-colorless background of the surrounding cells.



[243 Variegated Maize Kernels Result from "Loss" of the Ds Gene](#)

14.3.2 The Discovery of Mobile Genes: The Ac/Ds System

The experiments just described were reproducible using her original, single breeding stock of maize. But when McClintock tried to repeat the experiments by crossing the homozygous dominant males with homozygous recessive females from a *different* breeding stock, all the kernels of the progeny cobs were colorless, as if the Ds gene had not caused any chromosomal damage.

It seemed that the ***Ds*** genes contributed by the males were unable to function mutate (i.e., "break") chromosomes in females of this new breeding stock. McClintock hypothesized that the female in the original cross must have contributed a factor that could somehow activate the Ds gene to break, and that this factor, yet another gene, was absent or inactive in the females of the new breeding stock. McClintock called the new factor the **activator**, or **Ac** gene. Based on the dependence of ***Ds*** on the **Ac** locus, McClintock recognized that mosaicism in maize kernels was controlled by these two "genes" as part of a two-element, **Ac/Ds system**.

She then demonstrated that Ac-dependent Ds "breakage" was in some cases also associated with inactivation of a normal Bz gene, leading to a loss of purple kernels. At this

point McClintock concluded that, far from simply “breaking” the chromosome at a fragile Ds locus, *the Ds gene had moved to (or into) the Bz gene*, disrupting its function. Again, this could not happen in the absence of an active Ac gene.

McClintock had discovered the first transposon, earning her the 1983 Nobel Prize in Physiology or Medicine, albeit an honor belated by decades! View the homage to, and a brief history and summary of, McClintock’s work and impact at ^{14.1}[A short history of McClintock Science](#). With the advent of recombinant DNA technologies, we now know the following:

1. The **Ds** transposon lacks a gene for a *transposase* enzyme required for transposition.
2. The **Ac** element has this gene and is capable of independent transposition.
3. **Ac** provides the transposase needed to mobilize itself *and* the Ds element.
4. The sequence similarity of Ds and Ac elements supports their common ancestry.

The basic structural features of the maize Ac/Ds system are listed here:

- Ac is 4,563 bp long.
- Ds is a truncated version of Ac.
- There are eleven bp *inverted repeats* at either end of the Ac and Ds elements.
- There are eight bp *direct repeats of insertion site DNA* created at the time of transposition flanking each of the elements.

Look for these features as we describe prokaryotic and eukaryotic transposons.



[244 Discovery of Mobile Elements and the Ac-Ds System](#)



[245 The Ac-Ds System Today](#)



14.4 Since McClintock: Transposons in Bacteria, Plants, and Animals

Transposons exist everywhere we look in prokaryotes, and they account for much eukaryotic repetitive DNA. Sometimes called “jumping genes,” they can be a large proportion of eukaryotic genomes and include some “jumping genes” that no longer even jump (i.e., transpose). Transposons were also once considered useless; they were described as *junk DNA* containing *selfish genes*, whose only purpose was self-replication to copy useless, junk DNA sequences. But considering some new evidence, perhaps not!

As you will see, transposition shares many features with DNA replication, recombination, repair, and even viral infection. As you study how transposons move, keep in the back of your mind that transposition is often triggered by cellular stress. Let’s begin with a look at some bacterial transposons first, and then we’ll look at eukaryotic transposons.

14.4.1 Bacterial Insertion Sequences (IS Elements)

IS elements are the first mobile bacterial elements described. Discovered in the late 1960s, many were identified (IS1, IS2..., IS10, etc.). Some inserted into genes (e.g., in the lac operon), but most did not, likely because there is little noncoding ("extra") DNA in the compact bacterial genome. Without this extra DNA acting as a buffer against damaging mutations, few bacterial cells would live to tell a tale of transposition! Perhaps it should surprise us that IS elements can be made to transpose in the lab but are generally silent in nature.

Bacterial IS elements vary in length from about 750 to 1425 bp. One is shown below in Fig. 14.12. The entire IS element is flanked by **direct repeats** (i.e., repeated sequences facing the same direction) derived from the mechanism of transposition insertion. Within the element lie **transposase** and **resolvase** genes whose products are necessary for mobility.

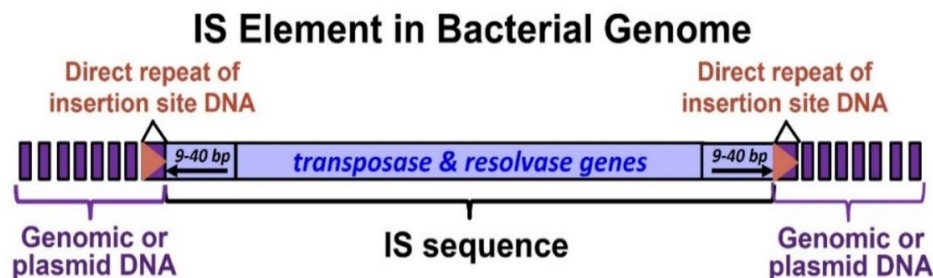


Fig. 14.12: Structure of a bacterial *IS* element

Flanking the IS element genes are a pair of repeat sequences facing in opposite directions. Because bacteria (and plasmids) only tolerate low copy numbers of IS elements, there are typically less than ten copies and as few as one!

14.4.2 Composite Bacterial Transposons: Tn Elements

If a pair of IS elements should lie close to each other, separated by a short stretch of genomic or plasmid DNA, they can transpose together, essentially as a 'double-element', carrying the DNA between them as part of a **composite transposon**, or **Tn element**. If some of the DNA between IS elements in a Tn element contains antibiotic-resistance genes, its transposition can carry and spread these genes to other DNA in the cell. Tn elements (like IS elements) are present in low copy number. Figure 14.13 illustrates a generic Tn element.

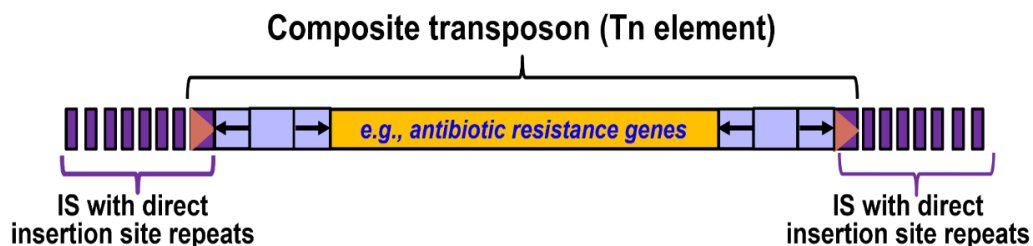


Fig. 14.13: Structure of a bacterial *Tn* element.

Antibiotic resistance genes have the medical community worried; their spread has led to antibiotic-resistant pathogens that cause diseases that are increasingly hard and even impossible to treat. Earlier we saw the genetic "transformation" of streptococcal cells that pick

up virulence genes in DNA from dead cells. We routinely transform cells with plasmids as part of recombinant DNA experiments, but bacteria can transfer plasmid DNA between themselves quite naturally. During bacterial conjugation, an *F* (*fertility*) plasmid normally transfers DNA between compatible bacterial mating types (review bacterial conjugation in chapter 8). An *F* plasmid containing a Tn element that harbors an antibiotic-resistance gene can thus be passed from donor to recipient during conjugation. The Tn element can then transpose into to the recipient's bacterial genome. In this way, transposition is a major pathway for the transfer and spread of antibiotic resistance.

14.4.3 Complex Transposons That Can Act Like Bacteriophage

Bacterial **complex transposons** also contain other genes in addition to those required for mobility. Some complex transposons resemble bacteriophage, and one type of transposon—the ***Mu* phage**—is in fact one! *Mu* can function either as an infectious phage (which reproduces in an infected cell) or as a transposon in the bacterial genome. Transposon genes in a *Mu* phage are illustrated in Figure 14.14.

Some Bacteriophage are Also Transposons

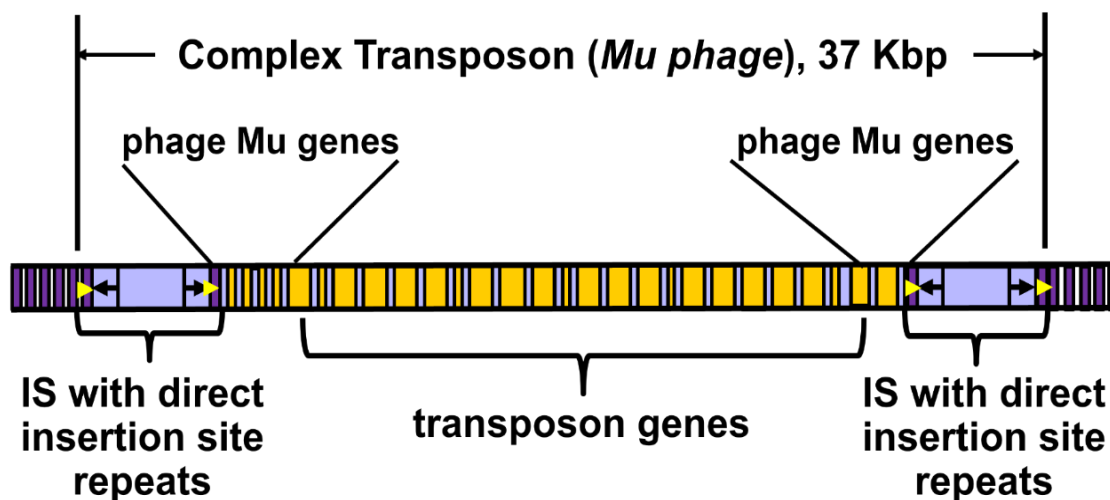


Fig. 14.14: Structure of *Mu* phage, a complex transposon.

After infecting a bacterium, *Mu* can enter the **lytic** phase of its life cycle, replicating its DNA, producing new infectious phage “particles,” ultimately releasing these by lysing the host bacterial cell. Alternatively, like other phage, *Mu* can undergo **lysogeny**, inserting its DNA into the host-cell chromosome. Integrated copies of *Mu* can excise and reenter the lytic phase to produce more phages, usually if some environmental stress threatens host bacterial survival. But a third lifestyle choice, transposition, is available to *Mu* once the phage integrates into the bacterial chromosome.

The three lifestyle options for *Mu* phages are illustrated in the next few pages. Figure 14.15 (below) illustrates the lytic and lysogenic lifestyle options for a bacterial virus, and the next illustration (Figure 14.16) shows the additional lifestyle options of *Mu* phage; the phage DNA can act as a transposable element while in the lysogenic pathway!

Structure & life cycle of a *bacteriophage* (bacterial *virus*)

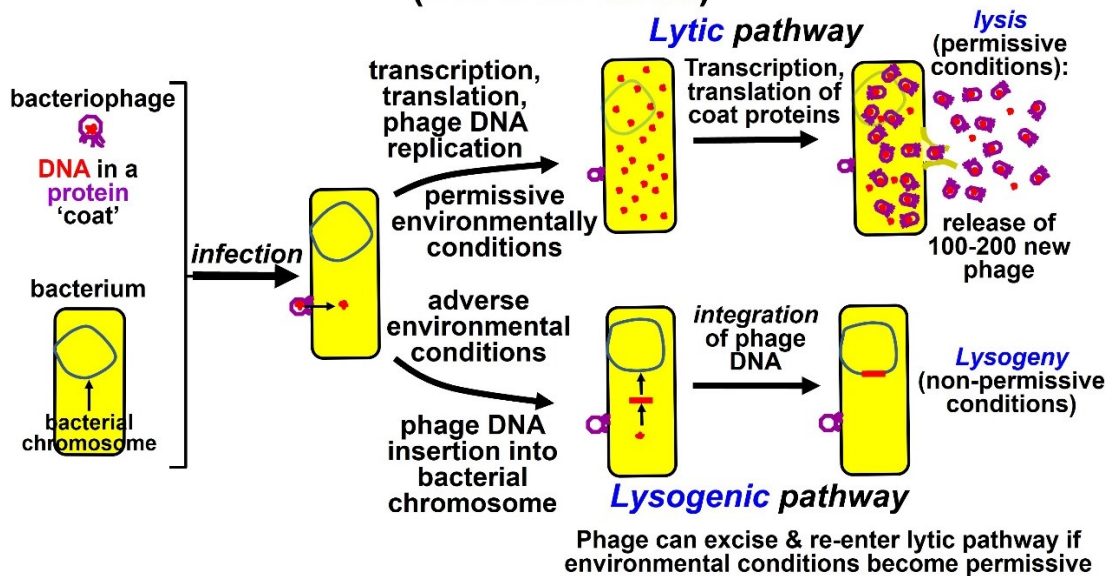


Fig. 14.15: Life-cycle options for a bacteriophage: The *lytic pathway* ends with cell lysis and phage release. In the *lysogenic pathway*, the phage DNA becomes part of the host-cell chromosome.

Structure & Life-Cycle Choices of *Bacteriophage Mu*

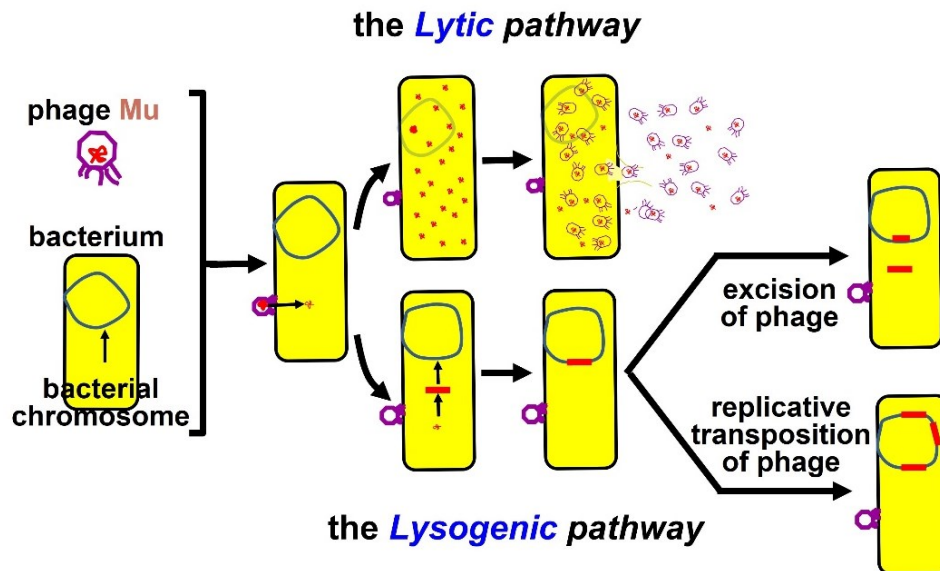


Fig. 14.16: The transposon (third) option for *Mu* phage: After lysogeny the phage DNA can replicate as part of the host-cell chromosome and excise to produce new phage, or it can transpose itself to new locations in the host cell chromosomal DNA circle.



[246 Bacterial Mobile Elements](#)



As we describe eukaryotic transposons, look for similarities to bacterial IS and Tn elements.

14.5 Overview of Eukaryotic Transposable Elements

There are two classes of transposons in eukaryotes:

1. **Class I (RNA) transposons** move (i.e., jump) by transcription of RNA at one locus, followed by reverse transcription and integration of the cDNA back into genomic DNA at a different location. Called **retrotransposons**, they may be derived from (or be the source of) retroviruses, since active retroviruses excise from and integrate into DNA much like retrotransposons. **Retroposons** are a subclass of retrotransposons (see Table 14.2).
2. **Class II (DNA) transposons** move by either of two mechanisms. In the **cut-and-paste pathway**, the transposon leaves one locus and integrates at another. In the **replicative pathway**, the original transposon remains in place, but new copies are mobile. Table 14.2 summarizes the distribution and proportion of genomes represented by different classes and types of transposable elements.

Table 14.2

Proportions of DNA and RNA transposons in Different Organisms

Typical Organism	Class I (RNA) Transposons as a % of all transposons	Class I (RNA) Transposons as a % of Genome	Class II (DNA) transposons as a % of all transposons	Class II (DNA) transposons as a % of Genome	All Transposons as a % of genome
Bacteria (<i>E. coli</i>)	-----	-----	100%	~3%	~3%
Yeast	100%	~3.5%	-----	-----	~3.5%
Corn (<i>Z. maize</i>)	>50%	~30-45%	<50%	~40%	~70-95%
Protozoa (<i>T. Vaginalis</i>)	-----	-----	100%	~66%	~66%
Frog (<i>R. esculenta</i>)	~25%	~19%	~75%	~58%	~77%
Mouse (<i>M. Musculus</i>)	~95%	~38%	~5%	~2%	~40%
Mosquito (<i>A. aegypti</i>)	~30%	~14%	~70%	~33%	~47%
Us! (<i>Homo sapiens</i>)	>90%	~40%	<10%	<5%	~42-45%
Flatworm (<i>C. elegans</i>)	~5%	~0.5%	~95%	~11-12%	~12%
Fruit fly (<i>D. Melanogaster</i>)	<80%	~3%	>20%	~1%	~4%
Rice (<i>O. sativa</i>)	~15%	~1%	~85%	~5%	~6%

The table confirms that bacteria contain relatively few transposons, in contrast to eukaryotes, which vary widely in *transposon load* (transposons as a percentage of genomic DNA). Transposon load can range from as low as 4% to more than 70% in different organisms.

Table 14.3 summarizes transposable elements by class, subtype, size, genomic distribution, mechanism of transposition, where they are found in nature, and other unique characteristics.

Table 14.3

Types, Some Characteristics, and Distribution of Transposable Elements

Class	Type	Sub-Types	Basis of Mobility	Organismic Distribution (e.g.)	Examples	All, as % of Genome	Length (bp)	Genomic Location; Special Features
Prokaryotic Mobile Elements	*DNA transposons	IS (Insertion Elements)	Cut-&Paste	<i>E. coli</i>	<i>IS1, IS2, IS3...</i>	3%	1000-2000bp	Intergenic DNA; usually insertion-site specific
		Composite Tn	Cut-&Paste		<i>Tn5</i>		~5900bp	Source of antibiotic resistance genes
		Complex Tn	Replicative, cointegrate formation		<i>Mu</i>		~37Kbp	Can function as bacteriophage or transposon
Eukaryotic Class II (DNA) transposons: **Move via DNA intermediates	*DNA transposons	-----	Cut-&Paste	<i>D. melanogaster</i>	<i>P-element, Mariner</i>	both = ~1%	1000-7000bp	Dispersed
				<i>C. elegans, Z. maize</i>	<i>Tc1 (Mariner), Ac, Ds</i>	~11-12%		
				<i>H. Sapiens (human)</i>	<i>Mariner</i>	~2-5%		
		-----	Replicative, cointegrate formation	<i>O. Sativa (rice) [also, plants, bacteria]</i>	<i>Miniature Inverted Repeat Transposable Elements (MITEs)</i>	~6% (rice)	<500bp	Mostly associated with genes; transcribed into small RNAs
Eukaryotic Class I (RNA) transposons: **Move via RNA intermediates	*retro-transposons, retroposons	<i>LTR retrotransposons</i>	Reverse-transcription & integration (original copy not excised)	<i>S. cerevisiae (yeast)</i>	<i>Ty</i>	~3%	Up to 7000bp	Dispersed; (retrovirus-like, but no envelope protein genes)
				<i>D. melanogaster</i>	<i>Copia</i>	~3%		
		<i>LINEs (NON-LTR retrotransposons)</i>		<i>H. Sapiens (human)</i>	<i>L1</i>	~5%	~6000bp	Interspersed
				<i>Z. maize</i>	<i>Cin-1</i>	~45%		
		<i>SINEs: NON-LTR retrotransposons, or retroposons</i>		<i>H. Sapiens (human)</i>	<i>Alu</i>	~40%	80-400 bp	

* **DNA transposons** move from one place to another in one of two ways. In **Cut & Paste Transposition**, the element excises and moves to another location in the genome. In **Replicative Transposition**, DNA transposons are copied and the copy transposes to a new location, leaving the original element in place. **Retrotransposons** are active if their transcripts are reverse transcribed into cDNAs as well as translated into the enzymes required to integrate their cDNA copies into genomic DNA.

** Many transposons are inactive, having been silenced by mutation or other factors. Active eukaryotic Class I or Class II transposons are either **autonomous** or **non-autonomous**. **Autonomous transposons** have all of the structural features necessary for their transposition (e.g., the maize *Ac element*). On the other hand, **non-Autonomous transposons** have all the structural elements of autonomous transposons (e.g., inverted repeats and other DNA necessary for transposition), except that they lack or can't transcribe one or more of the genes for enzymes needed for mobility (e.g., the maize *Ds element*). Nevertheless, they can be mobilized with the assistance of an actively transposing autonomous element that can provide the missing enzymes.

Between the two tables above, we can conclude the following:

- *Transposon load* is not correlated with the evolutionary complexity of organisms.
- Shared transposons have different evolutionary histories in different organisms.
- Where transposons remain active, they continue to shape genomic landscapes, especially in organisms with a high transposons load.

We will revisit some of these conclusions later, after looking at the structure and mechanism of mobility of different eukaryotic transposable elements.



[247 Introduction to Eukaryotic Transposons](#)

14.6 Structure of Eukaryotic DNA (Class II) Transposons

Active eukaryotic *DNA transposons* share structural features with bacterial mobile elements, including genes required for transposition, flanking inverted repeats, and flanking insertion-site direct repeats of host-cell DNA. As we'll see, Class II transposons can "jump" by **cut-and-paste** or **replicative** mechanisms.

Figure 14.17 illustrates the characteristic structure of a eukaryotic DNA transposon, including its "all-purpose" *transposase* gene.

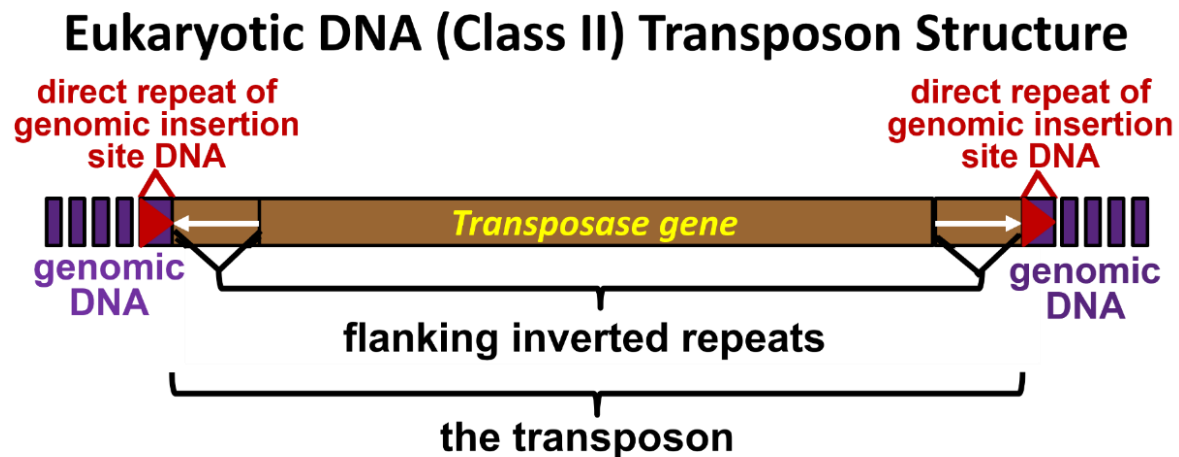


Fig. 14.17: Structure of a eukaryotic *Class II* DNA transposon.

14.6.1 Cut-and-Paste Transposition

This mechanism moves a copy from one location and *transposes* it to another location. As its name suggests, *replicative* transposition leaves a copy of the original transposon in place, while inserting a new copy elsewhere in the genome (Figure 14.18, below).

Cut & Paste Transposition of a DNA (Class II) Transposon: an Overview

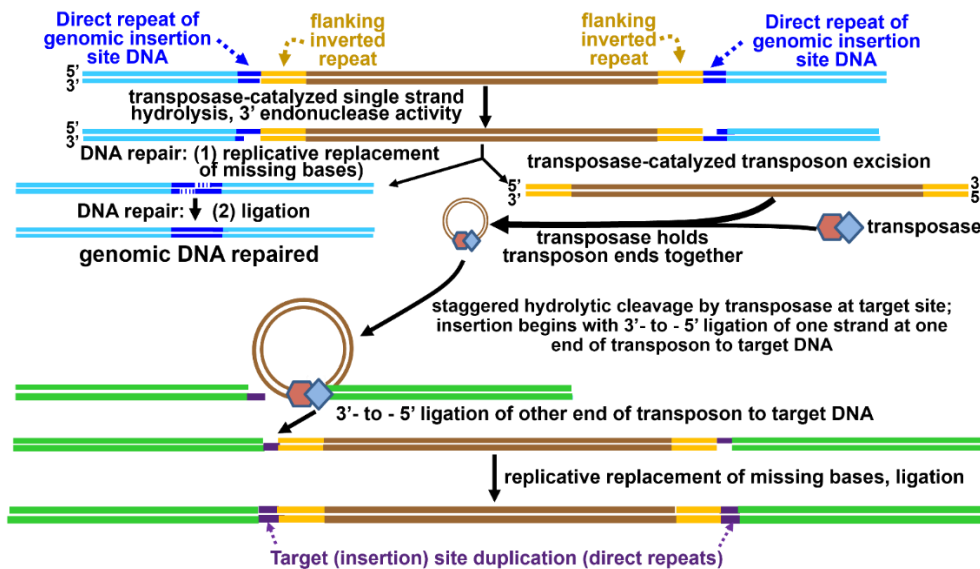


Fig. 14.18: Cut-and-paste transposition of a DNA transposon.

Transposase excises the transposon, trims its 3' OH ends to create a staggered cut, joins the transposon ends, and mediates its insertion at a new DNA site. After ligating the 3' OH transposon ends to the 5' OH at the insertion site, replication replaces missing bases to generate the direct repeats of host-cell genomic DNA. Ligation completes the transposition.

14.6.2 Replicative Transposition

Figure 14.19 details the steps of replicative transposition. Like the cut-and-paste mechanism, the transposon is nicked and trimmed at its source (i.e., original) insertion site.

Replicative Transposition of a DNA (Class II) Transposon: an Overview

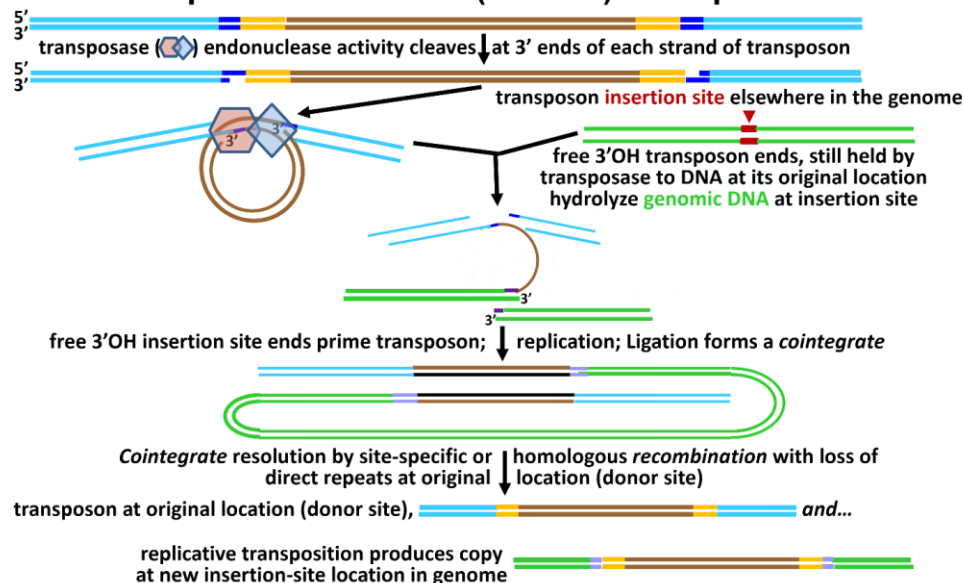


Fig. 14.19: Replicative transposition of a DNA transposon.

Unlike the cut-and-paste mechanism, transposons are not excised. Its transposase nicks the transposon's 3' ends at the insertion site, holding the transposon ends together while it cuts (hydrolyzes) DNA at a new insertion site. Replication from the 3' OH ends of the insertion-site DNA strands forms a *cointegrate* structure containing transposon copies. The *cointegrate* is resolved by one of two recombinational mechanisms, leaving transposon copies at the original site and the new insertion site.

Let's compare and contrast features of cut-and-paste and replicative transposition. The **common features** are that (first) a transposon-encoded *transposase* binds, brings transposon ends together, and catalyzes single-stranded cleavage (hydrolysis), leaving "staggered ends," and (second) that the transposase holds the transposon ends together for the remaining steps. The **differences** between the two mechanisms are that in *cut-and-paste* transposition, the transposon is completely excised and then transposed to a new site in genomic DNA. In contrast, after single-stranded cleavage in replicative transposition, the transposase-bound free 3' ends of the transposon hydrolyze both strands of double-stranded DNA at a new insertion site. After ligation of the 3' ends of transposon strands to the 5' ends of cut genomic DNA insertion-site ends, the remaining 3' ends of the insertion-site DNA will prime the replication of the transposon to form the *cointegrate*, the which is then resolved by one of two recombination pathways.



[248 Eukaryotic Class II \(DNA\) Transposition](#)

14.7 Structures of Eukaryotic RNA (Class I) Transposons

Like DNA transposons, all RNA transposons leave insert-site footprints (i.e., *direct repeats* of genomic DNA flanking the element). Unlike DNA transposons, active eukaryotic Class I transposons move via an RNA intermediate. Also unlike DNA transposons, RNA transposons lack terminal inverted repeats.

The mobility of the RNA intermediate of all retrotransposons requires a **promoter**, which recognizes a **reverse transcriptase** enzyme as well as endonuclease and integrase enzymes (to be described later). **Autonomous** Class I RNA transposons are characterized by **Long Terminal Repeats (LTRs)** as well as **Non-LTR retrotransposons** (that lack LTRs). Non-LTR retrotransposons include the **autonomous Long Interspersed Nuclear Elements (LINEs)** and the **non-autonomous Short Interspersed Nuclear Elements (SINEs)**. Both the *autonomous LTR* and *non-LTR LINEs* contain and express genes needed for the enzymes required for transposition. On the other hand, the *nonautonomous SINEs* (a subclass of non-LTR retrotransposons) lack genes for the enzymes required for transposition and therefore cannot transpose independently. *Nonautonomous* retrotransposons thus rely on "true" (*autonomous*) retrotransposon activity for mobility. SINEs are sometimes called **retroposons** to distinguish them from the autonomous retrotransposons.



[249 Introduction to Features of Retrotransposition](#)

14.7.1 LTR retrotransposons: The Yeast Ty element

Here we look at structures, genes, and enzyme activities required for *retrotransposition*. The Ty LTR retrotransposon in Figure 14.20 is an autonomous retrotransposon.

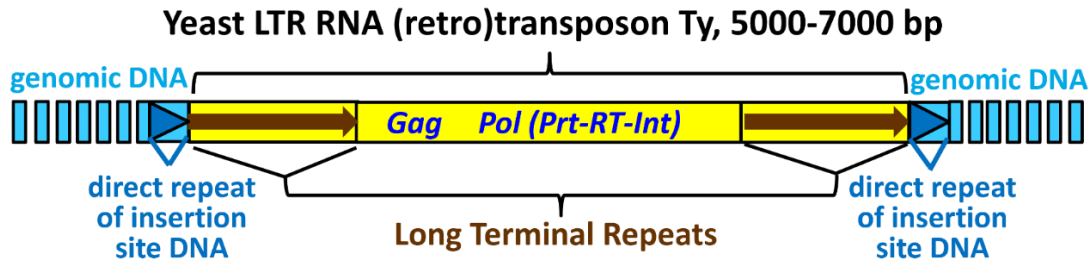


Fig. 14.20: Structure of yeast Ty, an RNA LTR retrotransposon

The Ty retroposon encodes several genes needed for transposition:

- **Gag** encodes *group-specific antigen*, a protein that forms a virus-like particle that will contain reverse-transcribed transposon DNA.
- **RT**, encodes *reverse transcriptase* to reverse-transcribe retrotransposon RNAs.
- **Prt** encodes a protease to break down the virus-like particle as the retrotransposon enters the nucleus.
- **Int** encodes *integrase*, needed to integrate the retrotransposon into a new insertion site. In fact, many of the events in Ty transposition occur in the cytoplasmic “virus-like particle” in yeast cells. To see more, see ^{14.2}[Virus-Like Particles in Ty Transposition](#). Note that the Pol region in Figure 14.20 consists of overlapping *open reading frames (ORFs)* that encode the Prt, RT, and Int genes. The ready-to-move transposon consists only of the bracketed region of DNA, including the long terminal repeat sequences.



[250 LTR Retrotransposons: the TY Element](#)

14.7.2 Non-LTR Retrotransposons: LINEs

Like LTR retrotransposons, **LINEs** (*long interspersed nuclear elements*) encode the enzymes needed for transposition. They also generate target-site direct repeats flanking the inserted element (Figure 14.21). But they do not have long terminal repeats! Instead, their ORFs (genes) are flanked by 5' and 3' *untranslated regions (UTRs)*.

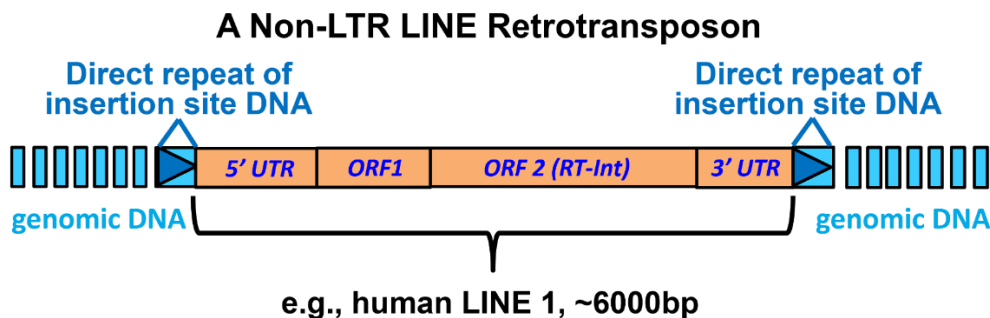


Fig. 14.21: Structure of a LINE, a non-LTR retrotransposon.

The 5' UTR contains a promoter from which cellular RNA polymerase II can transcribe the downstream genes (see chapter 10 on *transcription*). The second of these (*ORF2*) encodes the reverse transcriptase and an integrase activity that are essential for transposition of the LINE. All autonomous Class I (RNA-intermediate) retrotransposons share the following features:

- A *promoter* in the 5' UTR from which they can be transcribed
- A *reverse transcriptase* that generates a cDNA copy of the transposable element
- An *RNAse H* (an endonuclease) that degrades the transcript after reverse transcription
- An *integrase* (like a transposase) that catalyzes insertion of the retrotransposon copy at insertion sites



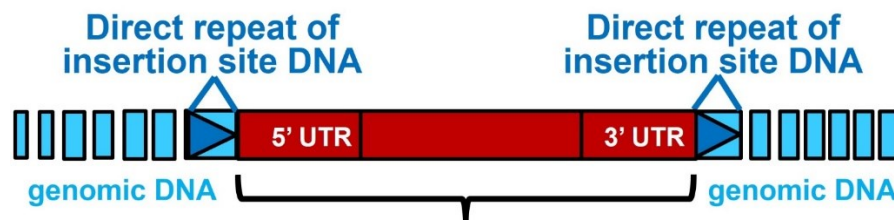
[251 Non-LTR Retrotransposons: LINEs](#)



14.7.3 Non-LTR Retrotransposons: SINEs

Non-LTR SINE retrotransposons are non-autonomous, typically lacking genes. However their nongenic DNA is flanked by 5' and 3' UTRs. RNA polymerase III can transcribe SINEs, but to transpose, SINEs rely on the concurrent activity of an autonomous, non-LTR transposon (a LINE) to provide the requisite enzymatic activities. The ~300bp human *Alu* element in Fig. 14.22 is a typical SINE (*short interspersed nuclear element*).

A SINE retrotransposon



Human *Alu* SINE flanked by direct repeats

Fig. 14.22: Structure of the *Alu* SINE—a *non-LTR retrotransposon*, or *retroposon*.



[252 Non-LTR Retrotransposons: SINEs](#)



14.8 Mechanisms of Retrotransposition

There are two mechanisms of retrotransposition. They are *extrachromosomally primed retrotransposition* (seen in e.g., LTR retrotransposons) and *insertion-target-site-primed retrotransposition* (seen in non-LTR retrotransposons like LINEs and SINEs). We will consider these two mechanisms next.

14.8.1 Extrachromosomally Primed Retrotransposition (e.g., of a LINE)

As its name suggests, in *extrachromosomally primed retrotransposition*, a separate circular reverse transcript of the retrotransposon attacks, nicks, and integrates into a genomic insertion site (Figure 14.23). In this mechanism, reverse transcriptase creates a cDNA copy of a transcribed retro-element. Integrase-endonuclease then binds the cDNA copy, holding the ends together, in effect circularizing it. This isolable ribonucleoprotein resembles an *intasome*, a structure similar to the nucleoprotein complex that catalyzes integration of retroviral cDNAs during lysogeny. The 3D structure of a retroviral *intasome* interacting with DNA and nucleosomes was recently determined (for more, see ^{14.3}[Retroviral Intasome 3D Structure](#)). In this form, the retrotransposon attacks DNA at an insertion site, creating staggered ends. After insertion, the gaps in the DNA are filled in. Ligation seals the retrotransposon into its new location, creating direct insertion-site repeats.

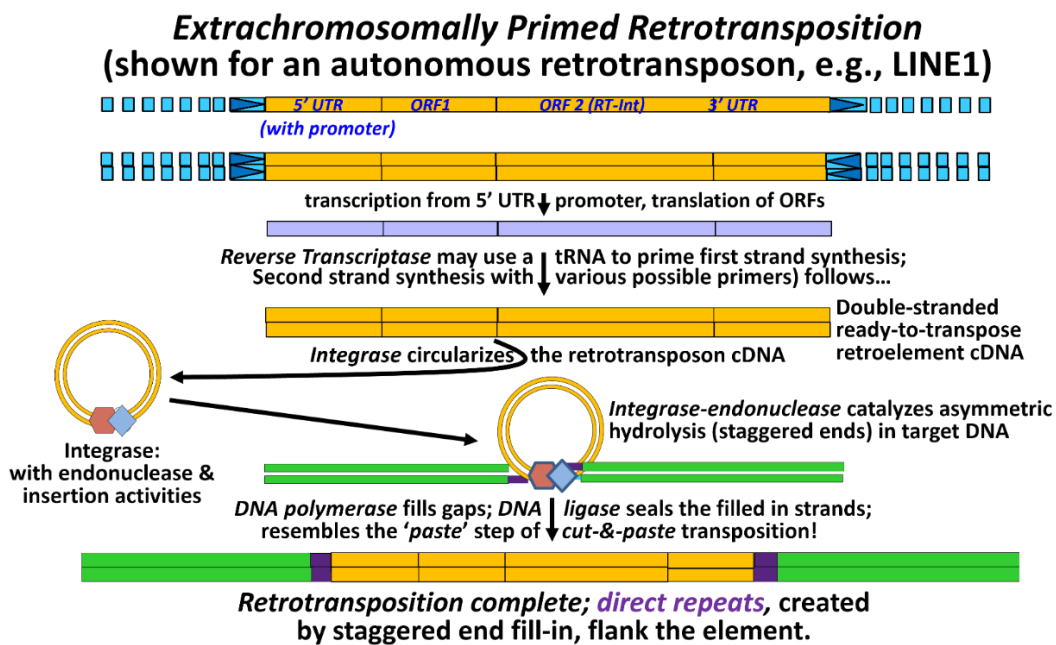


Fig. 14.23: Extrachromosomally primed transposition of a LINE.



[253 Extrachromosomally Primed Retrotransposition](#)

14.8.2 Target-Site Primed SINE Retrotransposition (e.g., of a SINE)

In *target-site primed retrotransposition* (*retroposition*) of a SINE, RNA polymerase III (the same enzyme that catalyzes tRNA and 5S rRNA transcription) transcribes the SINE. If a LINE is concurrently transcribed, its enzymes will be made. When its *integrase-endonuclease* catalyzes hydrolysis of one strand of DNA at a new insertion site, the 3' OH end of this strand can prime reverse transcription of the one SINE cDNA strand by the LINE *reverse transcriptase*. After hydrolysis of the second target-site DNA strand, its 3' OH end primes replication of the second strand of the SINE cDNA. Integrase then completes insertion of the copy-SINE into its new genomic location.

The key feature of *target site-primed retrotransposition* (*retroposition*) is the absence of an integrase-bound, separate, circular, double-stranded reverse transcript (Figure 14.24).

Target-primed retrotransposition (shown for a SINE)

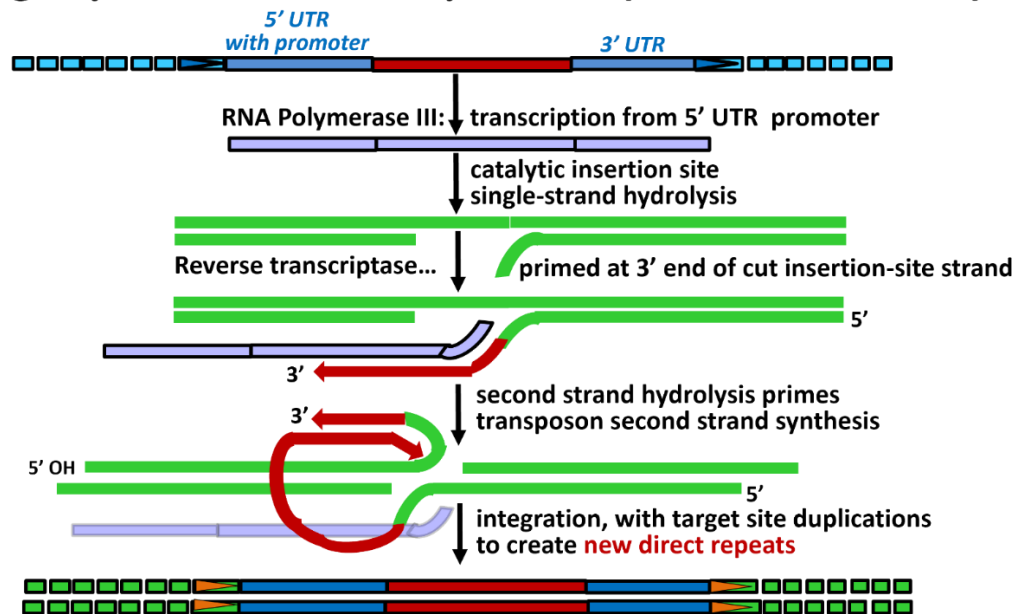


Fig. 14.24: Target-primed transposition of a LINE.



[254 Target-Primed Retrotransposition](#)



14.9 On the Evolution of Transposons, Genes, and Genomes

We noted that transposons in bacteria carry antibiotic resistance genes, a clear example of the benefits of transposition in prokaryotes. Of course, prokaryotic genomes are small, as is the typical bacterial transposon load. Yeast species also have low transposon load. But what can we make of the high transposon load in eukaryotes? To many geneticists, the fact that genes encoding proteins typically represent only 1–2% of a eukaryotic genome once suggested that the rest of the genome was informationally nonessential. Even though transposons turned out to be much of the noncoding DNA in some eukaryotic genomes, they seemed to serve no purpose other than their own replication. These large amounts of transposon DNA were dubbed “selfish DNA” and their genes, “selfish genes.”

So, are transposons just junk DNA—some kind of invasive or leftover *genomic baggage*? Given their propensity to jump around and their potential to raise havoc in genomes, how do organisms tolerate and survive them? Is the sole “mission” of transposons really just to reproduce themselves? Or are transposons in fact neither selfish nor junk? By their sheer proportions and activity in eukaryotic genomes, we will see that transposons have dispersed into and reshaped genomic landscapes. Does their relocation and dispersal of transposons in a genome (with the resulting disruption of genes and structural reorganization of genomes) have any functional or evolutionary value?

All of these questions are reasonable responses to the phenomena of jumping genes. A rational hypothesis might be that, like all genetic change, the origin of transposition was a random accident. But the spread and ubiquity of transposons in genomes of higher organisms must in the long term have been selected by virtue of some benefit they provide to their host cells and organisms.

Let's briefly look at the evolutionary history of transposons to see if this assumption has some merit.

14.9.1 A Common Ancestry of DNA and RNA Transposons

Transposases catalyze transposition of bacterial IS (and related) elements and eukaryotic Class II (DNA) transposons. They are structurally similar and may share a common ancestry. Figure 14.25 compares the amino-acid sequences of the so-called *integration domain* of transposases from different transposons.

Comparison of Aligned Transposase Sequences from a Bacterium, a Bacteriophage, and a Eukaryote

```
IS consensus  ...t--De--w---/-----tDd--r-----/-----s-----iEr-n-l---
Mu phage      ...t--D-----/-----tDn-r-----/-----iEr-----
Tc1 mariner   ...t--De--w---/-----Dn-----/-----s-----iE---l---
```

Letters identify evolutionarily conserved amino acids

Fig. 14.25: The alignment of *consensus* amino-acid sequences among bacterial *IS* transposases, *Mu phage*, and *Tc1 mariner* transposases reveals conservation of **D**, **D**, and **E** amino acids (uppercase) at key positions in the sequence. Other amino acids shared between some but not all of the sequences are in lowercase. Slashes denote variable-length gaps in the alignments.

A universally conserved alignment of **D...D...E** amino acids—at key positions in different transposase enzymes—defines a **DDE domain**, supporting the common ancestry of bacterial and eukaryotic transposases. This, along with shared structural features of these transposons (e.g., the flanking direct insertion-site repeats) further support a common ancestry of the transposons encoding the enzymes.

Other sequence comparisons reveal that the transposons themselves comprise distinct families of more closely related elements. This allows us to speculate on the origins of these families in different species. For example, the TC1/*mariner* (DNA) transposon is found in virtually all organisms that have been examined (except diatoms and green algae). Based on sequence analysis, there is even an insertion element in bacteria that is related to the *mariner* element. Clearly, *mariner* is an ancient transposon.

This amount and diversity of conservation bespeaks an early evolution of the enzymes of transposition and of transposition itself within (and even between) species. **Linear** or **vertical descent** (the “vertical” transmission of transposons from parents to progeny) is the rule. However, **horizontal transfer** best explains the presence of similar, related transposons in diverse species.

In contrast to linear descent of sequences within a species, **horizontal transfer** defines an interspecific sharing of DNA. That is, a transposon in one organism must have been the "gift" of an organism of a different species!

Wherever their origins, moveable *genes* have been a part of life for a long time, speaking more likely to their adaptive value for organisms than to the notion that they simply parasitic, selfish, rogue DNAs only out for their own replication!

14.9.2 Retroviruses and LTR Retrotransposons Share a Common Ancestry

The *integration domains* of retrotransposon transposases and retrovirus integrases also share significant sequence similarities, as shown in the amino-acid sequence alignment below (Figure 14.26).

Comparison of Aligned Transposase Sequences with Integrase Sequences

Transposases

IS consensus ...t--De--w---/-----tDd-r-----/------s-----iEr-n-l---
 Mu phage ...t--D-----/-----tDn-r-----/------iEr-----
 Tc1 mariner) ...t--De--w---/-----Dn-----/------s-----iE---l---

Integrases

COPIA ...v-sD-----/---lyiDngs-ft-----/------sErn-rti-e
 HIV-1 ...-qvD-----/---vhtDngr-y-----/------iEsmnk-lk-

Fig. 14.26: A comparison of amino acid sequences of the *COPIA* retrotransposon and a retroviral (HIV) integrase with typical transposase sequences: the alignments reveal conservation of the **D**, **D**, and **E** amino acids in the *DDE* domain of the enzymes. Other amino acids are shared between some but not all of the sequences (lowercase). Slashes denote variable gaps in the alignments.

Here the alignment of conserved **D...D...E** amino acids at key positions in *DDE domains* supports the common ancestry of Class I (RNA transposon) retroviral integrases and Class II (DNA transposon) transposases. In other words, Both DNA and RNA transposons may share a common ancestry.

The common ancestry of retrotransposons and retroviruses raises yet other questions: did transposons (specifically retrotransposons) arise as defective versions of integrated retrovirus DNA (i.e., reverse transcripts of retroviral RNA)? Or did the retroviruses emerge when retrotransposons evolved a way to leave their host cells? To approach this question, let's first compare mechanisms of retroviral infection and retrotransposition.

LTR retrotransposons and retroviruses both contain flanking long terminal repeats in addition to the structural similarities of the enzymes they encode. But retrotransposition occurs within the nucleus of a cell, while retroviruses must first infect a host cell before the retroviral DNA can be replicated and new viruses produced.

A key structural difference between retrotransposons and most retroviruses is the *ENV* gene-encoded protein envelope surrounding retroviral DNA. After infection, the incoming retrovirus sheds its envelope proteins, and viral RNA is reverse-transcribed.

After the reverse transcripts enter the nucleus, transcription of genes and translation of enzymes necessary for the replication of the viral cDNA leads to the production of new enveloped infectious viruses that will eventually lyse the infected cell. But here are two curious phenomena:

First, retroviral DNA, like any genomic DNA, is mutable. If a mutation inactivates one of the genes required for infection and retroviral release, it could become an LTR retrotransposon. Such a genetically damaged retroviral *integrate* might still be transcribed, and its mRNAs might still be translated. If detected by its own reverse transcriptase, the erstwhile viral genomes would be copied. The cDNAs, instead of being packaged into infectious viral particles, would become a source of so-called *endogenous retroviruses (ERVs)*. In fact, ERVs exist and make up a substantial portion of the mammalian genome (8% in humans)—and they do, in fact, behave like LTR retrotransposons!

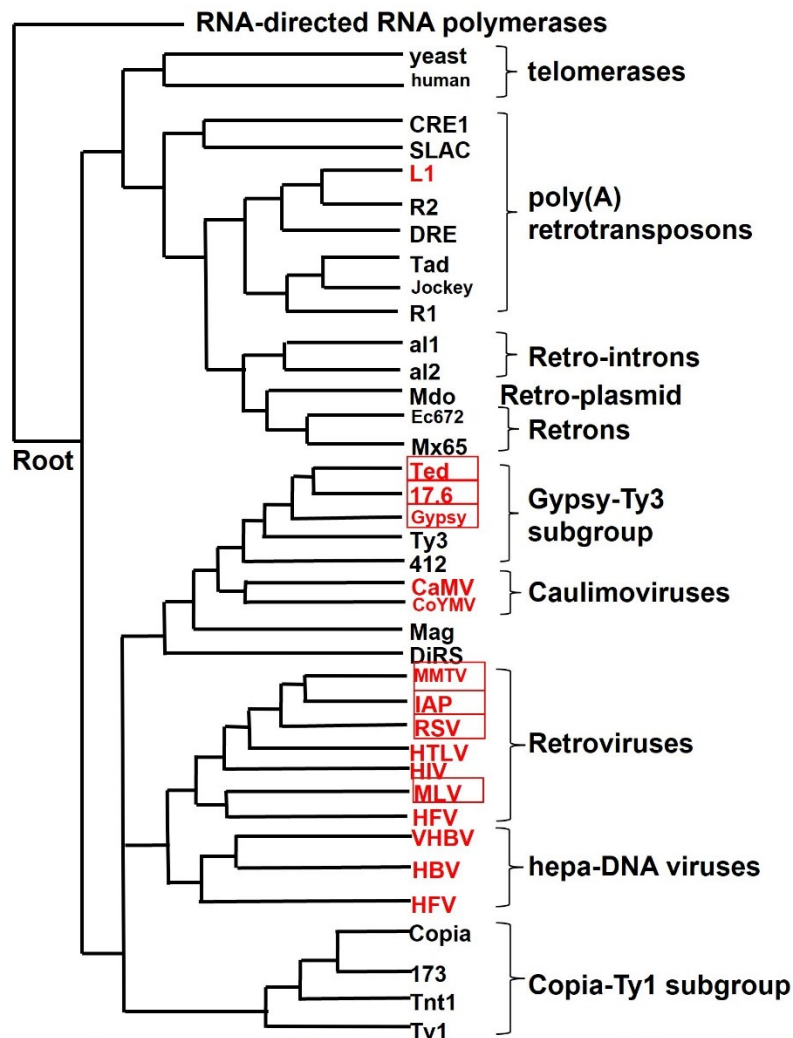
Second, yeast TY elements transcribe several genes during retrotransposition (see the list in *section 14.7.1* above) to produce not only a reverse transcriptase and an integrase but also a protease and a structural protein called **Gag** (*Group-specific antigen*). All the translated proteins enter the nucleus. Mimicking the retroviral ENV protein, the Gag protein makes up most of a coat protein called **VLP** (*virus-like particle*). VLP encapsulates additional retrotransposon RNAs in the cytoplasm, along with the other proteins. Double-stranded reverse transcripts (cDNAs) of the viral RNA are then made within the VLPs. But instead of bursting out of the cell, these encapsulated cDNAs (i.e., new retrotransposons) shed their VLP coat and reenter the nucleus, where they can now integrate into genomic target DNA. Compare this to the description of retroviral infection. During infection, retroviral envelope proteins attach to cell membranes and release their RNA into the cytoplasm. There, the reverse transcriptase copies viral RNA into double-stranded cDNAs that then enter the nucleus, where they can integrate into host-cell DNA. When transcribed, the integrated retroviral DNA produces transcripts that are translated in the cytoplasm into the proteins necessary to form an infectious viral particle. The resulting viral RNAs are encapsulated by an ENV (envelope) protein encoded in the viral genome.

Of course, unlike VLP-coated retrotransposon RNAs, the enveloped viral RNAs do eventually lyse the host cell, releasing infectious particles. Nevertheless, while VLP-coated Ty elements are not infectious, they sure do look like a retrovirus!

In much the same way as early biologists compared the morphological characteristics of plants and animals to show their evolutionary relationships, comparisons of aligned retroviral and retrotransposon reverse transcriptase gene DNA sequences reveal the phylogenetic relationships of these genes, and therefore of the elements themselves.

A phylogenetic tree of retrotransposon and viral genes is shown in Figure 14.27 (below). The data in this analysis supports the evolution of retroviruses from retrotransposon ancestors. From the "tree," TY3 and a few other retrotransposons share common ancestry with *Ted*, *17.6*, and *Gypsy* ERVs (shown boxed in red in Figure 14.27) in the Gypsy-TY3 subgroup. Further, this subgroup shares common ancestry with more-distantly-related retroviruses (e.g., *MMTV* and *HTLV*), as well as the even-more-distantly-related (older, longer-diverged!) *Copia-TY1* transposon subgroup.

Rooted Phylogenetic Tree of Reverse Transcriptase (RT) Gene Family



Viruses are shown in red. Boxed sequences include endogenous retroviruses (ERVs).

Conclusions:

- Reverse transcriptase and telomerase sequences share a common ancestry dating long before the appearance of RNA viruses.
- Reverse transcriptases sequences share common ancestry with RNA transposon sequences.

Therefore...,

- RNA viruses (L1 and other RNA retroviruses specifically), must have appeared after the evolution of RNA-directed RNA polymerase enzymes.

Fig. 14.27: Retroviral and retrotransposon reverse transcriptases share a common evolutionary ancestor.

This and similar analyses strongly suggest that retroviruses evolved from a retrotransposon lineage. For a review of retroposon and retrovirus evolution, see Lerat P. & Capy P., 1999. *Retrotransposons and retroviruses: analysis of the envelope gene*. Mol. Biol. Evol. 19: 1198-1207.

14.9.3 Transposons Acquisition by *Horizontal Gene Transfer*

As noted, transposons are inherited **vertically**, meaning that they are passed from cell to cell or parents to progeny by reproduction. But they also may have spread *between species, in which an individual of species A inadvertently picks up a transposon from species B* or even an individual of the same species), becoming *transformed* and adding a new transposon to their genome. Accidental mobility of transposons between species would have been rare, but when it occurred, an exchange of genes by horizontal gene transfer would have accelerated with the evolution of retroviruses. Once again, despite the potential to disrupt the health an organism, retroviral activity might also have supported a degree of genomic diversity useful to organisms.



[255 Transposon Evolution](#)

14.10 Evolutionary Roles of Transposition in Genetic Diversity

Here we'll see how transposition can affect genes and genetic diversity. We'll also look at parallels between transposition and the generation of immunological diversity, and we'll consider the provocative notion that our immune system owes at least some of its evolutionary history to transposons, or at least to mechanisms of transposition!

14.10.1 Transposons and *Exon Shuffling*

A role for unequal recombination in moving exons in and out of eukaryotic split genes was described earlier. This kind of **exon shuffling** could happen when short DNA sequences in two different introns misalign during meiotic synapsis, allowing unequal crossing over. Expression of a gene with a "new" exon produces a protein with a new domain and a new activity. If the new domain is not harmful (e.g., found in only one of two alleles of a gene), the mutation need not be lethal, and genetic diversity and the potential for evolution is increased!

Transposons embedded in introns are long regions of DNA similarity that can stabilize synapsis, increasing opportunities for unequal alignment of chromosomal DNA and therefore, chances for recombination and exon shuffling. For example, **Alu** (SINE) elements often integrate within introns with no ill effect. The similarity of Alu elements in the introns of unrelated genes seems to account for exon shuffling by unequal crossing over between the different genes that share domains and, thus, specific functions.

Another way in which transposons facilitate germline-cell exon shuffling is more direct. Imagine a pair of transposons in introns of a gene on either side of an exon. Should such transposons behave like the two outer IS elements in a bacterial *Tn element* (discussed earlier), they might be excised as a single, large transposon containing an exon.

The paired transposons flanking the exon might then insert in an intron of a completely different gene! The general pathway of exon shuffling involving paired proximal DNA transposons is illustrated in Figure 14.28.

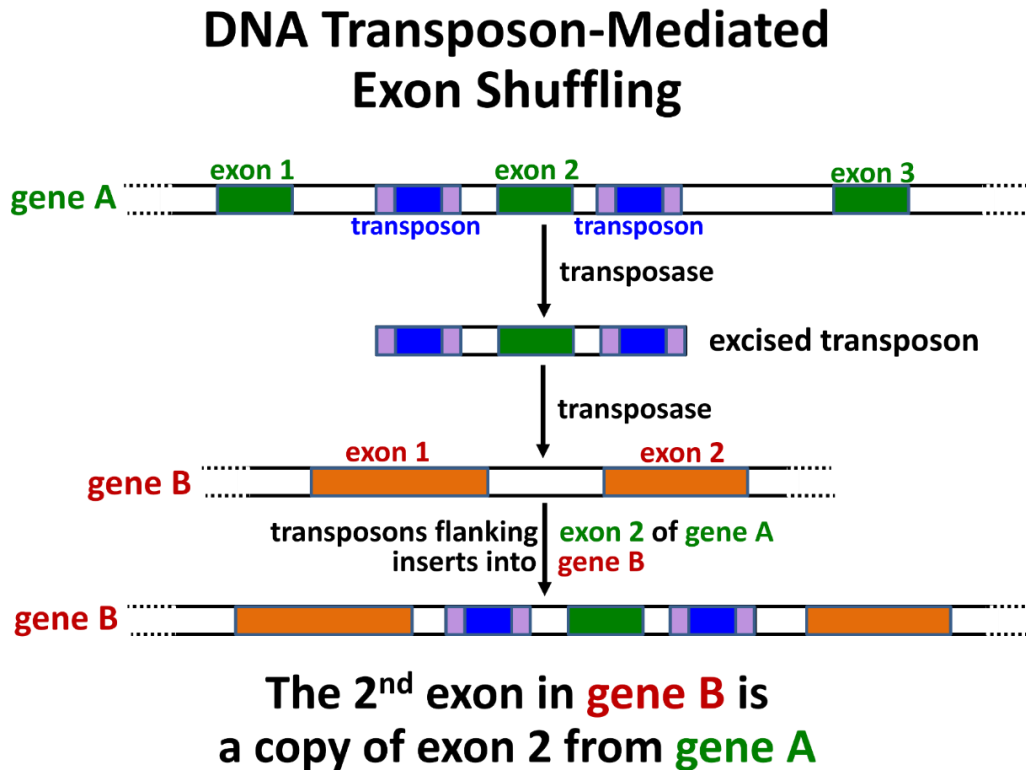


Fig. 14.28: Steps of paired DNA transposon-mediated exon shuffling.

In this generic example of exon shuffling, exon 2 of gene 1 has been inserted (along with flanking transposons) into another gene (gene 2). Transposon-mediated exon shuffling can explain the insertion of exon-encoded domains of *epidermal growth factor (EGF)* into several otherwise-unrelated genes. The mitogen EGF was discovered because it stimulated skin cells to start dividing. The gene for *TPA* (*tissue plasminogen activator*, a blood-clot-dissolving *protease*) shares domains with the EGF gene. TPA is a treatment for heart attack victims, and if it is administered rapidly after the attack, it can dissolve the clot and allow coronary artery blood flow to heart muscle to resume. Other genes that contain EGF domains include those for Neu and Notch proteins, both involved in cellular differentiation and development.

Some exon shuffling mutations may have been mediated by LINE transposition and by a special group of recently discovered transposons called *helitrons*. Helitrons replicate by a *rolling-circle* mechanism. If you are curious about helitrons, do a Google search to learn more about them and what role they may have had in refashioning and reconstructing genomes in evolution.

14.10.2 Transposon Genes and Immune-System Genes Have History

Several important eukaryotic genes may have been derived from transposons. Perhaps the most intriguing example of this is to be found in the complex vertebrate *immune system*. Our

immune system includes ***immunoglobulins (antibodies)***. You inherited your immunoglobulin genes for immunoglobulin proteins from your parents. These genes contain multiple variant ***V***, ***D***, and ***J*** regions linked to a ***C*** region. *V*, *D*, *J*, and *C* are defined as *variable*, *joining*, *diversity*, and *constant* DNA regions, respectively. These regions recombine to create many diverse *V-D-J-C* immunoglobulin antibody molecules (although the *D* region is not always included in the recombined gene). The recombinations occur during the maturation of bone marrow stem cells to become immune cells (***B*** or ***T lymphocytes***). In response to a challenge by foreign ***antigens***, (e.g., proteins on a bacterial surface or toxins released by invading cells), our immune system will select *B* or *T lymphocytes* that contain ***rearranged immunoglobulin genes***, which are able to make immunoglobulins that can recognize, bind, and eliminate the invading antigens.

A discussion of the molecular biology of the immune system is beyond our scope here. Suffice it to say that the recombinational immunoglobulin gene rearrangements include enzymatic activities very similar to those of transposition. In fact, the so-called "RAG1 enzyme" active in immunoglobulin gene rearrangement is closely related to genes in a family of transposons (*Transib*) found in invertebrates and fungi. It looks like immune-gene rearrangement might have origins in transposition!

14.11 Coping with the Dangers of Rampant Transposition

Most organisms do not have the degree of high transposon load that we humans have. Given a general tendency of transposons to insert at random into new DNA loci, the question is, why don't they often insert and inactivate important genes? In other words, how come we exist at all? Isn't the danger of transposition into essential gene sequences magnified by the possibility of multiple simultaneous transpositions of elements that have been generated by cut-and-paste and (especially) replicative mechanisms? An obvious explanation for how we have survived transposon activity is that most transposition occurs in the greater-than-90% of the genome that does not code for proteins. But transposons *have* been found in genes that are inactive as a result. In this case, the explanation is that eukaryotic organisms have two copies of every gene, so if one is inactive, the other may sustain us. Beyond this, several mechanisms exist to silence a transposon after transposition has occurred, mitigating the dangers of rampant transposition. If a transposition is not lethal (e.g., its integration does not disrupt an activity essential to life), the cell and organisms can survive the event. In time, mutations at the ends of transposons or in the genes needed for transposition would eventually render them inactive.

Finally, there are more direct curbs on the activity of transposons. Recall the ***piRNAs*** we encountered earlier that participate in silencing transposons by blocking their export from the nucleus and silencing transposon genes themselves. There is also evidence that ***siRNAs*** target transposon transcripts for destruction. In sum, transposition effects are limited by inherent mutations and, more directly, by mechanisms that limit transposition and thus genetic damage.

Even though transposon mobility can be restrained, a last question remains. Recall the table showing many species tolerate high transposon loads. If transposition was *ipso facto* deleterious, then how do we explain that a single organism should have so many? Setting aside the fact that a few transposons have evolved gene regulatory functions, a high transposon count in eukaryotes is largely neutral. Therefore, many transposons in a species

may simply increase a pool of diversity that, over time, supports the selection of new genotypes and phenotypic characteristics. Thus, mobile DNA elements may be neither selfish nor junk. Check out ^{14.4}[Not Junk After All?](#) and ^{14.5}[Eulogy for Junk DNA](#) and the link below for more.



[256 Transposons - Junk or Not?](#)

Some iText & VOP Key Words and Terms

Alu	fertility (F) plasmids	McClintock
anthocyanins	Gag	nonhomologous recombination
antibiotic-resistance genes	genomic complexity	Non-LTR transposons
antibodies	heterochromatin	piRNA
Argonaute proteins	immune system	PIWI-Interacting RNA
autonomous transposon	immunoglobulin	protease
bacterial composite transposons	integrase	protein coat
bacterial IS elements	inverted repeats	prt
bacterial Tn elements	jumping genes	renaturation kinetics
bacteriophage	L1	repetitive DNA
centromere	LINE	replicative transposition
chromatin	LTR (long terminal repeats)	resolvase
chromosomes	LTR transposons	retrotransposon
Class I transposon	lysis	retrovirus
Class II transposon	lysogeny	RNA transposon
cointegrate	lytic pathway of phage	satellite DNA
CoT curves	maize Ac (activator) gene	SINE
cut-and-paste transposition	maize Ds (dissociator) gene	transposase
density-gradient centrifugation	mosaicism	triploid endosperm
direct repeats	Mu phage	Ty
DNA sequence phylogeny	nonautonomous transposon	viral infection
exon shuffling	mariner	

CHAPTER 14 WEB LINKS



14.1



14.2



14.3



14.4



14.5

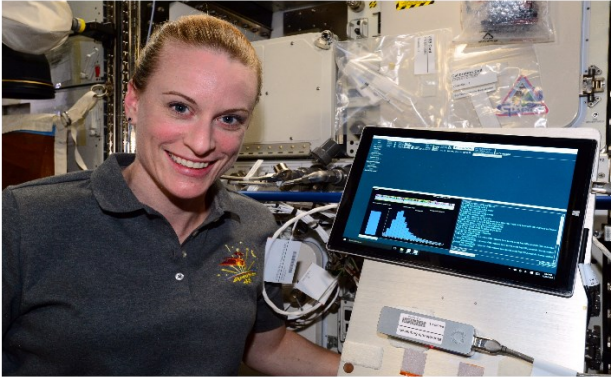
Chapter 15

DNA Technologies

Manipulating DNA; cDNA libraries, genomic libraries, DNA sequencing, PCR, microarrays, genomics, transcriptomics, proteomics

Reminder: For inactive *links*, google key words/terms for alternative resources.

Kathleen Rubins, Astronaut Sequencing DNA



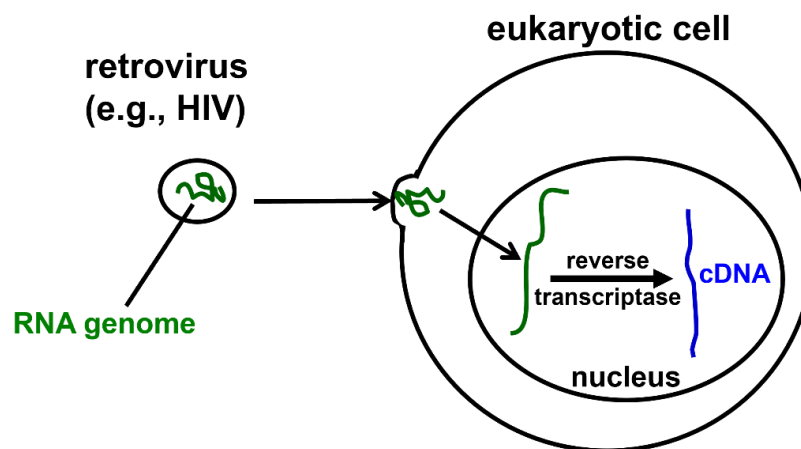
At the Space Station!

How about that?
15.1 [Sequencing DNA in Space](#)

15.1 Introduction

Let's start chapter by looking at technologies that led to **genetic engineering**. The ability to make **recombinant DNA** is such a seminal technology that just realizing it could be done and then doing it in a test tube for the first time earned Paul Berg a share of the 1980 Nobel Prize in Chemistry. First, we'll look at cDNA synthesis (the synthesis of DNA copies of RNA), something retroviruses routinely do as part of their reproductive pathway (Figure. 15.1).

Reverse Transcriptase



**Retrovirus finds cell; membranes fuse;
Retroviral RNA enters cell, then nucleus;
Viral RNA is reverse transcribed into cDNA.**

Fig. 15.1: Infection by a retrovirus (left) requires reverse transcriptase encoded in its own genome to make a copy (cDNA) of its genome leading to replication of viral cDNA and the reproduction of new viruses in the infected cell (right).

As shown, the retrovirus injects its RNA into a target cell. There it transcribes a **reverse transcriptase**. This enzyme *reverse-transcribes* a copy DNA (cDNA) complementary to the viral RNA. The same enzyme next makes a (ds)cDNA (double stranded cDNA), which then replicates. These cDNAs are transcribed into new viral RNA genomes and mRNAs for viral proteins. The latter encapsulate the RNA genomes into new viruses. Reverse transcriptases, along with many viral, bacterial, and even eukaryotic enzymes and biomolecules, are now part of our recombinant DNA and genetic engineering toolkit.

We will see how a **cDNA library** is made and screened for a **cDNA clone**, and how a cloned cDNA can fish an entire gene out of a **genomic library**. Then we look at **PCR** (the **polymerase chain reaction**) and how it can produce (*amplify*) millions of copies of a single gene (in fact DNA sequence) from as little DNA as is found in a single cell. In addition to its well-publicized use in forensics and more recently in definitive tests for the SARS-Cov-2 viruses, *PCR* is also a vital laboratory tool for fetching, amplifying, and studying sequences of interest. These venerable technologies illustrate important principles of cloning and sequence analysis. Of course, the analysis of traditionally cloned and amplified DNA sequences has been used to study the evolution and expression of individual genes.

A cautionary note: despite the realized and future promises of such powerful tools to clone recombinant cDNAs and genomic DNA, we are sometimes misled! For example, knowing that a genetic mutation is associated with an illness usually leads to a search for how the mutation might cause the illness. But as researchers in any discipline keep warning us, *correlation is not causation*! In fact, we know that many phenotypes, including genetic disease, are not the result of a single mutant gene. Autism is just one example. Nevertheless, newer fields of **genomics** and **proteomics** leverage a growing battery of tools to study many genes and their *regulatory networks* at the same time. The **molecular networking** made possible by genomics and proteomics (and other colorful, holistic terms we'll discuss later) promise to get us past naive and often incorrect notions of causation. We may soon be able to identify *many* correlations that might add up—if not to causation—at least to a propensity for genetic illness. We'll look at some of these *tools of leverage*!



[257 Overview of DNA Technologies](#)

Learning Objectives

When you have mastered the information in this chapter, you should be able to do the following:

1. Design a molecular experiment using *cDNA* or a *PCR* product to clone a gene?
2. Determine when to make or to use a *cDNA library* or a *genomic library*.
3. Outline an experiment to *purify rRNA* from eukaryotic cells.
4. Outline an experiment to isolate and then clone human cDNA for further study.
5. Explain why (you might want to *clone and to express* a human growth hormone gene).
6. List components needed to make a cDNA library, using purified *poly(A) RNA*.
7. List the components needed to make a genomic library from isolated *genomic DNA*.
8. Compare PCR and genomic cloning as strategies for isolating a gene.
9. Outline a strategy for using fly DNA to obtain copies of a human DNA sequence.

10. Ask a research question that requires screening a *genomic library* for a specific gene.
11. Ask a question that requires using a *microarray* to obtain a gene you want to study.

15.2 Make and Screen a cDNA Library

To make a **cDNA library**, first isolate cellular mRNA. This extract should represent all of the transcripts in the cells at the time of isolation, or the cell's **transcriptome**. This term is used in analogy to *genome*. While the genome is *all* of the genetic information of an organism, a transcriptome (usually eukaryotic) reflects all of the genes *expressed* in a given cell type at a moment in time. Reverse-transcribed cDNAs from an mRNA extract are *also* referred to as a transcriptome. A tube full of bacterial cells that were **transformed** with plasmids recombined with cDNAs is a cDNA library. cDNA libraries made from mRNAs from different cell types are in fact different transcriptomes; likewise, cDNA libraries from mRNAs taken from the same cells but grown under different conditions. A cDNA library reflects the mRNAs transcribed in cells at the moment of extraction. When cells in a cDNA library are spread out on a nutrient agar petri dish, each cell grows into a colony; each cell in the colony is a clone of the starting cell. cDNA libraries can be used to isolate and then to sequence the DNA encoding a polypeptide that you are studying.

Recall that the mature mRNA in eukaryotic cells has been spliced. Thus, cDNAs made from eukaryotic cells do not include introns. Introns, as well as sequences of enhancers and other regulatory elements in and surrounding a gene, must be studied in genomic libraries (to be discussed later). Here we look at how to make a cDNA library.

15.2.1 cDNA Construction

mRNA is only a few percent of eukaryotic cell RNA; most is rRNA. But that small amount of mRNA can be isolated from other RNAs because of their 3' *poly(A) tails*. Simply pass a total RNA extract over an **oligo(dT) column** like the one illustrated in Figure 15.2.

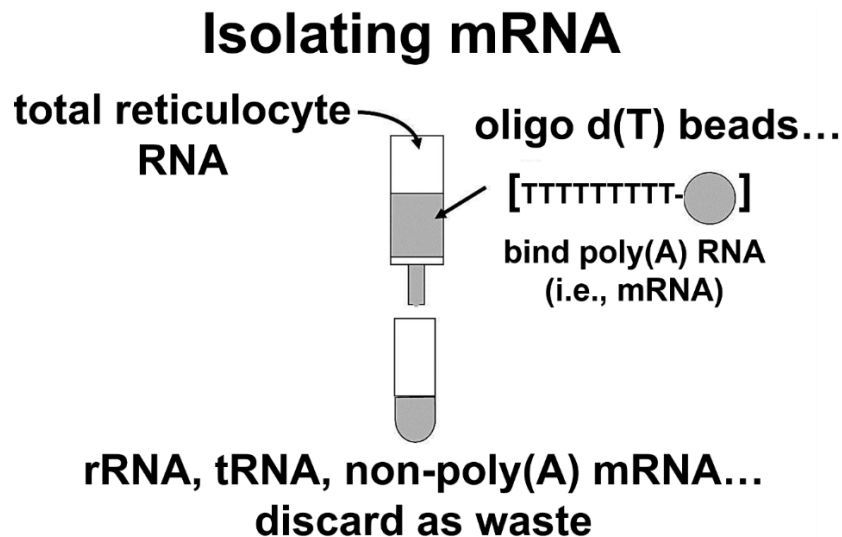


Fig. 15.2: Poly(A) tail at the 3' end of most eukaryotic mRNAs make them easy to separate from other cellular RNAs (e.g., rRNA, tRNA) by oligo(dT) column chromatography.

The strings of thymidine (T) in the oligo(dT) column form H-bonds with the *poly(A)* tails of mRNAs, tethering them to the column. All RNAs without a 3' poly(A) tail will flow through the column as waste. A second buffer is passed through the column to destabilize the H-bonds to allow **elution** of an mRNA fraction. If "free" oligo(dT) is then added to the eluted mRNA, it also forms H-bonds with the poly(A) tails of the mRNAs, where it can serve as a primer for the synthesis of cDNA copies of the poly(A) mRNAs originally in the cells. Adding four deoxynucleotide DNA precursors and **reverse transcriptase** (e.g., from chicken retrovirus-infected cells) will start reverse transcription. Figure 15.3 shows the synthesis of a cDNA strand complementary to an mRNA.

cDNA Synthesis; The First Strand

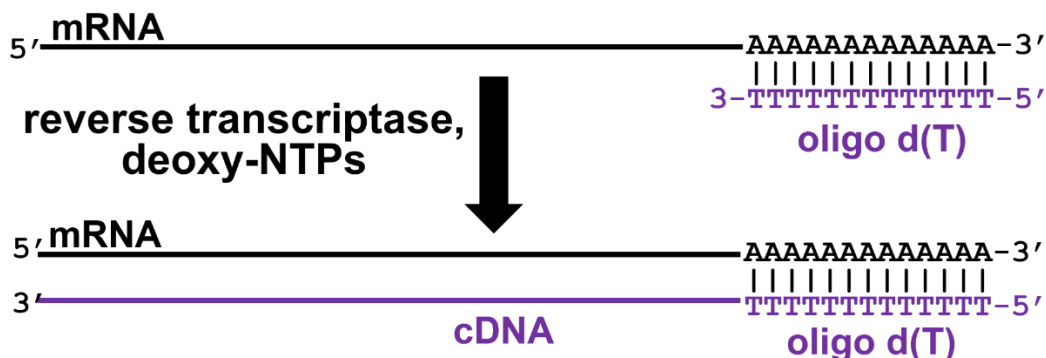


Fig. 15.3: Reverse transcriptase, supplied with deoxynucleotides, mRNAs, and an oligo(dT) primer, will catalyze synthesis 5' - 3' cDNA synthesis from an mRNA template in vitro.

After being heated to separate the cDNAs from the mRNAs, the cDNA is replicated to produce double-stranded cDNA, or (ds)cDNA (Figure 15.4).

Double-Stranded (ds) cDNA Synthesis

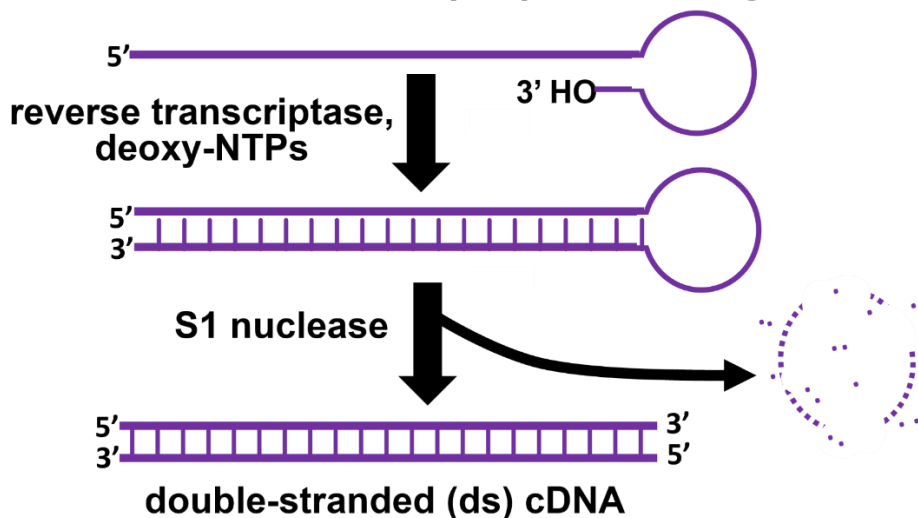


Fig. 15.4: A new cDNA strand can form a loop at its 3' end (upper) to act as a primer to synthesize a (ds)cDNA. Reverse transcriptase is also a DNA polymerase catalyzes second-strand synthesis (middle). The 'loop' can be removed (lower) by a single-stranded endonuclease (e.g. S1 nuclease).

Replication of the second cDNA strand is also catalyzed by reverse transcriptase. Second-strand synthesis is primed by the 3' end of a stem-loop structure that can form with most mRNAs. Reverse transcriptase is also a DNA polymerase, recognizing DNA as well as RNA templates, with the same 5'-to-3' DNA polymerizing activity as all DNA and RNA polymerases. After second cDNA strand synthesis, an ***S1 nuclease*** (a specifically *single-stranded endonuclease* originally isolated from an East Asian fungus!) is added. This nuclease will open the loop of the (ds)cDNA structure and trim the rest of the single-stranded DNA. What remains is the (ds)cDNA.



[258 Isolate mRNA and Make cDNA](#)



[259 Reverse Transcriptase](#)



15.2.2 Cloning cDNAs into Plasmid Vectors

To understand cDNA cloning and other aspects of making recombinant DNA, let's look again at what's in the recombinant DNA tool kit. In addition to *reverse transcriptase* and *S1 nuclease*, other enzymes in the kit include ***restriction endonucleases*** (***REs*** or restriction enzymes for short) and ***DNA ligase***. The natural function of *REs* in bacteria is to recognize and hydrolyze specific ***restriction site*** sequences in phage DNA, destroying the phage DNA and avoiding infection.

Some *restriction enzymes* cut through the two strands of the double helix to leave ***blunt ends***. Others make a ***staggered cut*** on each strand at their restriction site, leaving behind ***complementary*** ("sticky") ends for H-bond formation (Figure 15.5).

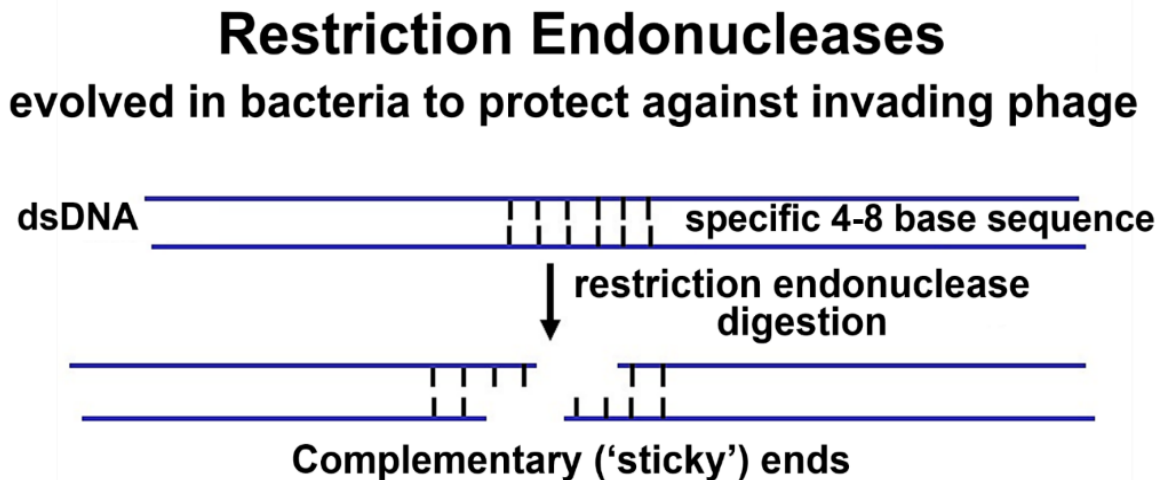


Fig. 15.5: Bacterial restriction endonucleases (*REs*) recognize and hydrolyze "foreign" DNA (e.g., phage DNA), blocking infection; most *REs* cut DNA at specific short DNA sequences, often at different positions along the double helix, leaving staggered (or "sticky") ends.

If two dsDNAs from different sources (e.g., different species) are digested with an *RE* that makes a staggered cut, the resulting fragments will have the same sticky ends. If you mix these DNAs, their complementary ends will form H-bonds linking the fragments. If the linked fragments are from different sources, we can add *DNA ligase* to covalently seal the fragments.

We say then that the fragments have been recombined. Such a protocol makes it relatively easy to recombine any two different DNAs at will. Let's look at how we recombine plasmid DNAs and cDNA—the first steps in cloning cDNAs.

15.2.3 Preparing Recombinant Plasmid Vectors Containing cDNA Inserts

Vectors (such as plasmids or phage DNA) are *carriers* engineered to recombine with foreign DNAs of interest. When a recombinant vector with its foreign **DNA insert** gets into a host cell, it can replicate many copies of itself, enough for easy isolation and study. cDNAs are typically inserted into *plasmid vectors* (usually “store-bought”). They can be *cut* with an RE at a suitable location, leaving those *sticky ends*. However, it would not do to digest (ds)cDNA with restriction endonucleases, since the goal is not to clone cDNA fragments but entire cDNA molecules. Therefore, it will be necessary to attach **linkers** to either end of the (ds)cDNAs. Plasmid DNAs and cDNA-linker constructs can then be digested with the same restriction enzyme to produce compatible *sticky ends*. Figure 15.6 shows the steps in the preparation of a vector and (ds)cDNA for recombination.

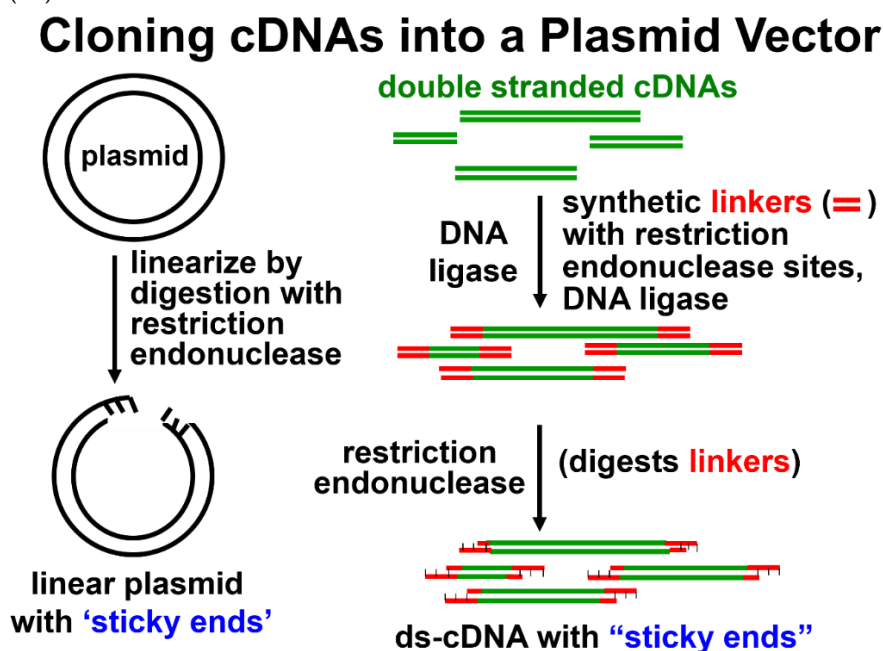


Fig. 15.6: Linearized plasmid with overlapping (“sticky”) ends is made by restriction digestion (left). Synthetic linkers with compatible restrictions sites are linked to cDNAs with DNA ligase (right).

To prepare for recombination, a plasmid vector is digested with a *restriction enzyme* to open the DNA circle. The (ds)cDNAs to be inserted into the plasmid vector are mixed with *linkers* and **DNA ligase** to attach a linker DNA at both ends of the (ds)cDNA. The *linkers* are short, synthetic (ds)DNA oligomers, containing restriction sites recognized by the same restriction enzyme that linearized the plasmid. Linkers attached to the (ds)cDNAs ends are digested with the appropriate restriction enzyme to give them “sticky ends” complementary to the staggered, plasmid ends.



15.2.4 Recombining Plasmids and cDNA Inserts and Transforming Host Cells

The next step is to mix the digested plasmids with the digested linker-ended cDNAs in just the right proportions so that most of the cDNAs form H-bonds with most of the sticky plasmid ends (and not with each other!). Once the H-bonds have formed between the cDNAs and plasmids, the addition of **DNA ligase** forms phosphodiester linkages between plasmid and cDNA insert to complete the recombinant circle of DNA (Figure 15.7).

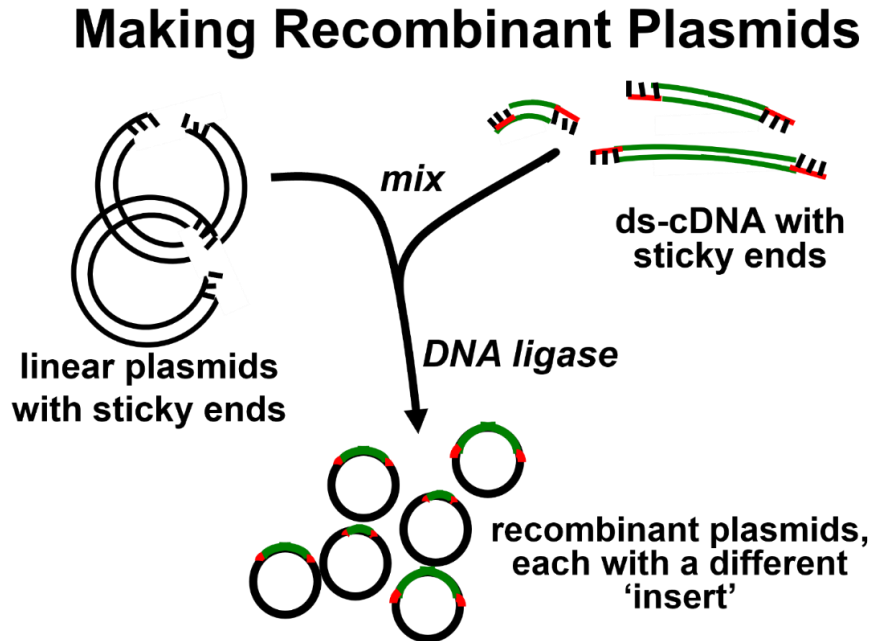


Fig. 15.7: Restriction-digested plasmid and linker-cDNAs with compatible “sticky ends” are mixed at appropriate concentrations (upper); the double-stranded cDNAs H-bond to the compatible sticky ends of plasmids. DNA ligase is added to covalently seal the *recombinant plasmids* with different cDNA inserts (below).

In early cloning experiments, an important consideration was how to generate plasmids with only one copy of a given cDNA insert (rather than how to make lots of re-ligated plasmids with no inserts or how to make lots of plasmids with multiple inserts). Improved, better-engineered vector and linker combinations made this issue less important.



[261 Recombine a cDNA Insert with a Plasmid Vector](#)



15.2.5 Transforming Host Cells with Recombinant Plasmids

The recombinant DNA molecules are now ready for the next step: the creation of a **cDNA library**. The recombinant DNAs are added to *E. coli* (or other) host cells, which have been *made permeable* so that they are easily **transformed**. Recall that transformation (as defined by Griffith) is the bacterial uptake of foreign DNA, leading to a genetic change. The **transforming principle** in cloning is the recombinant plasmid! The transformation step that results in a tube full of transformed cells (which is, in fact, the *cDNA library*) is shown in Figure 15.8.

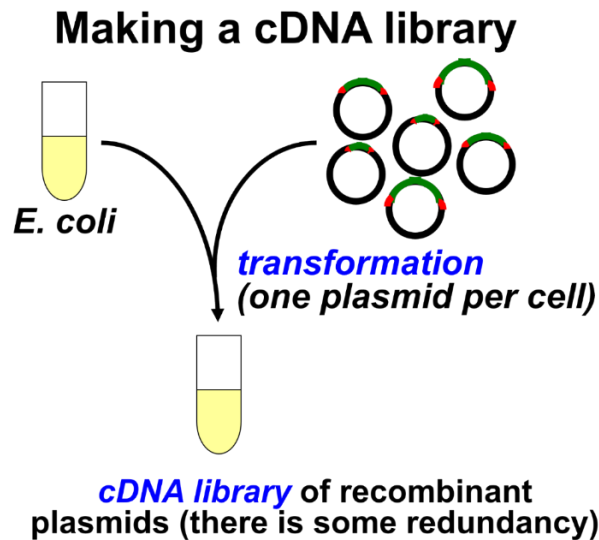


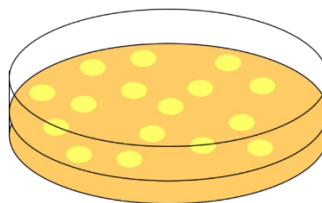
Fig. 15.8: To make a cDNA library, host cells are treated to make them permeable and then mixed with recombinant plasmids at a concentration that will favor transformation of each cell by only one plasmid (upper). The resulting tube of transformed cells (bottom) is the cDNA library.



15.2.6 Plating a cDNA Library on Antibiotic-Agar to Select Recombinant Plasmids

After all these treatments, not all plasmid molecules in the mix are recombinant; some cells in the mix haven't even taken up a plasmid. So, when the recombinant cells are plated on agar, how do you tell which of the colonies that grow came from cells that took up a recombinant plasmid? Both the *host strain* of *E. coli* and plasmid vectors used these days were further engineered to solve this problem. Specifically, the first practical plasmid vectors were designed to carry an *antibiotic-resistance gene*. In the following example, ampicillin-sensitive cells are transformed with recombinant plasmids containing the *ampicillin-resistance gene*. The cells are then plated on media containing *ampicillin* (a form of penicillin); the results are shown in Figure 15.9.

Spread Cells from cDNA Library on Agar Plate Containing Ampicillin, Incubate:



**Plasmids contain an ampicillin resistance gene.
Therefore, only transformed cells containing a
plasmid will grow and form colonies**

Fig. 15.9: Only cells containing plasmids will grow into colonies (yellow) on agar containing *ampicillin*, since the plasmids contain an ampicillin-resistance gene.

Transformed cells—that is, those that took up a recombinant plasmid carrying the ampicillin-resistance gene—can grow on the ampicillin-agar medium. Untransformed cells (cells that failed to take up a plasmid) lack the ampicillin-resistance gene and cannot grow on the ampicillin medium.

But there is still a question: How can you tell whether the cells that grew were really transformed by a recombinant plasmid containing a cDNA insert? It is possible that some of the transformants are nonrecombinant plasmids that still have the ampicillin-resistance gene! To address this issue, plasmids were further engineered to contain a ***streptomycin-resistance gene*** (as well as the ampicillin-resistance gene). But in this case, the restriction-enzyme sites in the plasmid that were to be used for recombination were placed in the middle of the gene. Thus, inserting a cDNA into *this* plasmid would disrupt and inactivate the ***streptomycin-resistance gene***. This second bit of genetic engineering enabled harvesting *only* cells transformed with recombinant plasmids (i.e., those containing a cDNA insert).

We can tell *transformants* with recombinant plasmids apart from those with non-recombinant plasmids by the technique of ***replica plating*** (Figure 15.10).

Replica Plating

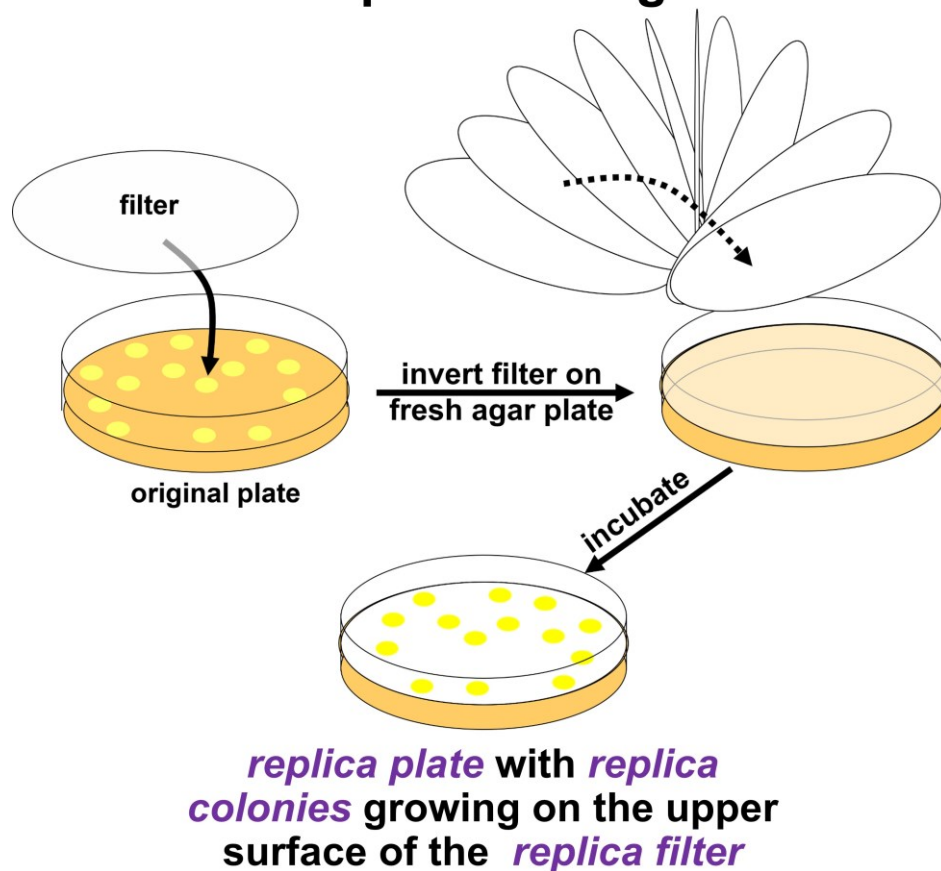


Fig. 15.10: In this example of replica plating, a filter picks up a few cells from colonies on an original plate (left). The 'replica filter' is inverted and placed on a fresh agar plate containing an antibiotic that selects for the growth of recombinant plasmids (right). Upon incubation, only cells transformed with recombinant plasmids grow on the filter in the 'replica plate' (bottom).

After colonies grow on the ampicillin-agar plate, lay a filter over the plate. The filter will pick up a few cells from each colony, in effect becoming a **replica filter** (i.e., a mirror image) of the colonies on the plate. Next, place the **replica filter** on a new agar plate containing *streptomycin*. Any colonies that grow on the filter must be streptomycin-resistant, containing only nonrecombinant plasmids. Colonies containing recombinant plasmids, those that did not grow in streptomycin, are easily identified on the original ampicillin-agar plate. In practice, highly efficient recombination and transformation procedures typically reveal very few streptomycin-resistant cells (i.e., colonies) after replica plating. When this happens, the ampicillin-resistant cells constitute a good cDNA library, ready for screening.



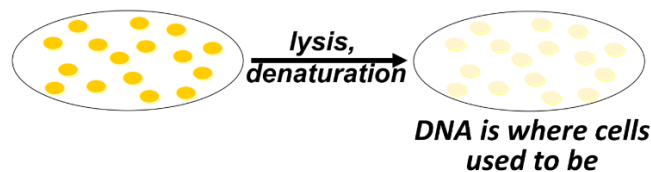
263 Making a Replica Plate Filter

15.2.7 Identifying Colonies Containing Plasmids with Inserts of Interest

The next step is to **screen** colonies in a cDNA library for those containing a specific cDNA. Since cells typically make thousands of proteins at the same time, a cDNA library should contain thousands of cDNAs made from thousands of mRNAs. Finding a single cDNA of interest can require plating the cDNA library in a tube on more than a few agar plates.

Continuing with the example above, actual screening would be done using multiple replica filters of ampicillin-resistant cells. The number of replica filters to be screened can be calculated from assumptions and formulas for estimating the number of colonies to be screened in order to represent an entire *transcriptome* (i.e., the number of different mRNAs in the original cellular mRNA source (including in this example, globin mRNAs). Once the requisite number of replica filters is made, they are subjected to *in situ lysis* to disrupt cell walls and membranes. The result is that the cell contents are released, and the DNA is denatured (unwound to become single-stranded). The DNA then adheres to the filter *in place* (in situ), where the colonies were. The result of *in situ lysis* is a filter with faint traces of the original colonies (Figure 15.11).

***In situ* Lysis Leaves Cellular Remains on Replica Filter**



All the DNA is denatured and bound to the filter...

Fig. 15.11: *Replica filters* are lysed *in situ* (in place), leaving partially denatured DNA (including recombinant plasmid DNA) from the colonies where the cells used to be. Filters can be probed for a sequence of interest.

Next, a molecular **probe** is used to identify DNA containing the sequence of interest. The probe is often a **synthetic oligonucleotide**, the sequence of which was inferred from

known amino acid sequences. These oligonucleotides are made radioactive and placed in a bag with the filter(s). DNA from cells that contained recombinant plasmids with a cDNA of interest will bind the complementary *probe*. The results of in situ lysis and **hybridization** of a radioactive probe to a replica filter are shown in Figure 15.12 (below).

Finding Recombinants of Interest by Exposing Replica Filters Containing Denatured DNA from *In situ* Lysed Colonies to a Radioactive *Probe

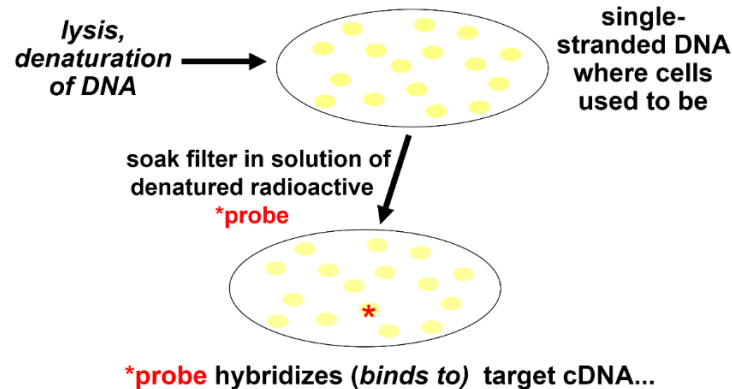


Fig. 15.12: Replica filters of denatured DNA are exposed to a radioactive nucleic acid probe that can hybridize to the target sequences on the filters. The asterisk on the lower filter is radioactive probe bound to complementary DNA released by a recombinant colony.



[264 Probing a Replica Plate Filter](#)

The filters are rinsed to remove the unbound radioactive oligomer probe and then placed on X-ray film. After a period of exposure, the film is developed. Black spots will form on the film from radioactive exposure, creating an autoradiograph of the filter. The black spots in the autoradiograph in Figure 15.13 correspond to colonies on a filter that contain a recombinant plasmid with the target cDNA sequence.

Make Autoradiograph:

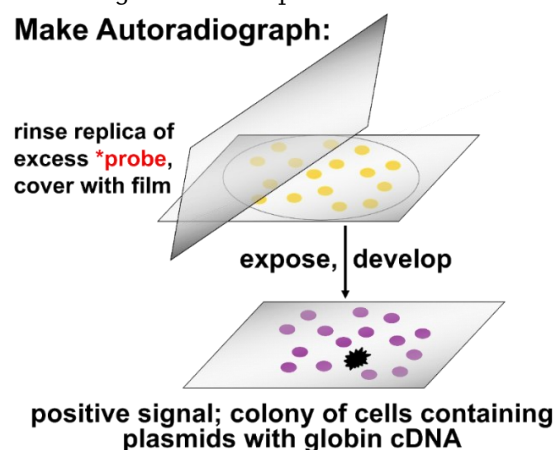


Fig. 15.13: After rinsing the replica filter to remove excess radioactive probe, an autoradiograph is made to detect a colony of interest. X-ray film is placed over the filter (upper image). After exposure, the 'X-ray' is developed. The resulting autoradiograph shows a dark spot where the probe hybridized to colony DNA.

From a positive on the film, recombinant colonies are located on the original plate and grown in liquid culture for plasmid DNA isolation. Next, cDNAs are sequenced and the amino acid sequences they encode are inferred from the genetic-code dictionary. Once verified, recombinant plasmids can be isolated, tagged (made radioactive or fluorescent), and used for various purposes:

- To probe for the genes from which they originated
- To identify and to quantitate the mRNA and even to locate the transcripts in the cells
- To quantitatively measure amounts of specific mRNAs

Isolated plasmid cDNAs can even be expressed in suitable cells to make the encoded protein. These days, diabetics no longer receive pig insulin; rather, they get synthetic human insulin made from expressed human cDNAs. Moreover, while the introduction of the *polymerase chain reaction* (**PCR**; see section 15.5) has superseded some uses of cDNAs, they still play a role in genome-level and transcriptome-level studies.



[265 Pick a Clone from a Replica Filter and Play with It!](#)



15.3 DNA sequencing

For some perspective, we should note that RNA sequencing came first, when Robert Holley sequenced a yeast transfer RNA (tRNA) in 1965. The direct sequencing of tRNAs was possible because they are small, short nucleic acids, and because many of the bases in tRNAs are chemically modified after transcription, making them easy to identify.

Twelve years later, in 1977, two different methods for sequencing DNA were reported. One method, developed by Allan Maxam and Walter Gilbert, involved fragmenting DNA, sequencing the small fragments, and then aligning the overlapping sequences of the short fragments to assemble longer sequences. This became known as **Maxam-Gilbert DNA sequencing**. Frederick Sanger and colleagues in England developed the other method: the DNA synthesis-based **dideoxy-DNA sequencing** technique. Sanger and Gilbert shared a Nobel Prize in Chemistry in 1983 for their DNA-sequencing work. However, because of its simplicity, Sanger's *dideoxy* method quickly became the standard for sequencing all manner of cloned DNAs, including the first complete sequencing of a genome—that of a bacterial virus called bacteriophage ϕ X174.

Early improvements in recombinant DNA technology were happening at the same time as advances in sequencing. More efficient, rapid cloning and sequencing of DNA from increasingly diverse sources led to *genome projects*. The first of these were on genomes of important model organisms—for example, *E. coli*, *C. elegans*, *S. cerevisiae* (yeast), and of course, us! By 1995 Craig Venter and colleagues at the *Institute for Genomic Research* completed an entire bacterial genome sequence (of *Haemophilus influenzae*); by 2001 Venter's private group, along with Francis Collins and colleagues at the National Institutes of Health, had published a first draft of the sequence of the human genome. Venter had proven the efficacy of a whole-genome-sequencing approach called *shotgun sequencing*, which was much

faster than the gene-by-gene, fragment-by-fragment, “linear” sequencing strategy being used by other investigators (more on shotgun sequencing later!). Since Sanger’s dideoxynucleotide-DNA-sequencing method remains a common, economical methodology, let’s consider the basics of the protocol.

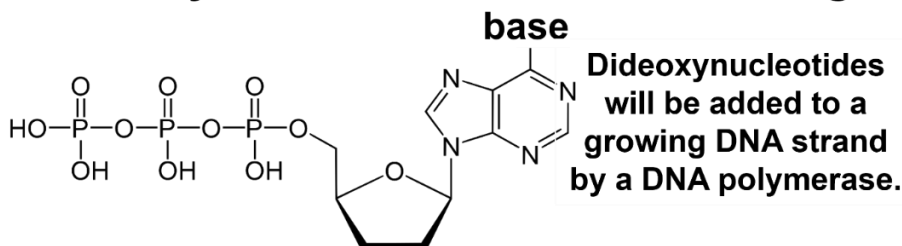
15.3.1 Manual DNA Sequencing

Given a template DNA (e.g., a plasmid cDNA), Sanger used in vitro replication protocols to demonstrate that he could do the following:

- Replicate DNA under conditions that randomly stopped nucleotide addition at every possible position in growing strands
- *Separate and then detect* these DNA fragments of replicated DNA

Recall that DNA polymerases catalyze the formation of *phosphodiester linkages* by linking the α phosphate of a **nucleotide triphosphate** to the free 3' OH of a deoxynucleotide at the end of a growing DNA strand. Recall also that the ribose sugar in the deoxynucleotide precursors of replication lack a 2' OH (hydroxyl) group. Sanger’s trick was to add **dideoxynucleotide triphosphates** to his in vitro replication mix. The ribose on a *dideoxynucleotide triphosphate* (ddNTP) lacks a 3' OH, as well a 2' OH (Figure 15.14).

Dideoxynucleotides: Anti-Cancer Drugs



But without a 3' oxygen on the sugar, no further nucleotides can be added, blocking replication in all (but preferentially cancer) cells.

Fig. 15.14: Chemical structure of a dideoxynucleotide.

Adding dideoxynucleotides to a growing DNA strand stops replication. Lacking an OH at the 3' end of the strand, DNA polymerase can't catalyze a dehydration synthesis to form the next phosphodiester linkage. Because they can stop replication in actively growing cells, ddNTPs such as *dideoxyadenosine* (tradename, **cordycepin**,) are anticancer chemotherapeutic drugs.



[266 Treating Cancer with Dideoxynucleosides](#)



A look at a manual DNA sequencing reveals what is going on in the sequencing reactions. Four reaction tubes are set up, each containing the template DNA to be sequenced. Each tube also contains a *primer* of known sequence and the required deoxynucleotide precursors necessary for replication. Figure 15.15 (below) shows the setup for manual DNA sequencing.

Dideoxynucleotide DNA Sequencing

Each tube below contains: DNA polymerase, dTTP, dCTP, dGTP, *dATP (**radioactive**), DNA to be sequenced, an oligonucleotide to prime DNA synthesis, and one of the di-deoxynucleotides as shown:

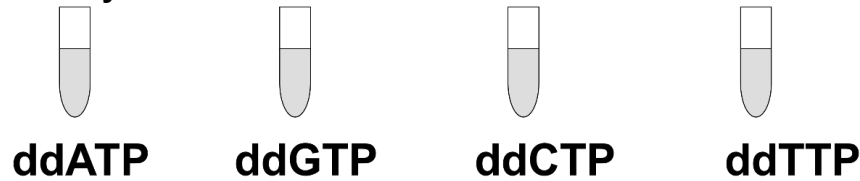


Fig. 15.15: Each tube in Frederick Sanger's manual DNA sequencing protocol contains the same DNA to be sequenced, a DNA polymerase (usually Taq polymerase), an oligonucleotide primer, and deoxynucleotide precursors, but each tube contains a different dideoxynucleotide.

A different ddNTP, (ddATP, ddCTP, ddGTP, or ddTTP) is added to each of the four tubes. Then, DNA polymerase is added to each tube to start the DNA-synthesis reaction. During DNA synthesis, different-length fragments of new DNA accumulate as ddNTPs are incorporated at random, opposite complementary bases in the template DNA being sequenced. A short time after adding the DNA polymerase, the mixture is heated to separate the DNA strands, and fresh DNA polymerase is added to replace the enzyme destroyed by heating and to repeat the synthesis reactions. These reactions are repeated as many as thirty times in order to produce enough radioactive DNA fragments to be detected. Expectations for these sequencing reactions in the four tubes are illustrated in Figure 15.16.

DNA Synthesis In a DNA Sequencing Reaction

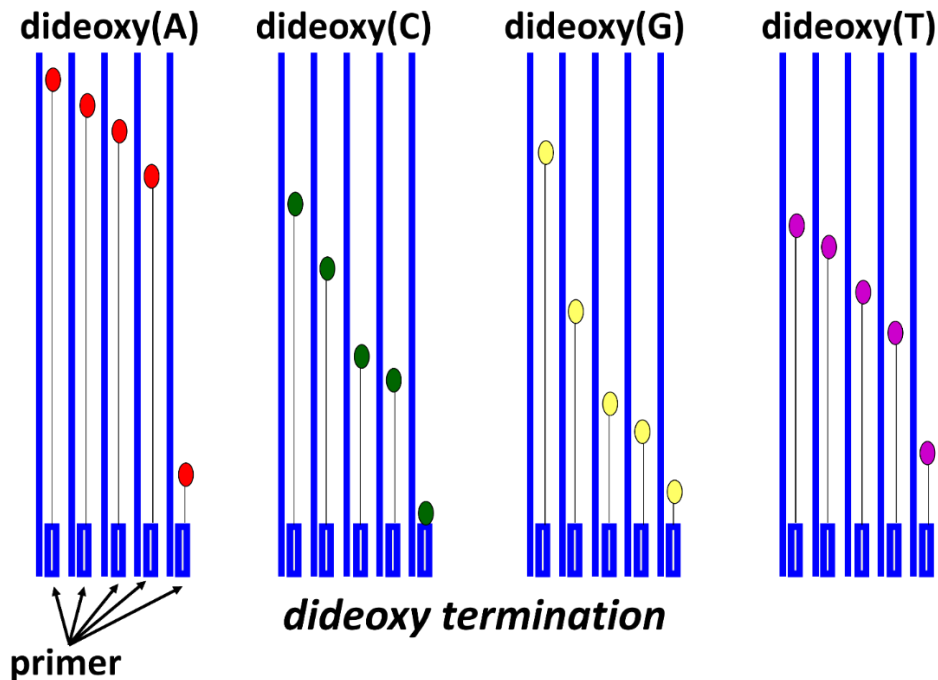


Fig. 15.16: In each of the 4 tubes in Fig. 15.15, different-length DNA fragments were synthesized until replication was terminated by incorporating a dideoxy nucleotide, as illustrated here. The fragments are separated by polyacrylamide gel electrophoresis to read a DNA sequence.

When the ***Taq DNA polymerase*** from the thermophilic bacterium *Thermus aquaticus* became available (more later!), it was no longer necessary to add fresh DNA polymerase after each replication cycle, because the heating step required to denature new DNA made in each cycle would not destroy the heat stable Taq DNA polymerase. Thanks to *Taq polymerase*, the many heating and cooling cycles required for what became known as *dideoxy-chain-termination DNA sequencing* were soon automated using inexpensive *programmable thermocyclers*. Since a small amount of a radioactive deoxynucleotide (usually ^{32}P -labeled ATP) was present in each reaction tube, all newly made DNA fragments are radioactive. After electrophoresis to separate the new randomly terminated DNA fragments in each tube, autoradiography of the electrophoretic gel reveals the position of each terminated fragment. The DNA sequence can then be read from the gel as illustrated in the simulated autoradiograph (Figure 15.17).

Simulated Results of Manual DNA Sequencing after Polyacrylamide Gel Electrophoresis

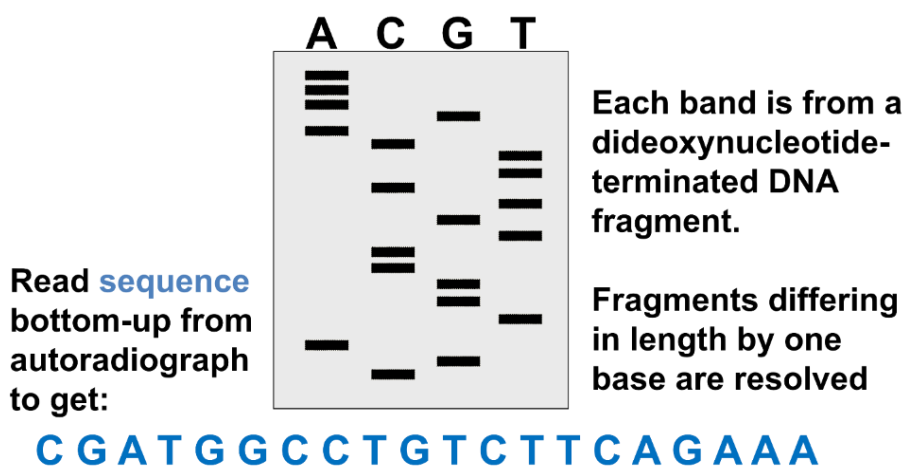


Fig. 15.17: The electrophoretic gel of dideoxy-terminated DNA fragments in Fig. 15.16 is exposed to X-ray film to create this autoradiograph. The bands are readable from the bottom up to reveal a DNA sequence based on the different length dideoxy-terminated DNA fragments.

The DNA sequence can be read by reading the bases from the bottom of the film, starting with the C at the bottom of the C lane. Try reading the sequence yourself!



 [267 Manual Dideoxy Sequencing](#)

15.3.2 Automated First-Generation DNA Sequencing

The first semi-automated DNA sequencing method was invented in Leroy Hood's California lab in 1986. Though the method still used Sanger's dideoxy sequencing, a radioactive phosphate-labeled nucleotide was no longer necessary. Instead, each dideoxynucleotide (ddNTP) in a sequencing reaction is tagged for detection with a different fluorescent dye.

Sequence reaction products are electrophoresed on an ***automated DNA sequencer***. The migrating dideoxy-terminated DNA fragments pass through a beam of UV light, and a *detector* "sees" the fluorescence of each DNA fragment. The color and order of the fragments is

sent to a computer, which generates a colored plot showing the length and order of the fragments, and thus their sequences (Figure 15.18).

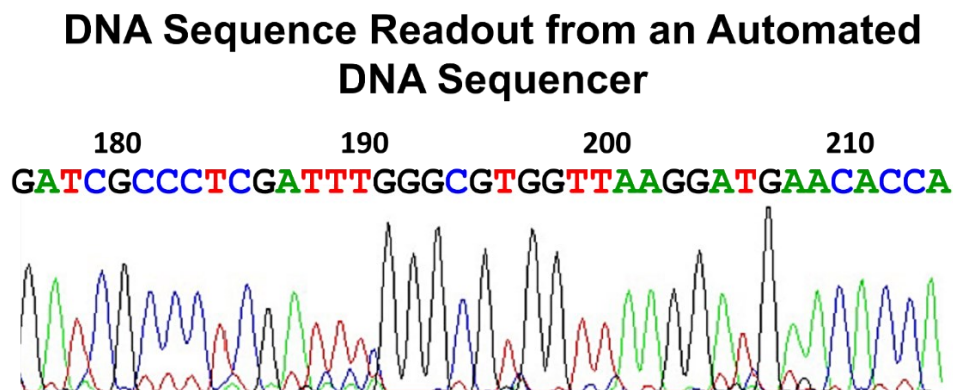


Fig. 15.18: An automated DNA sequence chromatograph; the sequence is read from left to right.

A most useful feature of this sequencing method is that a template DNA could be sequenced in a single tube containing all the required enzymatic components, as well as *all* four of the dideoxynucleotides! That's because the fluorescence detector in the sequencing machine separately sees all the short ddNTP-terminated fragments as they move through the electrophoretic gel. Hood's innovations were quickly commercialized, making major sequencing projects possible (including many genome projects). Automated DNA sequencing rapidly augmented large sequence databases in the United States and Europe. In the United States, the National Center for Biological Information (NCBI) maintains a sequence database and archives virtually all DNA sequences determined worldwide. New "tiny" DNA sequencers have made the requirements for sequencing DNA so portable that in 2016, one was even used in the *International Space Station* (see the photo at the top of this chapter!). New tools and protocols (some described in this chapter) are used to find, compare, and globally analyze DNA sequences almost as soon as they get into the databases.



[268 Automated Sequencing Leads to Large Genome Projects](#)



15.3.3 Shotgun Sequencing

Large-scale sequencing targets entire prokaryotic and (typically much larger) eukaryotic genomes. The latter require strategies that either sequence long DNA fragments or sequence many short DNA fragments more quickly (or both). We already noted the *shotgun sequencing* strategy used by Venter to sequence whole genomes (including our own—or, more accurately, his own!). In *shotgun sequencing*, cloned DNA fragments of 1,000 bp or longer are broken down at random into smaller, more easily sequenced fragments. The fragments are themselves cloned and sequenced. Nonredundant sequences are then assembled by the aligning of sequence-overlap regions.

When there are no gaps (which would be due to unrepresented DNA fragments that fail to be sequenced), today's computer software is quite adept at rapidly aligning the available overlapping sequences and connecting them to display long, contiguous DNA sequences.

Shotgun sequencing is summarized in Figure 15.19 (below).

Shotgun DNA Sequencing

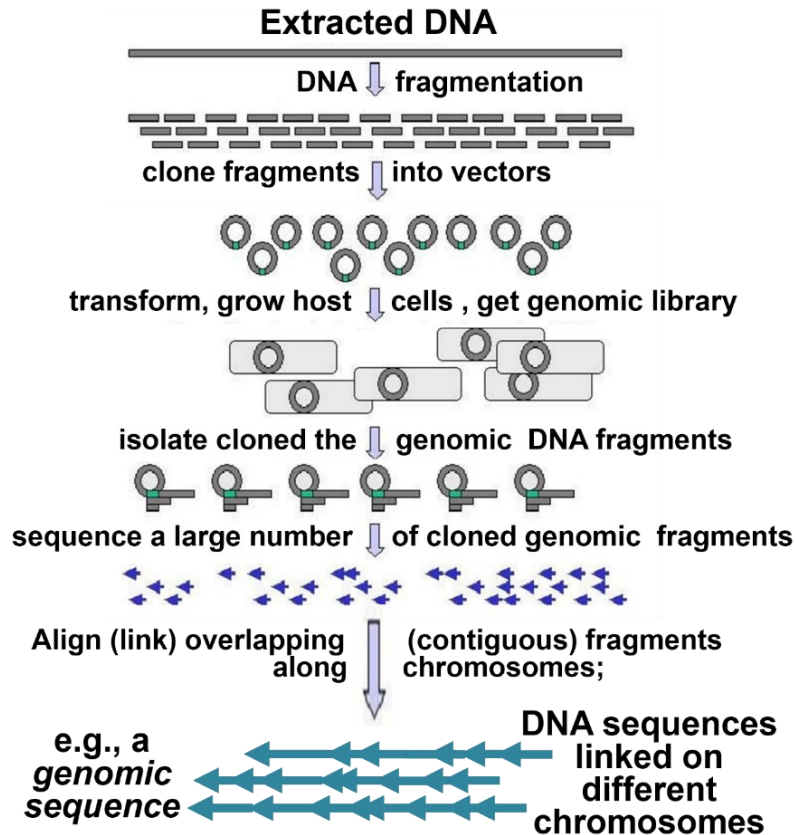


Fig. 15.19: Overview of shotgun sequencing of DNA. Randomly fragmented genomic DNA is recombined with vector (steps 1, 2). Host cells are transformed with recombinant DNAs and cultured to make a genomic library (step 3). DNA is extracted from many colonies and sequenced (steps 4, 5). Sequences are fed to a computer to find overlapping sequences and arrange sequenced fragments in order (step 6).

Sequence gaps that remain after shotgun sequencing can be filled in by **primer walking**. In this method, a sequencing primer (created based on a known sequence near the gap) "walks" into the gap region on an intact DNA (i.e., that has not been fragmented). Another "gap-filling" technique involves the *polymerase chain reaction (PCR)*. Two oligonucleotides are synthesized, based on sequence information on either side of a gap; then, PCR is used to synthesize the missing fragment, which is sequenced to fill in the gap (PCR will be described more fully in section 15.5).

15.3.4 Next-Generation DNA Sequencing

The goals of **next gen** sequencing methods are to get an accurate reading of millions or even billions of bases in a short time. Examples of these sequencing systems include *pyrosequencing*, *ligation-based sequencing*, and *ion-semiconductor sequencing*; these involve methods of **massively parallel sequencing reactions**. They differ in detail, but all require a library (e.g., genomic, or transcriptomic) attached to a solid surface, not unlike a microarray.

The attached DNAs are amplified in situ into separate DNA *clusters* derived from single library clones. The amplified, clustered DNAs then participate in the sequencing reaction. Sequences are analyzed and generated by computer. However, accuracy is a problem. Longer reads take more reactions and more time, and this leads to reduced accuracy. For a look at some of these technologies, see ^{15.2}[Next Generation Sequencing Video](#). A more accurate *new* technology is **nanopore**-based DNA sequencing, outlined in Figure 15.20. Nanopore sequencing was first used on the space station in 2016.

Principles of Nanopore Sequencing

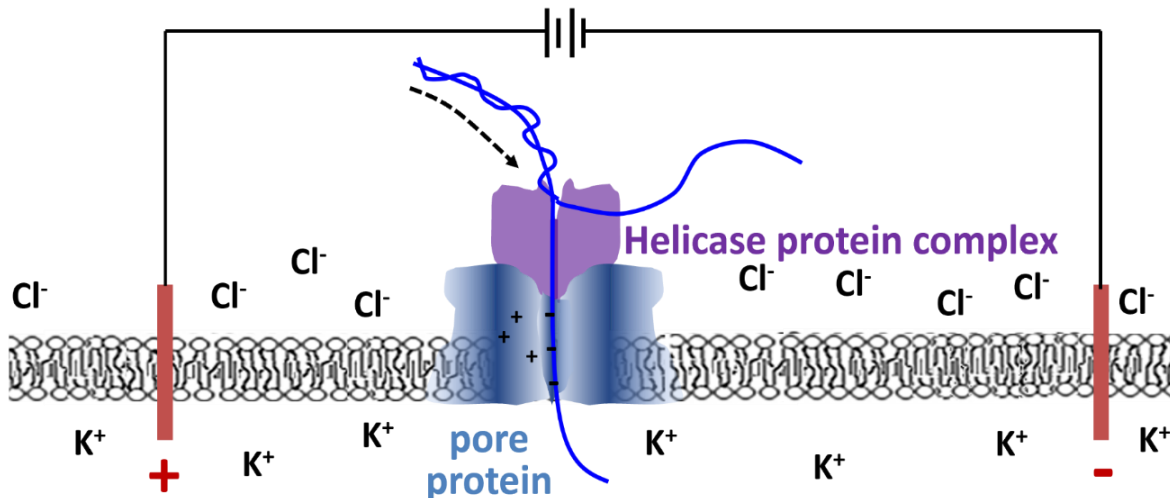


Fig. 15.20: Nanopore DNA sequencing is one of several “next-gen” sequencing technologies. A helicase-nanopore protein complex is embedded in a phospholipid bilayer membrane. An applied potential difference (voltage) across the membrane drives double-stranded DNA through the helicase-nanopore. As the DNA (now single-stranded) travels through the nanopore, disturbances of the voltage difference are detected and converted to sequence data.

This device uses an artificial phospholipid membrane, in which is embedded a synthetic, engineered **helicase/motor-pore-protein** complex. The nanopore is itself derived from the *Mycobacterium smegmatis* porin A (MspA). A voltage is applied across the membrane to create a membrane potential. Double-stranded DNA is propelled into the helicase, where it is unwound. The resulting single-stranded molecules then pass through a tiny hole (the *nanopore*) in the MspA. As the ssDNA (or RNA) passes through, its charge configurations disturb the electrical field in the porin. These disturbances are different for each nucleotide, generating a unique signal that is detected, recorded, and displayed in almost real time on a small computer. Next-generation sequencing protocols and devices are already providing long and increasingly more accurate data, which is required for research and for clinical applications. Check out ^{15.1}[Sequencing DNA in Space](#) (again) to see why one would want to do this!

While first-generation automated dideoxy-Sanger sequencing is still widely used to analyze individual cDNAs, genes or PCR products, *next-gen* DNA sequencing is more rapid, and it can also be used in the field, for research as well as for practical analyses. And, as if this weren't enough, now comes word of a marriage of a nano-sequencer and a cell phone analytical app to make immediate sense of a nano-sequence (^{15.3}[iGenomics: A Nanotech Marriage](#)). Check out

your Star Trek creds and ask can a real *tricorder* be on the way to a real Dr. McCoy even before we have a real *Enterprise*? Will this new married couple be able to identify microbial agents of disease *on location*?

15.4 Genomic Libraries

Any tube full of **recombinant bacteriophage** is basically a **genomic library**. Each phage DNA molecule in the library should contain a fragment of foreign cellular DNA. A good genomic library will contain a representation of all of the possible fragments of an organism's genome. Bacteriophage are often used as *vectors* to clone genomic DNA fragments because phage genomes are bigger than plasmids and can be engineered to remove large amounts of DNA that are not needed for infection and replication in host cells. The missing DNA can then be replaced by large foreign DNA inserts—fragments as long as 18-20 Kbp. That's nearly twenty times longer than cDNA inserts in plasmids. Purified phage coat proteins can then be mixed with the recombined phage DNA to reconstitute **infectious phage particles** (i.e., recombinant phage). Infection of host bacteria by these "particles" leads to replication of the recombinant phage DNA, new phage production, cell lysis, and the release of lots of new recombinant phage.

Consider the following bit of math: A typical mammalian genome consists of more than 2,000,000,000 bp (3,200,000,000 in humans!). Inserts in plasmids are very short, rarely exceeding 1,000 bp. If you divide two billion by one thousand, you get two million, the minimum number of recombinant plasmid clones that must be screened to find a sequence of interest. In truth, you would need much more than this number of clones to find the smaller parts of a gene that add up to a whole gene! Of course, part of the solution to this "needle in a haystack" dilemma is to clone larger DNA inserts in more accommodating vectors. And *that* is the value of a bacteriophage vector and even more accommodating vectors, including whole chromosomes! Consider the **Yeast Artificial Chromosome (YAC)**, hosted by (replicated in) yeast cells.

YACs can accept humongous foreign DNA inserts! To replicate in a yeast cell, a chromosome need have only **one centromere and two telomeres**, and little else! Recall that telomeres are needed in replication to keep the chromosome from shortening during replication of the DNA. The centromere is needed to attach chromatids to spindle fibers so that they can separate during *anaphase* in *mitosis* (and *meiosis*).

From this brief description, you may recognize a common strategy for engineering a cloning vector: determine the minimum properties that your vector must have and remove nonessential DNA sequences. Of course, for cloning, you must include some useful restriction sites. Then you have enabled recombination with inserts as long as 2,000 Kbp. That's a YAC... The tough part, of course, is keeping a 2,000 Kbp-long DNA fragment intact long enough to get it into the YAC!

Whatever the vector of choice, sequencing its insert can tell us many things. It can show us how a gene is regulated by confirming known and/or revealing new regulatory DNA sequences. It can show us neighboring genes, helping us to map them on chromosomes and reveal evolutionary genetic relationships. Genomic DNA sequences from one species can

probe for similar sequences in other species, allowing comparative sequence analysis that can tell us a great deal about gene evolution and the evolution of species. One early surprise from gene sequencing studies was that we share many common genes and DNA sequences with other (in fact most) species, like yeast, worms, flies, and (of course) vertebrates, including our more closely related mammalian friends. You may already know that chimpanzee and human genomes are 99% similar. And we have already seen comparative sequence analysis showing how proteins with different functions in the same (and even across) species nevertheless share structural domains.

Let's look at how we would make a phage genomic library that targets a specific gene of interest. As you'll see, the approach is like cloning a foreign DNA into a plasmid or in fact, into any other vector, but the numbers and details are shown here to exemplify cloning in phage.

15.4.1 Preparing Specific-Length Genomic DNA for Cloning: The Southern Blot

To begin with, high molecular weight (i.e., long molecules of) the desired genomic DNA are isolated, purified, and then digested with a restriction enzyme. Usually, the digest is partial, aiming to generate overlapping DNA fragments of random length. Before electrophoresis, the digested DNA is mixed with *ethidium bromide*, a fluorescent dye that binds to DNA. After electrophoresis on agarose gels and exposed to UV light, the DNA appears as a bright fluorescent smear. If we wanted to clone the complete genome of an organism, we could recombine all DNA with suitably digested vector DNA. But since we are only after one gene, we can reduce the number of clones required to be screened to find a sequence of interest.

In the early days of cloning, we would do this by creating a ***Southern blot*** (named after Edward Southern, the inventor of the technique). This technique lets us identify the size of genomic DNA fragments most likely to contain a desired gene. A summary of the steps to make a Southern blot follows:

1. Digest genomic DNA with one or more restriction endonucleases.
2. Run the digest products on an agarose gel to separate fragments by size (length). The DNA appears as a smear when stained with a fluorescent dye.
3. Place a filter on the gel. The DNA transfers (blots) to the filter for about 24 hours (or more).
4. Remove the blotted filter and place it in a bag containing a solution that can denature the DNA.
5. Add a radioactive probe (e.g., cDNA) containing the gene or sequence of interest. The probe hybridizes (binds) to complementary genomic sequences on the filter.
6. Prepare an autoradiograph of the filter and see a "band" representing the size of genomic fragments of DNA that include the sequence of interest.

The Southern-blot protocol is illustrated in Figure 15.21 (below).

Southern Blots Identify Genomic DNA Fragments to be Cloned & Sequenced

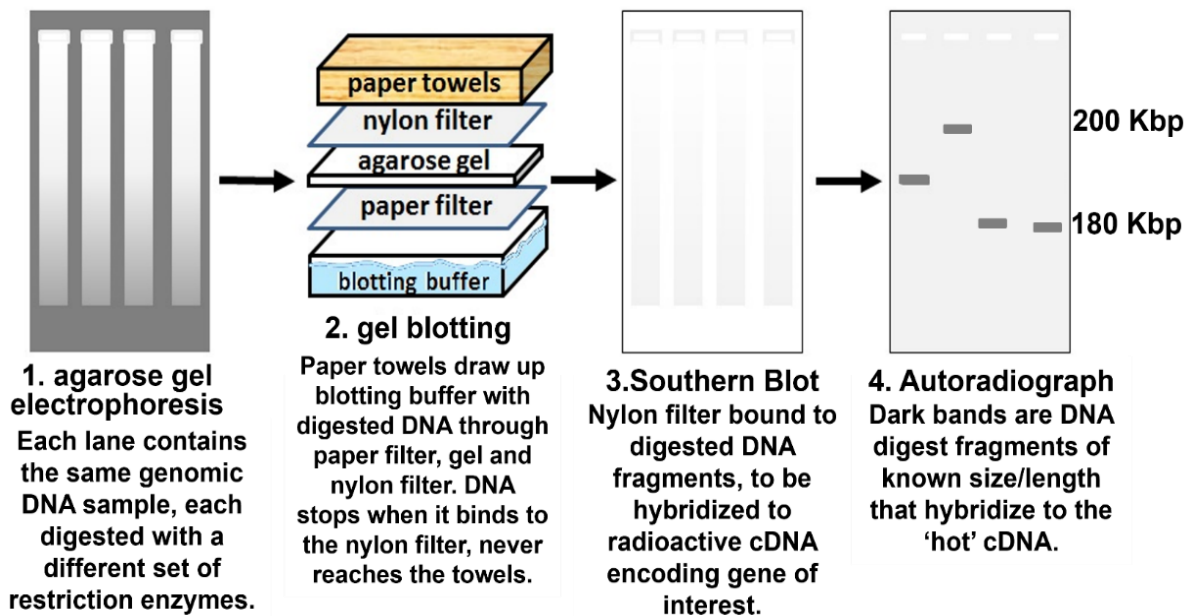


Fig. 15.21: Making a Southern blot: DNA fragments are electrophoresed on agarose gels (1). During 'blotting' (2), liquid and DNA fragments on the gel transfer and stick to a nylon membrane. After treating the blot to denature (separate) DNA strands (between steps 2 and 3), radioactive probe (e.g., cDNA, RNA) is allowed to hybridize to DNA on the filter (between steps 3 and 4). DNA fragments are detected and their sizes revealed by autoradiography (4).

Once you know the size (or size range) of the restriction-digest fragments that contain the DNA you want to study, you are ready to run another gel of digested genomic DNA, and then:

- cut out the piece of gel containing fragments of the size that "lit up" with your probe in the autoradiograph.
- Remove (elute) the DNA from the gel piece into a suitable buffer.
- Prepare these size-restricted DNA fragments for insertion into (i.e., recombination with) a vector for genomic cloning.

15.4.2 Recombining Size-Restricted Genomic DNA with Phage DNA

After eluting the restriction-digested DNA fragments of the right size-range from the gels, we mix the DNA with compatibly digested phage DNA at concentrations that favor the formation of H-bonds between the ends of the phage DNA and the genomic fragments (rather than with each other!). The addition of DNA ligase covalently links the recombined DNA molecules that will become a partial (i.e., DNA size-restricted genomic library). These steps are abbreviated in the illustration in Figure 15.22 (below).

Recombining Genomic DNAs with Phage vector DNA

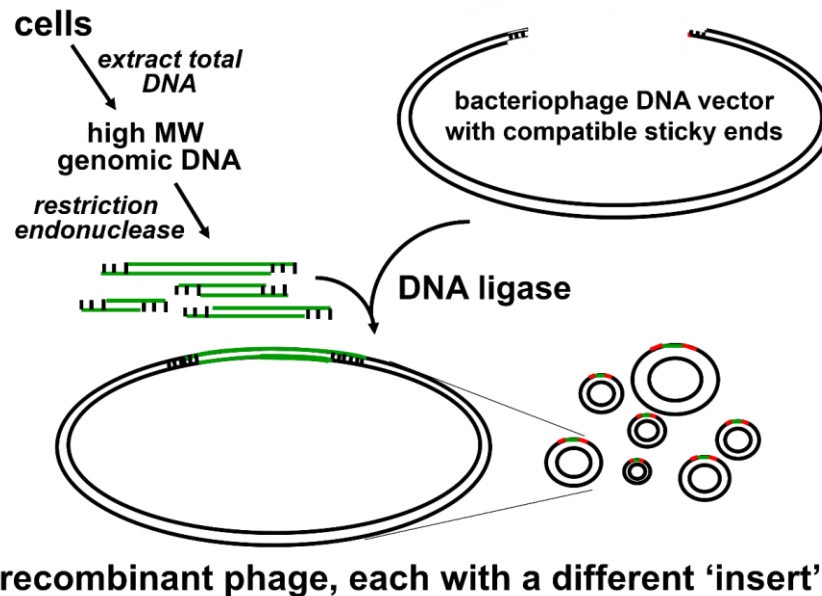


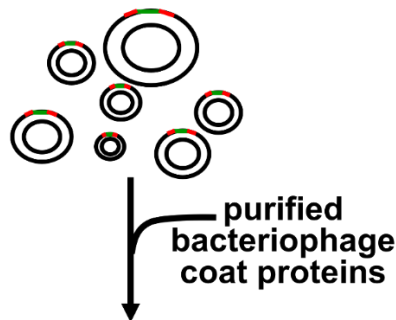
Fig. 15.22: Genomic DNA is digested with restriction enzymes (top left); the digest and phage vectors with compatible “sticky” ends are mixed (top right & center), and then ligated to make recombinant DNAs (bottom).

The recombinant phage to be made next will contain sequences that will become the genomic library.

15.4.3 Creating Infectious Viral Particles with Recombinant phage DNA

The next step is to **package** the recombined phage DNA by adding purified viral coat proteins to make infectious phage *particles* (Figure 15.23).

Packaging Viral DNA



Infectious, recombinant viruses!

Fig. 15.23: In this ‘packaging’ step, recombinant phage DNA are mixed with phage coat proteins to ‘package’ the DNA into infectious recombinant phage particles (viruses).



Packaged phages are added to a culture tube full of host bacteria (e.g., *E. coli*). After infection, the recombinant DNA enters the cells, where it replicates and directs the production of new phage eventually lyse the host cell (Figure 15.24).

Cloning Large Genomic DNA Fragments by Infection with Packaged, Recombinant Phage Vectors.

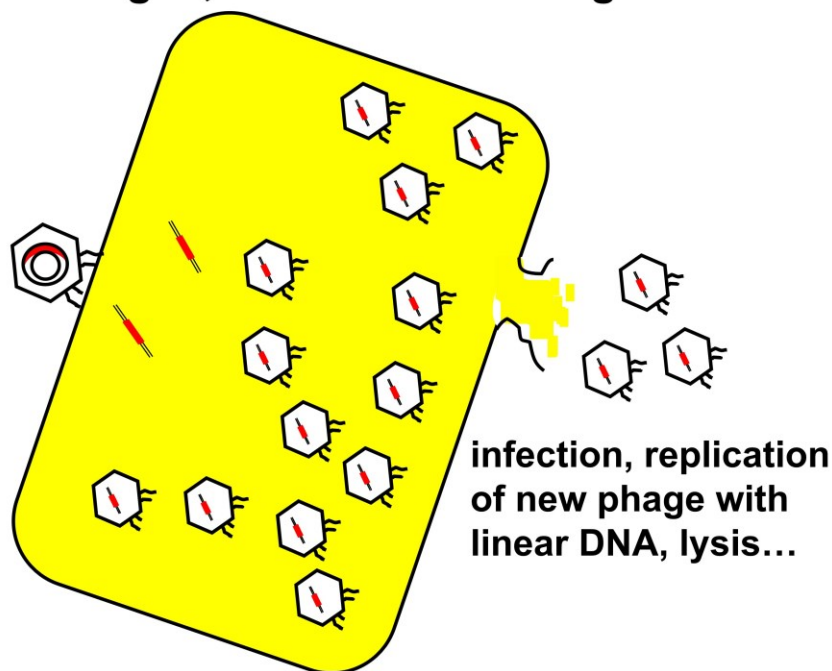


Fig. 15.24: Recombinant phage infect host cells and replicate their recombinant DNA to make new phage. The new phage accumulate, eventually lysing the host cells, forming plaques, allowing the collection of large amounts of new recombinant phage containing genomic DNA, in other words, a genomic library.

The recombined vector can also be introduced directly into the host cells by **transduction**, which is to phage DNA what transformation is to plasmid DNA. Whether by *infection* or *transduction*, the recombinant phage DNA ends up in host cells, which produce new phage that eventually lyse the host cell. The released phages go on to infect more host cells until all cells have lysed. What remains is a tube full of **lysate**, containing cell debris and lots of recombinant phage particles.



[270 Infect Host with Recombinant Phage to Make a Genomic Library](#)



Just a note on some other vectors for genomic DNA cloning: For large genomes, the goal is to choose a vector able to house larger fragments of “foreign” DNA, so that you end up screening fewer clones. We’ve seen that phage vectors accommodate larger foreign DNA inserts than plasmid vectors, and YACs even more. For a very large eukaryotic genome, it may be necessary to screen more than a hundred thousand clones in a phage-based genomic library. Apart from the size-selection of genomic fragments before you insert them into a vector, your selection of the appropriate vector is just as important. The following table lists commonly used vectors and the sizes of inserts they will accept.

Table 15.1

Cloning Vector type	Insert size (thousands of bases)
15.4 Plasmids	up to 15
15.5 Phage lambda (λ)	up to 25
15.6 Cosmids	up to 45
15.7 Bacteriophage P1	70 to 100
15.8 P1 artificial chromosomes (PACs)	130 to 150
15.9 Bacterial artificial chromosomes (BACs)	120 to 300
15.10 Yeast artificial chromosomes (YACs)	250 to 2000

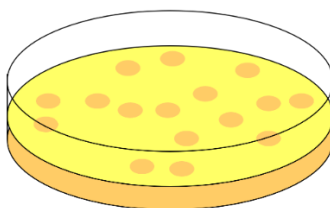
Open the links in the table or use the QR codes (at the end of the chapter) to learn more about these cloning vectors. We will continue this example by screening a phage-lysate genomic library for a recombinant phage with a genomic sequence of interest.

15.4.4 Screening a Genomic Library; Titering Recombinant Phage Clones

A **bacterial lawn** is made by plating so many bacteria on the agar plate that they simply grow together rather than as separate colonies. If a small number of phages are evenly plated over the bacterial lawn, each virus will infect one cell. Subsequent lysis of this cell releases many phage, each infecting a neighboring cell. After a day or so of repeated lysis and infection, **plaques** (clearings) appear in the lawn at the site of the first infection. Each plaque shown in the illustration in Figure 15.25 is a clone of the original phage particle.

Screening the genomic library for a globin gene clone

Spread library on bacterial lawn on agar plate;
phage lyse the bacterial cells leaving cleared areas
(**plaques**), each formed from a single phage:



The plaques represent most or all of the sequences in cloned genomic fragments.

Fig. 15.25: After spreading and incubating the genomic library on a 'lawn' of host bacterial cells, clearings form on the lawn; These are phage (viral) *plaques*, each of which is genomic clone.

To screen a phage genomic library, phage lysate is **titered** on bacterial lawns. **Titration** of a phage lysate typically consists of serial 10-fold dilutions (10X, 100X, 1000X, etc.) with a suitable medium. Each **serial 10X dilution** is then spread on a bacterial (e.g., *E. coli*) lawn, and the plaques formed on each lawn are counted. Remember, a plaque is a clone of a single virus. Let's say that you spread 10 μ l of an undiluted lysate on a bacterial lawn. After a day or

so, you see lots of tiny plaques, and after staring at the plate, you figure that there are thousands of them on the lawn (i.e., too many to count!). But among the serial 10× dilutions of the lysate on bacterial lawns you count 176 large, well-separated plaques on a 'lawn' that was plated with 10 ml of the eighth serial dilution of the lysate. Therefore, there must have been 1,408 (8×176) phage particles in 10 µl of the original undiluted lysate.

15.4.5 Screening and Probing a Genomic Library

To represent a *complete genomic library*, you will need many plates of a serial 10× dilution containing, say, about five hundred to a thousand plaques per plate. If size-selected fragments (e.g., identified by Southern blotting) were cloned to make a partial genomic library, then fewer plaques must be screened to find a sequence of interest. Since plaques contain a lot of phage DNA not yet packaged at the time of lysis, it is possible to transfer the unpackaged viral DNA directly to filters by replica plating (similar to replica plating of bacterial colonies). The replica filters are treated to denature the DNA and then hybridized to a probe with a known sequence. In the early days of cloning, the probes were often a sequenced cDNA previously isolated from libraries of the same or different (usually related) species. After soaking the filters in a radioactively labeled probe, you place an X-ray film over the filter to be exposed and developed. Black spots will form where the film is lying over a plaque containing genomic DNA complementary to the probe. In the example illustrated in Figure 15.26, a globin cDNA might have been used to probe our partial genomic library. (Globin genes were, in fact, among the first to be cloned!)

Screen a genomic library for e.g., a Globin Gene

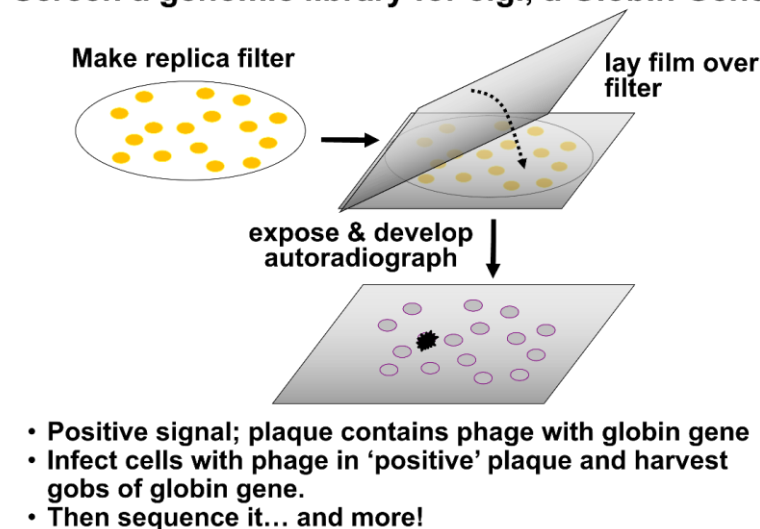


Fig. 15.26: Replica filters of plaques (upper left) are probed with (in this example) a globin DNA sequence). Next, exposure to Xray film and development of the film (right) reveals spots on the autoradiographs where the probe hybridized to recombinant phage plaques containing a globin gene sequence.

15.4.6 Isolating the Gene

Cloned genomic DNA fragments are much longer than any gene of interest, and always longer than any cDNA from a cDNA library. They are also embedded in a genome that is thousands of times as long as the gene itself, making the selection of an appropriate vector necessary. If the genome can be screened from a reasonable number of cloned phage (about a hundred

thousand plaques, for instance), the one plaque producing a positive signal on the autoradiograph would be further studied. This plaque should contain the gene of interest. At this point, we seem to have identified a clone containing a globin gene sequence. We can use this clone to infect yet more host cells and to grow up much more of the globin-gene-containing DNA for further study.



[271 Screen a Genomic Library; Pick and Grow a Phage Clone](#)

Given that a gene of interest might be a short sequence embedded in a large genomic insert that is often as long as 20 Kbp, we must further segregate the gene from the surrounding genomic DNA. Once again, the traditional strategy involves *Southern blotting*. The cloned DNA is purified and digested with restriction endonucleases, and the digest fragments are separated by *agarose gel electrophoresis*. A Southern blot is made on a filter (e.g., a nylon filter), which is soaked in a solution that denatures the DNA on the blot. The filter is then probed with the same tagged probe used to find the positive clone (plaque). The smallest DNA fragment containing the gene of interest will be typically *subcloned* and grown in a suitable vector to provide enough target-sequence DNA for further study. That suitable vector is often a plasmid easily grown in *E. coli*.

15.5 The Polymerase Chain Reaction

The polymerase chain reaction (PCR) can ***amplify*** a region of DNA from any source, like DNA fragments obtained from a fossil or residue left at a crime scene, or even DNA from a single cell. This amplification usually takes just a few hours, generating millions of copies of the desired target-DNA sequence. The result is the purification of a specific region of DNA from surrounding sequences in a single reaction! Kary B. Mullis was awarded a Nobel Prize in 1993 for his development of PCR, which is now the basis of innumerable research studies of gene structure, function, and evolution as well as applications in criminal forensics, medical diagnostics, and other commercial uses. PCR is described in detail next.



[272 PCR: Design and Synthesize Opposing Oligonucleotide Primers](#)



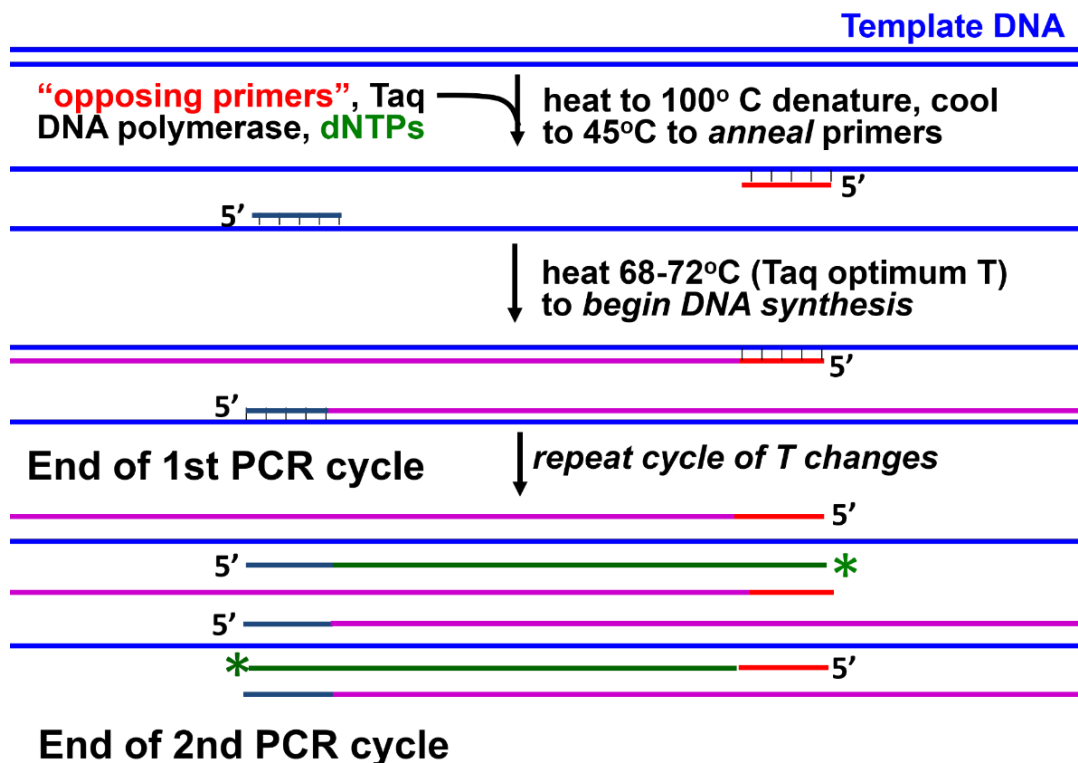
[273 PCR: The Amplification Reaction](#)

15.5.1 PCR—the Basic Process

Typical PCR relies on the investigator knowing just two bits of DNA sequence, which will be used to design and synthesize short oligonucleotides (*oligomers*). The oligomers must be complementary to sequences on opposite strands of double-stranded DNA containing the gene of interest, with their 3' ends facing (***opposing***) each other on either side of the sequence to be amplified. This way, the two oligomers can serve as *primers* for the replication of both strands of a double-stranded target-DNA sequence. Check out link #272 for further explanation.

The first step in PCR is to add oligomer primers to the target DNA from which a gene (or other genomic sequence) is to be amplified. The mixture is then heated to denature the target DNA. The mixture is cooled to allow the primers to H-bond to complementary target-DNA strands. Next, the four deoxynucleotide precursors to DNA (dATP, dCTP, dTTP and dGTP) are added, along with a small amount of a DNA polymerase. New DNA strands will now lengthen from the oligonucleotide primers on the template DNAs. To make lots of the PCR product, this reaction cycle must be repeated many times. Therefore, after allowing elongation, the mixture is heated to denature (separate) all the DNA strands. When the mixture is again cooled, the oligomers again find complementary sequences with which to H-bond. Early versions of PCR originally relied on an *E. coli* DNA polymerase, which is inactivated by heating, and so had to be re-added to the PCR mixture for each elongation cycle. When the heat stable *Thermus aquaticus* DNA polymerase (**Taq polymerase**) became available, it was adapted for PCR because it is active at high temperatures, eliminating the need to add fresh DNA polymerase after each PCR reaction cycle. It also allowed automation of PCR reactions with programmable *thermocyclers*, which raised and lowered the temperatures during PCR reactions. PCR amplification with *Taq polymerase* is shown in Figure 15.27.

The PCR Cycles



*** these strands double in each subsequent cycle.**

Fig. 15.27: In PCR, a template DNA to be amplified is mixed with a pair of primers whose sequences target opposite ends of the template. Addition of DNA polymerase (usually Taq Polymerase) initiates PCR. The actual sequence to be amplified (from primer to opposing primer) has been synthesized by the end of the second PCR cycle. Multiple cycles of PCR follow during which this 'primer-to-primer' DNA will be geometrically amplified.

You can see from the illustration that the second cycle of PCR has generated the two DNA strands that will be templates for doubling and redoubling the desired product after each subsequent cycle. A typical PCR reaction might involve thirty PCR cycles, resulting in a nearly exponential amplification of the desired sequence.

PCR amplification products are in such abundance that they can be seen under fluorescent illumination on an *ethidium bromide*-stained agarose gel (Figure 15.28).

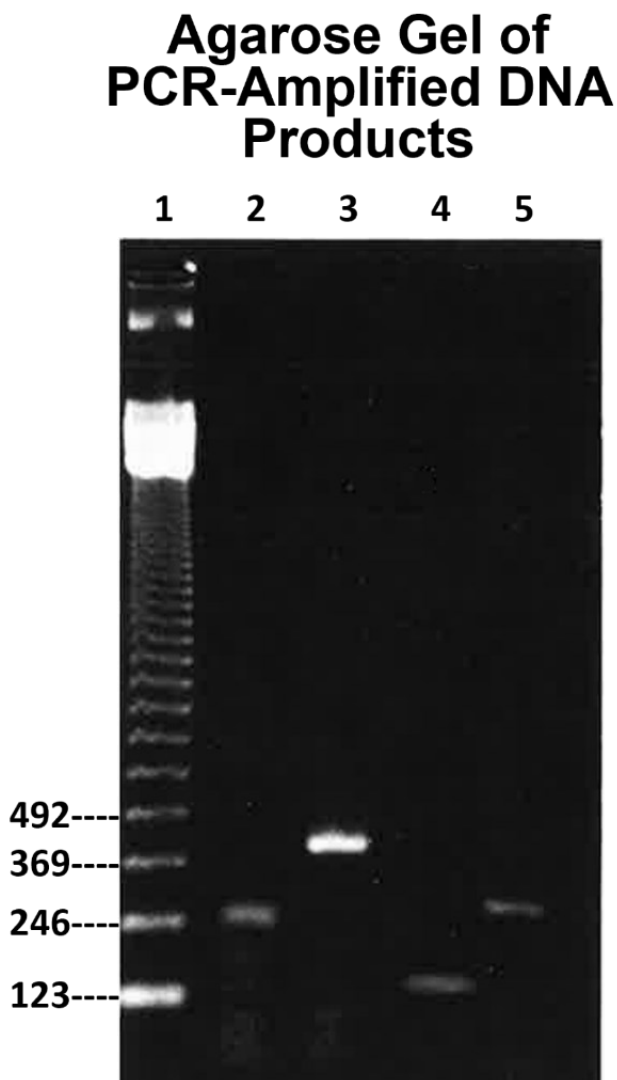


Fig. 15.28: Agarose-gel electrophoresis of PCR-amplified DNA is stained with ethidium bromide to make the DNA detectable under fluorescent light. PCR products of a globin gene clone from the insect *Chironomus thummi* are in lanes 2-5. Their sizes (lengths in base pairs) can be determined by comparison to the lengths of bands in a standard-DNA *size ladder* in lane 1.

Because they are so plentiful, PCR-amplified DNAs can be sequenced and used in many subsequent studies. In this gel, the first lane (on the left) contains a *DNA ladder*, a mixture of DNAs of known lengths that can be used to estimate the size of the bright bands (i.e., PCR products) in the remaining lanes. In this example, four PCR products were amplified from the same cloned genomic DNA preparation, using different combinations of oligonucleotide primers.

15.5.2 The Many Uses of PCR

PCR-amplified products can be labeled with radioactive or fluorescent tags to screen cDNA or genomic libraries, to determine where a DNA sequence of interest migrates on a Southern blot, or to determine where an RNA sequence of interest migrates on a Northern blot (a fanciful name for RNAs that are separated by size on gels and blotted to filter). In a major PCR advance, **Quantitative PCR** was developed to study differential gene expression and gene regulation. The technique allows cDNA amplifications from RNAs under conditions that detect *not only* the presence, but also the relative amounts of specific transcripts being made in cells. Other variant PCR protocols and applications are manifold and often quite inventive! For a list, see ^{15.11}[Variations on Basic PCR](#). A recent CRISPR/Cas9 variant amplifies DNA and may lead to PCR at 37°C, eliminating the need for a thermocycler altogether! This could prove especially useful for conducting PCR in the field. While not yet ready for prime time, check out ^{15.12}[Cas9 PCR-No Thermocycler!](#) for more. Future technological advances aside, PCR already has broad applications in research, forensic science, history, anthropology, ecological studies of species diversity, and more, and it may even reveal your own genealogy. Let's look at three of these.

15.5.2.a Forensics

PCR can be applied to identify a person or organism of interest by comparing its DNA to a standard (control) DNA. Figure 15.29 is an example of a polyacrylamide-gel DNA *fingerprint*.



Fig. 15.29: A Electrophoretic *DNA fingerprint*

Using this technology, it is now possible to detect genetic relationships between near and distant relatives (as well as to exclude such relationships), to determine paternity, to demonstrate evolutionary relationships between organisms, and to solve recent and even “cold-case” crimes using DNA left behind on surfaces at crime scenes. Recall that *Alu* sequences are ~300 bp *short-interspersed elements (SINES)* that are highly repeated throughout the human genome. DNA fingerprinting is possible in part because each of us has a unique number and distribution of *Alu* SINES in our genome. To read more about *Alu* sequences and human diversity, look at ^{15.13}[Alu Sequences and Human Diversity](#). Also see ^{15.14}[Sir Alec Jeffries](#) to learn about the origins of DNA fingerprinting in real life—and on all those TV CSI programs! For a video on DNA fingerprinting, check out ^{15.15}[Alu DNA Fingerprinting video](#).

15.5.2.b History and Society

DNA is a stable molecule (compared to RNA, which you may know is notably unstable!). This property made it the ideal molecular repository of genetic information for all life..., and for DNA fingerprinting. Not only is crime scene DNA stable, it can also survive in bones and teeth of long-dead organisms, allowing the identification of Egyptian mummies, the Russian Tsar and his family killed in 1918, and the unearthed English King Richard III.

Using DNA from the living, the missing African American ancestry is emerging, revealing the consequences of slavery in the United States, including its legacy of lingering racism (see ^{15.16}[Genetic Consequences of American Slave Trade](#), Micheletti *et al.*, 2020, *Amer. J. Human Genet.* 107, 265-277). Of more immediate impact on the living is the *Innocence Project*, started in 1992 in the US to add forensic DNA evidence in re-evaluating convictions (largely for capital crimes). The mission of the project is to use DNA fingerprinting evidence to exonerate persons wrongfully convicted of serious crimes, either on death row or serving long prison terms. As of this writing, the *Innocence Project* website reports 375 exonerations of mostly males. Of these, 225 (60%) are African American and 29 are LatinX.

15.5.2.c Who Are *Your* Ancestors?

Tracing your ethnic, racial, and regional ancestry is related to DNA fingerprinting in that it relies on PCR amplification of DNA and comparisons that distinguish *markers* in large sequence databases. Prices for commercial services have dropped, and their popularity has risen in recent years. You just provide spit or a salivary (buccal) swab. The service amplifies and sequences DNA in your samples and then compares your DNA sequences to database sequences for shared ethnic and regional markers. Based on these comparisons, you are provided with a (more or less) accurate map of your DNA-based ancestry. Folks spending about a hundred dollars (less when on sale!) often ask how accurate these analyses are and what they really mean. For example, what does it mean if your DNA says you are 5% Native American? In fact, different services can sometimes give you different results! Check out ^{15.17}[DNA Ancestry Testing](#) to get some answers and explanations.



[274 The Power of PCR: Some Examples](#)

15.6 Genomic Approaches: The DNA Microarray

Traditionally, if cellular protein levels changed in response to a chemical effector, molecular studies focused on the regulation of its gene. These studies often revealed that the regulation was at the level of transcription, turning a gene on or off through interactions of transcription factors with DNA. However, protein levels are also controlled posttranscriptionally, by the regulation of the rate of mRNA translation or degradation. Studies of transcriptional and posttranscriptional regulation mechanisms are seminal to our understanding of how the correct protein is made in the right amounts at the right time. We may have suspected, but now we know that control of gene expression and cellular responses can be more complex than increasing or decreasing the transcription of a single gene or translation of a single protein.

Newer technologies make possible the study of the expression of virtually all genes in a cell at the same time, a broadly defined field of investigation called **genomics**. As Plato might have said, necessity was 'the mother of invention' of the **microarray**, a technology that can reveal networks of regulated genes that must be understood to explain developmental and physiological changes in an organism. To hear how Patrick O. Brown conceived of microarrays as a tool to study whole genomes, and then how he made them, check out ^{15.18}[Using Microarrays to Study Cancer](#); ^{15.19}[What's On a Microarray](#); ^{15.20}[Making a DNA Microarray](#); ^{15.21}[The Genome is the Script](#); and ^{15.22}[From the Microarray to Impossible Foods](#). Figure 15.30 is a simulated microarray simultaneously probing multiple transcripts.

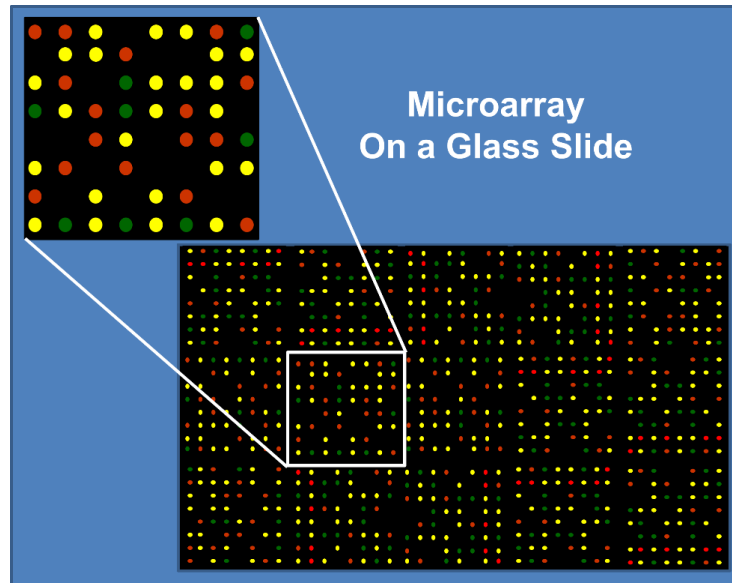


Fig. 15.30: This simulation of a glass-slide microarray shows multiple-color fluorescent spots (enlarged in the inset at the upper left), indicating a hunt for more than one DNA sequence at the same time.

Microarrays are typically made by “spotting” cloned DNA from a genomic or cDNA library, PCR products, or oligonucleotides on a glass slide, or *chip*. In microarray language, the slide is the **probe**. Spotting a chip is a robotic process. Since the DNA spots are microscopic, a cell-specific transcriptome (e.g., a cDNA library) can fit on a single chip. A small prokaryotic or viral genome microarray might also fit on a single chip. Larger genomes need several slides. In the simulated microarray above, the colored spots come from three different fluorescent tags on *specific* DNA sequences. If the spots are three specific cDNAs synthesized from cellular RNA, the microarray reveals that at least three different genes of interest were being actively transcribed in those cells at the time of RNA extraction.

In another use of microarrays, spotted genomic clones could be used to probe cDNAs made from total cellular mRNAs all of which attached to a single fluorescent tag. In this case, the question is *not* if as few specific genes were being specific genes are expressed, but rather which and how many genes are being expressed in the cell at the time of mRNA extraction. It is an approach to characterizing a cell's **transcriptome**. This is a more global question. Identifying the proteins encoded by all those genes would be the next step. Microarrays can be quantitative, so that the brightness (intensity) of the signal from each probe can be measured. Quantitative microarrays can be designed to show how global gene expression changes in cells during normal differentiation or in response to chemical signals.

DNA microarrays are also valuable for genotyping, (i.e., characterizing the genes in an organism). They are so sensitive that they can distinguish between genes or DNA regions that differ by a single nucleotide. See ^{15.23}[Single Nucleotide Polymorphisms](#) (SNPs) to learn more.

By analogy to genomics and transcriptomics, **proteomics** is the field of study of global protein interactions. Primary tools of proteomics include mass spectroscopy (which can distinguish slight differences in the mass of molecules), **Western blotting** (which can identify specific proteins immunologically after electrophoresis), and **protein microarrays**. While *Western blotting* has research and medical applications, the **protein microarray** is the most recent power tool of proteomics.

Because mass spectroscopy can detect the specific proteins and protein variants in samples, it was applied to an analysis of a **Leonardo Da Vinci** painting (*Donna Nuda*) to determine if it was actually painted by the renaissance master, an artist or artists of his school, or someone else entirely (take a look at ^{15.24}[Proteomics Determine Donna Nuda Provenance](#) to learn the verdict!).

Protein microarrays are in some ways similar to DNA microarrays, but they can look globally at protein-protein interactions, as well as the different states of proteins under different cellular conditions. Read even more about these exciting developments and their impact on basic and clinical research at ^{15.25}[Protein Microarrays from NCBI](#). Now think about this! Can we create a proteomic library analogous to a genomic library? This would seem a daunting prospect, but efforts are underway. Check out ^{15.26}[Trying to Map a Human Proteome](#) for original research leading to the sampling of a tissue-specific human proteome, and see ^{15.27}[Strategies for Approaching the Proteome](#) for more general information.

Microarrays and related technologies are powerful tools that can shift our focus from how single molecules influence events to how webs of biochemical interactions could more completely explain molecular and physiological causes and effects. Table 15.2 (adapted from Wikipedia) summarizes different applications of microarrays.

Table 15.2
The Power of Microarrays

Application or Technology	Synopsis
^{15.28} Gene-expression profiling	In a transcription (mRNA or gene expression) profiling experiment, the expression levels of thousands of genes are simultaneously monitored to study the effects of certain treatments, diseases, and developmental stages of gene expression.
^{15.29} Comparative genomic hybridization	Genome content is assessed in different cells or in closely related organisms, in which one organism's genome is the probe for a target genome from a different species.
GeneID	Small microarrays check the IDs of organisms in food and feed for genetically modified organisms (GMOs), mycoplasmas in cell culture, or pathogens for disease detection. These detection protocols often combine PCR and microarray technology.

ChIP (chromatin immunoprecipitation)	DNA sequences bound to a particular protein can be isolated by immunoprecipitating the protein. The fragments can be hybridized to a microarray (such as a tiling array) allowing the determination of protein binding-site occupancy throughout the genome.
^{15.30} DamID	Analogously to ChIP, genomic regions bound by a protein of interest can be isolated and used to probe a microarray to determine binding site occupancy. Unlike ChIP, DamID does not require antibodies but makes use of adenine methylation near the protein's binding sites to selectively amplify those regions, introduced by expressing minute amounts of the protein of interest fused to bacterial DNA-adenine methyltransferase.
^{15.31} SNP detection	Microarrays allow the detection of single-nucleotide polymorphism among alleles within or between populations. Some microarray applications make use of SNP detection, including genotyping, forensic analysis, measuring predisposition to disease, identifying drug-candidates, evaluating <i>germline</i> mutations in individuals or <i>somatic</i> mutations in cancers, assessing loss of heterozygosity, or analyzing genetic linkage.
^{15.32} Alternative splicing detection	An exon-junction array uses probes specific to expected or potential splice sites of predicted exons for a gene. Its density (coverage) is intermediate to gene expression arrays (one to three probes per gene) and genomic tiling arrays (hundreds or thousands of probes per gene). It assays the expression of alternative splice forms of a gene. Exon arrays employ probes designed to detect each individual exon for known or predicted genes, and they can be used for detecting different splicing isoforms.
^{15.33} Tiling array	Genome tiling arrays consist of overlapping probes designed to densely represent a genomic region of interest, sometimes as large as an entire human chromosome. The purpose is to empirically detect expression of transcripts or alternatively spliced forms which may not have been previously known or predicted.

Find QR codes for the numbered links in this table at the end of this chapter.



[275-2 The Power of Microarrays](#)

15.7 Ome-Sweet -Ome

Early technologies like the ones described in this chapter were applied to understanding the structure, function, and regulation of specific genes. Some more recent technologies (e.g., microarrays) are well adapted to a more holistic approach to studying cell function. Terms we have already seen (e.g., genome, epigenome, transcriptome) were coined in an effort to define different objects of study whose underlying network of molecular interactions can more accurately explain cell function. These *objects of study* often overlap and can be confusing. Here is a short compendium of *-omes*, with an attempt at clarification:

- **Genome:** total DNA content of a cell, identical in every cell of an organism
- **Proteome:** a cell's protein profile and steady state at any given moment
- **Exome:** A cell's total coding DNA (excluding noncoding DNA)
- **Epigenome:** A cell's total DNA-modification/chromatin topography
- **Methylome:** The pattern of methylation of DNA in the genome

- **Paleoproteome:** Profile of proteins found in ancient remains by mass spectroscopy
- **Transcriptome:** A cell's RNA-transcript profile and steady state at any given moment
- **Epitranscriptome:** A map of chemical modifications of RNAs that inform their function
- **Metabolome:** All of the small metabolites in a defined sample (cell, organelle, tissue, etc.)
- **Mechanobiome:** the molecular network cells use to generate, sense, and respond to intra- and extracellular forces
- **Regulome:** a cell's regulatory components, including metabolites, proteins, mRNAs, genes, cis DNA elements.

One might be excused for assuming that such cellular profiles would be the same for all cells in a tissue, only changing as gene expression is regulated during development or when signaled by extracellular events or chemical signals. But even **genomic** and **exomic** profiles can change! Recall somatic mutations in dividing cells, and genome loss in some cells (e.g., erythrocytes that emerge from our reticulocytes). Studies of large numbers of individual cells suggest profile variation even in cells of the same tissue or cell culture. How and why this is so is a rapidly growing new area of study, made possible by new tools for studying DNA, RNA and proteins molecules in single cells (see ^{15.34}[Mapping Protein Networks](#), ^{15.35}[The Dark Proteome](#), ^{15.36}[What on Earth is Paleoproteomics?](#), ^{15.37}[A Human Transcriptome Study](#), ^{15.38}[Nature vs Nurture: Diet, Behavior, & the Epigenome](#), ^{15.39}[Epitranscriptomics-Functional Modification of RNAs](#), ^{15.40}[Sex-Specific Human Transcriptomes](#) (^{15.41}[Sex-Specific Human Transcriptomes-full article](#)). Can you name the next...-ome? Maybe the **chondrome** (see ^{15.42}[Mining Mitochondrial Transcriptomes](#)).

15.8 From Genetic Engineering to Genetic Modification

Genomic, transcriptomic, and proteomic technologies have vastly increased our knowledge of how cells work at a molecular level and how genes and their regulation have evolved. We continue to add to our knowledge of the disease process and, in at least a few cases, how we can treat disease. The use of technologies to genetically engineer (i.e., modify) organisms is more controversial, despite the best of human intentions. The creators of some genetically modified organisms (GMOs) aim to rid the world of environmental waste and increase food productivity to better feed the world. The introduction of "beneficial" genes into microbes enabled the engineering of oil- and plastic-eating bacteria and food species, to create GMOs such as:

- drought-resistant crops to increase the range of land where major food crops can be grown.
- pest-resistant crops which reduce reliance on environmentally toxic chemical pesticides.
- herbicide-resistant crops which will survive the chemicals used to destroy harmful plants.
-

The quest for "improved" plant and animal varieties dates back to a time long before recorded history when farmers were crossbreeding cows, sheep, and dogs, as well as crop varieties from corn to wheat, in the hope that mating better with lesser varieties, all in the hope of producing faster-growing, larger, hardier, (you name it) varieties. It is the manipulation of DNA (the essence of the genetic material itself) to do the same that is at the root of the controversy. Controversy is reflected in opinions that GMO foods are potentially dangerous and that their cultivation should be banned. But the general consensus is that attempting to

ban GMOs is too late! In fact, you are probably already partaking of some GMO foods without even knowing it. Perhaps the good news is that after many years of GMO crops already in our food stream, the emerging scientific consensus is that GMO foods are no more harmful than unmodified foods. The current debate is whether to label foods that are (or contain) GMO ingredients as genetically modified. In an odd but perhaps amusing take on the discomfort some folks feel about GMOs, a startup company has genetically modified petunias. When grown in water, their flowers are white, but when “watered” with beer, they will produce pink or purple flowers, depending on how much beer they get. (Check it out at ^{15,46}[Can Beautiful Flowers Change Face?](#)) According to the company, it seeks “to bring what it sees as the beauty of bioengineering to the general public” (and perhaps some profit as well?).

Finally, new CRISPR and related tools can precisely edit any DNA sequences. While not yet ready for prime time, we already noted a CRISPR-based protocol even promises to amplify genes from DNA without prior denaturation at 37°C, and therefore without a thermocycler. Unlike the “*quack medicines*” of old, these tools have the real potential to cure disease, to destroy disease-carrying vectors, to cure cancer, to improve crops and possibly even alter the course of evolution. We may also need to mobilize such technologies to cope with the coming effects of climate change! The speed with which one can accomplish such good (or evil) is truly awesome.

Some iText & VOP Key Words and Terms

alternative splicing	genome	regulatory networks
automated DNA sequencing	genome projects	restriction endonucleases
autoradiography	genomic library	reticulocyte
BACs and YACs	insert DNA	reverse transcriptase
bacterial artificial chromosome vectors	library screening	RNA probes
blunt ends	linkers	RNAse
cDNA	Northern blot	shotgun sequencing
cDNA hairpin loop	oligo(dT) column	single-nucleotide polymorphisms
cDNA library	PCR	SNPs
cDNA probes	PCR primers	Southern blot
chemiluminescence	PCR steps	sticky ends
cosmid vectors	Phage-lambda vectors	systematics
dideoxy-chain termination	plasmids	Taq polymerase
dideoxy-sequencing method	poly(A) tail	thermophilic bacteria
DNA ligase	polymerase chain reaction	thermophilic DNA polymerases
DNA sequencing	primer	<i>Thermus aquaticus</i>
elution	primer walking	transcriptome
ethidium bromide	probe hybridization	transformation
fluorescence	proteome	vectors
forensics	recombinant vector	Western blot
genetic (DNA) fingerprint	recombination	yeast artificial chromosomes

CHAPTER 15 WEB LINKS





15.43



15.44



15.45



15.46



Chapter 16

Membrane Structure

Structure and function: a fluid mosaic, membrane proteins, glycoproteins, glycolipids

Reminder: For inactive links, google key words/terms for alternative resources.

You can create an art mosaic...



But can make it fluid?

16.1 Introduction

All cellular membranes share the same **phospholipid-bilayer** construction. All restrict the flow of substances from one side to the other (i.e., they are **semipermeable**). All are a **fluid mosaic** of proteins attached to or embedded in the phospholipid bilayer. Specific proteins and phospholipids structurally and functionally differentiate one kind of cellular membrane from another, both structurally and functionally. **Integral proteins** are held in the membrane by a **hydrophobic** domain, which anchors them to the hydrophobic interior of the membrane. Some integral membrane proteins span the phospholipid bilayer, with **hydrophilic** domains on either side of the membrane. In the case of the plasma membrane, the hydrophilic domains of **transmembrane proteins** interact with the watery extracellular fluid on one side and the aqueous cytoplasm on the other. Once embedded in the fatty acid innards of a cellular membrane, integral membrane proteins cannot escape! In contrast, **peripheral membrane proteins** are more loosely held in place by hydrophilic interactions with charged features of the membrane surface (e.g., phospholipid heads and hydrophilic surface-domains of integral proteins). Integral membrane proteins are often glycoproteins, the sugars of which face the outside of the cell. These cells thus present a sugar coating, or **glycocalyx**, to the outside world. As cells form tissues and organs, they become bound to extracellular proteins (which they themselves or other cells secrete) to form an **extracellular matrix**. We will spend much of this chapter looking at characteristic structures and biological activities of plasma membrane proteins and their functions.

Learning Objectives

When you have mastered the information in this chapter, you should be able to do the following:

1. Distinguish between membrane components that can move (diffuse) laterally in the membrane and those that *flip* from outer to inner surfaces of the phospholipid bilayer.
2. Compare the fluid mosaic model of membrane structure to other membrane models and argue pro and con for each.
3. Describe how cells might make their plasma membranes and suggest an experiment that would demonstrate your *hypothesis*.
4. Distinguish between *transmembrane* and *peripheral* membrane proteins, with examples.
5. Determine whether a newly discovered protein might be a membrane protein.
6. Predict the effect of *molecular* and *physical influences* on membrane fluidity.
7. Suggest how organisms living in *warm tropical waters* have adapted to the higher temperatures. Likewise, fish living under the *arctic ice*.
8. Explain how salmon can spend part of their lives in the ocean and another part swimming upstream in freshwater, without their cells shriveling or exploding.
9. List the diverse *functions* of membrane proteins.
10. Speculate on why only eukaryotic cells have evolved *sugar-coated* cell surfaces.
11. Compare and contrast the *glycocalyx* and *extracellular matrix* of cells.

16.2 Plasma Membrane Structure

In eukaryotic cells, the ***plasma membrane*** surrounds a cytoplasm that is filled with ribosomes, organelles, cytoskeletal (actin, intermediate) filaments and microtubules. Organelles are structures that are themselves encased in membranes. Some organelles (nuclei, mitochondria, chloroplasts) are even surrounded by double membranes. All cellular membranes are composed of two layers of phospholipids embedded with proteins. All are selectively permeable (***semipermeable***), allowing only certain substances to cross the membrane. The unique functions of cellular membranes are due to different phospholipid and protein compositions. Decades of research have revealed these functions. (See earlier discussions of mitochondrial and chloroplast functions, for instance.) Here we'll describe general features of membranes, using the plasma membrane as our example.

16.2.1 The Phospholipid Bilayer

In 1925, E. Gorter and F. Grendel reported red blood cell counts and the surface areas of samples of human, rabbit, dog, sheep, guinea pig, and goat red blood cells and used this data to predict a bilayer membrane structure as early as 1925. They knew that red blood cells (erythrocytes) have no nucleus or other organelles, and thus only a plasma membrane. They also knew that the membranes were mainly composed of ***phospholipids***. They then disrupted a known number of red blood cells and measured the number of phospholipids in their membrane extracts. They next calculated that there were enough phospholipid lipid molecules per cell to wrap around each cell twice. Since they knew that phospholipids had charged and hydrophobic regions, Gorter and Grendel predicted a ***phospholipid bilayer***, within which nonpolar fatty acid domains interact with each other. A space-filling model of a phospholipid in Figure 16.1 (below) highlights its ***hydrophilic*** (polar) head and ***hydrophobic*** (nonpolar) tail. Molecules with hydrophilic and hydrophobic domains are called ***amphipathic***.

Phospholipids are *Amphipathic*

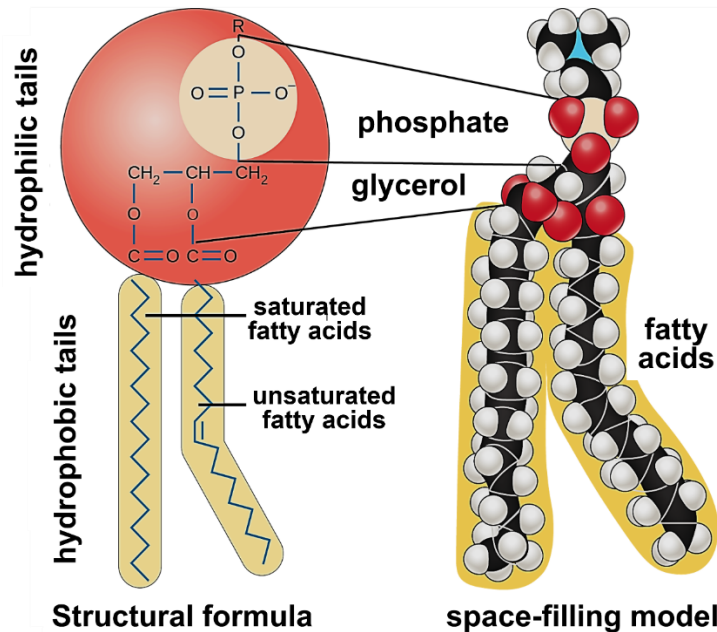


Fig. 16.1: *Structural* and *Space-Filling* models showing components of a phospholipid.

Curiously, Gorter and Grendel made two calculation errors in determining the amount of phospholipids per cells. Nevertheless, their errors compensated for each other, so—while not strictly speaking correct—their conclusion was prophetic! Figure 16.2 shows some common membrane phospholipids.

Structures of Some Common Phospholipids

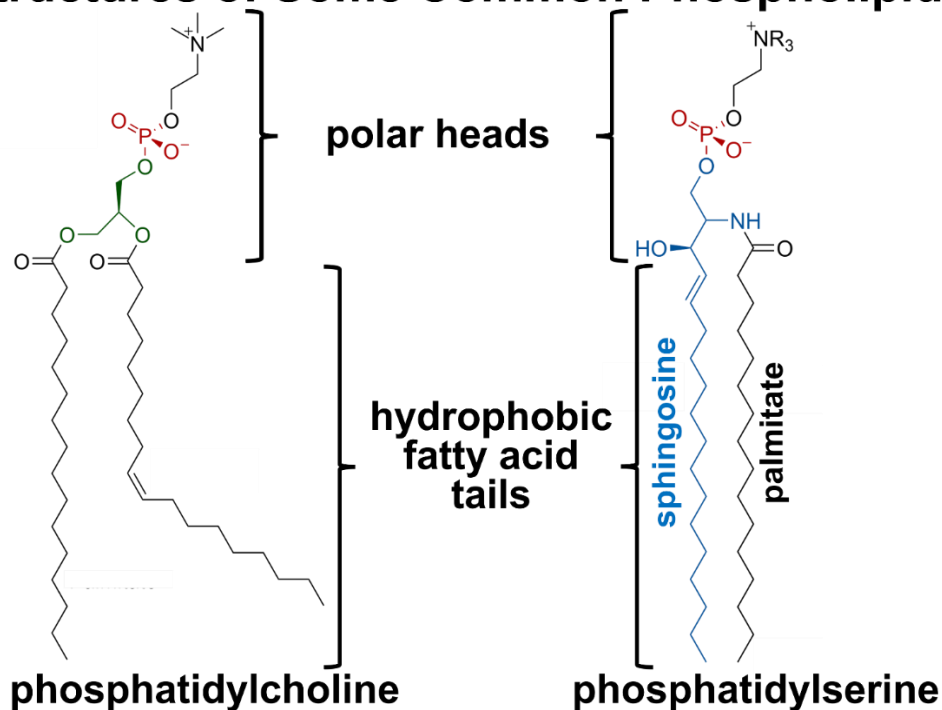


Fig. 16.2: Examples of common membrane phospholipids.

When mixed with water, the amphipathic phospholipid molecules spontaneously aggregate to “hide” their hydrophobic regions from the water, forming structures called *liposomes*, which sediment when centrifuged (see Section 16.2.5 and link 277-2 below).



[276-2 Membrane Lipids & Phospholipid Bilayer](#)



[277-2 Experiments with & Uses of Liposomes](#)



Liposome membrane structure is consistent with the predicted bilayer, in which the phospholipid hydrophobic tails interact with each other, and the polar heads face away from each other. This conceptual understanding led to a model of membrane architecture based on phospholipid interactions, in which the fatty acid tails make up a hydrophobic membrane interior, with the hydrophilic (polar) phosphates on the surfaces, facing the opposite, aqueous sides of the membrane. The iconic image of a phospholipid bilayer is illustrated in Figure 16.3

Structure of a Phospholipid Bilayer Membrane

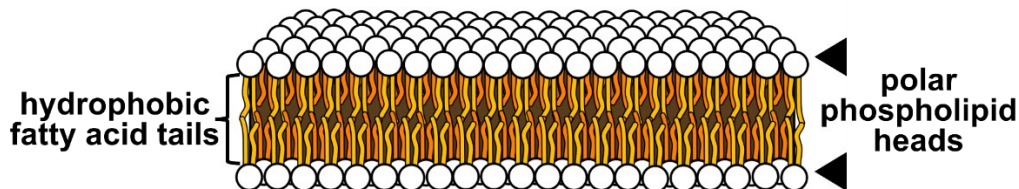


Fig. 16.3: Phospholipid bilayer membrane.

16.2.2 Models of Membrane Structure

In 1935, H. Davson and J. Danielli suggested that proteins might be bound to the polar heads of the phospholipids in the plasma membrane, creating a protein-lipid-protein sandwich. Decades later, J. D. Robertson observed membranes in the transmission electron microscope at high power, revealing that all cellular membranes had a *trilamellar* structure. The classic trilamellar a cellular membrane can be seen in the electron micrographs in Figure 16.4.

Trilamellar Membrane Structure

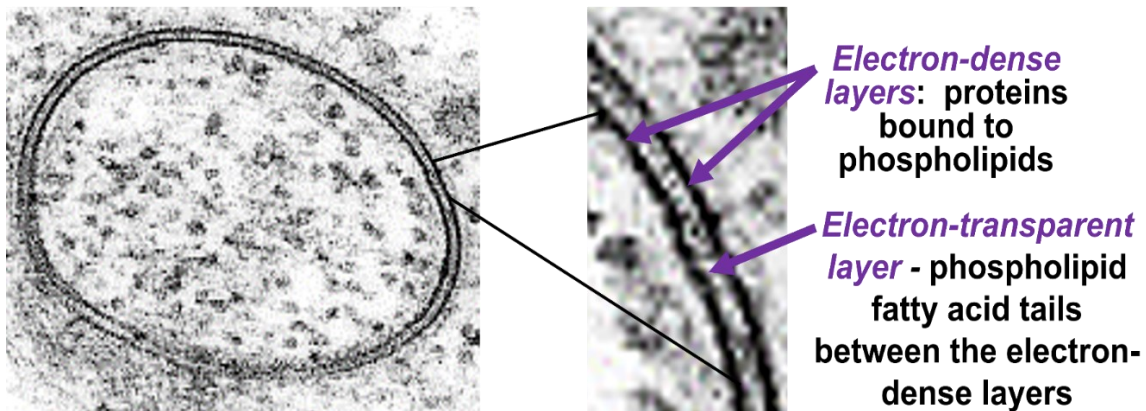


Fig. 16.4: Low- and high-magnification transmission electron micrographs of the *trilamellar* membrane.

The trilamellar structure is consistent with the protein-lipid-protein model proposed by Hugh Davson and James Danielli (since known as the **Davson-Danielli model**) of the phospholipid bilayer. Since *all* cellular membranes had this trilamellar structure when seen by electron microscopy, J. David Robertson further proposed his **Unit Membrane** model: *all membranes* consist of a clear phospholipid bilayer, coated with electron-dense proteins.

The static view implied by the Davson-Danielli or Robertson models of membrane structure was replaced in 1972 by Singer and Nicolson's **Fluid Mosaic** model (see S. J. Singer and G. L. Nicolson, *The fluid mosaic model of membranes*. Science 175:720-731). They suggested that, in addition to **peripheral proteins** (which bind to the membrane surfaces), many **integral membrane proteins** span the membranes. *Integral-membrane proteins* were imagined as a *mosaic* of protein "tiles" embedded in a phospholipid medium. But unlike a mosaic of glazed tiles set in a firm, cement-like structure, the protein "tiles" were predicted to *float in a phospholipid sea*.

In the model, these proteins are anchored in membranes by one or more *hydrophobic* domains, while *hydrophilic* domains face either or both aqueous environments (external and cytosolic). Thus, like phospholipids themselves, membrane proteins are *amphipathic*. We know that cells expose different structural and functional surface features to their aqueous environments. Since the exposed domains of membrane proteins would be different on opposite sides of a membrane, we say that cellular membranes are **asymmetric**. Figure 16.5 is a model of the plasma membrane of a cell and includes its integral proteins, peripheral proteins, and glycoproteins.

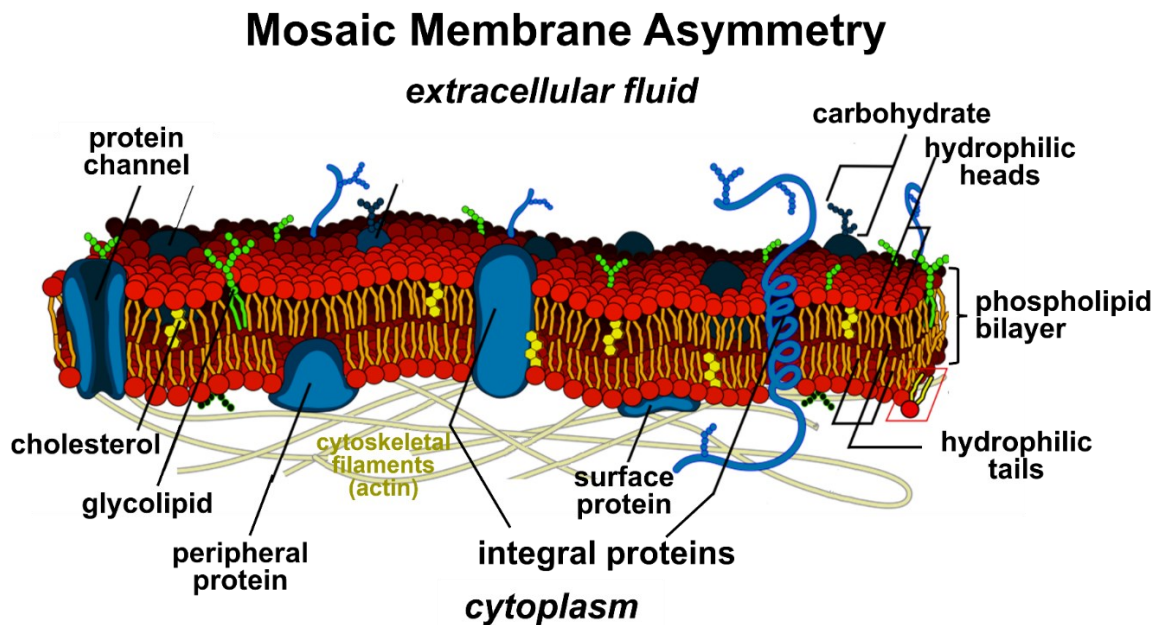


Fig. 16.5: This model of the eukaryotic plasma membrane demonstrates membrane asymmetry, with a sugarless cytoplasmic surface and a glycoprotein-rich extracellular surface.

In this model, some integral proteins have a hydrophobic domain that does not span the membrane but that anchors it to one side of the membrane. Peripheral (or so-called "surface") proteins are also shown. These are less firmly bound to the membrane via charge interactions with the polar phosphate groups of phospholipids or via the polar domains of the

integral membrane proteins. Because of their own aqueous, hydrophilic domains, membrane proteins present a natural barrier to the free passage of charged molecules across the membrane.

Membrane proteins are also responsible for the ***selective permeability*** of membranes, facilitating the movement of *specific* molecules into and out of cells (endocytosis and exocytosis, respectively). Membrane proteins also account for specific and selective interactions with their extracellular environment. These interactions include the adhesion of cells to each other, the cell's attachment to surfaces, communication between cells (both direct and via hormones and neurons), and more.

The "sugar coating" of the extracellular surfaces of plasma membranes comes from ***oligosaccharides*** that are covalently linked to membrane proteins (to form ***glycoproteins***) or to phospholipids (to make ***glycolipids***). Carbohydrate components of ***glycosylated*** membrane proteins inform their function. Thus, glycoproteins enable specific interactions of cells with each other to form tissues. They also allow interaction with the extracellular surfaces to which they must adhere. In addition, they figure prominently as part of receptors for many hormones and other chemical-communication biomolecules. The intracellular domains of plasma membrane proteins (as well as those regions crossing the lipoidal region of the membrane itself) are not glycosylated. The cytoplasmic domains of some membrane proteins become attached to components of the cytoskeleton, giving cells their shape and allowing them to change shape when necessary. Other intercellular membrane protein domains have essential enzymatic features, as we will see. Given the crucial role of proteins and glycoproteins in membrane function, it should come as no surprise that proteins constitute an average of 40-50% of the mass of a membrane. In some cases, proteins are as much as 70% of membrane mass (think cristal membranes in mitochondria!).



[278-2 Properties of Proteins Embedded in a Phospholipid Bilayer](#)



[279 Different Membrane Compositions](#)

16.2.3 Evidence for Membrane Structure

Membrane asymmetry, in which the features facing one side of the membrane differ from those facing the opposite side, was directly demonstrated by the scanning electron microscope technique of ***freeze-fracture***. The technique involves freezing isolated membranes in water, then chipping the ice. When the ice cracks, the encased membranes split along a *line of least resistance*, which turns out to be between the hydrophobic fatty acid opposing tails in the interior of the membrane. Scanning electron microscopy of a freeze-fractured membrane reveals features of the interior and exterior membrane surfaces that confirm membrane asymmetry. Among the prominent features shown by these micrographs are the *pits* and opposing *mounds* in the interior of the membrane. Figure 16.6 (below) illustrates the results of freeze-fracture electron microscopy of a plasma membrane.

Freeze-Fractured Plasma Membrane

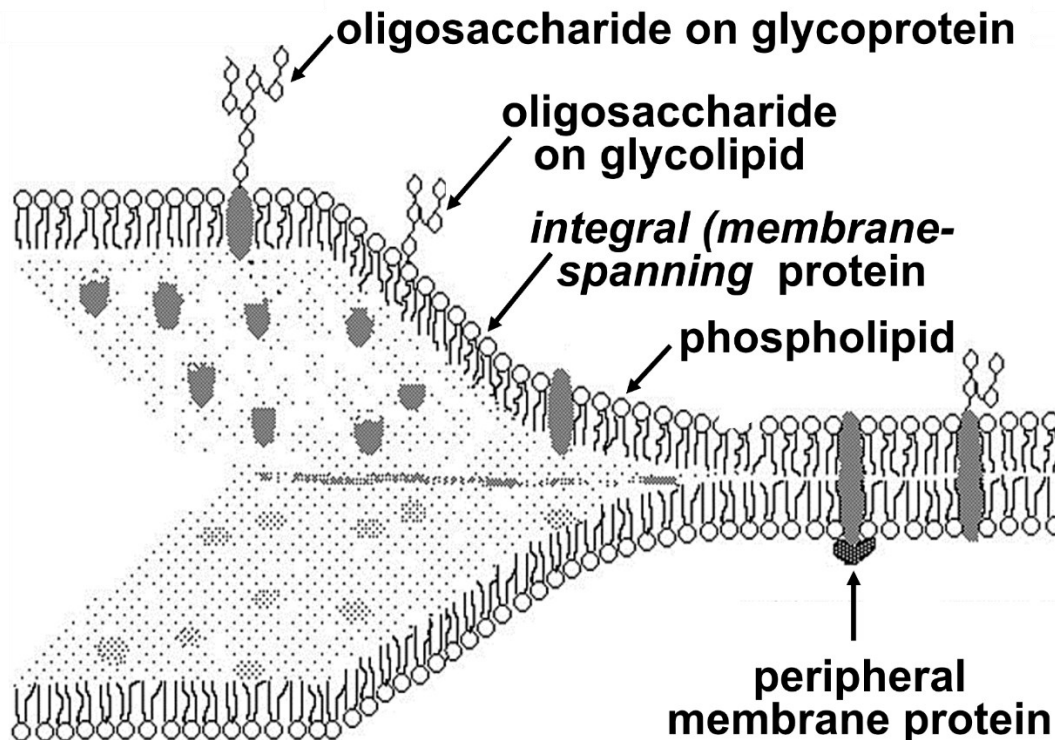


Fig. 16.6: In this drawing of a scanning-electron micrograph of a freeze-fractured plasma membrane, pits and mounds are seen on opposing phospholipid layers of the membrane.

Other features shown here are also consistent with phospholipid membrane structure.



[280-2 Freeze-Fracture Electron Microscopy of Cell Membranes](#)

Cytochemical studies also confirmed plasma membrane asymmetry, showing that only the external surfaces of plasma membranes are sugar coated. Check the following link for more detailed descriptions of the experiments.



[281-2 EM Cytochemical Demonstration of Membrane Asymmetry](#)

Finally, membrane asymmetry was also demonstrated biochemically. In one experiment, investigators treated whole cells with proteolytic enzymes, then extracted the membranes and isolated the membrane proteins. In a second experiment, investigators isolated plasma membranes from untreated cells first, *and then* treated the membranes with the enzymes. In a third experiment, investigators isolated plasma membranes from untreated cells, and then extracted the membrane proteins. Size separation of the three protein extracts

by electrophoresis demonstrated that different components of integral membrane proteins were present in the two digest experiments, confirming the asymmetry of the plasma membrane. For more details, check the following link.



[282-2 Electrophoretic Demonstration of Membrane Asymmetry](#)

The idea that membranes are *fluid* was also tested experimentally. Here is a description of the experiment:

1. Antibodies were made to mouse and human cell membrane proteins by isolating their membranes and injecting the isolates into a different animal (rabbits, most likely). Each rabbit's cells saw the membranes and their associated proteins as foreign and responded by making specific anti-membrane antibody molecules.
2. The rabbit antibodies against each membrane source were then isolated and separately tagged with colored fluorescent labels, so that each type would glow a different color when subjected to ultraviolet light.
3. Next, mouse and human cells were mixed under conditions that caused them to fuse, making human-mouse hybrid cells.
4. The tagged antibodies were added to the fused human-mouse cells and examined in a fluorescence microscope.

At first, the mouse and human antibodies were seen to bind to the mouse and human portions of the fused cell, as shown in Figure 16.7.

Fused Mouse & Human Cells After Exposure to Antibodies Made Against Human and Mouse Plasma Membrane Protein

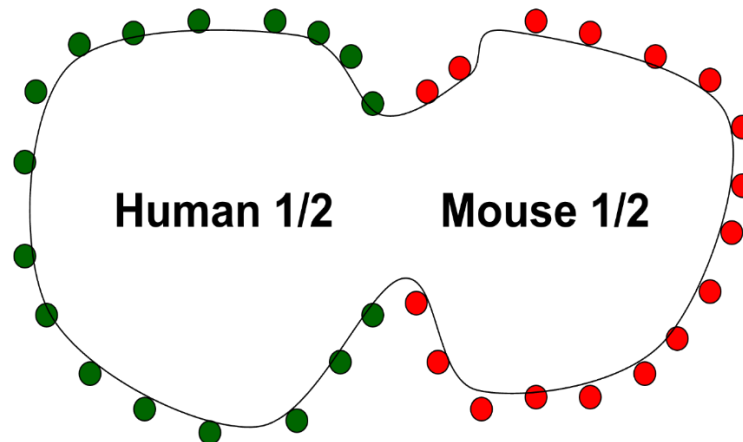


Fig. 16.7: Fluorescent antibodies against human and mouse membrane proteins localize their cell surface antigens (i.e., proteins) on opposite poles of recently fused cells.

After a short time, the different fluorescent antibodies were seen to mix under a fluorescence microscope under UV light. The fluorescent tags seemed to be moving from their original location in the fused membranes (Figure. 16.8).

Immunofluorescence Micrograph of Fused Human & Mouse Cells Incubated Over Time

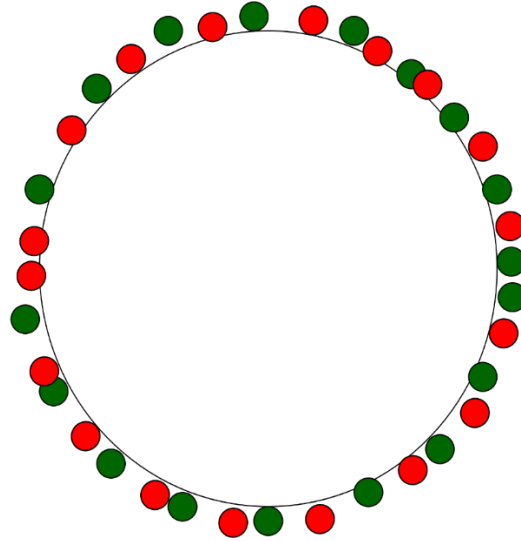


Fig. 16.8: Immunofluorescent human (green) and mouse (red) membrane proteins diffuse and mix over time in fused cells.

So, proteins embedded in the membrane are not static; rather, they are able to move laterally in the membrane, in effect floating and diffusing in a “sea of phospholipids.”




[283 Two Demonstrations of Membrane Fluidity: The Fluid Mosaic](#)



16.2.4 Chemical Factors Affecting Membrane Fluidity

As you might expect, the fluidity of a membrane depends on its chemical composition and the physical conditions surrounding the cell (e.g., outside temperature). Just as heating a solution causes dissolved molecules and particulates to move faster, so also at higher external temperatures, membrane phospholipids *and* proteins are more fluid. Also, the fatty acids of phospholipids that have more *unsaturated* (C=C) carbon bonds have more kinks, or bends. These tend to push apart the phospholipid tails. With more space between the fatty acid tails, membrane components can move more freely. Thus, membranes with higher levels of *polyunsaturated* fatty acids are more fluid. On the other hand, cholesterol molecules tend to fill the space between fatty acids in the hydrophobic interior of the membrane. This reduces the lateral mobility of the phospholipid and protein components in the membrane. By reducing fluidity, cholesterol reduces membrane permeability to some ions. Factors that affect membrane fluidity are summarized in Figure 16.9 (below).

Membrane Fluidity Depends on Temperature, Fatty Acid Saturation and Cholesterol

- higher T leads to increased fluidity
- more *unsaturated* fatty acids leads to more fluidity
- more *cholesterol* () stiffens membranes by filling in gaps between p-lipids, decreasing fluidity

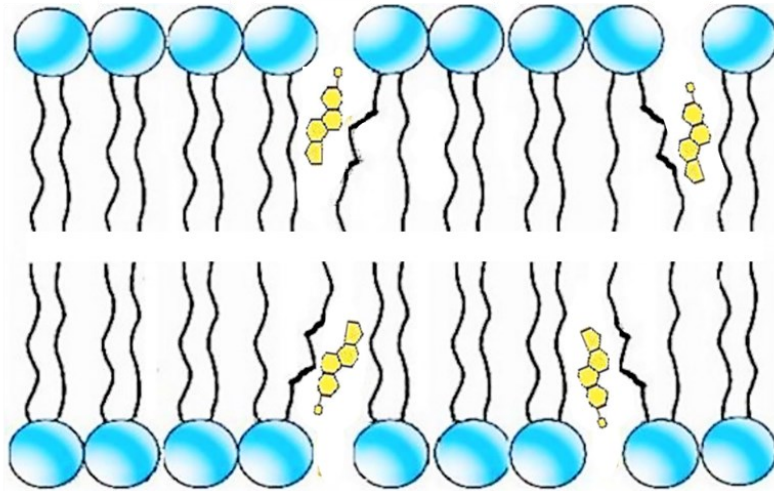


Fig. 16.9: Several factors affect the rate of diffusion (membrane fluidity) of components in a fluid mosaic phospholipid membrane. Cholesterol is one of these factors.

Evolution has adapted cell membranes to different and changing environments in order to maintain the fluidity necessary for proper cell function. **Poikilothermic** (i.e., cold-blooded) organisms, from prokaryotes to fish and reptiles, do not regulate their body temperatures. When exposed to lower temperatures, poikilotherms respond by increasing the *unsaturated* fatty acid content of their cell membranes. At higher temperatures, they increase their cell membrane's *saturated* fatty acid content. Thus, the cell membranes of fish living under the arctic ice maintain fluidity by having high levels of both monounsaturated and polyunsaturated fatty acids. What about fish species that range across warmer and colder environments (or that live in climates with changing seasons)? These fish regulate their cell membrane composition to adjust to the demands of their changing environment. The warm-blooded (**homeothermic**) mammals and birds maintain a constant body temperature. Thus, their membrane composition is also relatively constant. But there is a paradox! Their cell membranes are very fluid, with a higher ratio of *polyunsaturated* fat to *monounsaturated* fats than, say, in reptiles. But the paradox is resolved when we understand that this greater fluidity supports the *higher metabolic rate* of the warm-blooded species compared to poikilotherms. Just compare the lifestyles of almost any mammal to a lazy floating alligator or a snake basking in the shade of a rock!



[284-2 Factors Influencing Membrane Fluidity](#)

16.2.5 Making and Experimenting with Artificial Membranes

Membrane-like structures can form spontaneously. When phospholipids interact in an aqueous environment, they aggregate to exclude their hydrophobic fatty tails from water, forming **micelles**. Micelles are spherical phospholipid monolayer vesicles that *self-assemble*, a natural aggregation of the hydrophobic fatty acid domains of these amphipathic molecules. Figure 16.10 illustrates a micelle.

Micelles form when phospholipids are mixed and agitated in water

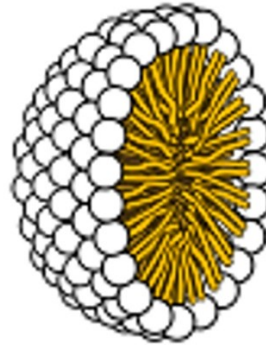


Fig. 16.10: Micelles are phospholipid monolayers with hydrophobic interiors that self-assemble in water.

Micelles can further form spherical phospholipid bilayer **liposomes** (Figure 16.11).

Micelles can be treated further to form phospholipid bilayer spheres called liposomes

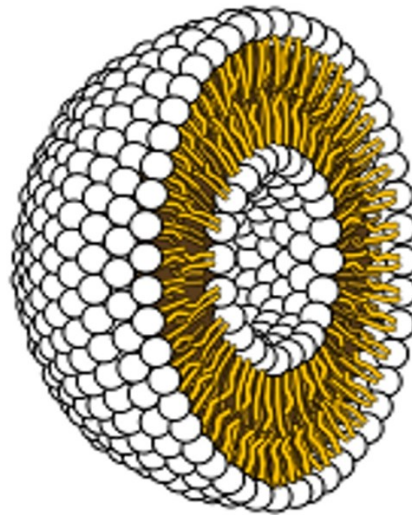


Fig. 16.11: A liposome is a synthetic lipid bilayer that can be formed from micelles.

Liposomes have a trilamellar membrane and behave somewhat like cells; for example, forming a pellet at the bottom of a tube when centrifuged. Liposomes can be custom designed from the different kinds of phospholipids and amphipathic proteins that have become integral to their membranes. When investigators prepare liposomes in the presence of specific proteins or other molecules that can't diffuse across the liposomes' membranes, these molecules become trapped in the vesicles. The trapped molecules cannot get out of this synthetic "organelle." Such were the studies that allowed the identification of the mitochondrial respiratory-chain complexes. The ability of investigators to manipulate liposome content and membrane composition also makes liposomes candidates for drug delivery to specific cells and tissues. (Google "liposome" for more information.)

16.2.6 Separate Regions of a Plasma Membrane with Unique Fluidity and Permeability Properties

As we'll see shortly, fluidity *does not* result in an equal diffusion of all membrane components around the cell-membrane surface. Instead, extracellular connections between cells, along with intracellular connections of the membrane to differentiated regions of the cytoskeleton, effectively compartmentalize the membrane into subregions. To understand this, imagine an epithelial cell sheet like the one below based on electron micrographs in Figure 16.12.

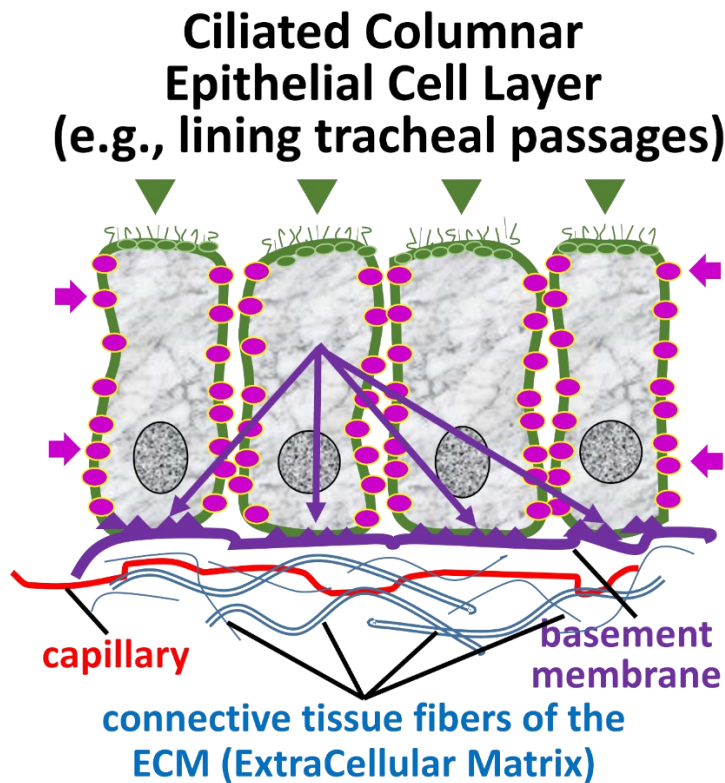


Fig. 16.12: Properties of plasma membranes may differ in ways determined by cell-to-cell associations during tissue development. The lavender (short arrows), green (arrow heads), and purple (long arrows) indicate the differentiation of a sheet of epithelial cells into specific regions.

This sheet of cells exposes a domain (surface) with unique functions to the inside of the organ it lines. It exposes the opposite sheet surface, one with a quite different function, to fluid surrounding the organ. The lateral cell surfaces comprise yet another membrane compartment, one that functions to connect and communicate between the cells in the sheet. Membrane proteins, illustrated here with different symbolic shapes and colors, may remain fluid within a compartment. Of course, tissue-level *macro-differentiation* of cell membranes enabling cell-cell and cell-environmental interactions make intuitive sense. The possibility that cellular membranes are even more compartmentalized was perhaps less anticipated. In fact, membranes are further divided into micro-compartments. Within these compartments, components are fluid but seldom move between compartments. Studies indicate that cytoskeletal elements maintain these micro discontinuities. For example, integral membrane proteins are immobilized in membranes if they are attached to cytoskeletal fibers (e.g., actin) in the cytoplasm. Furthermore, aggregates of these proteins can line up due to similar interactions, forming a kind of *fence*, inhibiting other membrane components from crossing. By

analogy, this mechanism of micro-compartmentalization is called the *Fences and Pickets* model: proteins attached to the cytoskeleton serve as the pickets (i.e., the usually white, pointy boards attached to a literal picket fence). The movement across the fences (i.e., from one membrane compartment to another) is infrequent. Extra kinetic energy is presumably needed for a molecule to “jump” a fence between compartments. Hence, this kind of motion (*hop diffusion*) differs from the Brownian motion implied by the original fluid mosaic model.



285-2 Regional Differentiation of a Plasma Membrane Domains

16.3 Membrane Proteins

Of course, membrane proteins themselves have domains. These provide catalytic and other activities inside and outside of cells and organelles, and keep them attached to the membrane. Like phospholipids, membrane proteins are amphipathic, with hydrophobic domains that *noncovalently* interact strongly within the fatty-acid interior of membranes. Some integral-membrane proteins span the entire membrane, with hydrophilic domains facing the cytosol or cell exterior. Peripheral proteins bind to a membrane surface through noncovalent interactions. The different interactions of membrane proteins with e.g., a plasma membrane are shown below in Figure 16.13.

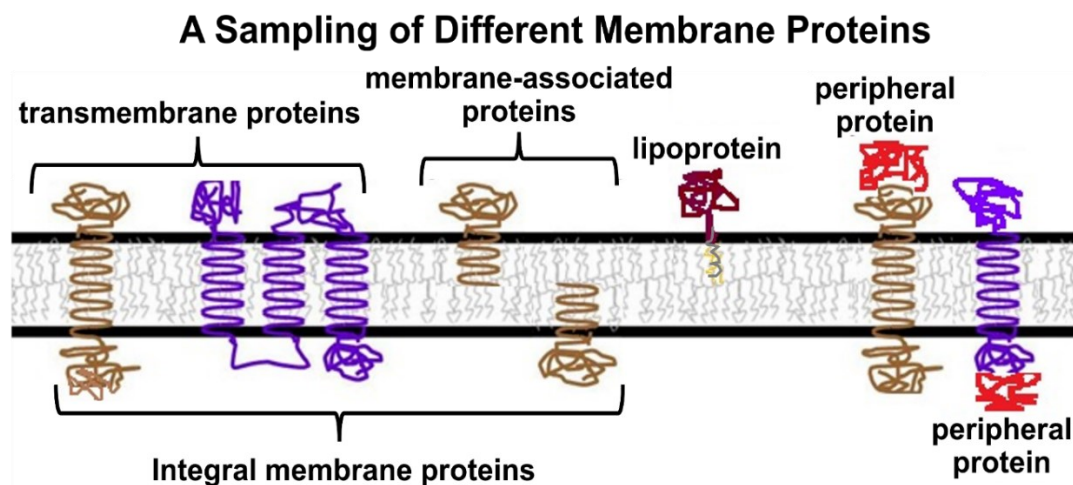


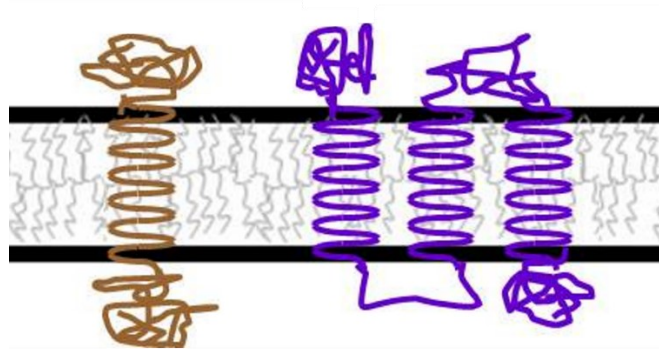
Fig. 16.13: Integral-membrane proteins may penetrate or span the membrane; also shown are peripheral proteins and lipoproteins.



286-2 Domains of Membrane Proteins

Hydrophobic amino acids of membrane proteins are organized into functional regions. These consist of one or more nonpolar alpha-helical domains that interact with the fatty-acid interior of the membranes. Hydrophilic domains tend to have a more tertiary structure. These domains face the aqueous cytosol and cell exterior. Two transmembrane proteins are cartooned in Figure 16.14.

Transmembrane Proteins



**...span the membrane
one or more times**

Fig. 16.14: Integral transmembrane proteins cross the membrane one or more times.

The protein on the left in Figure 16.14 crosses the membrane once, while the one on the right crosses the membrane three times. Regardless of the number of times a polypeptide crosses the membrane its C-terminus always ends up on the extracellular surface of the cell. Alpha-helical domains that anchor proteins in membranes are mostly nonpolar and hydrophobic themselves. For example, consider the amino acids in the alpha-helical domain of the red blood cell-membrane protein *glycophorin A*, a protein that prevents red blood cells from aggregating or clumping in the circulation. One glycophorin A polypeptide with its hydrophobic transmembrane alpha helix is cartooned in Figure 16.15 (below).

**Glycophorin A, showing mostly
hydrophobic amino acids in its
transmembrane domain:**

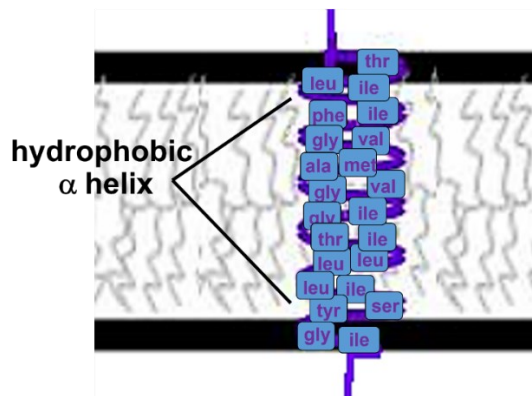


Fig. 16.15: Hydrophobic amino acids in glycophorin form a helix spanning a red blood cell plasma membrane.

Glycophorin A monomers pair to form dimers in the plasma membrane (not shown above). Proteins that span membranes multiple times may include amino acids with charged polar side chains, provided that these side chains interact between helices, so that they are shielded from the fatty acid environment in the membrane. Because of these hydrophilic

interactions, such proteins can create **pores** for the **transport** of polar molecules and ions. (We'll see some of these proteins later).

Integral membrane proteins that do not span the membrane also have a hydrophobic helical domain to anchor them in the membrane, while their hydrophilic domains typically interact with intracellular or extracellular molecules that can hold cells in place, give cells and tissues their structure, and the like.

The very presence of hydrophobic alpha-helical domains in transmembrane proteins makes them difficult if not impossible to isolate from membranes in a biologically active form. But the peripheral polypeptide *cytochrome c* readily dissociates from cristal membranes, making it easy to purify. The inability to purify other biologically active cristal membrane electron carriers is what slowed our understanding of the structure and function of the mitochondrial electron transport system.

It is possible to determine the primary structure of a polypeptide encoded by a gene before the protein itself has been isolated. Just by knowing the DNA sequence of a gene, we can infer the amino acid sequence of the protein encoded by the gene. Then we can identify all of the hydrophobic amino acids in the inferred sequence and generate a **hydrophobicity** (or **hydropathy**) plot, such as the one below, in Figure 16.16.

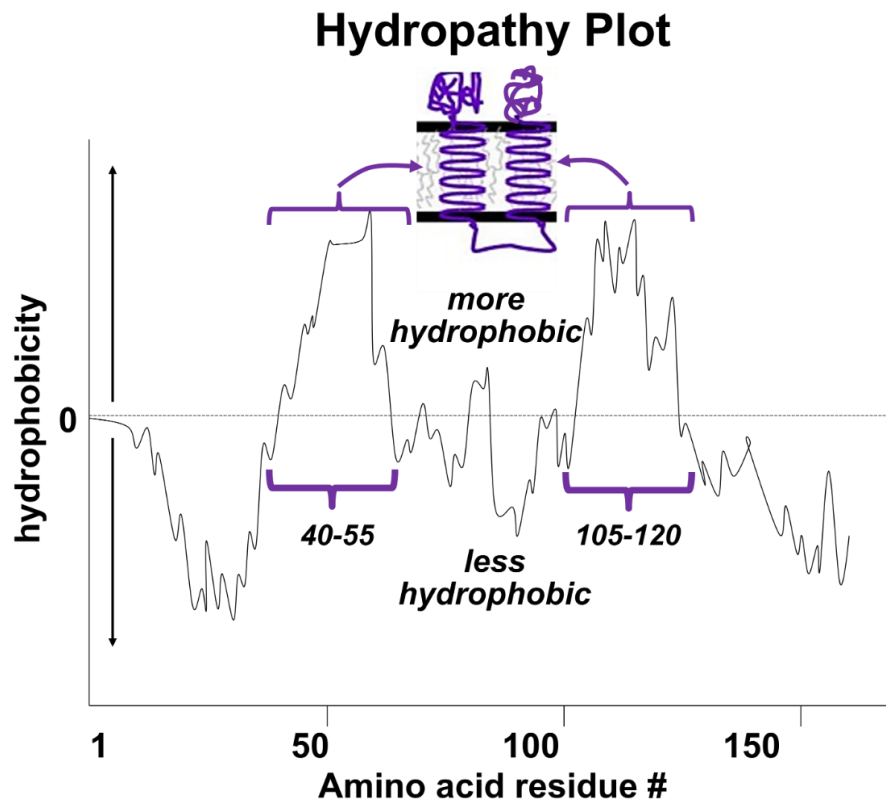


Fig. 16.16: Hydropathy plots correlate amino acid hydrophobicity with its position in a polypeptide. Long regions of hydrophobic amino acids suggest a possible membrane-protein domain.

The protein in this hypothetical example has two regions, or domains of hydrophobic amino acids. They are identified as stretches of uninterrupted *hydropathy* (hydrophobicity)

above the 0 level along the X-axis. The protein shown here is probably a transmembrane protein. To see how a hydropathy plot can predict whether a protein is a membrane protein, check out the following link.



 [287-2 Hydropathy Predicts Hydrophobic-Membrane Protein Domains](#)

16.4 A Diversity of Membrane Protein Functions

Membrane protein functions include doing the following:

- serving as receptors for hormones or neurotransmitters.
- working as immune-system antibodies (immunoglobulins) to recognize foreign substances (antigens).
- acting as cell-recognition molecules that bind cells together.
- forming cell membrane structures that directly pass chemical information between cells.
- anchoring cells to extracellular surfaces like connective tissue.
- anchoring the plasma membrane to proteins of the cytoskeleton (e.g., actin).
- facilitating molecular transport (entry or exit of substances into or out of cells).
- working as enzymes to catalyze crucial reactions in cells.
- electron transport (in mitochondria and chloroplasts).

Figure 16.17 summarizes some membrane protein functions.

Membrane Protein Functions

- **anchor** extracellular matrix and cytoskeletal proteins
- **enzymes** (e.g., mitochondrial respiratory (ET) complexes, F1 ATP synthase)
- **receptors** for chemical signals (e.g., hormones)
- **passive transport** of solutes into or out of cells (facilitated diffusion)
- **active transport** of solutes into or out of cells, requiring energy
- cell-cell recognition, communication

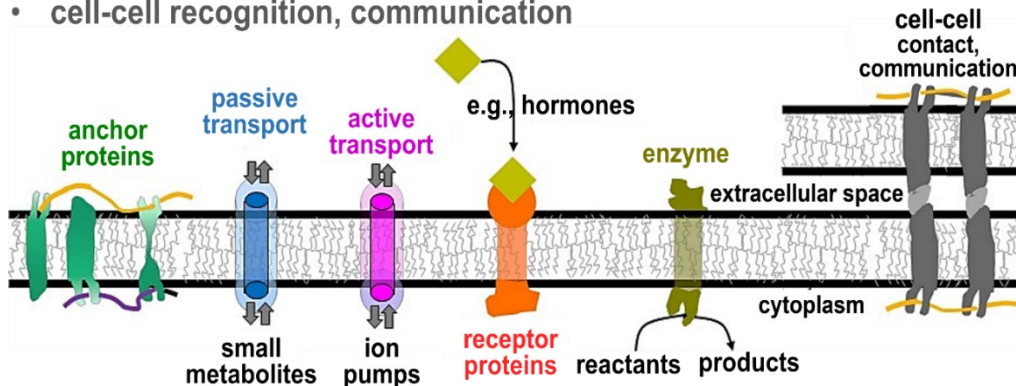


Fig. 16.17: Examples of the many functions of membrane proteins.

Transmembrane proteins perform most of the functions illustrated in this illustration. However, peripheral membrane proteins also play vital roles in membrane function. *Cytochrome c* is an example. It is a redox component loosely bound to the rest of the electron transport system in the mitochondrial cristal membrane. Other peripheral membrane proteins may serve to regulate transport or signaling activities of transmembrane-protein complexes or may mediate connections between the membrane and cytoskeletal elements. As shown here, peripheral proteins do not penetrate membranes. They bind reversibly to the internal or external surfaces of the biological membrane with which they are associated. Shortly, we'll take a closer look at what holds membrane proteins in place and how they perform their unique functions. Check out major membrane-protein functions, actions, and cellular locations in Table 16.1, below.

Table 16.1
Some Functions of Membrane Proteins

Basic Function	Specific Actions	Examples
Electron transport	Oxidizes reduced energy substrates (e.g., NADH, FADH ₂ , NADPH)	Mitochondrial respiration, photosynthesis
Facilitated transport, facilitated diffusion	Regulate diffusion of substances across membranes along a concentration gradient	Ca ⁺⁺ and other ion channels, glucose transporters
Active transport	Use energy to move ions from low to high concentration across membranes	Mitochondrial protein pumps, the Na ⁺ /K ⁺ ion pump in neurons
Signal transduction	Conveys information from molecular signals to cytoplasm (e.g., for hormones that can't enter cells), leading to a cellular response	Protein hormone and growth factor signaling, antibody/antigen interactions, cytokine mediation of inflammatory responses.
Cell-cell interactions	Cell-cell recognition and binding to form tissues	Formation of desmosomes, gap junctions, and tight junctions
Anchors to cytoskeleton	Link membrane proteins to cytoskeleton	Cell shape, cell movement, and cell response to molecular signals
Enzymatic	Usually multifunctional proteins with enzymatic activities	F1 ATP synthase; uses proton gradient to make ATP; adenylyl cyclase, which makes cAMP during signal transduction; note that some receptor proteins are linked to enzymatic domains in the cytoplasm.



[288-2 Diversity of Membrane Protein Structure & Function](#)



[289-2 Pore Proteins May Cross the Membrane Many Times](#)



[290-2 Red Blood Cell \(Erythrocyte\) Membrane Protein Functions](#)

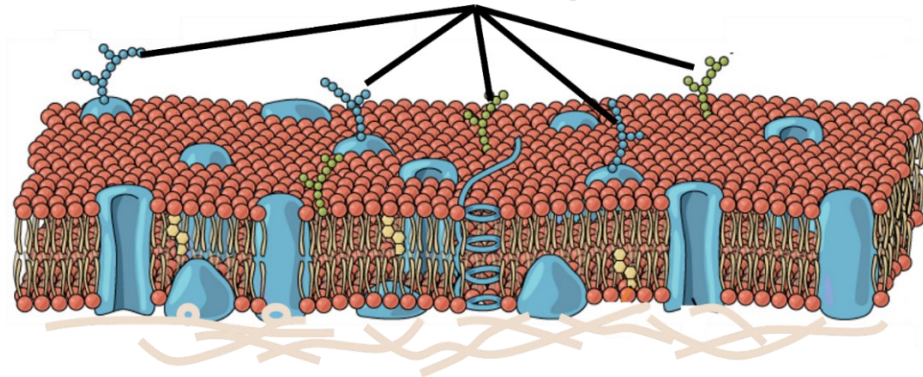


16.5 Glycoproteins and Glycolipids

Membrane proteins are often covalently linked to **oligosaccharides**, which are branched **glycoside-linked** sugars (averaging around fifteen sugar residues). As **glycans**, they are the sugars linked to **glycoproteins** and glycolipids.

Glycoproteins are rare in the cytosol but common on secreted and membrane proteins. Oligosaccharides are typically linked to proteins via the hydroxyl group on serine or threonine amino acids (*O-glycosylation*) and are occasionally to modified amino acids (e.g., hydroxylysine, hydroxyproline) or to the *amide nitrogen* on asparagine (*N-glycosylation*). As a major feature of the **glycocalyx**, oligosaccharide domains of glycoproteins and glycolipids are the “face” of a plasma membrane that communicates with the extracellular world. Figure 16.18 illustrates the glycocalyx.

The **glycocalyx**: the covalently sugar-coated extracellular surface of the plasma membrane



Glycocalyx proteins and sugars articulate with components of the extracellular matrix (ECM) and participate directly or indirectly in cell-cell interactions.

Fig. 16.18: The glycocalyx is the sugar-rich region on the extracellular surface of cells, formed by covalently bound sugars on glycoproteins and glycolipids. It is the basis of many cell functions and is associated with other macromolecules to form an extracellular matrix.

Oligosaccharides begin their synthesis in the rough endoplasmic reticulum (*RER*), with the creation of a **core glycoside**. These partial **glycans** are enzymatically linked to compatible amino acids of a membrane protein. As these proteins travel through the *Golgi vesicles* of the *endomembrane system*, **terminal glycosylation** attaches more sugars to the core glycoside to complete glycoprotein synthesis. When vesicles budding from the trans-Golgi vesicles fuse with the plasma membrane, the sugars on the glycoproteins end up on the exterior cell surface. This is illustrated in the following link.



[291 The Path to Sugar-Coated Cells](#)



Like glycoproteins, glycolipids are only found on the extracellular surface. Glycolipids are synthesized in much the same way as glycoproteins. Specific enzymes catalyze initial

glycosylation of phospholipids and polypeptides, followed by terminal glycosylation reactions. Glycoproteins, glycolipids, and proteoglycans on cell surfaces play critical roles in cell-cell recognition and the formation of tissues. They recognize and bind to carbohydrate receptors (*lectins*) on adjacent cells, leading to cell-cell attachment as well as intracellular responses in the interacting cells.

Glycoproteins and glycolipids also mediate the interaction of cells with extracellular molecular signals and with chemicals of the *ECM*, or *extracellular matrix* (Figure 16.19). The *ECM* includes components of connective tissue, basement membranes, and, in fact, any surface to which cells attach.

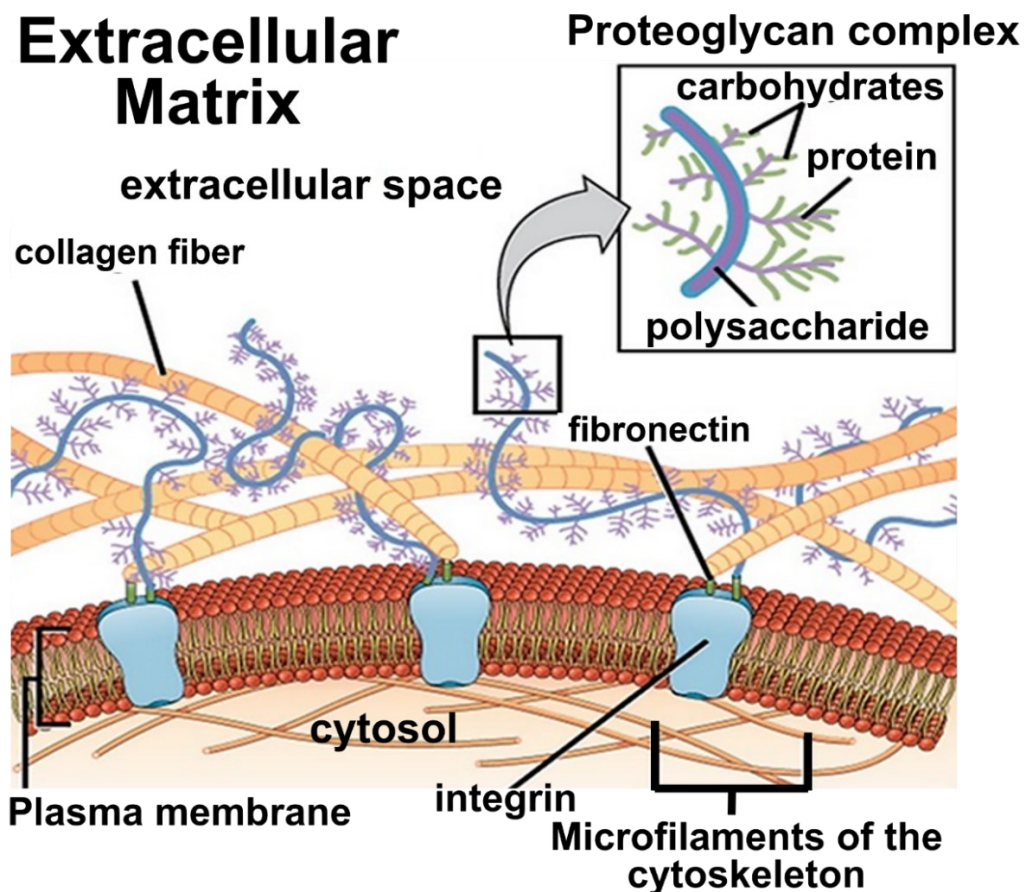


Fig. 16.19: The extracellular matrix (ECM) forms by a noncovalent association of e.g., fibronectin, collagen, proteoglycans, and other macro-molecules with membrane proteins and elements of the glycocalyx.



[292-2 The Extracellular Matrix](#)



16.6 Glycoproteins and Human Health

We'll close this chapter with a few examples of glycoproteins that play crucial roles in human physiology. Let's look first at the major human A, B, AB, O, and Rh blood groups that result from the presence or absence of glycoprotein *antigens* embedded in red blood cell membranes

and the presence or absence in the blood of **antibodies** against the antigens. Typically, an exposure to *antigens* (e.g., foreign substances like bacteria, viruses, and toxins) generates **immunoglobulins**, the *antibody* molecules of our immune system. The *immunoglobulins* are glycoproteins. Since blood-group antibodies already in the blood of a healthy person *are not* a response to foreign antigen invasion, they are something of a paradox!

You probably know that blood groups must be compatible for a successful blood transfusion. A mismatch between donor and recipient can be devastating. An interaction of the red blood cell (erythrocyte) antigens of one blood group with antibodies in another blood group will cause the red cells to clump, restricting blood flow, and ultimately killing the transfusion recipient. Figure 16.20 summarizes why transfusions with mismatched A, B, AB, and O blood groups must be avoided.

The Major Blood Groups

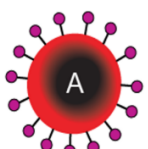
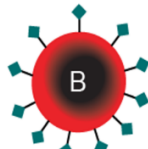
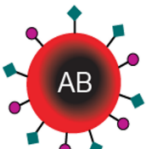



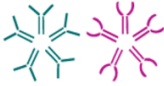
	Group A	Group B	Group AB	Group O
Cell – surface antigens				
Antibodies in the blood	 Anti-B	 Anti-A	None	 Anti-A and Anti-B
Acceptable donors	Group A or Group O donors	Group B or Group O donors	Universal Recipient (Groups AB, A, B or O donors)	Only Group O donors
Why red cells clump in mismatched blood	Anti-A binds & aggregates recipient red cells; Anti B binds & aggregates donor red cells.	Anti-B binds & aggregates recipient red cells; Anti A binds & aggregates donor red cells.	Recipients have no antibodies to attack donor red cells – neither recipient nor donor cells clump	Antibodies in Group O blood will bind any donor red cell antigens and cause the cells to clump

Fig. 16.20: Blood groups with membrane proteins and elements of the glycocalyx.

From the table above, you can see why type O blood is called the *universal donor* and is thus in high demand by clinics and hospitals.

Now, thanks to enzymes produced by members of our *gut microbiomes*, the current absolute requirement, that blood groups must be matched for safe transfusions, may someday be a thing of the past! These enzymes include glycosidases that catalyze glycoside bond hydrolysis. Thus, glucosidases can digest the sugars on A and B antigens on the surfaces of A, B, and AB blood cells. Removal of these antigenic sugars effectively converts these red blood cells to type O cells! Check the Red Cross website (^{16.1}[About A, B, AB, & O Blood Types-Red Cross](#)) or Wikipedia (^{16.2}[Blood type - Wikipedia](#)) for more details about blood groups. See ^{16.3}[Conversion of A, B & AB to O](#) for a summary of recent research on enzymatically converting all blood types to type O.

Another red blood cell antigen is the Rh (rhesus) factor. Rhesus factors are antigens originally found to define blood types in rhesus monkeys, but humans have equivalent antigens, so human Rh blood groups have nothing to do with the monkeys! People either have their own Rh factor and are Rh⁺, or they don't have the factor and are Rh⁻.

When an Rh⁻ recipient is transfused with blood from an Rh⁺ donor, the recipient's immune system makes anti-Rh antibodies in the usual way. This too can cause blood-cell clumping with bad consequences. A well-known example is when an Rh⁻ pregnant woman is carrying an Rh⁺ fetus; if mother and baby's blood should mix, the mother will make antibodies to the Rh factor that could cause fetal hemolytic disease, damaging and destroying fetal red blood cells.

A word to the wise: While you should be "typed" in the event you need a transfusion, it's always a good idea to know your own ABO and Rh blood groups!

The last example here involves cell surface antigens of the **MHC** (*Major Histocompatibility Complex*) glycoproteins that distinguish *self* from *nonself* in body tissues and organs. Transplantation of liver, kidneys, heart, and other major organs, from donors into patients with failing organs has become if not routine, then at least increasingly common. Before a transplant, *MHC tissue typing* determines donor and recipient compatibility to reduce the chances of the rejection of the transplanted organ. Since available donors are few and good matches even fewer, patients wait on prioritized lists for a matched organ. Even when *MHC* typing is a match, the transplant recipient's immune system must be suppressed with hormones to further reduce chances of rejection. Unlike the limited number of blood groups, there are many MHC proteins that must be analyzed to determine a match. Thus, it is not practical (or routinely necessary) to "know" your MHC type! In the next chapter, we look at membrane functions intrinsic to cellular existence itself.

16.7 A Brief Summary of Membrane Surface Functions

As the ultimate barrier and interface between the cell and its environment, the plasma membrane has a number of easily defined functions. In anticipation of the next chapter, here are some of them:

- Cell-cell recognition during development to form tissues and then organs.
- Cell-cell recognition leading to cell junctions during tissue repair following wound healing.
- Communication by selective ion channels and energy-dependent ion pumps.
- Facilitated diffusion of molecules (e.g., glucose) that can't diffuse into or out of cells.
- Recognition of signaling molecules in the environment or even on other cells leading to the transduction of chemical information from outside to inside the cell.
- Participation in the immune recognition of humoral (blood- or lymph-borne) or cellular components of the immune system.

Some iText & VOP Key Words and Terms

amphipathic molecules	glycocalyx	N-glycosylation
asparagine	glycolipids	O-glycosylation
cell membrane	glycosylation	Peripheral membrane

		proteins
Cell-cell attachment	Golgi vesicles	phospholipid bilayer
cytoskeleton	hydropathy plot	plasma membrane
Davson-Danielli membrane model	hydrophilic phosphate heads	poikilothermic organisms
endomembrane system	hydrophobic fatty acid tails	rough endoplasmic reticulum (RER)
exocytosis	hydrophobicity plot	saturated fatty acids
extracellular matrix (ECM)	hydroxyproline	serine
fluid mosaic	hydroxylysine	temperature effects on membranes
Freeze-fracture method	integral-membrane proteins	threonine
membrane evolution	membrane asymmetry	transmembrane proteins
glycan	membrane proteins	unsaturated fatty acids

CHAPTER 16 WEB LINKS



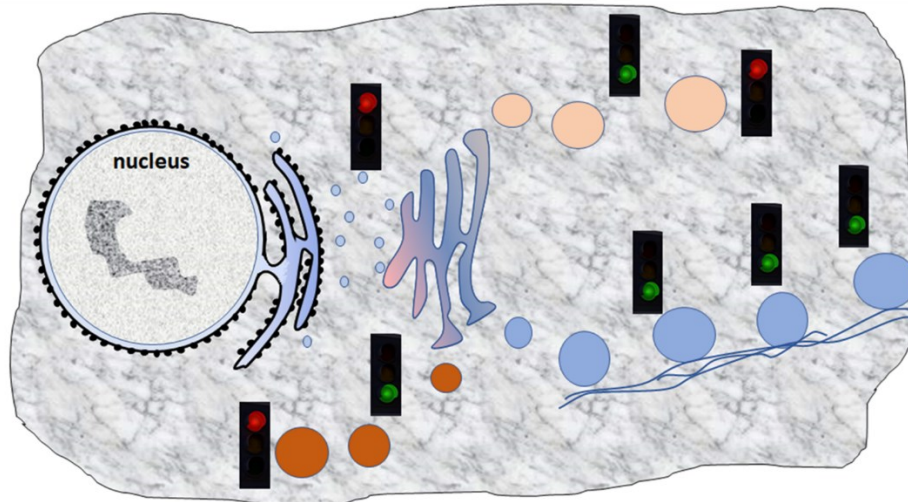
Chapter 17

Membrane Function

Passive diffusion, facilitated diffusion, and active transport; the traffic of proteins in cells; cell-cell interactions, excitability, and signal transduction

Reminder: For inactive links, google key words/terms for alternative resources.

**GPS? maybe not,
but cells have a Cell Positioning
System (call it a CPS)...**



**telling proteins where to go
(nicely of course).**

17.1 Introduction

Small molecules like O_2 or CO_2 can cross cellular membranes unassisted, since neither the hydrophilic surfaces nor the hydrophobic interior of the phospholipid bilayer are barriers to their transit. On the other hand, most molecules (even water) need the help of *membrane-transport proteins* to get into, or out of cells and organelles.

Transport proteins can act as *gates*, which may be open or closed. When open, these gates permit the diffusion of molecules into or out of cells along a concentration gradient, so that their concentrations equalize across the membrane. Like the ***passive diffusion*** of small gases, ***facilitated diffusion*** by membrane proteins does not require an input of energy.

Unlike *passive* or *facilitated* diffusion, ***active transport*** uses chemical energy to transport substances across a membrane *against* a concentration gradient. Membrane transport proteins that do this are in fact, *pumps*. We already saw one in the cristal membrane of the mitochondrion. This pump uses the energy of electron transport to actively push protons

across the cristal membrane to create a concentration gradient; the free energy in *this* gradient is coupled to *ATP synthesis* by oxidative phosphorylation. Other membrane pumps use chemical energy (typically from *ATP hydrolysis*) to move ions or molecules across membranes and to create or to maintain chemical gradients. For example, pumps that create sodium- and potassium-***ion gradients*** are responsible for the *excitability of cells*. Recall that this is a fundamental property of life: the ability of cells and organisms to respond to stimuli. As you read this chapter, look for how allosteric changes regulate membrane function. We'll consider the following:

- How membrane ***gates*** and ***pumps*** work
- How membrane protein interactions allow cells to ***self-assemble*** into tissues and organs
- How cells direct protein ***traffic*** (e.g., to the cytoplasm, into membranes, into organelles, or out of the cell)
- How membrane proteins participate in ***direct communication*** between adjacent cells
- How membrane proteins are receptors for ***long-distance communications***, responding to neurotransmitters, hormones, and other external chemical signals

Learning Objectives

When you have mastered the information in this chapter, you should be able to do the following:

1. Explain how and why one cell type's plasma membrane differs from that of another cell type.
2. Explain how and why the plasma membrane differs from other membranes in the same cell.
3. Predict and then determine if solutes will cross a plasma membrane by passive or facilitated diffusion.
4. Explain how *salmon* can spend part of their lives in the ocean and part swimming upstream in freshwater to spawn, without their cells shriveling or bursting.
5. Explain how *active transport* stores *chemical energy*. (Recall electron transport and oxidative phosphorylation.)
6. Explain the role of active transport in maintaining/restoring a cell's *resting potential*.
7. Compare and contrast different kinds of *gated channels*.
8. Describe the order of ion movements that generate an *action potential*.
9. Define and compare *exocytosis*, *pinocytosis*, *phagocytosis*, and *receptor-mediated endocytosis*.
10. Distinguish between *signal molecules* that enter cells to deliver their chemical message and those that deliver their message only as far as the plasma membrane.
11. Trace an intracellular response to a *steroid hormone*, from its beginning to its likely *cellular effect*.
12. Trace a liver-cell response to *adrenalin*, from the plasma membrane to glycogenolysis (*glycogen breakdown*) in the cytoplasm of the cell.
13. Compare the *signal-transduction* activities of different *G-protein receptors* leading to the first active kinase enzyme in a *phosphorylation cascade*.
14. Explain how a liver cell can respond the same way to two different hormones (e.g., adrenalin and glucagon) and why this should be possible.
15. Describe/explain how a phosphorylation cascade *amplifies* the cellular response to a small amount of an *effector* (signal) molecule.

16. Discuss the differences and interactions between the *glycocalyx*, *basement membrane*, and *extracellular matrix* (ECM).
17. Explain *ECM* functions and identify components involved in those functions.
18. Describe how the molecular structure of *fibronectin* supports its different functions.
19. Describe some structural relationships and interactions between cell surfaces and the *cytoskeleton*.
20. Compare and contrast the structures and functions of the different cell junctions.
21. Distinguish between the structures and functions of *cadherins*, *clathrin*, *COPs*, *adaptin*, *selectins*, *SNAREs*, and *CAMs*.
22. State a hypothesis to explain why some cancer cells divide without forming a tumor.

17.2 Membrane Transport

The first control on the passage of molecules across membranes is the semipermeable character of the membrane itself. Molecules move in and out of cells in one of three ways: ***passive diffusion***, ***facilitated diffusion***, and ***active transport***.

Only a few small, relatively uncharged molecules can cross a membrane unassisted (i.e., by *passive diffusion*). Hydrophilic molecules that must enter or leave cells require help (i.e., from *facilitated diffusion*). Passive and facilitated diffusion release the free energy inherent in concentration gradients as molecules diffuse across a membrane.

In contrast, active transport (i.e., membrane pumps) consumes energy to create concentration gradients of specific solutes. The specificity of *facilitated diffusion* and *active transport* lies in the integral membrane proteins that recognize and bind specific solutes for transport. As you may predict, allosteric regulation of these proteins controls the movement of their target molecules into or out of cells.

Despite water's polarity, many believed that the small water molecules crossed membranes without help. Indeed, they do to a limited extent. However, others suspected that given its highly charged *polar covalent* bonds relative to its small size, a water molecule would require an assist to get across membranes efficiently. Let's begin with a closer look at passive diffusion and diffusion by facilitated diffusion, followed by osmosis (a special case of facilitated diffusion of water), and finally active transport.

17.2.1 Passive Diffusion of Solutes

Diffusion across membranes does not require energy. In fact, diffusion to relieve a solute concentration gradient can release energy—recall the movement of protons through the F1 ATP synthase proton gate that synthesizes ATP during mitochondrial oxidative phosphorylation. Passive diffusion in solution is the movement of molecules over time by random motion (also called *Brownian motion*) from regions of higher to regions of lower concentration. Significant passive diffusion across cellular membranes is limited to a few molecules, mostly gases like O₂, CO₂, and N₂, that can freely cross the hydrophobic phospholipid barrier. The rapid diffusion of gases is essential for O₂ and CO₂ exchange between the alveolar capillaries and the cells of the lungs during physiological respiration. O₂ and CO₂ exchange also occurs in mitochondria during cellular respiration.

The rate of diffusion of a molecule is dependent only on its own concentration. It is unaffected by the concentration of other molecules. Over time, the random motion of solutes within and across compartments results in a **dynamic equilibrium** for each different solute. At equilibrium, solute molecules continue to diffuse across the membrane, but for each molecule moving across in one direction, another molecule of the same solute crosses in the other direction.

17.2.2 Facilitated Diffusion of Solutes and Ions

Like the passive diffusion of those gasses, *facilitated diffusion* (e.g., of ATP through an F1 ATP synthase) is the spontaneous (downhill) passage of molecules or ions across membranes, but with the help of specific *transmembrane proteins*. The kinetics of passive and facilitated diffusion reveals the differences between the two processes. To understand the latter, recall that the rate of enzyme catalysis is *saturable*. That is, as the concentration of substrate is increased, the rate of the catalyzed reaction approaches a maximum (V_{max}). This occurs when all enzyme molecules in solution are bound to substrate molecules. The same saturation phenomenon applies to facilitated diffusion—the rate of diffusion of a solute across a membrane is directly proportional to the concentration of the solute but is limited by the number of transport proteins in the membrane. Figure 17.1 plots the kinetics of passive and facilitated diffusion.

Kinetics of passive vs facilitated diffusion

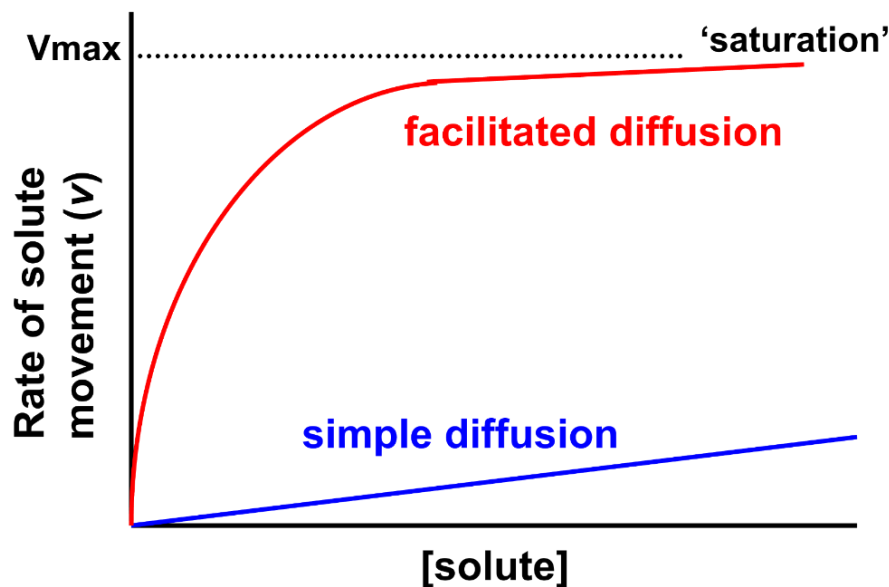


Fig. 17.1: The graphed rates of passive and facilitated diffusion show that facilitated diffusion is a saturable process.

Perhaps you see another similarity between facilitated diffusion and enzyme catalysis in this graph! Relative rates of facilitated diffusion are typically rapid, compared to those of passive diffusion. This is because the allosteric changes that accompany facilitated diffusion are rapid, just as they are during enzyme catalysis. This should suggest yet another similarity between enzymatic catalysis and facilitated transport: both can be regulated!

Figure 17.2 (below) illustrates three kinds of *facilitated transport* of solutes.

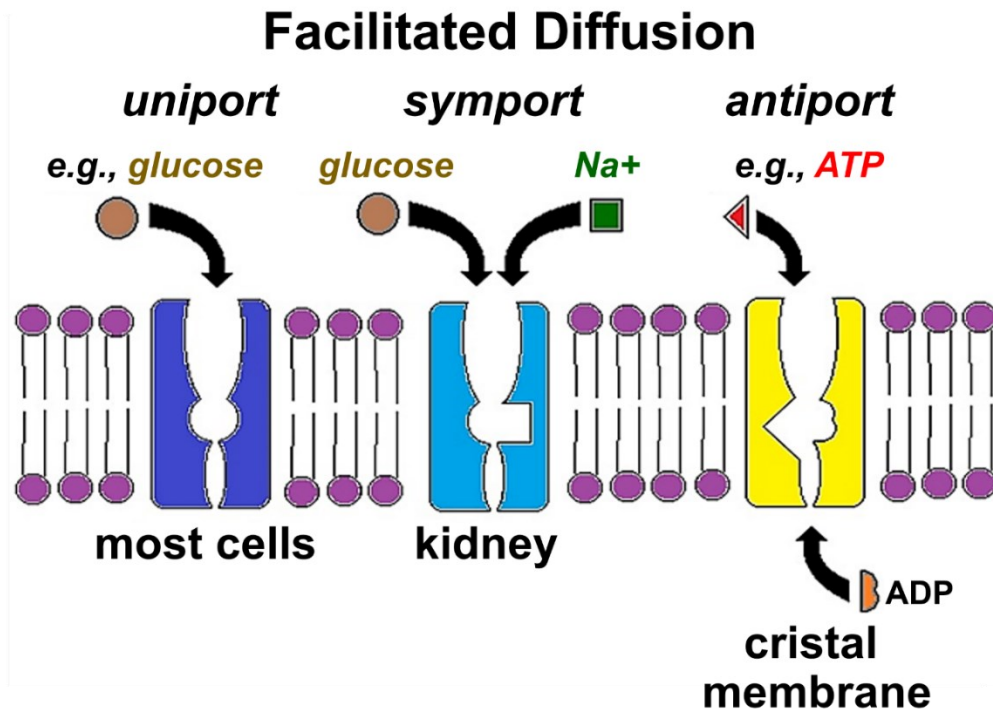


Fig. 17.2: Three kinds of facilitated diffusion: *left*, transport of a single molecule; *middle*, co-transport of two molecules; *right*, exchange of molecules by opposing transport.

The **GLUT** (**GLU**ucose **T**ransporter) protein (Figure 17.2, left) allows glucose **uniport**, the specific transport of a single substance in or out of cells. In **symport** (Figure 17.2, middle) glucose transporters couple the simultaneous movement of glucose and sodium ions (e.g., in kidney cells). The **SGLT** (Sodium-**GLU**ucose Transporter) serves a similar function in small intestine cells, enabling absorption of dietary glucose and sodium. **Antiport** (Figure 17.2, right) allows the specific exchange of molecules across a membrane. In the example shown, ATP leaves the mitochondrial matrix, crossing the cristal membrane at the same time as ADP enters the matrix.

Whether by uniport, symport, or antiport, each solute will independently cross a membrane down its concentration gradient, moving from regions where it is at a higher concentration to regions of lower concentration. Recall that the direction of diffusion depends on relative concentrations of the solutes, and that diffusion along (i.e., down) a gradient releases free energy.

Proteins mediating facilitated diffusion are of two kinds: **carrier proteins** and **channel proteins**. **Carrier proteins** allow diffusion of large solutes across the membrane. While small, ions also need help to cross the hydrophobic membrane barrier because they have a *high charge-to-mass ratio*. This is the job of **channel proteins**, which essentially serve as ion pores.

Like all transporter proteins, both *carrier* and *channel* proteins undergo allosteric change during transport. They are also typically subject to allosteric regulation, rather than being in a constant "open" state. Examples of facilitated diffusion are considered in more detail next.

17.2.2.a Carrier Proteins

When a carrier protein binds a solute that must cross the membrane, it undergoes the first of a series of allosteric changes, as shown in Figure 17.3.

Multiple allosteric changes in transporter proteins accompany facilitated diffusion of a solute. Here a *carrier protein* recognizes, binds, and transports a solute from one to the other side of a membrane with 4 allosteric changes!

e.g., **glucose**

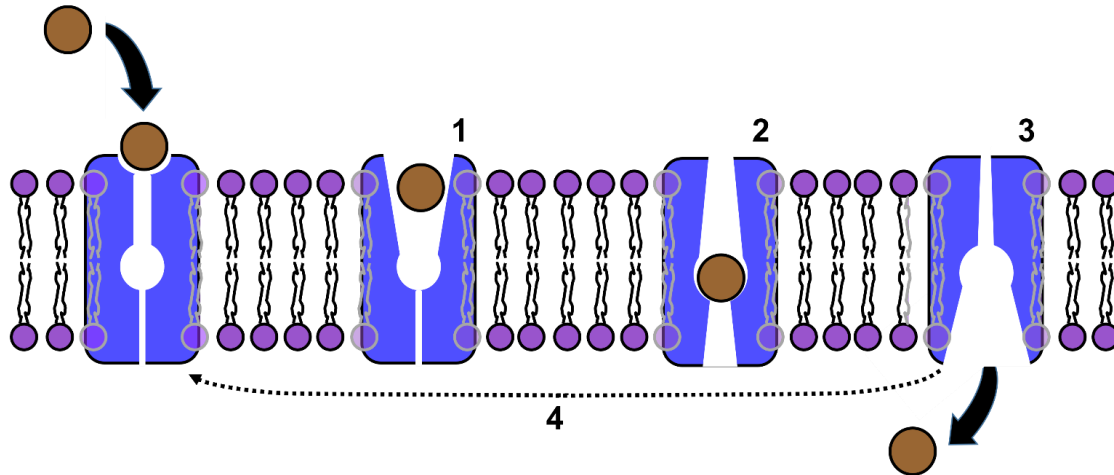


Fig. 17.3: *Transporter proteins (e.g., a glucose transporter) enable facilitated diffusion do the following: (1) bind the solute, (2) undergo allosteric change to initiate transfer, (3) change shape again to release the solute on the other side of the membrane, and (4) finally return to the starting shape after releasing the solute (4).*

During transport itself, the carrier protein undergoes another change in shape. When the solute reaches the other side of the membrane, it no longer has a high affinity for the carrier protein. After release of the solute, a final allosteric change restores the original conformation of the transport protein.

A given carrier protein is specific for a single solute, or at most a single family of closely related solutes. Thus, *GLUT1*, a glucose uniport transporter, allows glucose (but not fructose or ribose!) to cross membranes. Other specific carrier proteins facilitate the transport of amino acids or charged solutes across cell membranes. Once again, molecules that indicate cell status (i.e., a need to import or export solute) are *allosteric effectors* that regulate carrier proteins. Insulin is a perfect example of the regulation of solute transport, specifically glucose transport into cells. One consequence of insulin being released during a meal (or in anticipation of a meal!) is the stimulation of glucose transporters to take up glucose. An inability of those transporters to respond to insulin accounts, in part for type-2 (adult-onset) diabetes.

Water gets across membranes by osmosis; we'll look more closely at how osmosis affects cells in a moment. But first, recall that small amounts of water can cross the phospholipid bilayer unassisted. Water can also cross a membrane incidentally when ions flow

through their channel proteins. But most osmosis involves facilitated diffusion mediated by **aquaporins**. Some aquaporins only transport water. Others have evolved to co-facilitate the transport of glucose (see above), glycerol, urea, ammonia, carbon dioxide, and even ions (protons) along with water.

Like other carrier proteins, aquaporins are allosterically regulated to allow cells to meet specific water balance requirements. Understanding how cells managed their water balance was so fundamental that the discovery of aquaporins earned Peter Agre a Nobel Prize in Chemistry in 2003.

Since Agre's discovery (in 1992), several genetic diseases have been linked to aquaporin gene mutations. Kidney cells are critically involved in vertebrate water balance and have many aquaporins in their membranes. In a rare form of diabetes, abnormal aquaporins cause the kidneys to excrete unusually large volumes of water. In another example, aquaporin gene mutations lead to the development of cataracts in both eyes. Since their initial discovery, aquaporins have also been described in bacteria and plants. To learn more, see ^{17.1} [Aquaporins](#).

17.2.2.b Ion Channels

Allosteric regulation of ion channel proteins controls ion *homeostasis* in blood and extracellular fluids within narrow limits. Often, multiple integral proteins contribute to the formation of an ion channel. When stimulated, channel proteins rearrange to open a pore, allowing specific ion transport. Some ion channels, like the glucose-sodium ion symport system noted earlier, mobilize the energy of diffusion of one solute (the ion in this case) to rapidly transport another solute through the same channel (acting like an ion channel *and* a carrier protein). Finally, ion channels are responsible for the excitability of cells, where Na⁺, K⁺, and Ca⁺⁺ channels collaborate in ion movements into and out of cells leading to neuronal or muscle cell responses (more shortly!).



[293-2 Passive & Facilitated Diffusion \(Transport\)](#)



17.3 Osmosis

Osmosis is the diffusion of water across membranes from low to high solute concentrations, an essential cellular activity. It allows cells to use water to maintain cellular integrity or to adapt to changes in the solute composition in the extracellular environment. Osmosis relies on the transport of water by **aquaporins**. The passive diffusion of water molecules can be demonstrated with artificial (e.g., dialysis) membranes. If solute concentrations are higher on one side of the membrane, **free water** will cross the membrane "trying" to equalize the solute concentrations on both sides of the membrane.

In effect, water movement is away from the side of a membrane where the *free water* concentration is higher (i.e., where the concentration of solute is lower) and goes to the side where the concentration of *free water* is lower (basically, the side where the concentration of solute is higher).

17.3.1 Osmosis in Plant and Animal Cells

We could present this section in the context of free-water concentrations, but we will do so in the more familiar terms of solute concentrations. Osmosis affects plant and animal cells according to the same principles, but with different effects. The effects of different experimental solute concentrations on animal cells are illustrated in Figure 17.4.

Osmosis (Diffusion of Water) Across Membranes in Animal Cells

Relative to the cell, these solutions are:

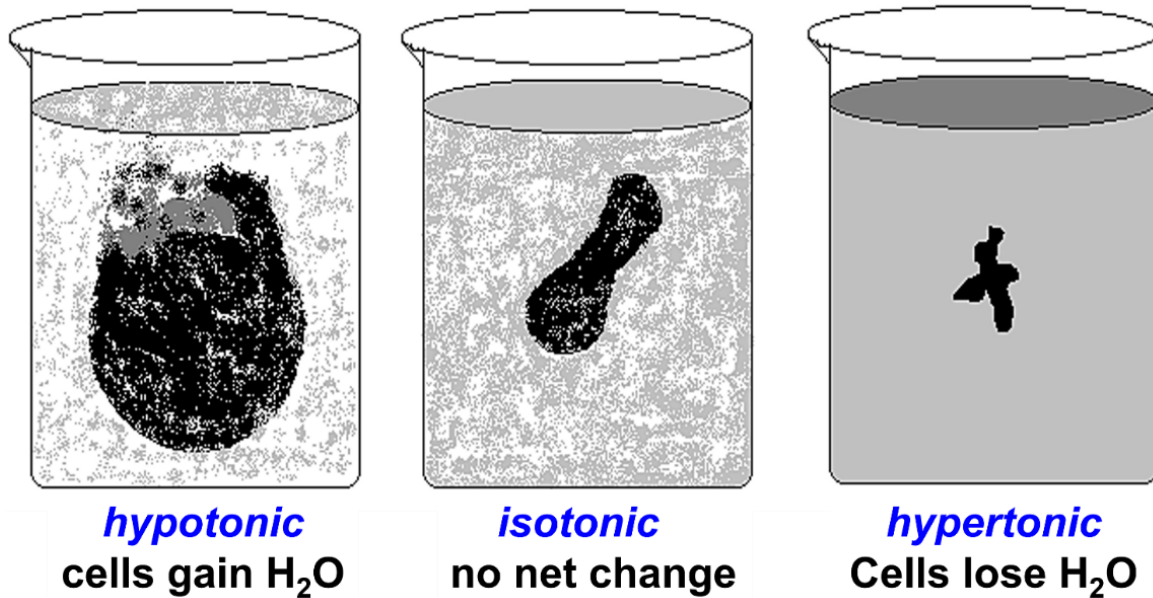


Fig. 17.4: Differences in solute concentrations inside and outside *animal* cells affect the movement of water into or out of the cells (i.e., *osmosis*). Cellular *solute concentrations* are higher than (left), similar to (middle) or lower than (right) than the surrounding solution.

If the solute concentration inside and outside the cell is the same, there is no net water movement into or out of the cells. The extracellular medium and cytosol are said to be **isotonic** to each other.

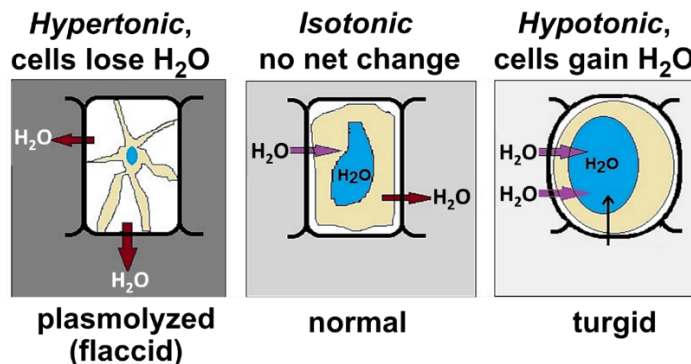
When water diffuses into the cells from a low-solute medium, the medium is said to be **hypotonic** to (less concentrated than) the cytosol. In this case, movement of water into a cell lowers the cytosol-solute concentration. Animal cells swell and burst in a hypotonic solution.

Animal cells in **hypertonic** solutions (with higher solute concentrations than the cytosol) shrivel up as water leaves the cell. From this brief description, you should conclude that *water crosses from the hypotonic to the hypertonic side of a membrane*.

As with animal cells, exposure of plant cells to hypotonic or hypertonic solutions causes the same directional water movements, but with some key differences due to their cell walls. Figure 17.5 (below) shows the effects of different solutions on plant cells.

Osmosis (Water Diffusion) Across Membranes in Plant Cells

Relative to the cell, these solutions are:



**Unlike animal cells, plant cells don't
burst in a hypotonic environment.
Swelling is restrained by their cell walls.**

Fig. 17.5: Differences in solute concentrations in and out of *plant cells* affect water movement into or out of the cells, (i.e., *osmosis*), with different effects than in animals because (a) plants have a rigid cell wall, and (b) plant cell membranes are attached to the cell wall. During transient changes in extracellular solute concentrations, a *tonoplast* accumulates or releases water to achieve correct cytoplasmic solute.

In *hypotonic* solutions, water enters plant cells, moving into the cytosol and then into water vacuoles called *tonoplasts*. This results in higher *osmotic pressure* (water pressure) in the tonoplasts. The expanding *tonoplast* creates *turgor pressure*, compressing the cytosol against the cell wall. Rather than swelling excessively and bursting like animal cells, plant cells and therefore the plant tissues, stiffen and become *turgid*. Since water cannot enter plant cells indefinitely, water stops entering the cells when the *osmotic pressure* outside the cells and the *turgor pressure* inside the cells are at equilibrium. You've would have seen this phenomenon if you ever overwatered houseplants. Turgor pressure stiffens the leaves and stems that then become brittle and are easily snapped and broken.

In hypertonic medium, plant cells (like animal cells) lose water. The resulting shrinkage of the plasma membrane away from the cell walls is called *plasmolysis*, in which bits of plasma membrane remain tightly attached to the plant cell wall at several points, limiting shrinkage (especially if the water loss is transient). You may have seen under-watered plants with floppy or droopy stems and leaves. These have become *flaccid* due to loss the loss of turgor pressure needed to keep leaves and stems upright. Formally, *osmotic* or *turgor pressure* is defined as the force per unit area (i.e., *pressure*) required to prevent the passage of water across a semipermeable membrane from a hypotonic to a hypertonic solution.

17.3.2 Osmosis in Plant Life

While individual plant cells respond to changes in solute concentrations, these changes are rapidly communicated to adjacent cells through *plasmodesmata*. These structures connect the plasma membranes of adjacent cells through their cell walls, allowing rapid, direct sharing of physical and chemical information. A *plasmodesma* is illustrated in Figure 17.6.

Plasmodesmata Connect Cytoplasm of Adjacent Cells

direct cell-to-cell movement of solutes or water through the plasmodesmata

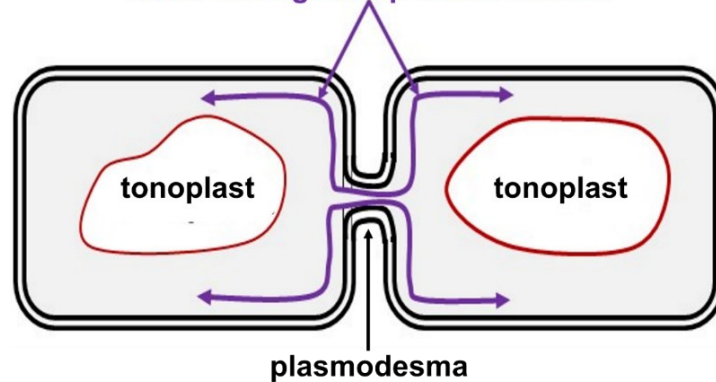


Fig. 17.6: Plasmodesmata are cell wall “tunnels” connecting plant cells and allowing direct movement of water between cells to transmit changes in osmotic pressure in one part of a plant (e.g., roots).

In this way, the effects on osmotic pressure in a few cells—effects created by changes in water availability—are transmitted to adjacent cells, affecting turgor pressure in those cells and, ultimately, in plant tissues. Finally, plant life depends on water! Recall that plant cells require a continual supply of water for use in photosynthesis, to provide hydrogen to reduce CO_2 to glucose. Photosynthesis, as well as the loss of excess water from plant tissues (especially leaves) by **transpiration**, lowers cellular osmotic pressure. As water moves up from the roots to replace water used and lost by leaf cells, the osmotic pressure drops in the *fine root-hair cells* (with their high surface area). This draws water into the root cells by osmosis. Thus, osmotic pressure is the main force driving water into plants and, defying gravity, moving it up from the roots to the rest of the plant.

17.3.3 Osmosis in Animal Life

Changes in osmotic environment can stress or kill an organism. For example, freshwater organisms (protozoa or fish) placed in sea water will die. Likewise, salt-water fish placed in freshwater. But cells and organisms can **osmoregulate** (i.e., control the osmotic pressure in their cells), at least to a point. For example, as a *paramecium* takes on water, a **contractile vacuole** expels excess water to prevent the cell from bursting (Figure 17.7, below).

Paramecium

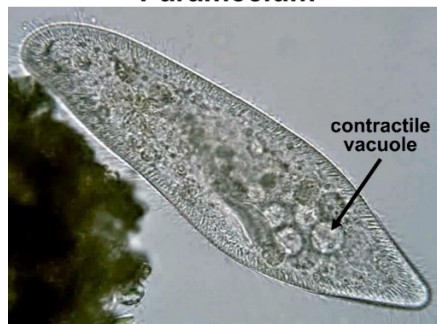


Fig. 17.7: The light micrograph of a paramecium highlights a contractile vacuole, an organelle that pumps out excess water from the cell, protecting it from osmotic shock.

Water constantly enters these freshwater protists because the solute concentration in the cytosol is always higher than the fresh water they live in. To cope with a constant uptake of water, their *contractile vacuoles* collect excess water and then contract to expel the water. At a high energy cost, *paramecia* constantly pump water out of the cell to maintain water balance (i.e., correct osmotic pressure). Another protist strategy to cope with changes in environmental solute concentrations (e.g., salinity) is to pump salts (or suitable salt-solute substitutes into or out of the cell, as needed. (For some details, see ^{17,2}[Protist Osmoregulation Genes from Bacteria?](#).)

Larger organisms, like freshwater fish, cope with their hypotonic environment by urinating a lot! At the other end of the spectrum, salt-water fish cope with the high solute concentration of solutes (salts) in their environment by excreting excess salt. Salmon spend time in seawater growing to maturity and later swim upstream in fresh water to spawn. You can imagine how salmon and similar organisms have to **osmoregulate** to adapt to their changing, very different environments. In this case, **osmoregulation** begins when hormonal changes respond to changes in living circumstance and dictate a compensatory response.

17.3.4 Summing Up Osmosis

Osmosis is the movement of water across membranes to the side where solutes are at a higher concentration. At the same time, solutes diffuse across membranes, moving into or out of cells toward the side where they are at lower concentration, either passively or with the help of transporter proteins. We've evolved different facilitated diffusion proteins specific for the cross-membrane transport of different solutes. And finally, most water crosses membranes by facilitated diffusion through aquaporin proteins that serve as pores in cellular membranes.



[294-2 Osmosis](#)



17.4 Active Transport

Excitability (adaptation) is another of the defining properties of life. This property of all cells is based on chemical and electrical reactivity. Neurotransmitters released at a synapse cross the synaptic cleft from a "sending" neuron to a responding cell (another neuron or a muscle cell). The neurotransmitter binds to receptors on the responding cell resulting in a **membrane depolarization**, a rapid change in electrical potential difference across the cell membrane. While responses to neurotransmitters occur in fractions of a second, all kinds of cells are responsive, albeit not always as fast as neurons or muscle cells. Changes in the membrane polarity of any cell depend on *unequal* concentrations of ions inside and outside cells. These ionic differences across membranes enable all cells, but especially neurons and muscle cells, to respond to chemical and other (e.g., electrical) signals. Thus, cells have a **resting potential** due to a higher $[K^+]$ in the cytosol and higher $[Cl^-]$ and $[Na^+]$ outside the cell. The measured resting potential (difference in charge or **potential difference**) of most cells is typically between -50 mv to -70 mv, negative because negative (i.e., Cl^-) ions are at higher concentration inside the cell than outside. Such ion concentration gradients permit

physiological response to chemical or other signals. They change when cells are excited when they (quite normally) leak ions. Whether incidental or intentional, the correct ion balance must be restored to maintain excitability. Relative concentrations of K^+ , Cl^- , and Na^+ ions accounting for a cell's resting potential are shown in Figure 17.8.

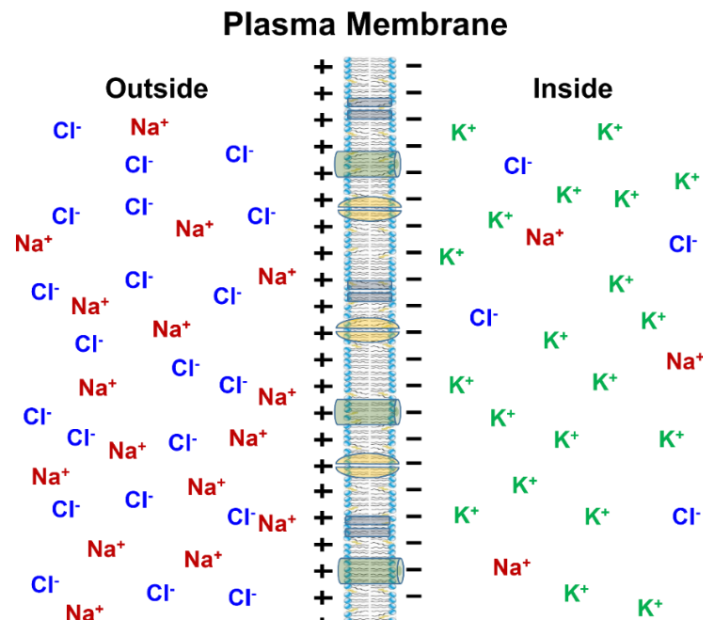


Fig. 17.8: The *resting potential* of a cell results from an ion-concentration imbalance across the plasma membrane. Sodium and chloride ion concentrations are normally higher outside cells while potassium ion levels are higher inside the cell. Thus, the cytoplasm is slightly negative compared to the extracellular fluid.

Maintaining a cell's resting potential requires the energy of an ATP-dependent Na^+/K^+ pump, an *active-transport* protein complex. Follow the action below in Figure 17.9.

Allosteric Changes (blue asterisks) During Operation of a Na^+/K^+ Pump

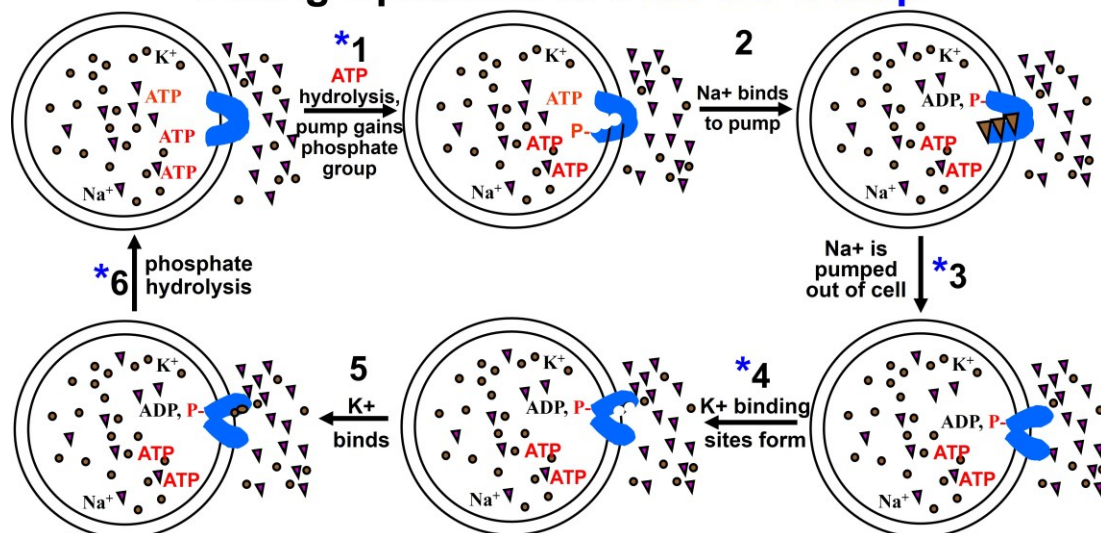


Fig. 17.9: After a change in the resting potential (e.g., depolarization) of a cell, Na^+/K^+ ion balance across the cell membrane is restored by an ATP-powered sodium/potassium pump. Three Na^+ ions are pumped out of the cell for every two K^+ ions that enter the cell.

In operation, the ATPase domain of the Na^+/K^+ pump protein hydrolyzes ATP, leaving a phosphate attached to the pump and inducing the first of several allosteric changes in the pump proteins (*step 1* in the illustration). In its new conformation, the pump binds three Na^+ ions, causing a second conformational change that in turn releases the Na^+ ions into the extracellular fluid (*step 2*). The release of Na^+ ions outside the cell causes a third allosteric change (*step 3*), after which two K^+ ions from the extracellular fluid can bind to the pump protein. K^+ binding causes the hydrolysis of the phosphate from the pump protein, returning it to its original conformation (*step 4*) and releasing the two K^+ ions into the cytosol. The Na^+/K^+ pump is ready for action again!



[295 Potassium Leakage Helps Maintain Cellular Resting Potentials](#)



[296 Active Transport by the Sodium/Potassium Pump](#)

For his discovery of the ATPase-powered sodium/potassium pump and for his studies of how it works to maintain intracellular ion balance, Jens C. Skou earned a share of the Nobel Prize in Chemistry in 1997. Read more at ^{17.3}[1997 Chemistry Nobel prize-Skou](#).

17.5 Ligand- and Voltage-Gated Channels in Neurotransmission

Ligand- and voltage-gated channels play a major role in neurotransmission and muscle contraction by regulating a flow of ions into and out of responding cells. With the advent of the *patch-clamp* device, it became possible to correlate ion flow with measurements of membrane potential and to define the sequence of electrical and chemical events leading to muscle contraction (Figure 17.10).

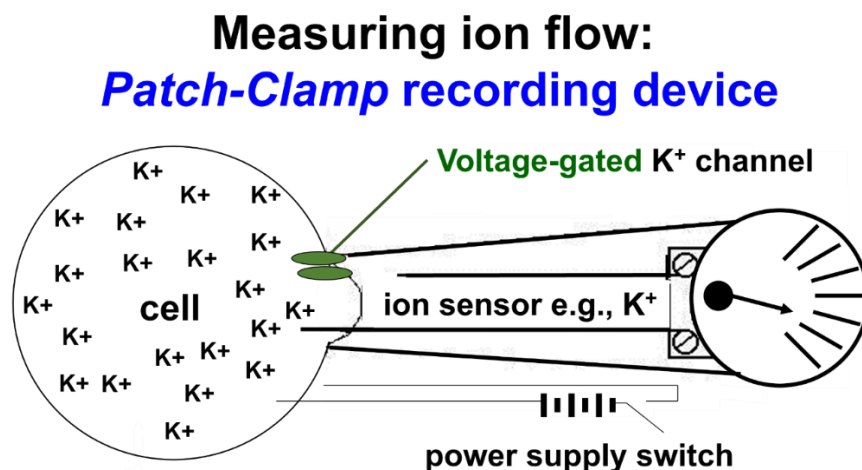


Fig. 17.10: A patch-clamp device can measure ion flow through voltage-gated channels in a membrane during a depolarization event. After a voltage is applied across the plasma membrane, specific ion sensors measure ion flow into or out of the cell.

17.5.1 Measuring Ion Flow and Membrane Potential with a Patch-Clamp Device

When neurotransmitters bind to their receptors, *ion channels* in responding neuron or muscle cells open. The resulting influx of Na^+ ions disrupts the *resting potential* of the target cell. The effect is transient if the membrane potential remains negative. But if enough Na^+ ions enter the cell, the membrane becomes depolarized. If the cell experiences *hyperpolarization*, a localized *reversal of normal membrane polarity* (say from -70 mV to $+65$ mV or more) will generate an *action potential*. This action potential will travel like a current along the neural or muscle cell membrane, eventually triggering a physiological response (e.g., the excitation of the next nerve cell in a neuronal pathway or the contraction of the muscle cell).

In this example of patch-clamp measurements, closing the power supply switch sends an electrical charge to the cell, opening a *voltage-gated ion channel*. A potassium sensor in the device then detects a flow of K^+ ions through the channel out of the cell; at the same time, a voltmeter registers the resulting change in membrane potential. In addition to voltage-gated ion channels, the patch-clamp device can measure ion flow through *ligand-gated ion channels* and *mechanically gated ion channels*. The former channels are *receptor-ion gates* that open when they bind an effector molecule.

Mechanically gated ion channels detect physical *pressure* or *stress* that cause a local membrane deformation, and then open the channel (recall *piezoreceptors* discussed earlier).



[297 Patch-Clamp Device Records Membrane Potential and Ion Flow](#)



[298 Patch-Clamp Measures Resting Potential and Depolarization](#)



[299 Gated Ion Channels](#)



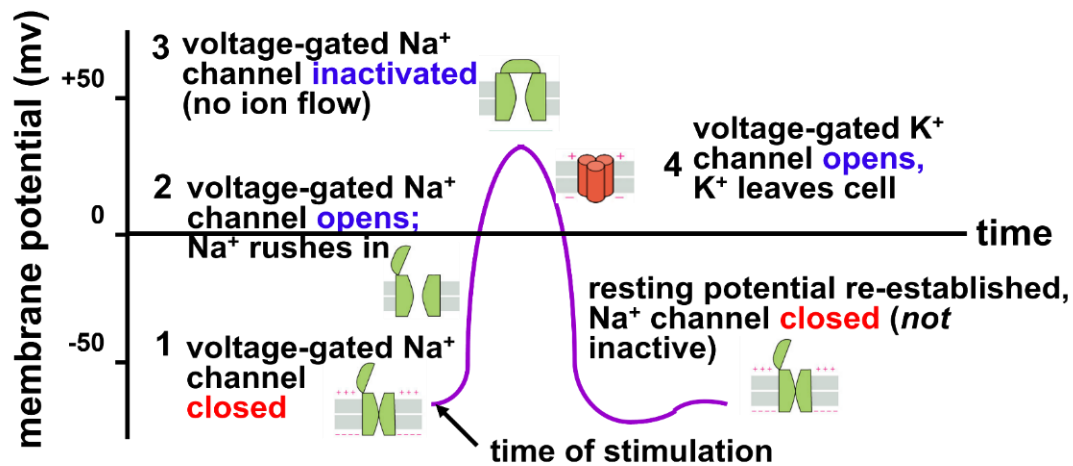
[300 Two Types of Gated Ion Channels](#)



Finally, cells maintain a high intracellular concentration of K^+ ions, causing K^+ ions to slowly leak from the cell, a phenomenon detectable by a patch-clamp. Cl^- ions, as well as other, organic, anions inside a cell, limit the leakage, helping to create the relatively electronegative interior of a cell that is its resting potential. The patch-clamp technique has been used to correlate the flow of ions and changes in membrane potential when a neuron fires, causing an action potential in a responding cell (e.g., Figure 17.11). In the illustration, follow the opening and closing of ion channels and the resulting flow of ions along their concentration gradients. A shift from resting potential and possibly an action potential will result from facilitated diffusion of specific ions into or out of the cell through *gated ion channels* that must open and close in sequence.

Patch Clamp Experiments Reveal Ion Flow in an *Action Potential*

Upon stimulation, an influx of Na^+ ions that *depolarizes* the membrane, followed by K^+ efflux to re-establish *resting potential*.



The Na^+ channel protein can exist in 3 conformational states.

Fig. 17.11: With the patch-clamp device, follow changes in ion flow through gated channels and resulting changes in potential (voltage) across e.g., a muscle cell membrane during an action potential.

The behavior of two different *voltage-gated ion channels* is illustrated in the graph. Electrical stimulation opens Na^+ channels, allowing Na^+ ions to rush into the cell. This reduces the membrane potential from the resting state to zero. If the Na^+ influx continues, it can make the cytoplasm more positive than the extracellular fluid, which may lead to an *action potential*. If the reversal in polarity is high enough, a voltage-gated K^+ opens, and potassium ions rush out of the cell, restoring the *resting potential* of the cell.

A cell can continue to respond to stimuli with action potentials for as long as there is sufficient Na^+ outside the cell and K^+ inside the cell. While active transport of Na^+ and K^+ is not required to reestablish the resting potential, it will eventually be necessary to restore the balance of the two cations in the cell. If a nerve or muscle cell fires several times (or even if it just leaks ions), the $[\text{K}^+]$ inside the cell and the $[\text{Na}^+]$ outside the cell can drop to a point where the cell can't generate an action potential when stimulated. The role of ATP-dependent Na^+/K^+ pump is, ultimately, to restore the Na^+/K^+ balance across the responding cell membrane. As we have seen, each cycle of pumping exchanges three Na^+ ions from the intracellular space for two K^+ ions from the extracellular space. Operation of this ion pump has two effects:

- It restores Na^+ concentrations in the extracellular space relative to the cytoplasm.
- It restores K^+ concentrations in the cytoplasm relative to the extracellular space.



[301 Gated Ion Channels Open & Close in Order during an Action Potential](#)



Together with the higher negative-ion concentrations in the cytosol, the unequal exchange of Na^+ for K^+ ions maintains the resting potential of the cell over the long term and ensures that nerve and muscle cells remain excitable. Next, we will take a closer look at the role of both *ligand-gated* and *voltage-gated* ion channels in neurotransmission.

17.5.2 Ion Channels in Neurotransmission

Action potentials result in an orderly, sequential opening and closing of *voltage-* and *ligand-gated* channels along the neuronal axon. In the following link, you can see the sequential cycles of voltage-gated channels that propagate a localized ***action potential*** (membrane depolarization) along an axon toward a synapse.



[302 Propagating an Action Potential along an Axon](#)

When a propagated depolarization reaches a synapse, gated ion channels either open or close in the neuron and the responding cell. The cooperation of voltage- and ligand-gated channels at a neuromuscular junction is illustrated in Figure 17.12 (below).

Ligand (l)- and voltage (v)-gated channels cooperate in neurotransmission:

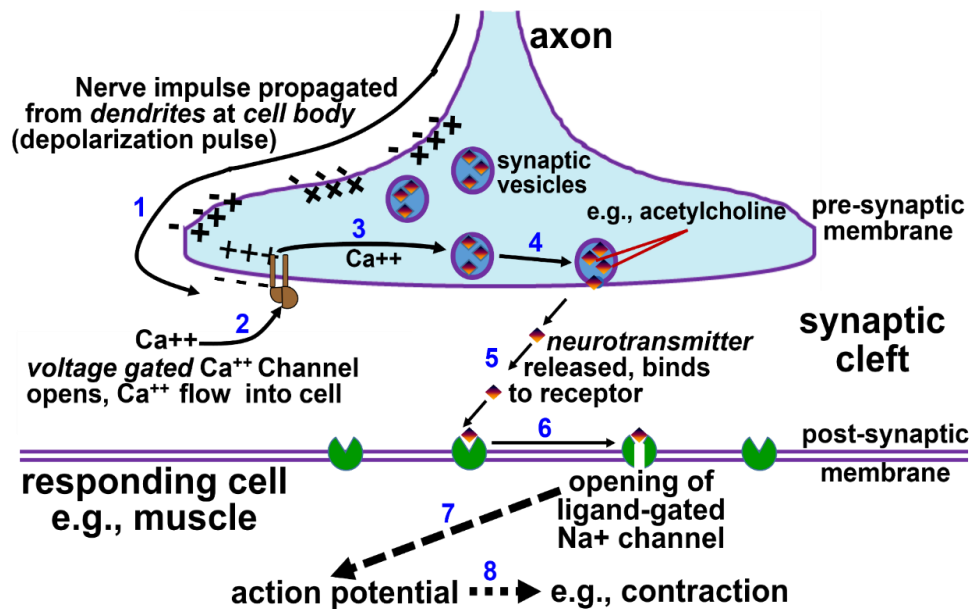


Fig. 17.12: Neurotransmission starts with an *action potential* in the cell body of a neuron. The action potential is propagated along the axon to the nerve terminal *synapse*. The *depolarization pulse* initiates Ca^{++} flow into the neuron through *voltage-gated channels*. This causes a release of neurotransmitters into the *synaptic cleft*. The neurotransmitter binds to a *ligand-gated channel* on the responding cell, causing Na^+ ions to flow into the responding cell and leading to an action potential.

As you can see from the illustration, when a neuron fires, a region of *hyperpolarization* (an electrical impulse) travels down the axon to the nerve ending. There, the traveling charge difference across the cell membrane (electrical potential) stimulates a Ca^{++} -specific *voltage-gated channel* to open. Ca^{++} ions, at higher concentration in the *synaptic cleft* than in the

cytoplasm, flow into the neuron. The Ca^{++} ions cause synaptic vesicles to fuse with the pre-synaptic membrane at the nerve ending, causing the release of neurotransmitters into the synaptic cleft. *Acetylcholine*, the neurotransmitter in this example, crosses the synaptic cleft and binds to receptors on the responding (e.g., muscle) cell plasma membrane. This receptor is a **ligand-gated channel** (also called a *chemically gated channel*). Binding of the acetylcholine ligand opens the gated channel. A rapid diffusion of Na^+ ions into the muscle cell creates an action potential leading to the response, in this case, muscle contraction. We have already seen the role of K^+ channels in restoring the membrane resting potential after an action potential, and we have also seen the role of the sodium/potassium pump in restoring the cellular Na^+/K^+ balance.



[303-2 The Role of Gated Ion Channels at a Neuromuscular Junction](#)

17.6 Endocytosis and Exocytosis

Endocytosis internalizes extracellular molecules (e.g., proteins), insoluble particles, or even microorganisms. The main pathways of endocytosis are *phagocytosis*, *pinocytosis*, and *receptor-mediated endocytosis*. Pinocytosis of molecules is nonspecific. Phagocytosis is also nonspecific, internalizing large structures (e.g., bacteria and food particles). In contrast, receptor-mediated endocytosis is specific for substances recognized by a cell-surface receptor. Some endocytotic processes are even used to recycle membrane components. *Exocytosis* is the secretion of large molecules like digestive enzymes and peptide or polypeptide hormones, each of which must exit the cell to the extracellular fluid or circulation. Exocytotic pathways also deliver membrane proteins to the cell surface, either new or to replace older, worn-out proteins.

17.6.1 Endocytosis

The three main kinds of endocytosis are summarized in Figure 17.13.

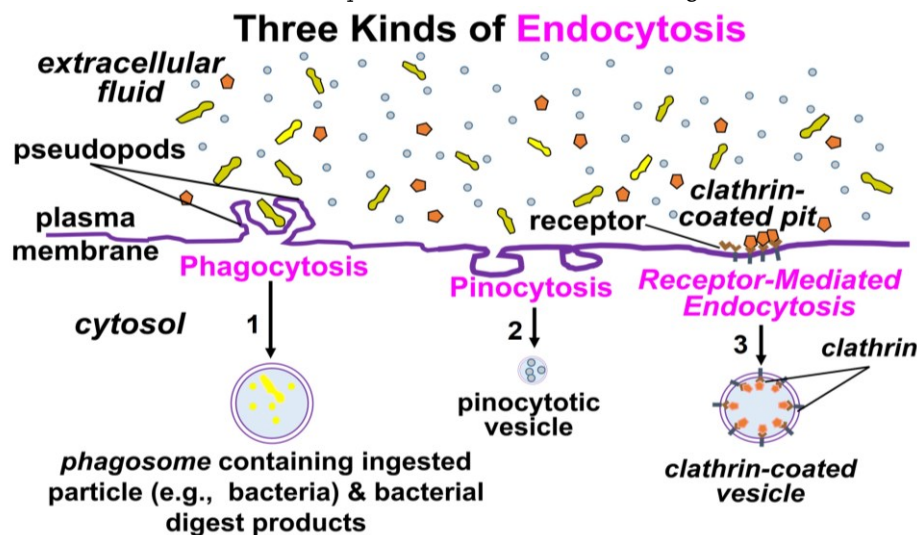
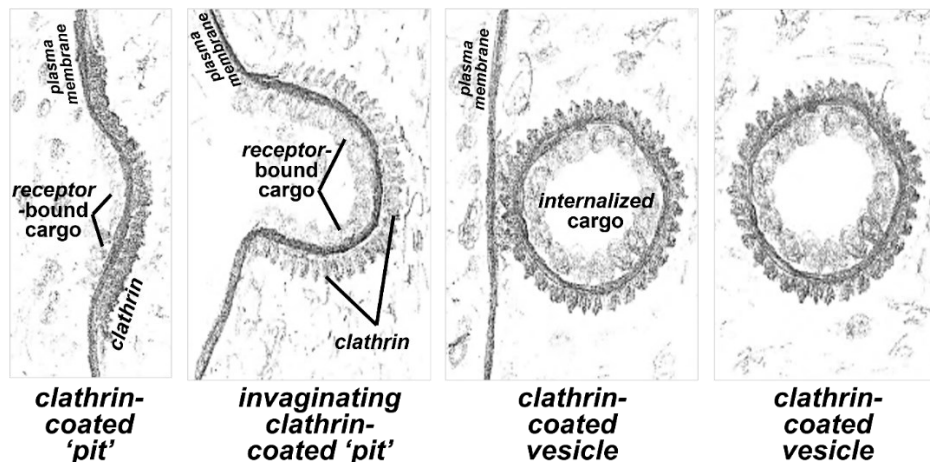


Fig. 17.13: Mechanisms of *endocytosis* import of extracellular materials into cells with different specificities.

From left to right in Figure 17.13:

1. **Phagocytosis** (above left): *phagocytes* extend *pseudopodia* by *evagination* of the cell membrane. The pseudopodia of amoeba (and amoeboid cells generally) engulf particles of food, which end up in digestive vesicles (*phagosomes*) inside the cytosol. Phagocytes are a class of white blood cells that are part of our immune system. They engulf foreign particles that must be eliminated from the body. A *lysosome* fuses with the phagosome, after which inactive hydrolytic enzymes stored in the lysosomes are activated. The result is the digestion of the engulfed particles. Phagocytosis begins upon contact between the outer cell surface and those particles.
2. **Pinocytosis** (above center): pinocytosis is a nonspecific, more-or-less constant pinching off of small vesicles that engulf extracellular fluid containing solutes; they are too small to include significant particulates.
3. **Receptor-mediated endocytosis** (above right): this kind of endocytosis relies on the affinity of *receptors* for specific extracellular substances. Upon binding their ligands, the membrane receptors aggregate in differentiated regions of plasma membrane called *coated pits*. The coated pits then *invaginate* and pinch off, forming a *coated vesicle*, bringing their extracellular contents into the cell. After the coated vesicles deliver their contents to their cellular destinations, their membranes are recycled to the plasma membrane. Receptor-mediated endocytosis is the best understood mechanism for bringing larger substances into cells. The drawings in Figure 7.14 are taken from electron micrographs that illustrate the *invagination* of coated pits to form *clathrin-coated vesicles*.

The formation of Clathrin-Coated Vesicles...



begins with receptor recognition of specific cargo in clathrin coated pits (far left).

Fig. 17.14: Stages of *receptor-mediated endocytosis*. Receptors at coated pits bind substances to be imported into a cell, triggering the invagination of the pit to form a cytoplasmic vesicle with imported contents (shown from left to right). Invagination is mediated by several proteins, including clathrin.

Receptor and coat proteins are seen as larger structures on the inner surfaces of the pits and on outer surfaces of clathrin-coated vesicles. *Clathrin*, the principal protein on surfaces of invaginated, coated pits, is linked to specific integral membrane proteins by **AP1** (*adaptor protein 1*, one of several *adaptins*). AP1 recruits specific *cargo proteins* to bring into the cell as coated pits invaginate. Some details of receptor-mediated endocytosis are illustrated below in Figure 17.15.

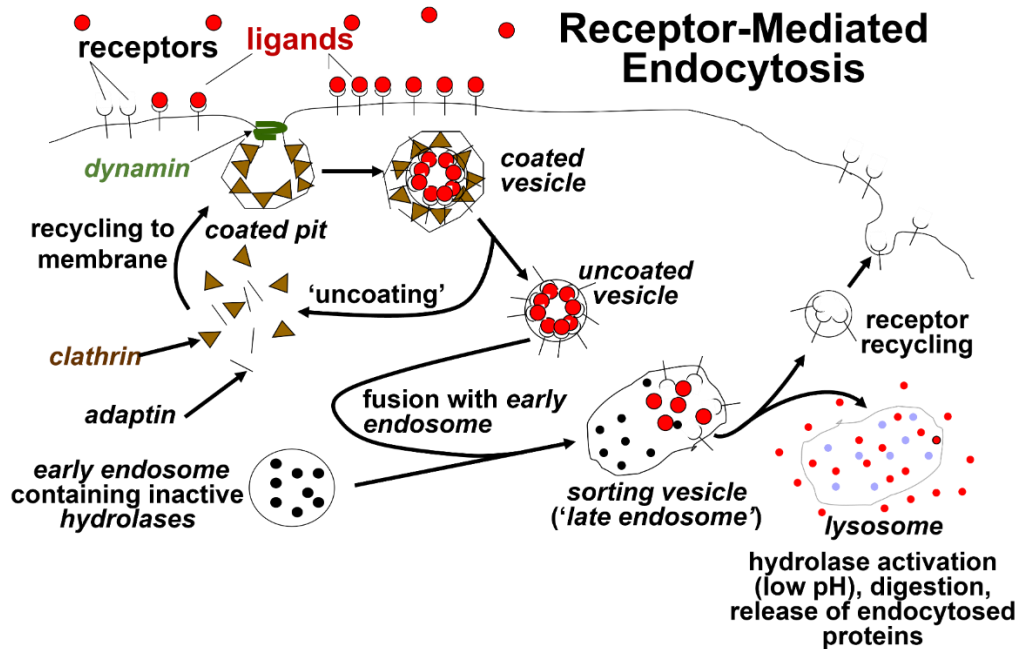


Fig. 17.15: Follow *ligand* import into a cell by *receptor-mediated endocytosis*, from receptor binding at, and invagination of a coated pit, to fusion with an *early endosome*, formation of a *sorting vesicle* (*late endosome*), finally becoming a *lysosome* that digests and release imported material to the cytoplasm.

In the illustration, substances to be internalized bind to membrane receptors that then cluster to form a **coated pit**. Assisted by **dynamin** (a GTPase), the coated pits invaginate. The final pinch-off of a **coated vesicle** requires GTP hydrolysis (not shown). Once internalized, coated vesicles lose their clathrin and associated adaptor-protein coat. Uncoated vesicles fuse with **early endosomes** to form **sorting vesicles** (i.e., **late endosomes**). Sorting vesicles separate imported content from their receptors. The latter are recycled to the membrane. The vesicles that remain are **lysosomes**, with digestive enzymes that digest (hydrolyze) vesicle contents that are released and recycled in the cytoplasm. Watch live video of fluorescently labeled proteins, the bright spots, entering cells at ^{17.4}[Receptor-Mediated Endocytosis](#) (watch the two left panels). The uptake of cholesterol bound to **low-density lipoprotein (LDL)** is a well-known example of receptor-mediated endocytosis (Figure 17.16).

Low Density Lipoprotein (LDL, or 'Bad Cholesterol')

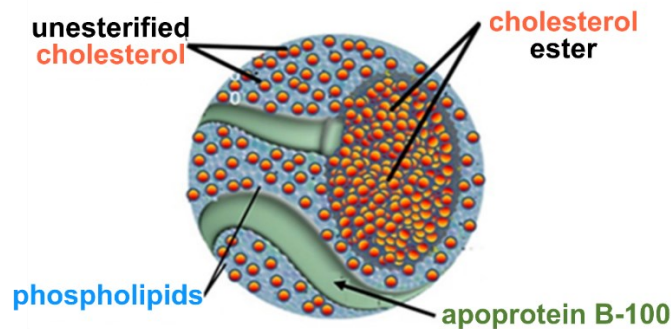


Fig. 17.16: A low-density lipoprotein coated with cholesterol

A single LDL phospholipid-protein complex carries as many as fifteen thousand molecules of cholesterol. LDL, sometimes called “bad cholesterol,” is not good for you at high levels, whereas high-density lipoprotein (HDL) is “good cholesterol.” As one gets older, it is important to monitor one’s HDL/LDL ratio; the higher it is, the better!

17.6.2 Exocytosis and the Formation of Protein-Storage Organelles

Maintaining cell size or volume seems to be a built-in task of the receptor-mediated endocytosis machinery, which balances endocytosis with membrane recycling. Exocytosis is also necessary to restore plasma membrane internalized by pinocytosis and phagocytosis, and for eliminating cellular waste products. Exocytosis is also the end point of the process of packaging proteins for secretion, for intracellular storage (e.g., in lysosomes, peroxisomes) and for insertion into the membrane itself. *Endocytotic* and *secretion* vesicles form in “opposite directions,” but both share common structural features with the cell membrane, from which one vesicle type is derived and with which the other fuses (*respectively*). The formation of lysosomes and secretion vesicles starts in the *rough endoplasmic reticulum (RER)*, followed by passage and maturation through Golgi vesicles, as shown in Figure 17.17.

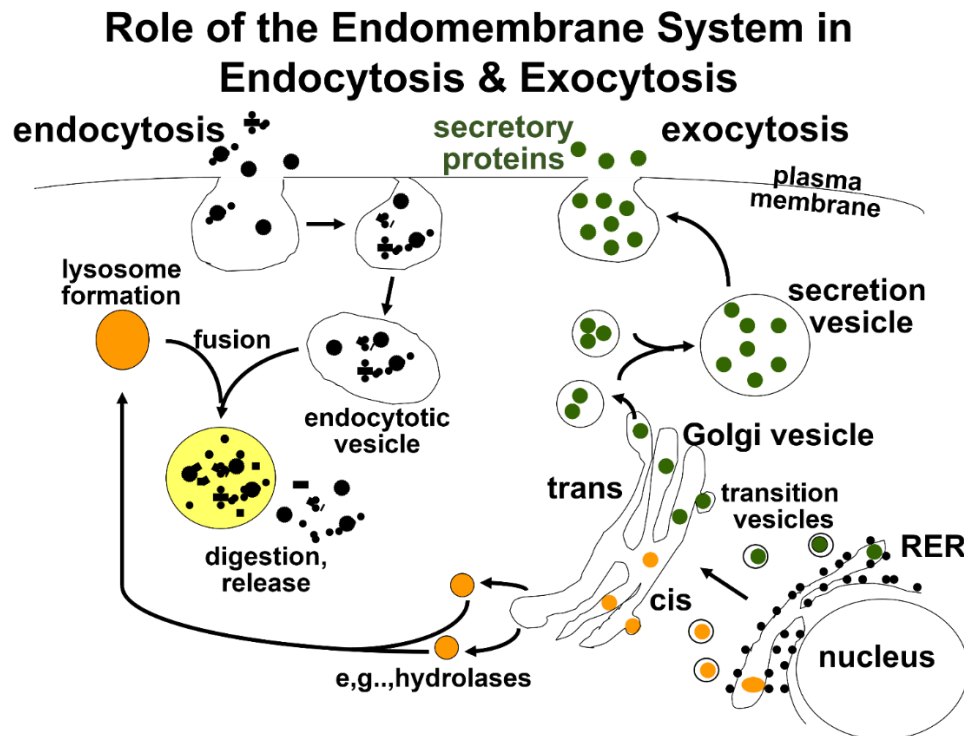


Fig. 17.17: Follow protein sorting (traffic) from translation on RER to-and-through Golgi vesicles (lower right and center-right) to either secretion vesicles (upper right) or storage vesicles (e.g., lysosomes, upper left), and the roles of each pathway in exocytosis and endocytosis.

As we have seen, many secretory and membrane proteins are glycoproteins, and their glycosylation begins in the RER. Check the following link to review the process.



[291 The Path to Sugar-Coated Cells](#)



A cell often produces many packaged proteins at the same time that must be sorted to the correct place—extracellular fluids, lysosomes, peroxisomes, other “microbodies,” and of course, membranes. How do they do it? *Table 1* lists some representative packaged proteins that must be segregated from the cytoplasm.

Table 17.1
Cellular Proteins Segregated from the Cytoplasm

Hormones	Immune-System Proteins	Neurotransmitters	Other
insulin	immunoglobulin G (IgG), a class of circulating antibodies	acetylcholine	epidermal growth factor (EGF)
growth hormone	IgM and other cell membrane antibodies	dopamine, adrenaline, noradrenaline, and other monoamines	neural growth factor (NGF)
follicle-stimulating hormone (FSH)	major histocompatibility complex (MHC) proteins	serotonin	fibrinogen (and other blood-clotting factors)
oxytocin		some amino acids (glutamate, aspartate, glycine)	fibronectin (and other extracellular matrix proteins)
prolactin			plant cell wall components
adrenocorticotrophic hormone (ACTH)			trypsin, pepsin (gut digestive enzymes)
			chloroplast and mitochondrial proteins
			nuclear proteins

17.7 Directing the Traffic of Proteins in Cells

All polypeptide proteins have been translated by ribosomes from a sequence of bases in an mRNA, and each has a specific functional location, whether in the cytoplasm, on cellular membranes, inside organelles, or in extracellular fluids. In this section we consider the movement and sorting of proteins from RER through vesicles of the *endomembrane system*, as well as the transport of proteins into and out of organelles.

17.7.1 Proteins Packaged in RER Are Made as Larger Precursor Proteins

All protein synthesis begins with the formation of an initiation complex and subsequent elongation cycles of peptide-bond formation and carboxyl-terminal amino acid addition. But proteins to be packaged for secretion or into lysosomes, peroxisomes, or other microbodies will complete translation elongation directly into the cisternae (enclosed spaces) of the RER. What *traffic signal* led some proteins to the RER and others elsewhere in the cytoplasm? A model system for studying secretory protein synthesis is *mouse myeloma cells*, cancerous lymphocytes of the immune system. Normal lymphocytes make immunoglobulin G (IgG) with

light- and heavy-chain polypeptide subunits. IgGs are circulating **antibodies**. The myeloma cells make only IgG light chains, which are easily isolated from a cell culture medium.

In an early experiment, mouse myeloma cells were cultured under different conditions, and the secreted IgG light chain that was produced was analyzed. The results revealed that the secreted mouse light-chain IgG proteins that had been made in an in vitro translation system were in fact larger than proteins naturally secreted by the cells. The experimental protocol is illustrated in Figure 17.18.

Cell-Free Translation of Secretory Proteins Evidence of Something Different About Secretory Protein Synthesis

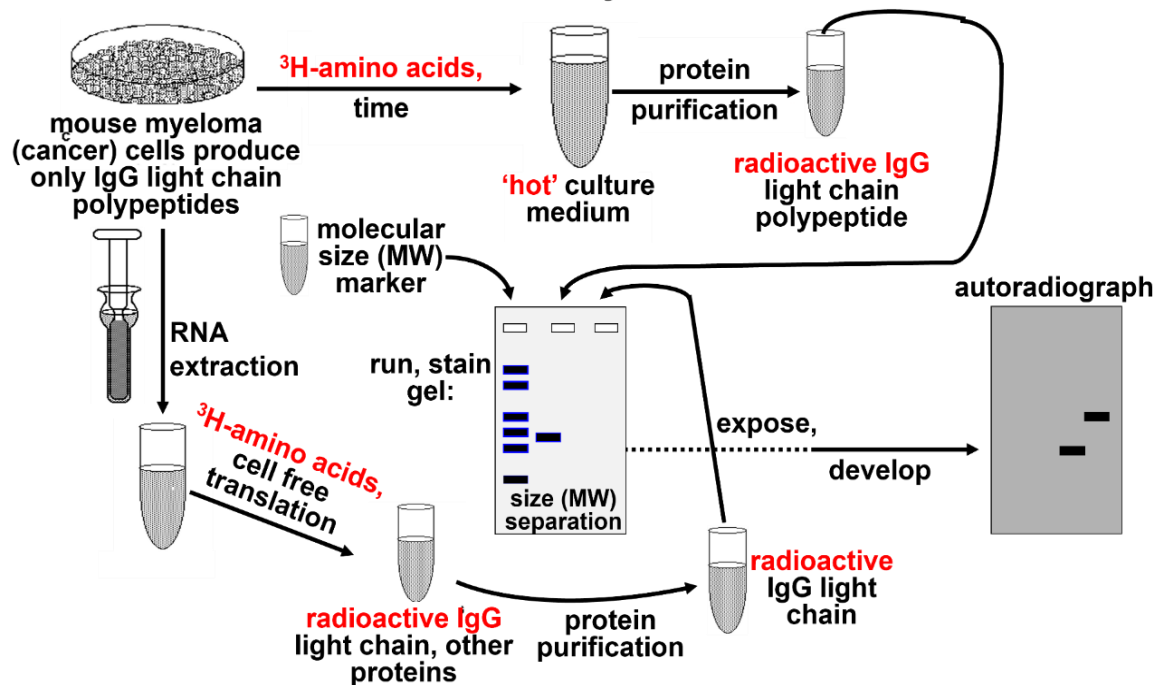


Fig. 17.18: Experimental strategy for determining the role of RER in the synthesis of secretory proteins.

In one part of this experiment, myeloma cells were grown in the presence of radioactive amino acids. The resulting radioactive IgG light-chain polypeptides secreted by the cells were isolated (follow the arrows across the top of Figure 17.18). In another part of the experiment, mRNA was isolated from a different batch of the myeloma cells and added to an in vitro translation system containing radioactive amino acids (follow the arrows down the left and across the bottom of the figure).

The radioactive polypeptides made *in vivo* and *in vitro* were electrophoresed (follow the arrows to the gel) and the gels then autoradiographed. From the results at the lower right, you can see that the secreted polypeptides made *in vivo* had migrated faster on the gel than those translated *in vitro*. So, the cell-free translation product was indeed larger than the *mature* secreted polypeptide. To explain these results, Günter Blobel and colleagues extended the "Signal Hypothesis," proposing that the signal was a short N-terminal **signal peptide** that directs a growing secretory polypeptide to the RER. They further proposed that the **signal**

peptide is a temporary "traffic" signal, removed by an RER-associated enzyme as the polypeptide crossed the RER membrane into the cisternal space.



[304 Formulating the Signal Hypothesis: Early Experiments](#)

17.7.2 Testing the Signal Hypothesis for Packaging Secreted Proteins in RER

In the test of the *Signal Hypothesis* (which earned Blobel the 1999 Nobel Prize in Physiology or Medicine), isolates of RER membranes were included with mouse myeloma cell mRNA in cell-free protein-synthesis systems. This time, when secreted and cell-free synthesized IgG light chain polypeptides were electrophoresed and the gel was autoradiographed, both polypeptides were the same size as the mature, secreted polypeptides. Therefore, as predicted, RER contains *processing* activity—that is, a **signal peptidase** that removes the signal peptide! The steps of the signal hypothesis that emerged from the experiments of Blobel and his colleagues are illustrated in Figure 17.19.

The Signal Hypothesis, Step-By-Step (counterclockwise from bottom)

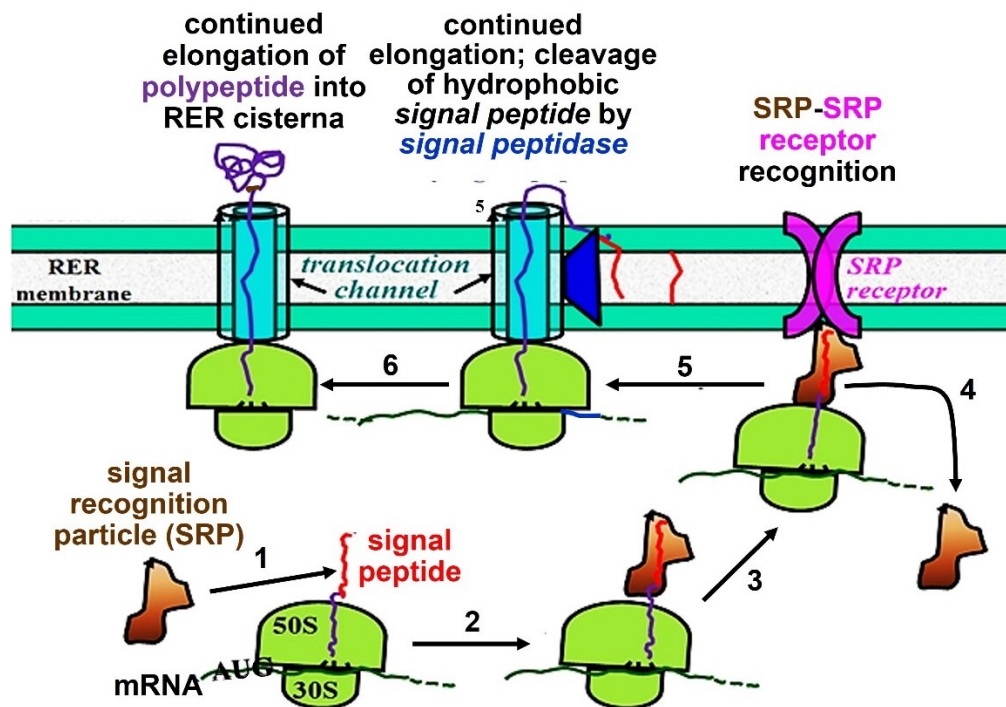


Fig. 17.19: Molecular details (steps) of the Signal Hypothesis leading to protein secretion

Recall that in translation, during polypeptide elongation, the growing polypeptide moves through and emerges from a channel or groove in the large ribosomal subunit. As the N-terminal **signal sequence** (i.e., the signal peptide) of a secretory polypeptide emerges from this groove, it interacts with the RER membrane.

Beginning at the lower left in Figure 17.19, the steps of protein packaging for the secretion process are these:

1. An **SRP** (**signal recognition particle**) binds to the hydrophobic **signal peptide**.
2. Elongation stops until the SRP-ribosome complex finds the RER membrane.
3. The ribosome-SRP complex binds to an **SRP receptor** on the RER membrane.
4. The SRP detaches from the growing polypeptide chain and is recycled.
5. Translation elongation resumes through a **translocation channel**, and a **signal peptidase** in the RER membrane catalyzes **cotranslational** hydrolysis of the signal peptide, which remains embedded in the RER membrane.
6. Elongation continues, and the growing polypeptide begins to fold in the RER.



[305 Testing the Signal Hypothesis](#)



[306-2 Details of the Signal Hypothesis](#)

Step 2 requires that the SRP find and bind to the signal peptide before the nascent polypeptide gets too long and starts folding into a 3D (tertiary) conformation that would hide the signal peptide. The ribosome itself may keep the signal peptide available by destabilizing electrostatic interactions that would otherwise lead to premature folding and an undoubtedly incorrect conformation. For more on ribosome involvement in protein folding, see ^{17.5}[Protein-Folding starts Near The Ribosome Making It](#).

Bacterial secretions include proteins involved in nutrient scavenging as well in cell-wall synthesis, using a secretory mechanism like that of eukaryotes, with obvious differences in detail. Partially elongated signal peptides guide mRNA-bound ribosomes to the cytoplasmic side of the plasma membrane, where the ribosomes bind and then pass elongating proteins through the plasma membrane into the space between the cell membrane and wall. As the protein exits the cell, a bacterial signal peptidase (**SPase**) cleaves the signal peptide. Apparently, the basic mechanism for the secretion of proteins evolved early and is conserved in prokaryotes. Early on, we knew that some antibiotics stop bacterial growth by disrupting the cell wall or by killing the cells outright. Others (e.g., *arylomycins*) disrupt plasma membrane **SPase** function, preventing proteins needed in the space between the cell wall and membrane from ever making it out of the cell. Once widely used against *S. aureus*, arylomycins are no longer effective because many *Staphylococcus* strains have become resistant to these antibiotics. Check out ^{17.6}[Bacterial SPase and Antibiotic Resistance](#) to read about the mechanism of *arylomycin* resistance. As you may already know, *S. aureus* is now resistant to most antibiotics, and illness from untreatable infections has its own name, **MRSA** (Methicillin-Resistant *S. Aureus*—dig on your own to see more about methicillin resistance). While named for *methicillin* resistance, the MRSA acronym has also come to describe untreatable *S. aureus* infections in general.

We turn now to well-studied mechanisms for packaging proteins into eukaryotic organelles and into the membranes themselves.

17.8 Synthesis of Integral Membrane Proteins

Integral membrane proteins may span a membrane one or more times or may simply be anchored (embedded) in the membrane at one end (Figure 17.20, below).

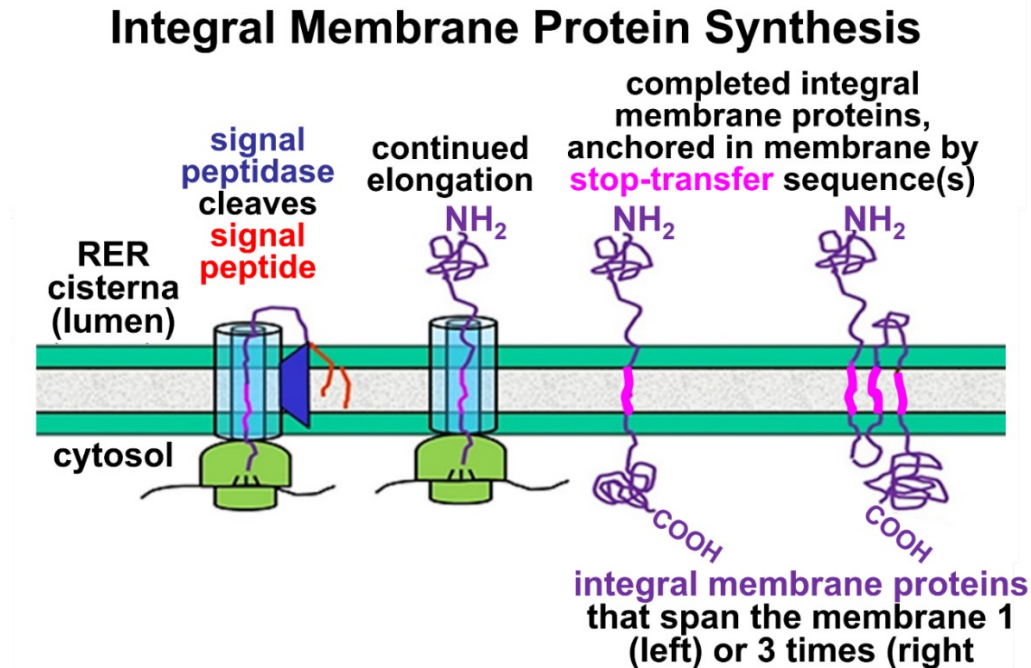


Fig. 17.20: Integral membrane proteins that span the entire membrane have one or more very hydrophobic *stop transfer* signals, in addition to their signal sequence. Such membrane proteins cross the membranes one or more times during their synthesis.

N-terminal signal sequences guide ribosomes translating *integral membrane proteins* to the RER. But before such a protein can pass completely into the RER cisternae, a *stop-transfer* sequence (a hydrophobic domain within the polypeptide chain) traps the protein in the fatty acid interior of the membrane. Multiple stop-transfer sequences account for transmembrane proteins that span a membrane more than once.



[308-2 Integral Proteins Have Stop-Transfer Sequences](#)



17.9 Moving and Sorting Proteins to Their Final Destinations

Like proteins packaged in RER, those made in the cytoplasm go to different destinations before they become functional. What's more, cells can truly multitask (unlike most of us!), handling the synthesis and distribution (or secretion) of many proteins at the same time! Let's begin with a look at the sorting mechanisms for proteins sequestered by the endomembrane system.

17.9.1 Traffic on the Endomembrane Highway

We have already seen that, once packaged in the RER cisternae, proteins begin posttranslational modification (e.g., by "core glycosylation"). Transport vesicles that bud off

from the RER carry packaged and membrane proteins to the *cis* vesicles of the Golgi apparatus. There, vesicle fusion is mediated by the recognition of complementary integral membrane proteins embedded in the two membranes. Later, such packaged proteins are sorted to different organelles or to the plasma membrane. Protein sorting starts as proteins move from the *cis* to the *trans* face of the Golgi vesicles, where specific sorting proteins associate with different packaged proteins in the trans Golgi vesicles. The packaged proteins then sort to vesicles that bud off from trans Golgi stacks. These vesicles move to their final destinations, recognizing and then fusing with appropriate membranes. James E. Rothman, Randy W. Schekman, and Thomas C. Südhof won the 2013 Nobel Prize in Physiology or Medicine for their studies of the regulation of vesicle traffic (check out ^{17.7}[2013 Physiology or Medicine Nobel Prize](#) for more information). Let's follow some proteins in and on RER membranes through the cell (also animated at ^{17.8}[Events of Protein Trafficking](#)) from the RER to their cellular destinations, as summarized in Figure 17.21.

Sorting Pathways for Packaged Proteins

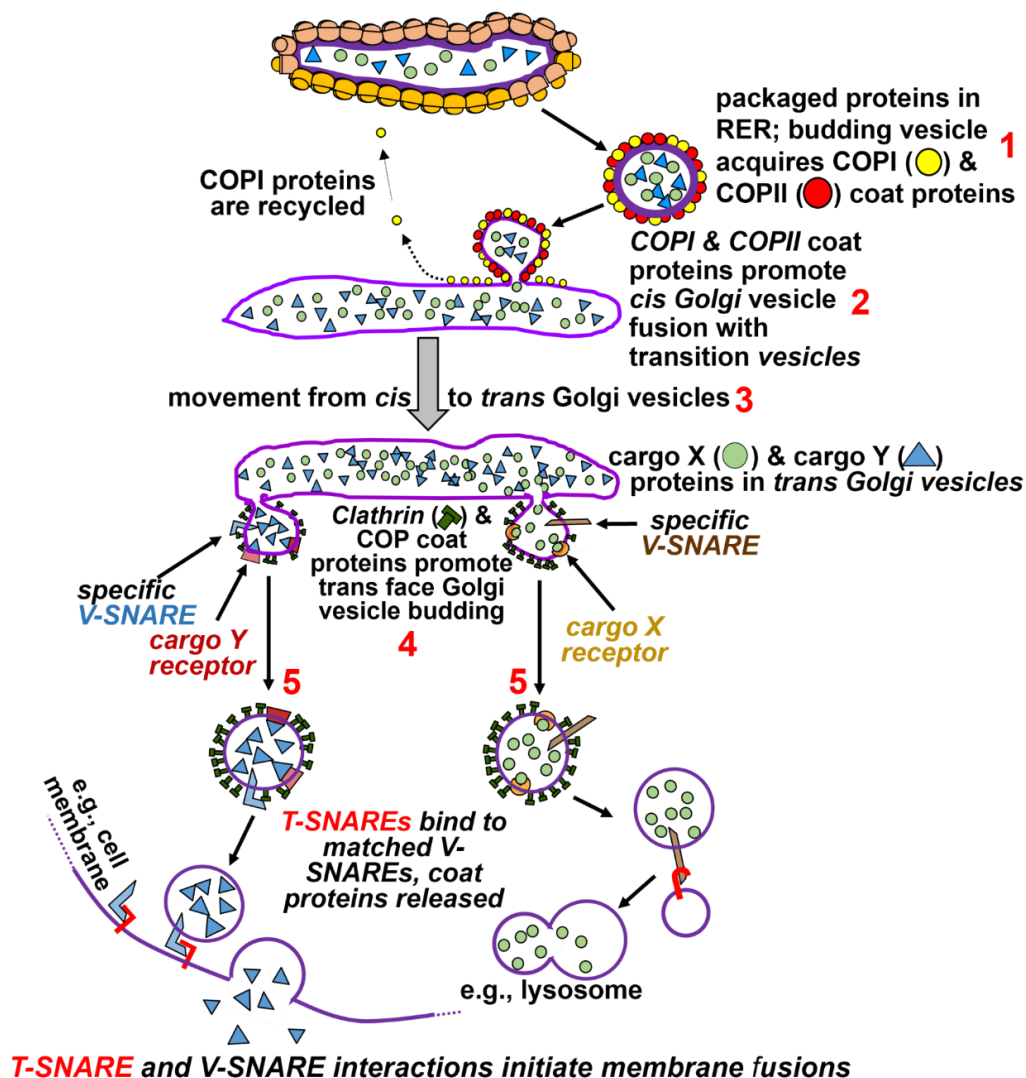


Fig. 17.21: Alternate pathways for **sorting** of cellular (e.g., secreted, vesicular, other) proteins.

Here are the steps to follow the traffic of protein sorting shown in the illustration:

- **Transition vesicles**, carrying their mix of packaged proteins, bud off from the RER with the help of **COPI** and **COPII** coat proteins, dissociating from ribosomes. Transition vesicles, however, remain associated with the **COP** proteins.
- These vesicles fuse with *cis* Golgi vesicles, a process also mediated by **COP** proteins. **COPI** proteins detach during or after fusion, recycling back to the RER.
- Packaged proteins and membrane proteins are further processed as they pass through the Golgi vesicle stack (e.g., undergoing terminal glycosylation).
- A key step in sorting packaged proteins occurs as vesicles bud off from the trans Golgi face with the help of **clathrin**, other **COP proteins**, and specific **cargo-receptor** proteins. The **cargo-receptor proteins** recruit specific packaged proteins (now called **cargo proteins**) and enrich them in their nascent vesicles. **V-SNARE** (for vesicle-SNARE) proteins also associate with the vesicles, guiding them to their ultimate destination.

When **V-SNARE** proteins on their vesicles next bind to complementary **Target-Snare (T-SNARE)** proteins on receiving membranes, the two membranes fuse, at which point the different types of vesicle part ways.

Some vesicles follow a pathway to organelles, fusing with **lysosomes** or similar vesicles along the way to stock up on appropriate enzymes and other required protein content. Coat proteins come off the fusing vesicle and are recycled, while vesicle contents are transferred into the next generation vesicle. Those containing secretory proteins typically fuse to form larger **secretory vesicles** (or secretion vesicles). These will next fuse with the plasma membrane, releasing their contents to the extracellular fluid. Once again, coat proteins and clathrin dissociate from the secretory vesicle during this fusion. At some point, secretory vesicles will release their contents, but until they are signaled to do so, they will remain resident in the cell cytoplasm, not unlike the other vesicles formed in these sorting pathways.

Other players have been left out of this discussion, notably those that hydrolyze nucleotide triphosphates to provide the energy for protein trafficking. In addition, you may recognize **clathrin** as a molecular player that also functions in receptor-mediated endocytosis. Maybe that's not a surprise! After all, endocytosis is at least partly molecular traffic in the opposite direction of vesicle formation and secretion.

17.9.2 Nuclear Protein Traffic

We saw earlier that large molecules (mRNAs, tRNAs) and even whole particles (i.e., ribosomal subunits) can cross the nuclear envelope through nuclear pores. But what about those large proteins that are destined to enter the nucleus?

It turns out that these proteins have **nuclear localization signals** rich in positively charged amino acids (lysine, proline), which enable binding to the negatively charged domain of a **nuclear transport receptor** protein in the cytosol (below, Figure 17.22).

Importing Proteins into the Nucleus: Requires a *nuclear localization signal* (e.g., **-pro-pro-lys-lys-arg-lys-val-**) and ATP-mediated active transport

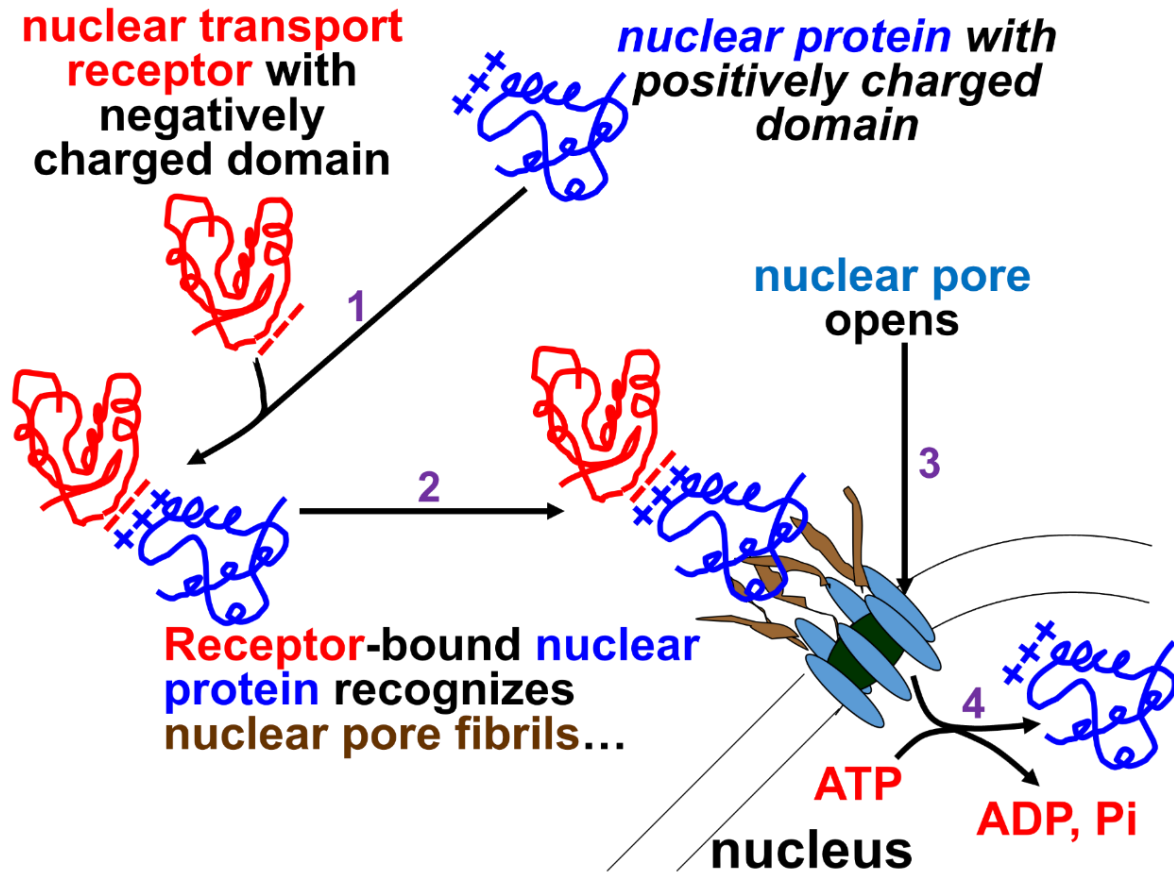


Fig. 17.22: Nuclear proteins made in the cytoplasm contain a positively charged *nuclear localization signal*. This signal binds to the electronegative region of a *nuclear transport receptor* (1) which enables the nuclear protein to bind nuclear pore fibrils (2), guiding the protein into the nucleus (3), a step that requires energy supplied by ATP hydrolysis.

As the complex of the two proteins approaches a *nuclear pore*, it interacts with ***nuclear pore fibrils***, causing the pore to open. The two bound proteins then cross the double membrane of the nuclear envelope, where they accumulate against a concentration gradient. This *active transport* comes from ATP hydrolysis as the nuclear proteins enter the nucleus.

17.9.3 Mitochondrial Protein Traffic

Recall that mitochondria contain their own genome and translational machinery. Thus, they transcribe RNAs and translate proteins of their own. But genes in the nucleus also encode many of the proteins found in mitochondria. Figure 17.23 (below) illustrates how the mitochondria import these proteins.

Mitochondrial Protein Import

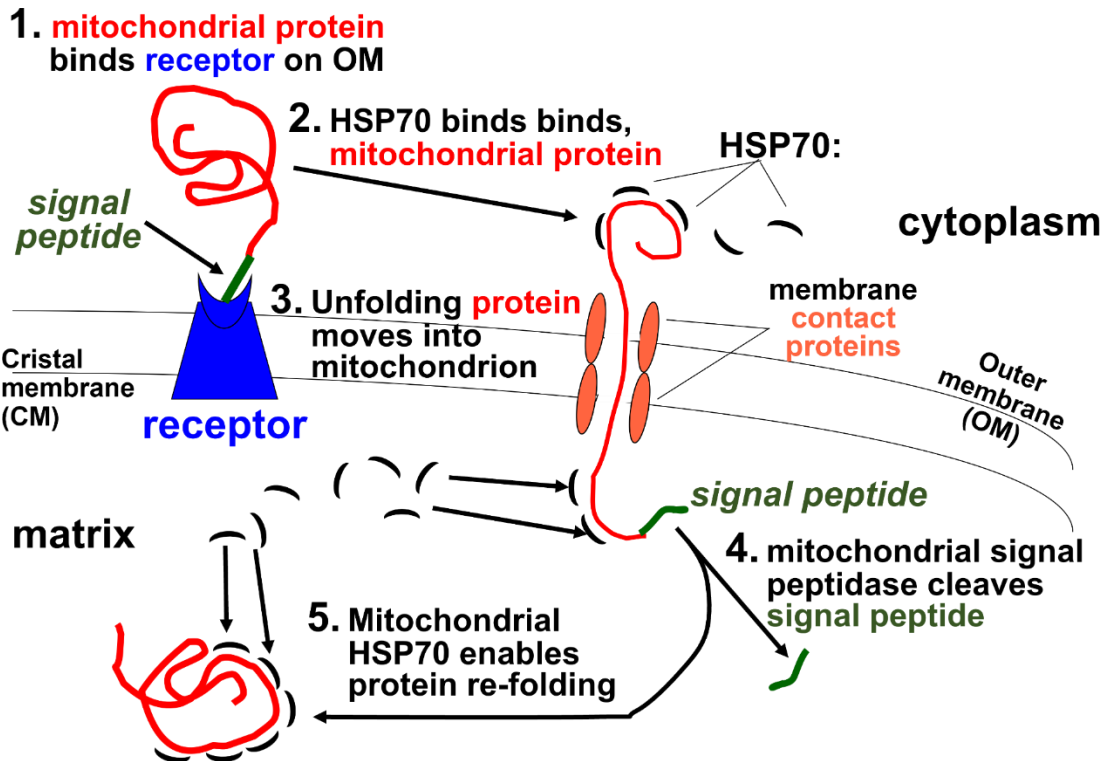


Fig. 17.23: Nuclear-encoded proteins destined for mitochondria are synthesized with a mitochondrial N-terminal *signal sequence* that is removed by a *mitochondrial signal peptidase*. Polysomes don't attach to mitochondria. A different mechanism engages the signal sequence with membrane proteins for transport.

Unlike the cotranslational packaging of proteins by the RER, mitochondrial protein transfer is posttranslational. This means that mitochondrial proteins formed in the cytoplasm have already folded into a tertiary structure. However, the folded protein exposes an N-terminal **signal peptide** on its surface that recognizes and binds to a **receptor protein** at the outer mitochondrial membrane (1). The **receptor protein** spans both the mitochondrial **outer membrane (OM)** and the **cristal membrane (CM)**. The receptor protein delivers the protein to be imported to **membrane-contact proteins**, which also span both mitochondrial membranes (2). The **membrane-contact proteins** act as channels or pores, through which a mitochondrial protein made in the cytoplasm will cross into the mitochondrial matrix. But there is a problem: the folded protein *cannot* cross the membrane by itself, precisely *because* it is folded!

The entry of completed proteins into a mitochondrion requires **chaperone** proteins—in this case, the **HSP70** (*heat-shock 70*). **HSP70** controls *unfolding* of the mitochondrial protein as it traverses the membranes into the matrix (3). Upon removal of the signal peptide by a mitochondrial **signal peptidase** (4) another HSP70 molecule resident in the matrix facilitates *refolding* of the protein into a biologically active shape (5). Recall that HSPs were initially discovered as 70 Kd proteins that accumulated in heat-stressed organisms!



17.10 How Cells Are Held Together and How They Communicate

Proteins and glycoproteins on cell surfaces play a major role in how cells interact with their surroundings and with other cells. We'll look first at some of the proteins in the **glycocalyx** of adjacent cells that interact to form different kinds of cell-cell junctions. Then we'll see how some of these proteins interact with extracellular proteins and carbohydrates to form the **extracellular matrix (ECM)**. Still other membrane proteins are part of receptor systems that bind hormones and other signaling molecules at the cell surface, conveying information into the cell by **signal transduction**.

17.10.1 Cell Junctions

Cell junctions serve different functions in cells and tissues. In healthy cells they serve to bind cells tightly, to give tissues structural integrity, and to allow cells in contact with one another to pass chemical information directly between them. Electron micrographs and illustrations of different cell junctions are shown in Figure 17.24.

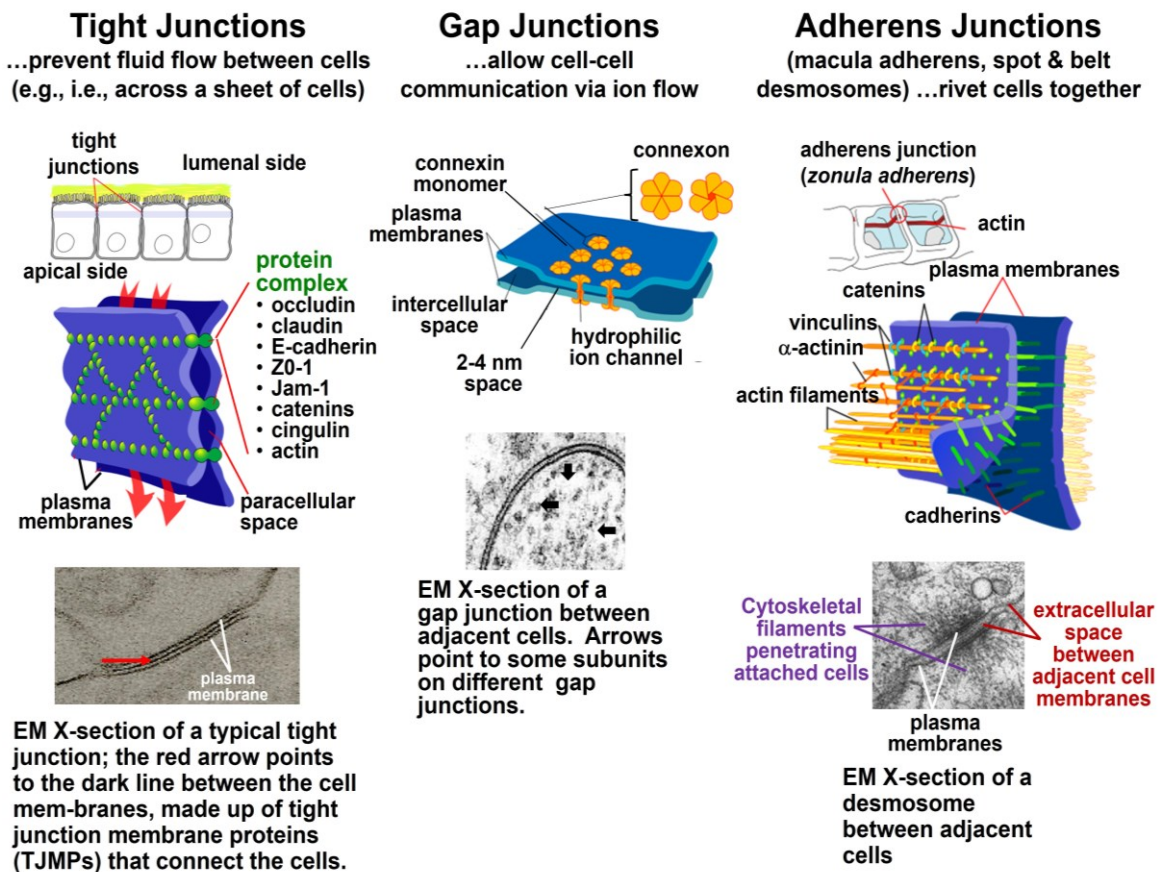


Fig. 17.24: **Tight, gap, and adherens** junctions of animal cells involve different membrane proteins.

Tight Junctions (*zonulae occludentes*) are typical in sheets of epithelial cells that line the *lumens* of organs (e.g., intestines and lungs). *Zonulae* (singular: *zonula*) refers to the fact that these structures form a band encircling an entire cell, attaching it to all surrounding cells. *Occludentes* (singular: *occludens*) refers to the "watertight" seal, or *occluding barrier* of tight

junctions that stops extracellular fluids from passing between cells to cross to the other side of a sheet of cells. **Tight-junction membrane proteins (TJMPs)** create this waterproof barrier.

Gap junctions enable chemical communication between cells. **Connexon** structures made of **connexin** proteins act as pores that open to allow direct movement of ions and small molecules between cells. Communication by ion or molecular movement is quite rapid, ensuring that all cells in a sheet or other tissue in one metabolic state can respond to each other and switch to another state simultaneously. In plants we have seen the plasmodesmata, which perform functions similar to those of gap junctions of animal cells.

Desmosomes (*adherens junctions*) essentially glue (adhere) cells together, giving tissues their strength. **Belt desmosomes** (*zonula adherens*) surround entire cells, strongly binding them to adjacent cells. Spot desmosomes (*macula adherens*) act like rivets, attaching cells at *spots*. In both cases, **cadherin proteins** cross cell membranes from intracellular **plaque** proteins, spanning the intercellular space to link adjacent cell membranes together. Plaques are connected to intermediate filaments (keratin) in the cytoskeleton, strengthening intercellular attachments and, thus, the tissue cell layer.



[310-2 Cell-Junction Structure and Function](#)

Many glycocalyx proteins that interact to form junctions between cells are glycoproteins. Generally, proteins that interact to bind cells together are called **ICAMs** (Intercellular Cell Adhesion Molecules), or just **CAMs**.

- **Selectins** are one kind of ICAM. During blood clotting, *selectins* on one platelet recognize and bind to specific receptors on other platelets, contributing to the clot.
- **NCAMs** are another kind of ICAM, ones with sugary immunoglobulin domains that interact specifically to enable neural connections.
- We've already seen the calcium-dependent **cadherins** involved in forming *adherens junctions* (desmosomes). These are essentially the "glue" that binds cells together to form strong cohesive tissues and sheets of cells. Cell-cell binding begins with recognition and connection via the glycocalyx and extracellular matrix.



[311-2 Glycocalyx Sugars Covalently Link to Plasma Membrane Proteins](#)



[312 Cell-Adhesion Molecule Functions in the Glycocalyx](#)

Figure 17.25 (below) illustrates several examples of membrane proteins that enable cell-cell recognition and adhesion.

The Glycocalyx: Cell-Cell Recognition, Adhesion

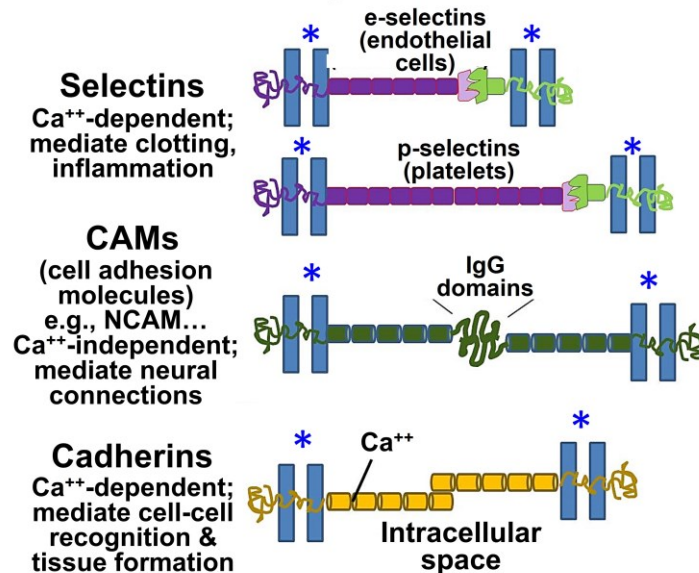


Fig. 17.25: Glycocalyx protein interactions play roles in cell-cell recognition, attachment, and junctions. Asterisks indicate plasma membrane bilayers between attached cells.

17.10.2 Microvesicles and Exosomes

Of the several ways cells take up material from their environment, *pinocytosis*, or bulk transport, seems to be a random process in which tiny bits of plasma membrane engulf small amounts of extracellular fluids along with ions, particles, and incidental solutes. Pinocytotic vesicles range from 0.5 to 5 μm in diameter. On the other hand, **exosomes** and **microvesicles** are spherical bits of plasma membrane and endosomes (respectively) that are shed by cells into the extracellular space. Known by several names, these extracellular vesicles (**EVs**) were first reported in the 1980s as small vesicles released by reticulocytes. At around 1,000 nm (1 μm), *microvesicles* are similar in size to pinocytotic vesicles. *Exosomes* are about a tenth the size of microvesicles, ranging in size from 40–100 nm in diameter (see Figure 17.26).

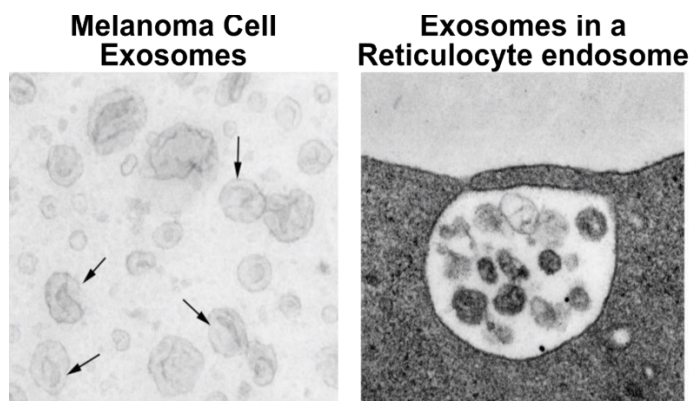


Fig. 17.26: Left: Transmission electron micrograph of melanoma cell exosomes. Some of the larger vesicles show an irregular, cup-like shape (arrows). Right: Transmission electron micrograph of a reticulocyte endosome containing exosomes about to fuse with the plasma membrane.

There is much evidence that microvesicles and exosomes are not artifacts. For example:

- Microvesicles, such as those from melanoma cells (Figure 17.26, left panel), may be released normally, suggesting that they are not cellular waste products.
- Reticulocyte endosomal vesicles were caught in the act of fusing with the cell membrane, releasing exosomes with their contents (Figure 17.26, right panel).
- Microvesicles released by *dendritic cells* can stimulate *T cells*, both of which are cells of the immune system.

We can conclude that microvesicles are physiologically significant structures. We know that cells talk to each other by releasing chemicals that can act over short or long distances. Familiar examples include information transfer by hormones from endocrine glands, chemicals released into the intercellular space, and the neurotransmitters of the nervous system. Microvesicle activity may be part of intercellular communication pathways, and they are clearly involved in normal reticulocyte maturation to erythrocytes. Perhaps exosomes are yet another unique mechanism of intercellular communication. For a review, see ^{17.9}[Exosomes & Intercellular Communication](#).

17.10.3 Cancer and Cell Junctions

During embryogenesis, cells migrate from a point of origin by attaching to and moving along an **extracellular matrix (ECM)**, which acts as a path to each cell's destination. This ECM (or basal lamina) is made up of secretions from other cells..., or from the migrating cells themselves! One major secretion is **fibronectin**, one of whose functions is to bind to **integrins**, integral membrane proteins that attach cells to the ECM.

During development, integrins respond to fibronectin by signaling cell and tissue differentiation, complete with the formation of appropriate cell junctions. An orderly sequence of gene expression and membrane protein syntheses enables developing cells to recognize each other as different or the same (summarized in the drawing below (Figure 17.27)).

Role of Cell Surfaces in Tissue Differentiation

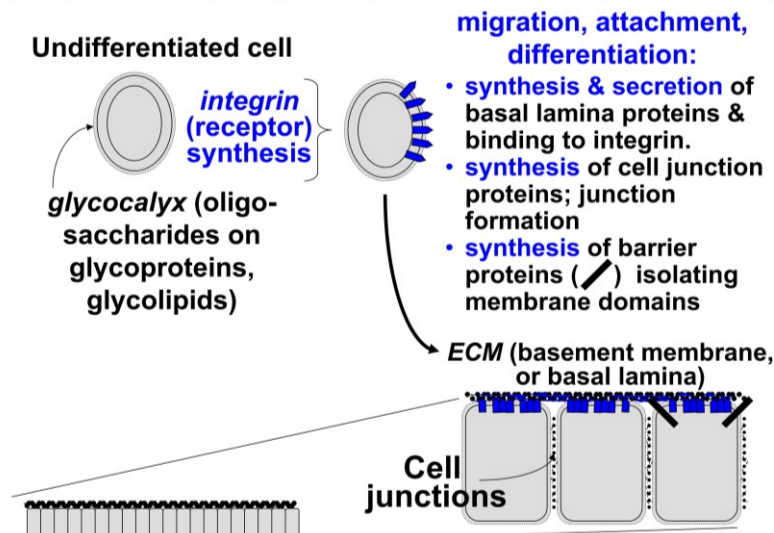
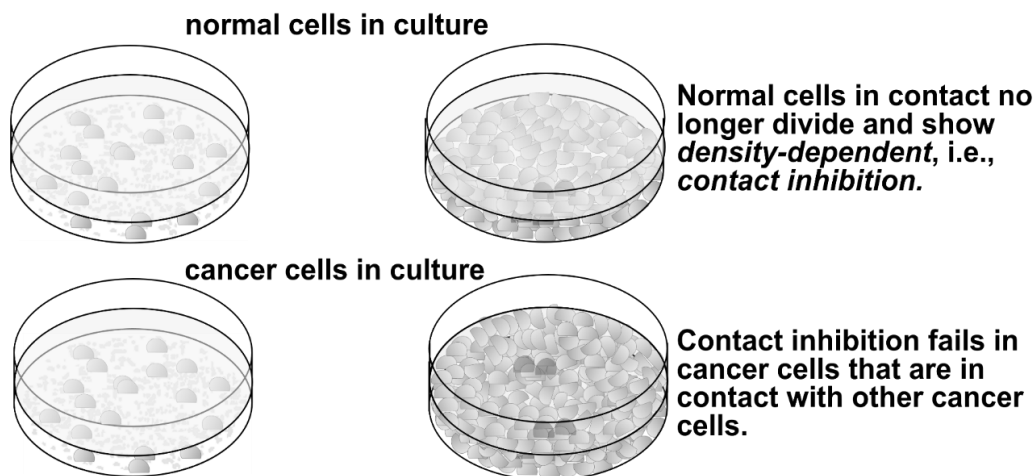


Fig. 17.27: Migration and attachment of cells during development requires the synthesis of membrane proteins that recognize and bind proteins on the membranes and glycocalyx of other cells, forming cell junctions that enable tissue formation.

An early difference between eukaryotic normal and cancer cells is how they grow in culture. Normal cells settle to the bottom of a culture dish when placed in growth medium. Then they grow and divide, increasing in number until they reach **confluence**, when a single layer of cells completely covers the bottom of the dish. The cells in this monolayer seem to “know” to stop dividing, as if they had completed formation of a tissue (e.g., a layer of epithelial cells). This phenomenon, originally called **contact inhibition**, implies that cells let each other know when they have finished forming a tissue and can stop cycling and dividing. In contrast, cancer cells do not stop dividing at confluence. Instead, they continue to grow and divide, piling up in multiple layers.

Among other deficiencies, cancer cells do not form *gap junctions* and typically have fewer *cadherins* and *integrins* in their membranes. Thus, cancer cells cannot inform each other of when they reach confluence. Neither can they form firm *adherens junctions*. In vivo, a paucity of integrins would inhibit cancer cells from binding and responding to *fibronectin*. As a result, they also have difficulty attaching firmly to an extracellular matrix, which may explain why many cancers **metastasize**, or spread from their original site of formation. These differences in growth in culture between normal and cancer cells are shown in Figure 17.28.

Contact Inhibition The growth of Cancer vs. Normal Cells



Upon contact, cancer cells fail to form *gap junctions*, some have defective or fewer *integrins* or *cadherens* on their plasma membranes.

Fig. 17.28: Density-dependent inhibition (*contact inhibition*) occurs when cells multiplying and spreading on a surface cease dividing when there is no more room on the surface. Cancer cells have lost the property of contact inhibition and keep growing over one another in layers. Loss of contact inhibition is correlated with an absence of gap junctions in cancer cells.



[313 Formation of a Glycocalyx: Normal Development and Cancer](#)



[314-2 Role of the Extracellular Matrix in Cell Migration & Development](#)



17.11 Signal Transduction

When hydrophobic chemical **effector** molecules, such as steroid hormones, reach a target cell, they can cross the hydrophobic membrane and bind to an intracellular receptor to initiate a response. When large **effector** molecules (e.g., protein hormones) or highly polar hormones (e.g., adrenalin) reach a target cell, they can't cross the cell membrane. Instead, they bind to transmembrane protein receptors on cell surfaces. A conformational change initiated on the extracellular domain of the membrane receptor protein induces further allosteric change on its cytoplasmic domain. A sequential series of molecular events then converts information delivered by the external effector into intracellular information, a process called **signal transduction**. A general outline of signal-transduction is illustrated below in Figure 17.29.

Signal Transduction Can Lead to Many Different Effects

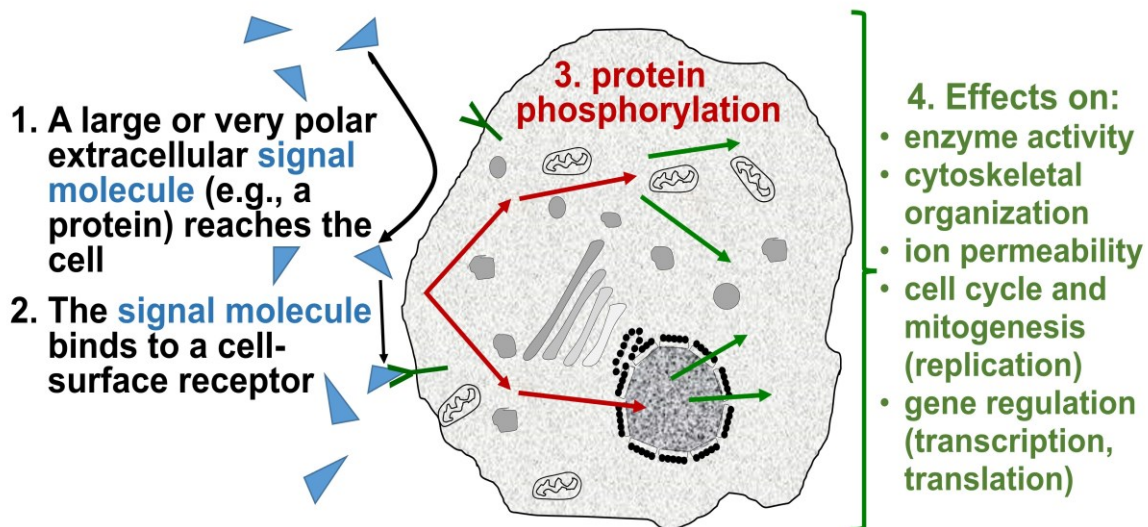


Fig. 17.29: Signal transduction by *effectors* (e.g., hormones) can lead to many different effects in the cytoplasm as well as in nuclei.

Many effects of signal transduction are mediated by a sequence, or *cascade* of protein phosphorylations catalyzed by *protein kinases* inside the cell, all of which share some common features as described in the elegant YouTube video at ^{17.10}[Cell Messaging by Signal Transduction](#). Here we will consider *G-Protein-linked* and *enzyme-linked receptors*.



[315-2 Introduction to Signal Transduction](#)

17.11.1 G-Protein-Mediated Signal Transduction by PKA (Protein Kinase A)

GTP-binding proteins (**G-Proteins**) transduce extracellular signals by inducing synthesis of **second-messenger** molecules in the cells. Figure 17.30 illustrates seven steps of G-protein-mediated signal transduction.

G-protein Mediated Signal Transduction and cAMP synthesis

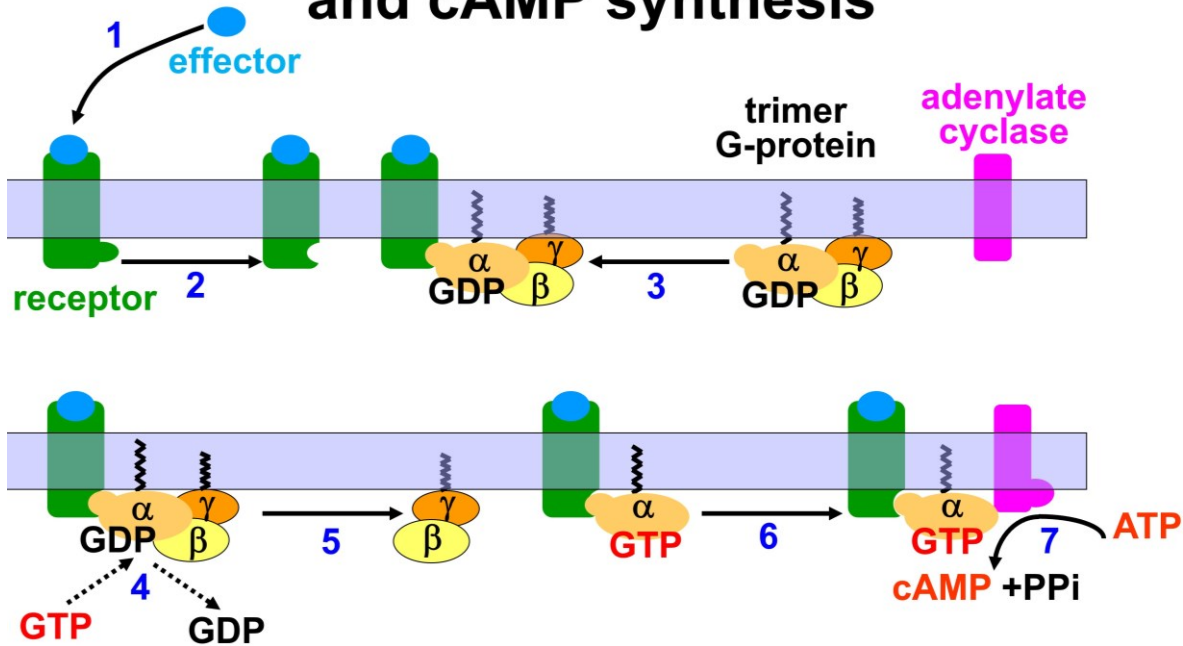


Fig. 17.30: G-proteins cycle between bound GTP and bound GDP. They are intermediates in signal transduction between an active, effector-bound membrane receptor and an adenylate cyclase.

When hormones or other effector (signal) molecules bind to a membrane receptor, an allosteric change on the *extracellular domain* of the receptor is transmitted to the cytoplasmic domain of the receptor, increasing its affinity for trimeric G-proteins embedded in the cytoplasmic surface of responsive cell membranes. G-protein *trimers* consist of α , β and γ subunits. The receptor changes shape upon binding its effector signal molecule (steps 1 and 2). In this altered conformation, the receptor recognizes and binds to the G-protein trimer (step 3). Upon the binding of the trimer to the receptor, GTP displaces GDP on the **α subunit** of the G-protein (step 4). After a conformational change in the **α subunit**, it dissociates from the β and γ subunits (step 5). In this illustration, the GTP- α subunit can now bind to a transmembrane **adenylate cyclase** enzyme (step 6). Finally, the initial extracellular chemical signal is **transduced** to an intracellular response, involving a second-messenger molecule (step 7). In this case, the second messenger is **cAMP**.



[316-2 G-Protein Signal Transduction](#)



The well-known **fight-or-flight** response to adrenalin in the liver cells of higher animals is a good example of a cAMP-mediated cellular response. Once formed, cAMP binds to and activates **protein kinase A (PKA)**, setting off a **phosphorylation cascade** that leads to a physiological response. Details of a G-protein-mediated fight-or-flight signal amplification cascade are shown in Figure 17.31.

G-protein Mediated Signal Transduction and the Fight-or-Flight Response: cAMP and an Amplification Cascade

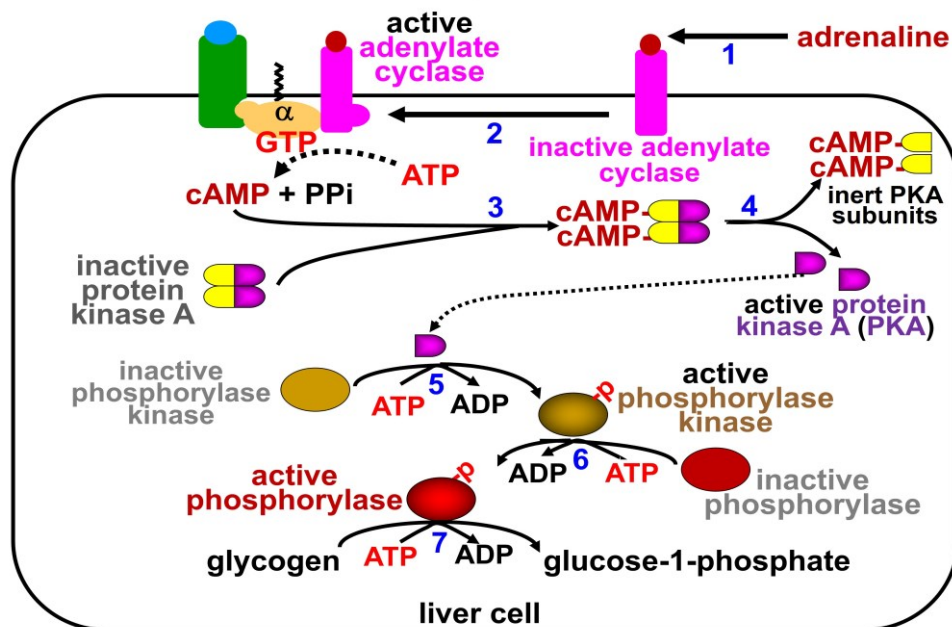


Fig. 17.31: In the *fight-or-flight* response to adrenalin (adrenalin rush), cAMP activates protein kinase A, and a phosphorylation cascade leads to the release of glucose into the circulation. cAMP is a second messenger of signal transduction for many different cellular responses.

After activation of adenylate cyclase (steps 1 and 2 above), cAMP is synthesized. A cAMP binds to each of two of the four subunits of an *inactive PKA* (step 3). An allosteric change dissociates the tetramer into two cAMP-bound inert subunits and two *active PKA* subunits (step 4). Each *active PKA* enzyme catalyzes the phosphorylation and activation of the enzyme *phosphorylase kinase* (step 5). In step 6, phosphorylase kinase catalyzes *glycogen phosphorylase* phosphorylation.

In the cascade, activation of just a few enzyme molecules at the top of the sequence results in the activation of multiple subsequent enzymes, resulting in a much-amplified cellular response, referred to as an **amplification cascade**. The result can be an immediate dramatic response. At the end of the fight-or-flight *phosphorylation cascade*, a now-active glycogen phosphorylase catalyzes the hydrolysis glycogen to glucose-1-phosphate (step 7). This results in a rapid retrieval of free glucose from liver cells into the circulation.

See again how this works by reviewing the conversion of glucose-1-phosphate (G-1-P) to G-6-P in glycolysis and its fate in gluconeogenesis. Of course, the increase in circulating glucose provides the energy for the *fight-or-flight* decision.



In addition to activating enzymes that break down glycogen, cAMP-activated PKA mediates cellular responses to different effectors, resulting in a phosphorylation cascade leading to several results:

- Activation of the enzymes catalyzing glycogen synthesis
- Activation of the *lipases* that hydrolyze fatty acids from triglycerides
- Microtubule assembly
- Microtubule disassembly
- Mitogenic effects (activation of the enzymes of replication)
- Activation of transcription factors, which increases or decreases gene expression

Of course, when the organism no longer needs a cellular response, it must stop producing the signal molecules (hormones or other effectors). As their levels drop, effector molecules dissociate from their receptors and the response stops. This is all possible because the binding of signals to their receptors is freely reversible!

17.11.2 Signal Transduction using PKC (Protein Kinase C)

A different G-protein-mediated signaling pathway activates a *protein kinase C* (*PKC*). The activation of *PKC* is illustrated in Figure 17.32 (below).

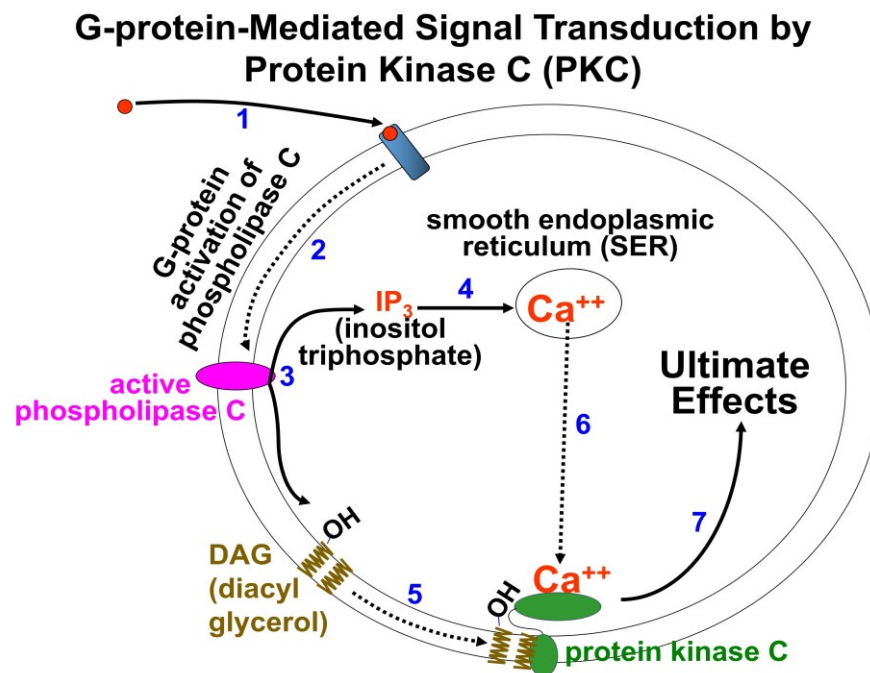


Fig. 17.32 G-proteins mediate signal transduction through different membrane receptors and enzymes. Interactions shown here generate second messengers, which activate *protein kinase C*, leading to a phosphorylation cascade with different responses in target cells.

The role of G-proteins is similar for *PKA*- and '*PKC-signal transduction*', but *PKC* activation does not involve cAMP and requires two intervening cellular responses to an effector. The net result is to generate different second messengers. These in turn activate different *phosphorylation cascades*. Like *PKA*, *PKC*-mediated signal transduction also *amplifies* the cell's first molecular response to the effector.

Here are the steps of PKC signal transduction:

1. An effector-signal molecule binds to its receptor.
2. The effector-bound receptor now binds the G-protein trimer, the GTP for GDP swap on the alpha subunit of the trimer, dissociating the beta-m and gamma-subunits from the trimer. The receptor-bound GTP-alpha- subunit activates an inactive membrane-bound phospholipase C enzyme.
3. Active *phospholipase C* catalyzes formation of cytosolic '*inositol triphosphate*' ('*IP₃*') and of membrane-bound '*diacyl glycerol*' ('*DAG*'), two of those other intracellular second messenger molecules.
4. *IP₃* interacts with receptors on smooth endoplasmic reticulum, causing the release of sequestered *Ca⁺⁺* ions into the cytoplasm.
5. *DAG* finds and binds PKC in the membrane.
6. The *Ca⁺⁺* ions also fin and bind PKC, completing activation of the '*protein kinase C*' ('*PKC*') enzyme.
7. Activated PKC initiates a phosphorylation-amplification cascade leading to cell-specific responses.

PKC-mediated effects include the following:

- Neurotransmitter release
- Hormone (e.g., growth hormone, luteinizing hormone, and testosterone) secretion, leading to cell growth, division, and differentiation
- Glycogen hydrolysis, fat synthesis

Additional independent *phospholipase C* effects include these:

- Liver glycogen breakdown
- Pancreatic amylase secretion
- Platelet aggregation

PKA and PKC are both *serine-threonine kinases*, which place phosphates on serine or threonine in target polypeptides. Let's consider tyrosine kinases next.



[318-2 G-Protein Activation of Protein Kinase C and Phospholipase C](#)

17.11.3 Receptor Tyrosine Kinase-Mediated Signal Transduction

The intracellular activity of *tyrosine kinase receptors* is in the cytoplasmic domain (and is part) of the receptor itself. When bound to its effector molecule, these receptor kinases catalyze the phosphorylation of specific tyrosine amino acids in target proteins. While studying the actions of *nerve growth factor (NGF)* and *epidermal growth factor (EGF)* in stimulating growth and differentiation of nerves and skin, Stanley Cohen and Rita Levi-Montalcini discovered the *EGF receptor*, the first *enzyme-linked tyrosine kinase*, and they won the 1986 Nobel Prize in Physiology or Medicine for their discovery! Watch the animation of receptor-kinase signal transduction at the following link (a description is provided in the next few paragraphs).



[319 Receptor-Kinase Signal Transduction](#)

Monomer membrane receptor kinases dimerize when they bind effector ligands, activating kinase domains that catalyze multiple cross-phosphorylation of the monomers. The phosphorylated dimer kinases recruit and phosphorylate yet other cytoplasmic proteins to continue the response pathway. A typical response to EGF and NGF signaling is **mitogenic**. Not surprisingly, the mutations correlated with cancer cells often lie in mitogenic signaling pathways that lead to cell proliferation (growth and division). Cancer-causing genes, or **oncogenes**, were first discovered in viruses. But as humans, we own our own **oncogenes**! For being the first to show that eukaryotic cells were the origin of a chicken retrovirus (the Rous sarcoma virus), J. Michael Bishop and Harold Varmus earned the 1989 Nobel Prize in Physiology or Medicine. *Oncogenes* turn out to be mutations of animal genes for proteins involved in mitogenic signal transduction pathways.

Under normal circumstances, mitogenic chemical signals binding to a receptor tyrosine kinase induces target cells to begin dividing (see the YouTube video link above). One of the end-products of a receptor kinase-mediated phosphorylation cascade is **MAP kinase** (*mitogen-activated protein kinase*). MAP kinase enters the nucleus where it catalyzes the phosphorylation of *transcription factors* and other nuclear proteins which lead to cell proliferation and differentiation, as shown in Figure 17.33.

Some Effects of Map Kinase

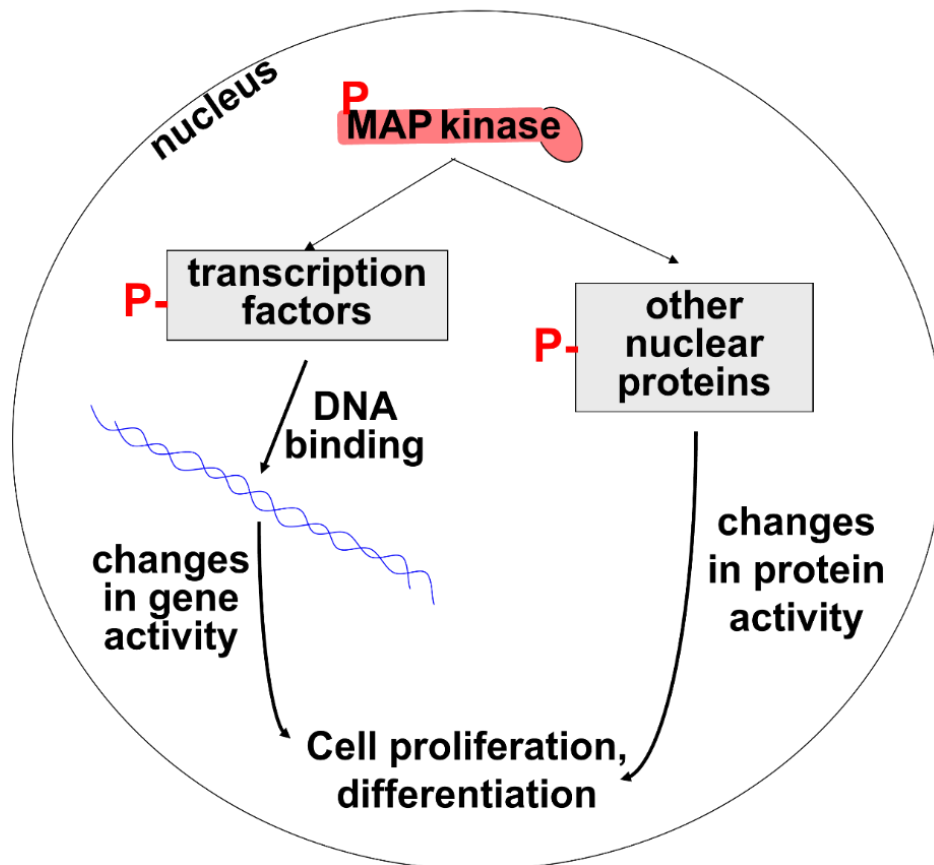


Fig. 17.33: The effects of *MAP kinase* include the phosphorylation and activation of DNA-binding transcription factors and other nuclear proteins, which leads to cell proliferation.

The **Ras** protein-mediated activation of a phosphorylation cascade leading to the MAP kinase is an example of such a signal transduction pathway. It plays a central role in many receptor-kinase signaling events. The *Ras* gene was one of those originally discovered as an oncogene, the mutation of which leads to uncontrolled cell division (i.e., cancer). Ras gene protein activity may in fact be responsible for up to 30% of all cancers!



[320-2 The RAS Oncogene: Its Normal Mitogenic Effects and Cancer](#)

17.12 Signal Transduction in Evolution

We saw that signal transduction typically requires just a few signal molecules interacting with a few cell-surface receptors start a dramatic amplification cascade of enzyme-catalyzed reactions (typically phosphorylations) to get a strong cellular response that will ultimately activate or inactivate target proteins. Amplification cascades can take a single effector-receptor interaction and magnify its effect in the cell by orders of magnitude, making the signaling systems rapid and highly efficient. The range of cellular and systemic (organismic) responses to the same chemical signal is broad and complex. Different cell types can have receptors for the same effector but may respond differently. For example, adrenalin targets cells of liver and blood vessels, among others, with different effects in each. As it happens, adrenalin is also a neurotransmitter.

As organisms evolved, they must have responded to environmental imperatives by adapting and co-opting already existing signaling systems in the service of new pathways. Just as the same signal transduction event can lead to different pathways of response in different cells, evolution has allowed different signal transduction pathways to engage in what has come to be called ***crossstalk***. This occurs when two different signal transduction pathways intersect in the same cell. For example, cAMP produced at the front end of the PKA signaling pathway can activate (or under the right circumstances, inhibit) enzymes in the MAP kinase pathway. These effects result in changes in the levels of active or inactive transcription factors and can therefore modulate the expression of a gene using two (or more) signals. We are only beginning to understand something that looks less like a linear pathway and more like a web of signal transduction.

Some iText & VOP Key Words and Terms

action potential	fight-or-flight	peroxisomes
active transport	flaccid	phagocytosis
adaptin	free energy	phospholipase C
adenylate cyclase	G-protein subunits	phosphorylase kinase
adherens junctions	gap junctions	pinocytosis
adrenaline	gluconeogenesis	PKA
allosteric change regulates transport	GLUT1	PKC
antiport	glycolysis	plasmodesmata
aquaporins	good cholesterol	plasmolysis
bad cholesterol	G-protein-linked receptors	poikilothermic organisms

basal lamina	heat shock protein	potential difference
belt desmosomes	HSP70 protein	protein kinase A
Ca ⁺⁺ ions	hydrophilic corridor	protein kinase C
cadherin	hypertonic	protein packaging
cargo receptor	hypotonic	protein phosphorylation
carrier proteins	IgG light chain	proton gate
Cell adhesion molecules	inositol triphosphate	receptor-mediated endocytosis
Cell-cell attachment	integrin	RER membrane
Cell-cell recognition	ion channels	resting potential
cell-free translation	ion flow	secondary active transporters
channel proteins	ion pumps	serine-threonine kinases
chaperone proteins	IP ₃	signal peptide
cholesterol effects in membranes	isotonic	Signal recognition particle
clathrin	LDL (low-density lipoprotein)	signal sequence
coated pits	Ligand- (chemically) gated channels	signal transduction
coated vesicle	lysosome	smooth endoplasmic reticulum
connexins	MAP kinase	sodium-potassium pump
contact inhibition	mechanically gated channels	solute concentration gradients
contractile vacuole	membrane depolarization	solute transport
COP	membrane hyperpolarization	sorting vesicle
cotransport	membrane invagination	spot desmosomes
coupled transport	membrane potential	stop-transfer sequence
cytoskeleton	microbodies	symport
DAG (diacylglycerol)	mitochondrial membrane contact proteins	tight junction membrane proteins (TJMPs)
diffusion kinetics	mitogenic effects	tight junctions
early endosome	Nerve growth factor	tonoplast
ECM	neurotransmitters	T-SNARE
effector molecules	NGF	turgid
EGF	nuclear envelope	turgor pressure
endocytosis	Nuclear pore fibrils	
endomembrane system	Nuclear transport receptor	

CHAPTER 17 WEB LINKS



17.1



17.2



17.3



Chapter 18

The Cytoskeleton & Cell Motility

Microfilaments, intermediate filaments, and microtubules—roles in cell structure, secretion, cell migration, and organelle and muscle movements

Reminder: For inactive *links*, google key words/terms for alternative resources.

Take a Load off Harry



...Cell Motility at Work

18.1 Introduction

To early microscopists the cell looked like a sack of liquid containing a nucleus. Recall Nikolai Koltsov, the Russian biologist who nearly 35 years before Watson and Crick had already predicted that the genetic material would be a "large genetic molecule" composed of "two mirror-image strands" that would replicate semi-conservatively? Well, in 1903 Nikolai also predicted the existence of a network of tubes in cells to account for cell shape (^{18.1}[The Remarkable Nikolai Koltsov](#)). He called this network a **cytoskeleton**! His prediction predates the mid-twentieth century electron microscope observations of the **cytoskeleton** composed of rods, tubes, and filaments which are responsible for maintaining cell shape and for cell motility. We will see that intracellular structures and organelles are enmeshed in these **microfilaments**, **intermediate filaments** and **microtubules**. To start this chapter, we revisit these structures and more closely examine how they work to support cell shape, motility, and overall organization. Remember that cell motility includes the movement of cells and organisms, as well as the movements of organelles (e.g., vesicles) and other structures inside the cell. Of course, these movements are not random, and they require chemical energy. After a close look at movements in and of cells, we'll look at the interaction of **actin** and **myosin** in **skeletal muscle contraction**. We'll look at a famous **paradox** arising from early studies that showed that ATP was required for muscle contraction *but also* for relaxation. Then we'll look at experiments that resolve the paradox. Animals control skeletal muscle contraction, but some muscles contract rhythmically or with little or no control on the part of the animal—think cardiac muscles of the heart or smooth muscles like those in the digestive and circulatory systems. We focus here on the regulation of skeletal muscle contraction by calcium ions and

regulatory proteins and at responses of *voluntary* (i.e., skeletal) muscle to neural commands. Finally, we'll look at skeletal muscle **elasticity** and its contribution to skeletal muscle function.

Learning Objectives

When you have mastered the information in this chapter, you should be able to do the following:

1. Compare and contrast roles of cytoskeletal structures in different kinds of *cell motility*.
2. Distinguish between the roles of microfilaments, microtubules, and intermediate filaments in the *maintenance and alteration of cell shape* and structure.
3. Consider how and why microtubule length in cilia is stable but dynamic in spindle-fibers.
4. Propose an experiment to ask what part of a *motor protein* (e.g., dynein) has *ATPase activity*.
5. Describe the key experiments that led to the actin-myosin *contraction paradox*.
6. Outline the steps of the *microcontraction cycle* involving myosin and actin.
7. Compare and contrast muscle and flagellar structure and function.
8. Explain why *smooth muscles* do not show striations in the light microscope.
9. Outline the structure of a skeletal muscle, from a whole muscle down to a sarcomere.
10. *Hereditary muscle weakness* involves specific genes and proteins. Propose alternate hypotheses to explain how gene action or a particular protein could cause muscle weakness and suggest how you could test one of your hypotheses.

18.2 Overview of Cytoskeletal Filaments and Tubules

In early microscopy, eukaryotic cells looked like membrane-bound sacs of cytoplasm containing nuclei and assorted organelles. By the late nineteenth century, microscopists were describing the dynamic fibers that accompanied dramatic changes in the structure of dividing cells. No doubt you recognize this as **mitosis**, in which duplicated chromosomes (**chromatids**) materialize just as the nuclear membrane dissolves, and **spindle fibers** form. Over time the fibers seem to pull the chromatids apart to opposite poles of the cell. Spindle fibers turn out to be bundles of **microtubules**, each of which is a polymer of **tubulin** proteins. Let's look again at a fluorescence micrograph of a mitosing **metaphase** cell (Figure 18.1). Most of the cell, other than what is fluorescing, is not visible in the micrograph.

Chromatids Aligned on Spindle Apparatus in Metaphase of Mitosis

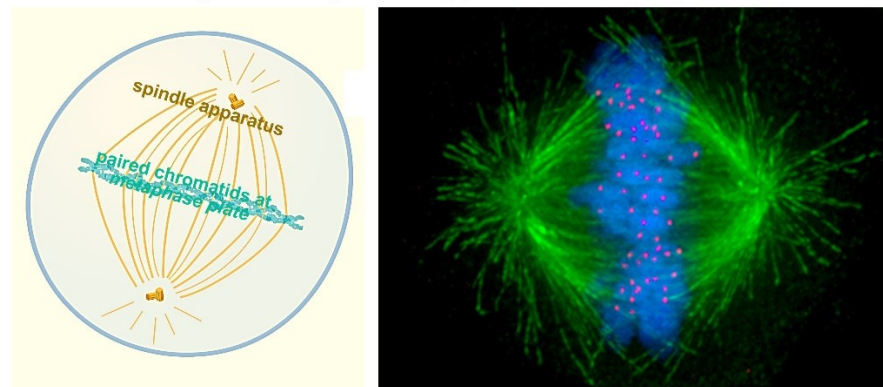


Fig. 18.1: Drawing (left) and fluorescence micrograph (right) of a cell in metaphase of mitosis: aligned chromosomes (chromatids) at the center of the cell (blue in the micrograph) are just about to be pulled apart by microtubules of the spindle apparatus (green) extending from the poles to the center of the cell.

To get the micrograph in Fig. 18.1, *antibodies* were made against purified microtubule, kinetochore, and chromosomal proteins (or DNA), and then linked to different **fluorophores** (organic molecular fluorescent tags). When the tagged antibodies were added to dividing cells in metaphase, they bound to their respective fibers. In a fluorescence microscope, the *fluorophores* emit different colors of visible light. Microtubules are green, metaphase chromosomes are blue, and kinetochores are red in the micrograph. Both mitosis and meiosis are very visible examples of movements *within* cells.

As for muscle movement in whole organisms, early-to-mid-twentieth-century studies asked what the striations (stripes) seen in skeletal muscle by light microscopy might have to do with muscle contraction. The striations turned out to be composed of an isolable protein complex that investigators called **actomyosin** (*acto* for "active"; *myosin* for "muscle"). Electron microscopy later revealed that actomyosin (or **actinomyosin**) is composed of thin filaments (**actin**) and thick filaments (**myosin**) that slide past one another during muscle contraction.

Electron microscopy also hinted at a more complex cytoplasmic structure of cells in general. The **cytoskeleton** consists of fine rods and tubes, in more or less organized states, that permeate the cell and in which organelles are embedded. As noted, the most abundant of these are **microfilaments**, **microtubules**, and **intermediate filaments**. Though myosin is less abundant, it is nonetheless present in nonmuscle cells. Microtubules account for chromosomal movements of mitosis and meiosis; and together with microfilaments (i.e., actin) they enable organelle movement inside cells. (You may have seen **cytoplasmic streaming** of *Elodea* chloroplasts in a biology lab exercise.) Microtubules also underlie the movements of the **cilia** and **flagella**, which power the movement of whole cells like paramecia, amoebas, phagocytes, and the like. Actin and myosin enable muscle contraction and thus, higher animal movement.

Finally, the cytoskeleton is a dynamic structure. Its fibers not only account for the movements of cell division but also give cells their **shape** and **mechanical strength**. All the fibers can disassemble, reassemble, and rearrange, allowing cells to change shape. These changes range from the *pseudopodia* extended by amoeboid cells to the spindle fibers that stretch cells in mitosis and meiosis, to the constriction of a dividing cell that eventually pinches off daughter cells, and more! In this chapter we look in some detail at the roles of these tubules and filaments in cell structure and in different forms of cell motility.

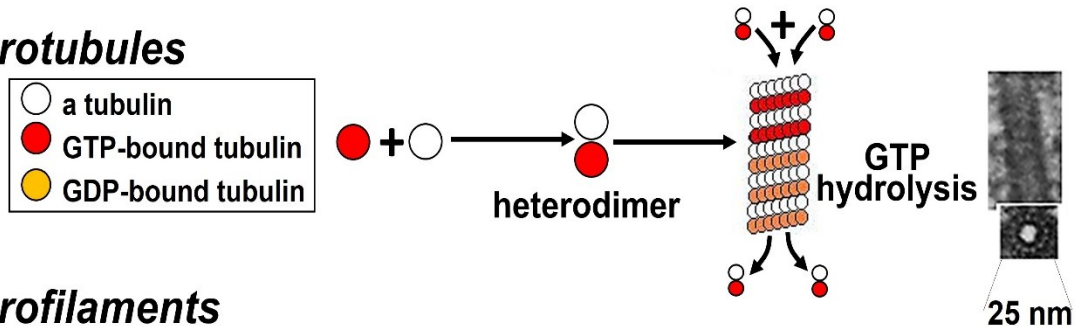
18.3 Molecular Structure and Organization of Cytoskeletal Components

Of the three main cytoskeletal fibers, intermediate filaments serve a mainly structural role in cells. Microtubules and microfilaments have dual functions, dynamically maintaining cell shape and enabling cell motility. For example, when attached to the plasma membrane, actin microfilaments maintain cell shape. And when they interact with **motor proteins** (e.g., *myosin*), they can pull or push against a muscle-cell membrane, changing the shape of the cell. Likewise, motor proteins such as *dynein* and *kinesin* can move *cargo* back and forth along microtubule tracks from one point to another in the cell.

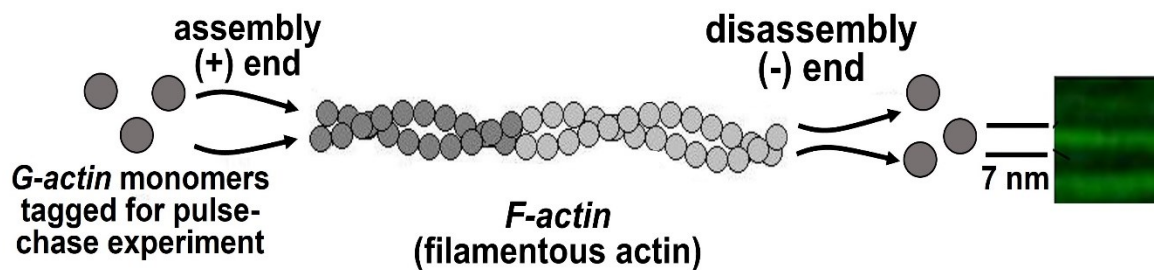
We will look at how motor proteins interact with microtubules and microfilaments shortly. For now, let's take another look at the drawings and micrographs of the three main cytoskeletal filaments of eukaryotic cells, which we also saw earlier in the text (Figure 18.2, below).

Structure, Polarity & Assembly of Cytoskeletal Elements

Microtubules



Microfilaments



Intermediate Filaments (IFs)

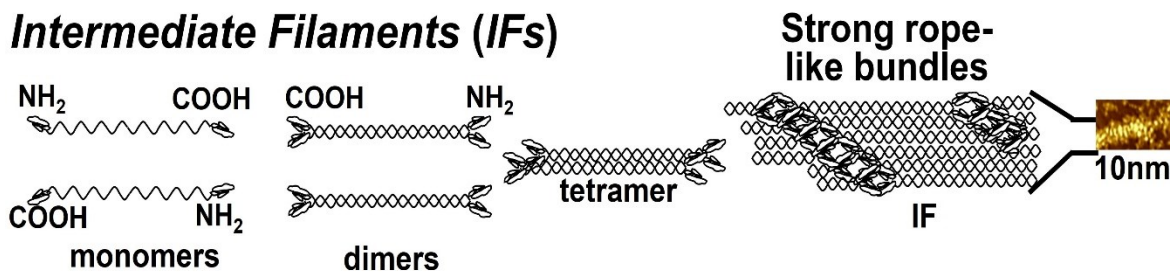


Fig. 18.2: The assembly, structure, and polarity of microtubules, microfilaments (actins), and intermediate filaments, alongside immunofluorescence micrographs made using fluorescent antibodies against isolated microtubule (top), microfilament (middle), and intermediate-filament (bottom) proteins.



[321-2 Introduction to the Cytoskeleton](#)

We also saw earlier the location and general functions of microtubules, microfilaments, and intermediate filaments. Let's look at those fluorescence micrographic localizations of fibers in cells, shown again in Figure 18.3 (below). They are consistent with known functions of the major cytoskeletal component filaments in cell structure and motility. Despite the small size of prokaryotic cells, they too were recently found to have previously unsuspected cytoplasmic structures that could function as a cytoskeleton. For more about such cytoplasmic structures, see ^{18.2} [A Prokaryotic Cytoskeleton](#). So perhaps not just eukaryotic cells, but *all* cells are more than an unorganized bag of fluid sap!

Cellular Localization of Cytoskeletal Fibers

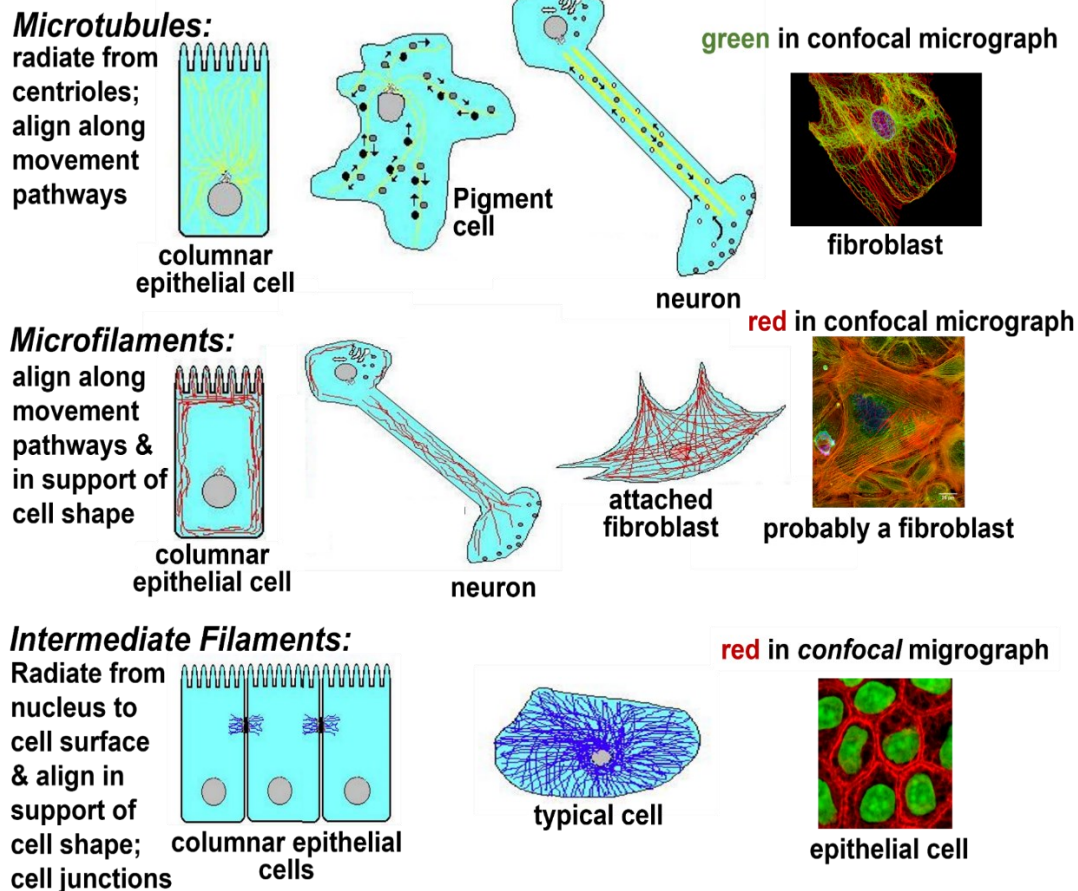


Fig. 18.3: Illustration and immunofluorescence and confocal microscope localization of microtubules, microfilaments (actins), and intermediate filaments in cells.



[322-2 Microtubules, Microfilaments, and Intermediate Filaments in Cells](#)



Next, we'll consider the specific roles of microtubules, microfilaments, intermediate filaments, and related proteins in the structure of the eukaryotic cytoskeleton and how cytoskeletal interactions enable cell motility.

18.4 Microtubules: Dynamic Structures Composed of Tubulin Monomers

Microtubules assemble from dimers of α -tubulin and β -tubulin monomers. After their formation, α/β -tubulin dimers add to a growing, or *plus end* (*+end*), fueled by *GTP* hydrolysis (see Figure 18.2). Disassembly at the *-end* of microtubules powers the changes in cell shape or in the separation and movement of chromatids to opposite poles of cells during mitosis or meiosis.

Isolated single microtubules were shown to grow by addition to one end and to disassemble at the opposite end, thus distinguishing the *+ends* and *-ends*. Find a summary of evidence for microtubule *polarity* in the following link.



[323-2 Demonstration of the Polarity & Dynamics of Microtubules](#)

Microtubules in most cells (other than when they are dividing) can seem disordered. In nondividing (*interphase*) animal cells, they tend to radiate from *centrioles* without forming discrete structures. But as cell division approaches, microtubules reorganize to form spindle fibers. The reorganization is *nucleated* from *centrioles* in animal cells and from a more amorphous *MicroTubule Organizing Center (MTOC)* in plant cells. A typical centriole (or *basal body*) has a "nine-triplet" microtubule array, as seen in electron-micrograph cross sections (Figure 18.4).

Cross Section of 9-Triplet Microtubule Ring, Typical of a Centrioles and Basal Bodies

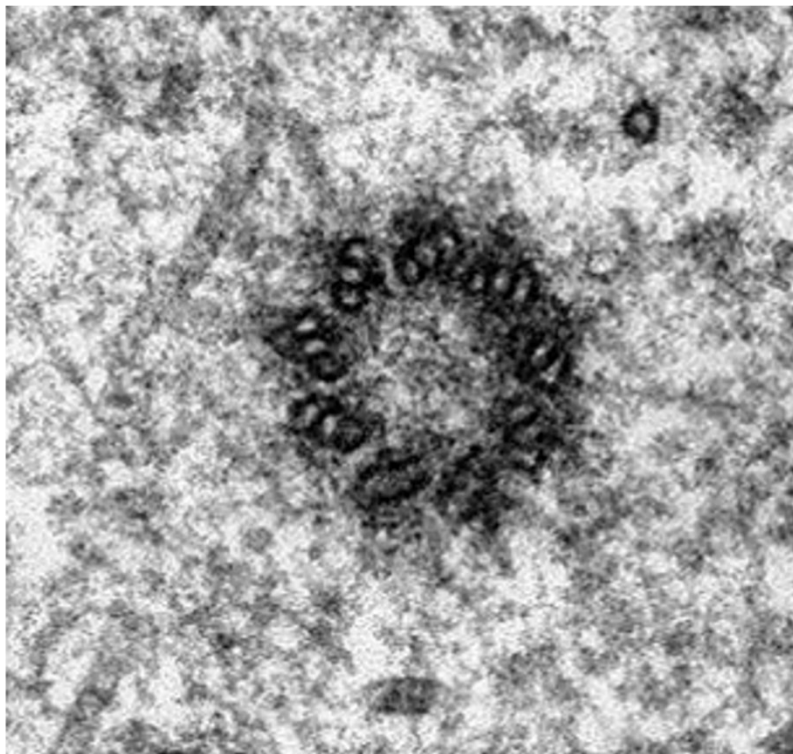


Fig. 18.4. Transmission-electron micrograph of the characteristic "nine triplet microtubule" array of *centrioles* and *basal bodies*.

18.4.1 The Two Kinds of Microtubules in Spindle Fibers

The spindle fibers of mitosis and meiosis are made up of *kinetochore microtubules* and *polar microtubules*. The former use chemical energy to pull duplicated chromatids apart; the latter use chemical energy to separate daughter cells during *cytokinesis*.

18.4.1.a Kinetochore Microtubules

Duplicated chromosomes condense in prophase of mitosis and meiosis, forming visible paired **chromatids** attached at their **centromeres**. Specific proteins associate with centromeres to make a **kinetochore** during condensation. As the spindle apparatus forms, some spindle fibers attach to the kinetochore; these are the **kinetochore microtubules**. By **metaphase**, bundles of kinetochore microtubules stretch from the kinetochores at the cell center to the centrioles at opposite poles of a dividing animal cell, as drawn in Figure 18.5.

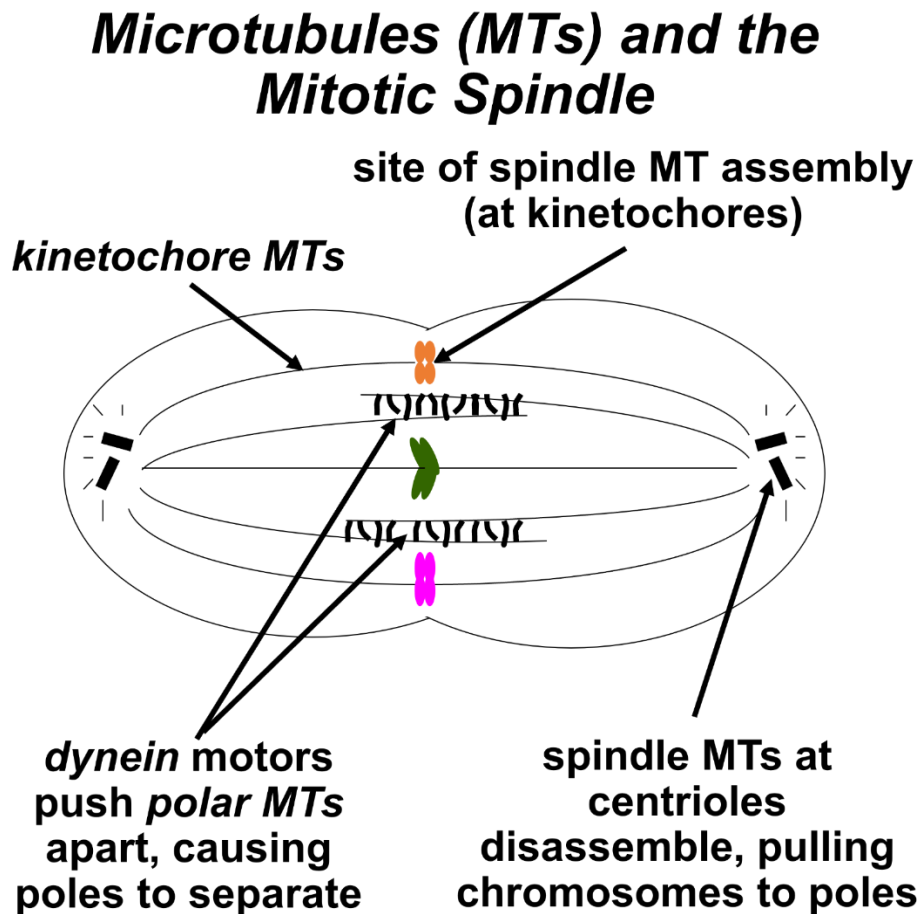


Fig. 18.5: As a cell starts its mitotic anaphase, spindle-fiber microtubules exert forces that separate and pull chromatids apart and push the poles of the cell apart.

The **+ends** of kinetochore microtubules are in fact at the kinetochores, where these fibers assemble! During **anaphase**, kinetochore microtubules disassemble and shorten at their **-ends** (at the centrioles in animals or at the MTOCs in plant cells). The forces generated by these activities pull against the centromeres of the chromatids, separating them and then, drawing what are now daughter chromosomes to the opposite poles of the dividing cell.

The role of kinetochore microtubule disassembly at the centrioles (i.e., at their **-ends**) was shown in a clever experiment in which a tiny laser beam was aimed into a cell at the spindle fibers attached to the kinetochore of a pair of chromatids. (See Figure 18.6 below as well as an animation of these events at the link that follows.)

The Dynamic Mitotic Spindle

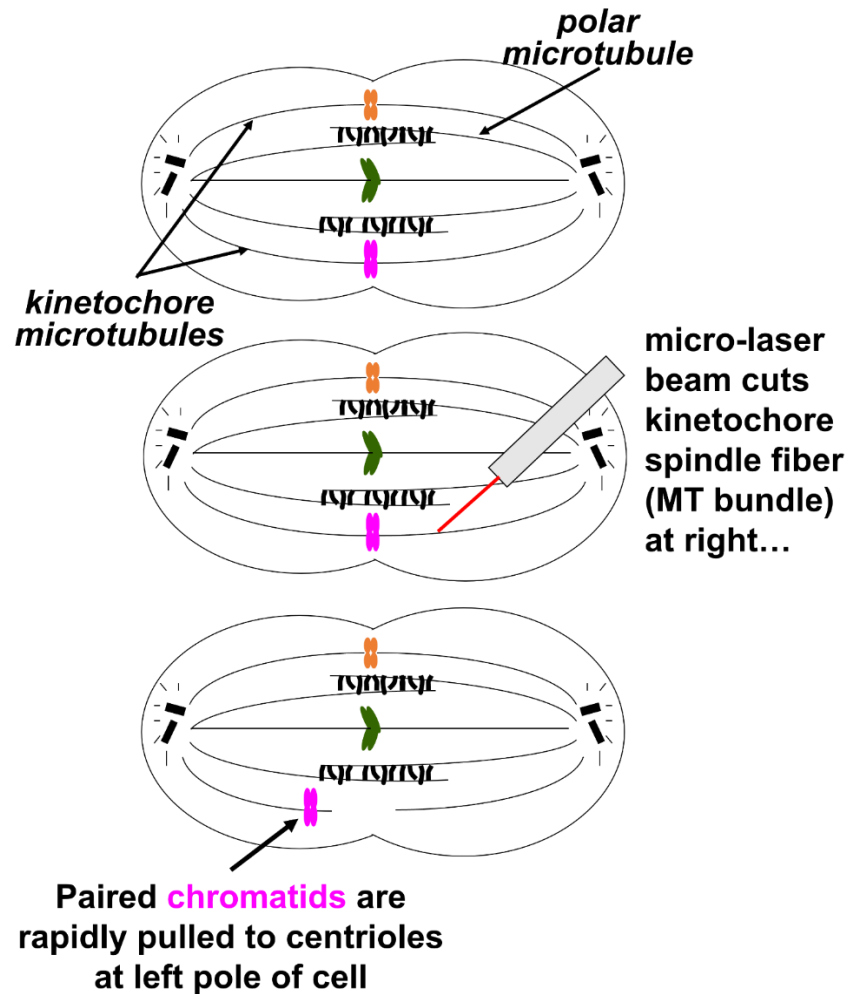


Fig. 18.6: Disrupting a kinetochore spindle fiber demonstrates the presence of a strong force from microtubule disassembly, which quickly pulls a pair of chromatids toward the centrioles.



[324 Spindle-Fiber Microtubules Generate Force on Chromatids](#)

18.4.1.b Polar Microtubules

Spindle-fiber **polar microtubules** extend from centrioles/MTOCs at opposite poles, toward the center of dividing cells. But instead of binding to kinetochores, they overlap at the center of the dividing cells. While the kinetochore microtubules are tugging apart the paired chromatids at the center of the cell, the **polar microtubules** are sliding past one another in opposite directions, pushing apart the poles of the cell. In this case, **dynein motor proteins** attached to microtubules (illustrated in Figs. 18.5 and 18.6) hydrolyze ATP to power microtubule sliding. The dynein motors on the microtubules at one pole of the cell in effect “walk” along overlapping microtubules extending from the opposite pole.

18.4.2 Microtubules in Cilia and Flagella

The microtubules of cilia or flagella emerge from a *basal body*, shown in the electron micrograph at the left in Figure 18.7.

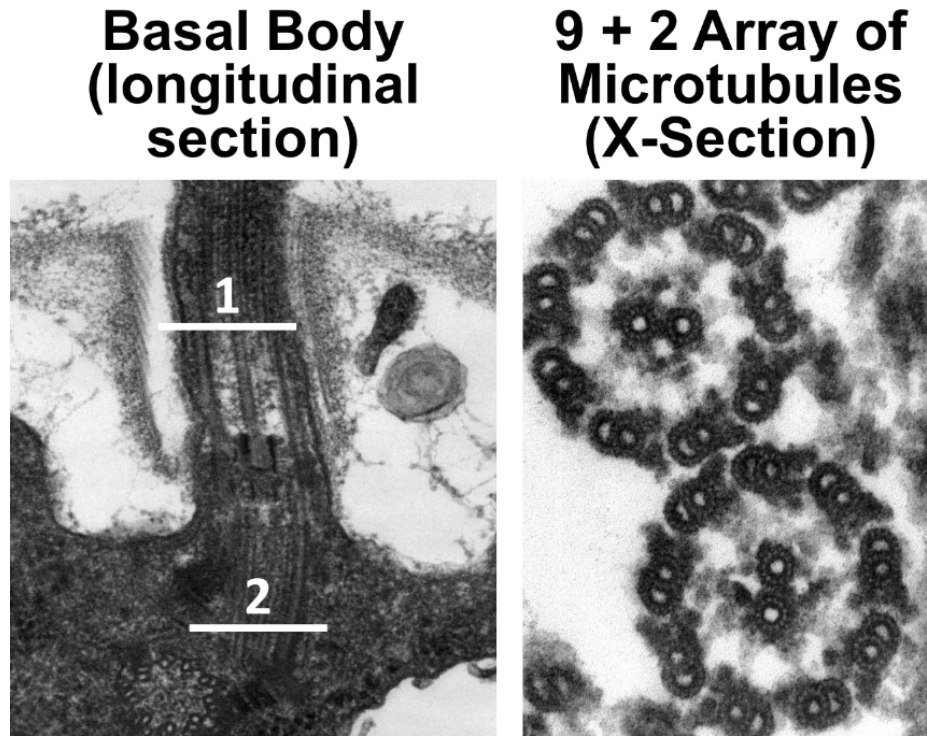


Fig. 18.7: Transmission-electron micrographs of a basal body (left) and of cross sections of a pair of 9 + 2 microtubule arrays (right), each of which would emerge from a basal body (white line #1). The position of the 9-triplet microtubule array of microtubules seen in Fig. 18.4 is indicated by white line #2.

Basal bodies, organized as *nine-triplet* microtubule rings, are structurally similar to centrioles. It turns out that eukaryotic flagella and cilia can be stripped from their cells in a very high-speed blender (not the kind you may have in your kitchen!). Treatment with detergents dissolves their membranes, leaving behind *axonemes* with a 9 + 2 array of microtubules. While the formation of cilia and flagella begins at the basal bodies, both structures soon show a typical **9 + 2 arrangement** (i.e., nine outer *doublets* plus two central *singlet* microtubules) seen in the cross section at the right in the micrograph). Axonemes from cilia and flagella are virtually identical.

Microtubules and axonemes are shown in greater detail in Figure 18.8 (below). In the inset at the upper left, it is possible to see the tubulin subunits that make up a microtubule polymer in cross section. Each tubule is made up of a ring of thirteen tubulin subunits. In the high-magnification, high resolution cross section micrograph of an axoneme at the right in Figure 18.8, the microtubules in the *doublets* share tubulins but are also composed of thirteen tubulins. When fully formed, the 25 nm diameter microtubules appear to be hollow cylinders. When isolated, they typically come along with the dynein motor proteins and other *Microtubule-Associated Proteins (MAPs)*, indicated in the micrograph on the right (Figure 18.8). These proteins hold microtubules together in an axoneme and play a role in motility.

Microtubules and Axonemes

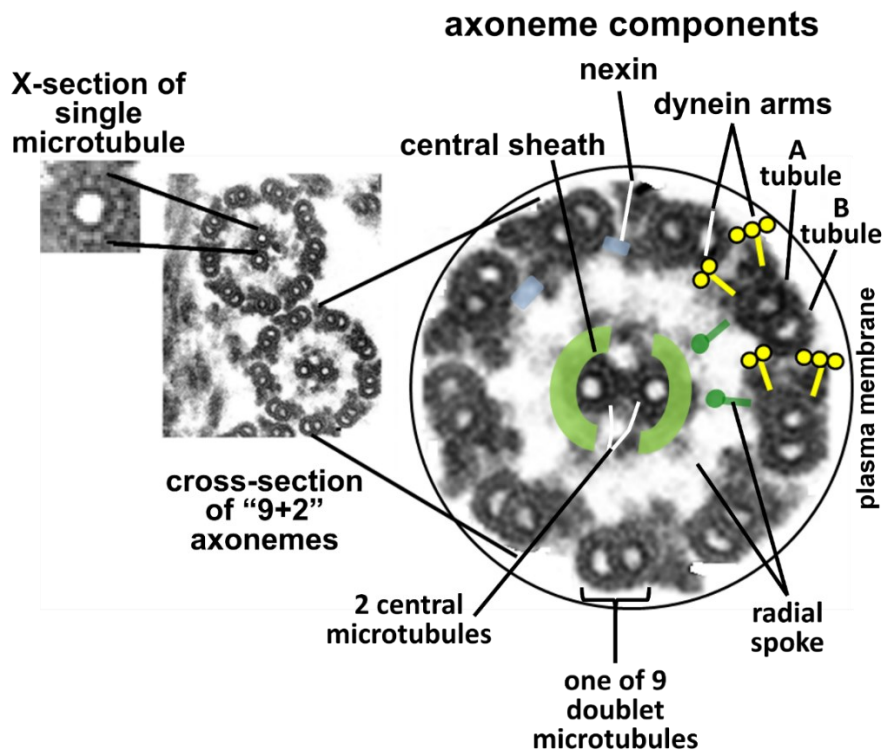


Fig. 18.8: Transmission-electron micrographs of *axonemes* show that a microtubule is made up of a ring of thirteen tubulins (left) and that those microtubules in cilia or flagella are arranged in typical 9 + 2 arrays (middle). The enlarged cross section illustrates specific microtubule-associated proteins (*MAPs*) that maintain the 9 + 2 structure, and it shows the dynein motors that power motility.

When fully formed, the 25 nm diameter microtubules appear to be hollow cylinders. When isolated, they typically come along with the dynein motor proteins and other *MAPs* labeled in the micrograph at the right. These proteins hold microtubules together in an axoneme and play a role in motility.

18.4.3 Microtubule Motor Proteins Move Cargo from Place to Place in Cells

Motor proteins, such as dynein and kinesin, are *ATPases*; they use the free energy of ATP hydrolysis to power intracellular motility. Let's take a closer look at how these two major motor proteins carry *cargo* from place to place inside of cells. Organelles are a typical *cargo*. Examples include *vesicles* formed at the *trans-Golgi face*, which contain secretory proteins, pigments, or neurotransmitters. *Secretory vesicles* move along microtubule tracks to the plasma membrane for *exocytosis*. Vesicles containing *neurotransmitters* move from the cell body of neurons, along microtubule tracks in the axons, and reach the nerve ending where they become *synaptic vesicles*. In a chameleon, *pigment vesicles* in skin cells disperse or aggregate along microtubule tracks to change skin color to match the background.

Vesicle transport in neurons is well understood. Neurotransmitter vesicles arise from the endomembrane system in neuron *cell bodies*. ATP-dependent *kinesin* motor proteins

power **anterograde** movement of the vesicles (from the cell body to the nerve endings). In contrast, ATP-dependent **dynein** motors (part of a **dynactin** complex) power **retrograde** movement of empty vesicles back to the cell body. Motor protein structure and action are shown in Figure 18.9.

Motor Proteins Move Cargo on Microtubules in Axons

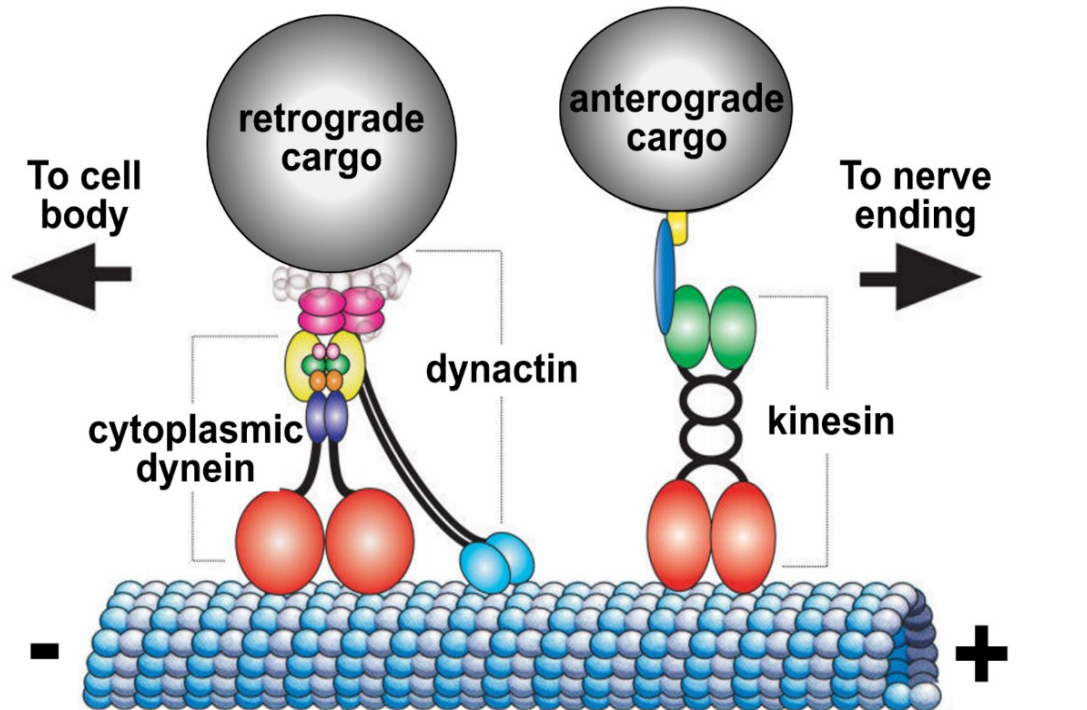


Fig. 18.9: Roles of *dynein* and *kinesin* in retrograde (backward) and anterograde (forward) movement of cargo vesicles along microtubules.



[325-2 Microtubule Motor Proteins](#)

A fanciful (and *not too inaccurate!*) animation of a motor protein in action on an axonal microtubule is at this link: ^{18.3}[Kinesin 'Walks' An Organelle Along a Microtubule](#). Next let's look at some elegant studies of isolated axonemes and see what they tell us about microtubule-based cell motility.

18.4.4 Demonstrating Sliding Microtubules

Experiments on *axonemes* that have been isolated from demembranated cilia or flagella confirm the sliding-microtubule mechanism of ciliary and flagellar motility. For example, adding ATP to detached cilia or flagella makes them beat in much the same way as they do when attached to their cells. The phenomenon is easily seen in a light microscope. But isolated axonemes (with their original 9 + 2 microtubule arrangement) also “beat” (after a fashion!) in the presence of ATP! The experiment is illustrated in Figure 18.10 (below).

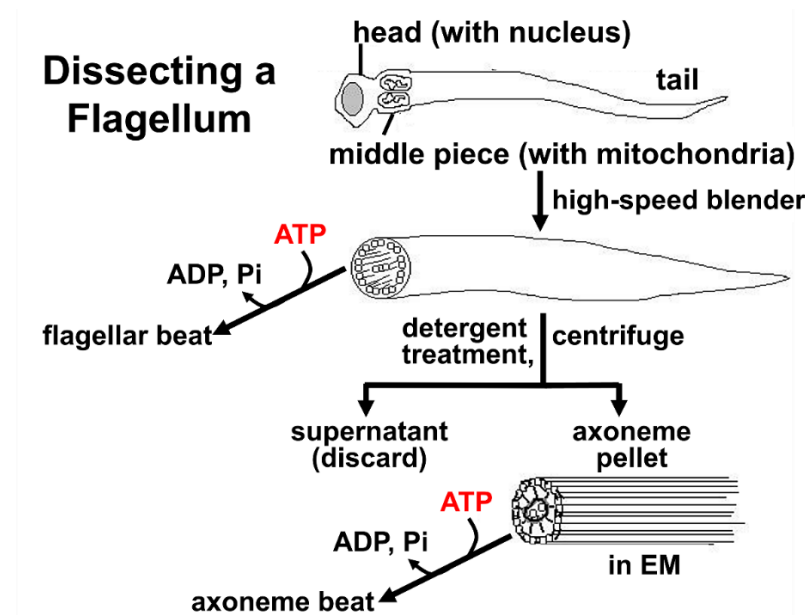


Fig. 18.10: Removing membranes from isolated cilia or flagella leaves behind the *axoneme*. In the presence of ATP, the isolated cilia, the flagella, and the axonemes will “beat” as they hydrolyze the ATP.



[326-2 9+2 The Microtubule Array in Axonemes that Beat](#)

Selective addition of different detergents removes *radial spokes*, *nexin*, and other proteins from the axoneme, causing the microtubules to separate. Dissociated microtubule doublets and central “singlets” can then be observed in the electron microscope. When such separated microtubules are dialyzed to remove the detergents, doublet microtubules re-associate to form sheets, as shown in Figure 18.11.

Separation and Reconstitution of Axoneme Microtubule Doublets

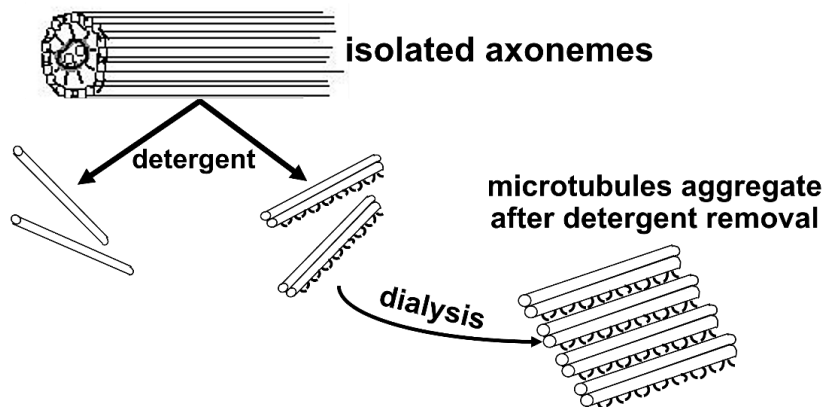


Fig. 18.11: Detergents can dissociate axonemes into individual and doublet microtubules. Dialysis of separated microtubules to remove the detergent will cause the microtubules to re-associate (i.e., to *reconstitute*) into a sheet, with connections resembling those seen in intact axonemes.

ATP added to *reconstituted* microtubule doublets causes the microtubules to separate as the ATP is hydrolyzed. When such preparations are fixed for electron microscopy *immediately after* adding the ATP, the microtubules are caught in the act of sliding. See this animated in the following link (#327-2).



[327-2 Proof of Sliding Microtubules in Flagella and Cilia](#)



[328 Bacterial Flagella are Powered by a Proton Gradient](#)



[329 The Effects of Different Drugs on Microtubules...and Cancer](#)



18.4.5 The Motor Protein Dynein Enables Axonemes to Bend

Take another look at the cross section of axonemes in Figure 18.9. In the 9 + 2 *axoneme* of cilia and flagella, dynein arms attached to the **A tubules** of the outer doublets walk along the **B tubules** of the adjacent doublet. If the doublets on one side of an axoneme take a walk while those on the other side hold still, the microtubules will slide past one another, and the axoneme (and therefore a cilium or flagellum) will bend. This microtubule sliding is constrained by flexible *nexin* and *radial spoke* attachments. Figure 18.12 compares the movements of cilia and flagella.

The Motion of Cilia and Flagella

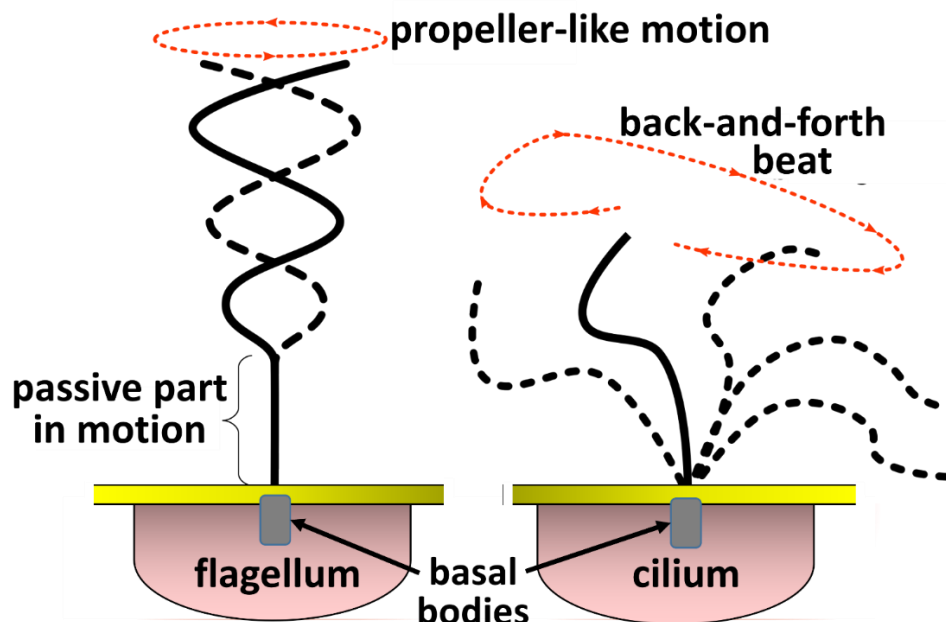


Fig. 18.12: Comparison of the ciliary beat with the propeller-like flagellar wave motion.

The differences in flagellar motion (wave-like propeller) and ciliary motion (back-and-forth beat in a single plane) result in part from which microtubules are sliding at a given moment and the nature of their restraint by axoneme proteins.

18.5 Microfilaments—Structure and Role in Muscle Contraction

At 7 nm in diameter, **microfilaments** (actin filaments) are the thinnest cytoskeletal component. Globular actin (**G-actin**) monomers polymerize to form linear **F-actin** polymers. Two polymers then combine to form a twin-helical actin microfilament. As with microtubules, microfilaments have a **+end**, to which new actin monomers are added to assemble *F-actin*, and a **–end**, at which they disassemble when they are in a dynamic state, such as when a cell is changing shape. When one end of a microfilament is anchored to a cellular structure, (e.g., to **plaques** in the cell membrane, motor proteins like myosin can use ATP to generate a *force* that deforms the plasma membrane and, thus, the shape of the cell. One of the best-studied examples of myosin/actin interaction is in skeletal muscle, where the sliding of highly organized thick myosin rods and thin actin microfilaments results in muscle contraction.

18.5.1 The Thin (Micro) and Thick Filaments of Skeletal Muscle

Bundles of parallel muscle cells make up a skeletal muscle. Thin sections of skeletal muscle cells, called **myocytes**, appear **striated** in the light microscope (Figure 18.13).

Striated Skeletal Muscle

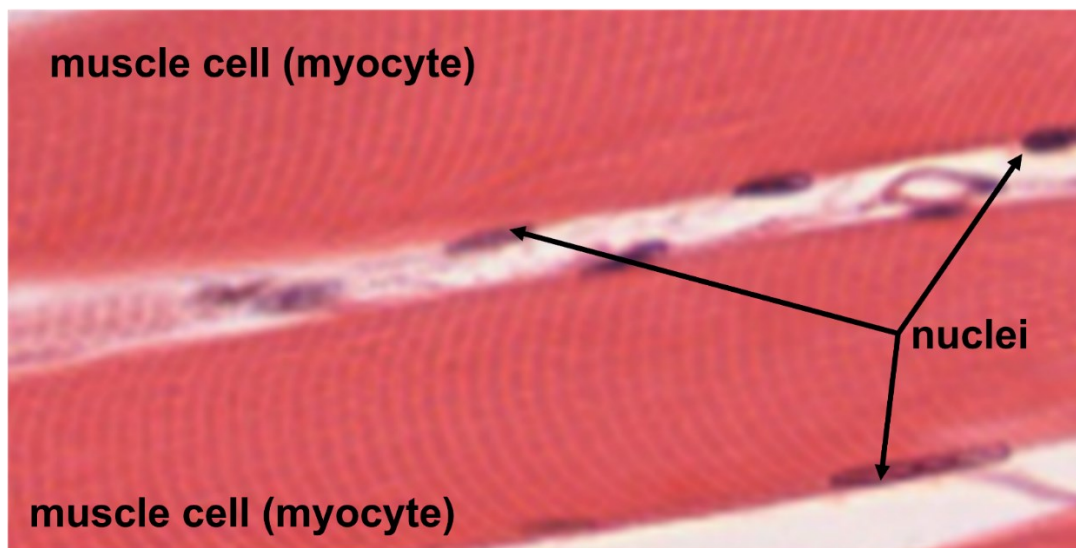


Fig. 18.13: Light micrograph of skeletal muscle, stained to show characteristic striations.

During development, myoblast cells (muscle cell precursors) fuse to form a myocytes (myofibers). The dark structures surrounding the striations are the multiple nuclei that end up in the mature myocyte. Multinucleate cells resulting from such fusions are called syncytia (singular: syncytium). Each syncytial myocyte also contains many *mitochondria* to provide ATP to fuel contraction. Skeletal muscle is made up of aligned, bundled myocytes (the *myofibers*), which are in turn, further organized into fascicles, which are finally bundled to form a muscle.

Figure 18.14 shows the anatomical organization and fine structure of a muscle.

Gross Anatomy of Skeletal Muscle and Structure of a Sarcomere

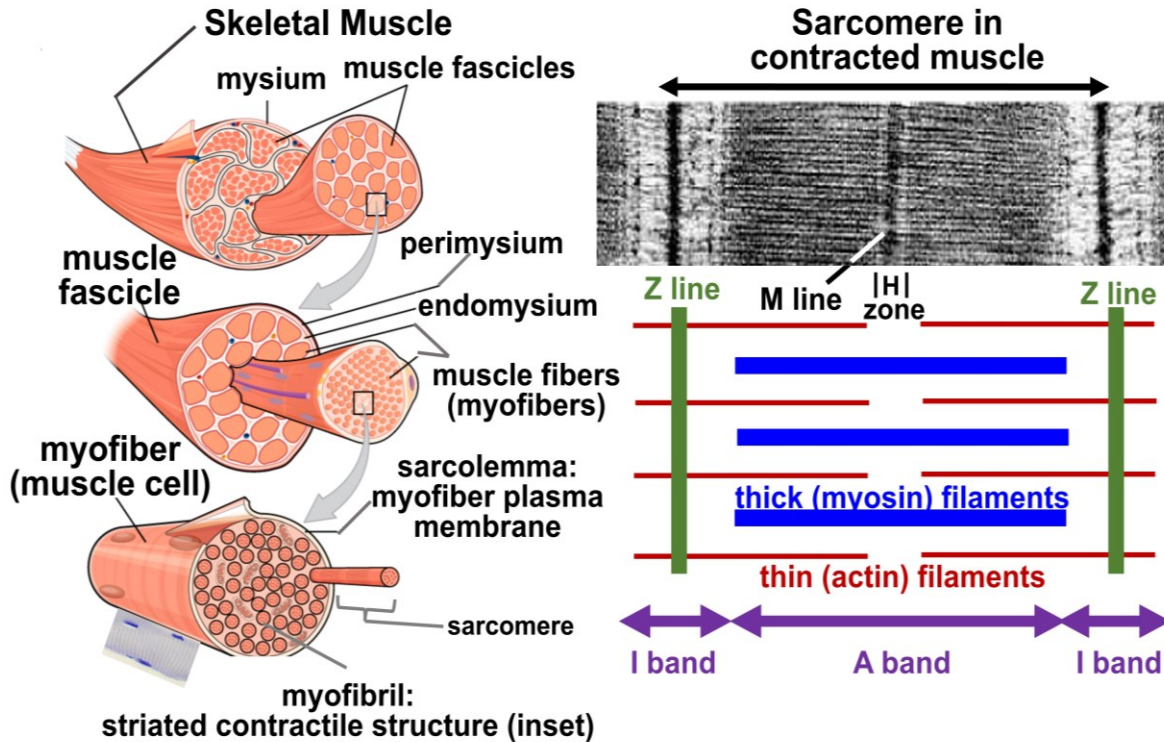


Fig. 18.14: Skeletal muscle organization and the anatomy of a muscle cell sarcomere.

Light passing through the more or less ordered regions of the sarcomere will bend to different degrees because each region has a different *refractive index*. Polarizing light microscopy detects these differences, enhancing the contrast between the regions of the sarcomere and defining them as either *isotropic* (i.e., with a low refractive index) or *anisotropic* (with a high refractive index). High-resolution electron microscopy from the 1940s revealed the fine structure of skeletal muscle (right panel). Paired, dark, vertical *Z-lines* define a *sarcomere* in the electron micrograph. In the illustration of the myofiber (lower left) the *Z-lines* (shown in gray) are aligned in register in bundled myofibers as well as in the fascicles (upper left); these are what appear as the striations characteristic of skeletal muscle seen in Figure 18.13. Based on light and electron microscopy, we can define regions within a sarcomere:

- The *A band (anisotropic band)* of overlapping, aligned actin and myosin filaments, which runs down the middle of the sarcomere, is more ordered than the I-band, so it has a higher refractive index.
- The *I-band (isotropic band)* has a low refractive index compared to the *A band*. It is largely made up of thin (actin) microfilaments.
- The paired *Z-lines* demarcate the *sarcomere* (Z for *zwischen*, German for *between*).
- The *H zone* is a region where myosin does not overlap actin filaments.
- An *M-line* lies at the center of the H zone.

18.5.2 The Sliding Filament Model of Skeletal Muscle Contraction

Electron microscopy of relaxed and contracted muscle is consistent with the sliding of thick and thin filaments during contraction (Figure 18.15).

Sarcomeres from *Relaxed* and *Contracted* Skeletal Muscle

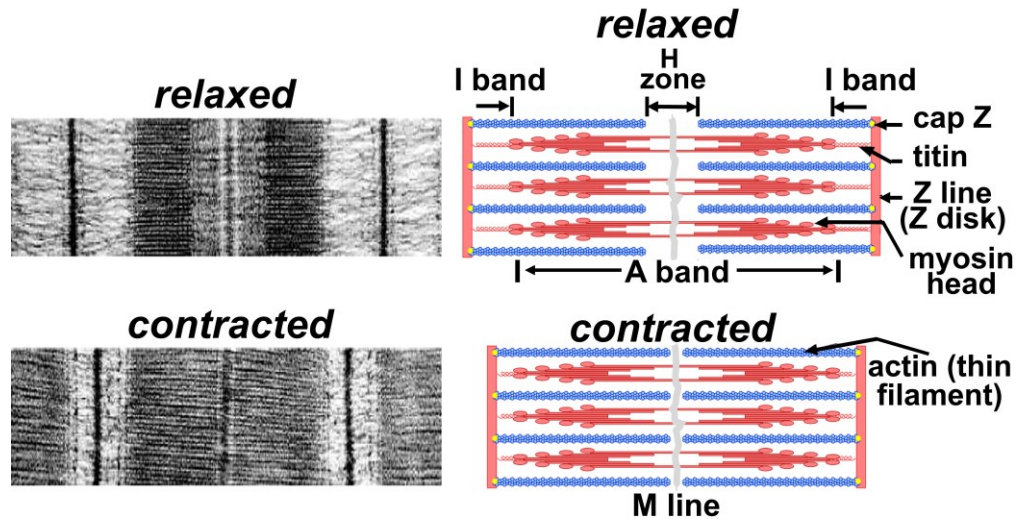


Fig. 18.15: The transmission-electron micrograph and its corresponding illustration show a sarcomere shortening, consistent with the sliding of filaments during skeletal muscle contraction.

Additional key structures of the sarcomere can be seen in the interpretive illustration at the right. Note that in the sarcomeres of a contracted muscle cell, the H zone has virtually disappeared. The width of the A band (essentially, the length of the myosin filaments spanning the sarcomere) has not changed after contraction, but the length of the I-bands from the Z-lines has decreased, bringing the Z-lines of the sarcomere closer. The best explanation here was the *Sliding Filament Hypothesis* (or *model*) of skeletal muscle contraction.



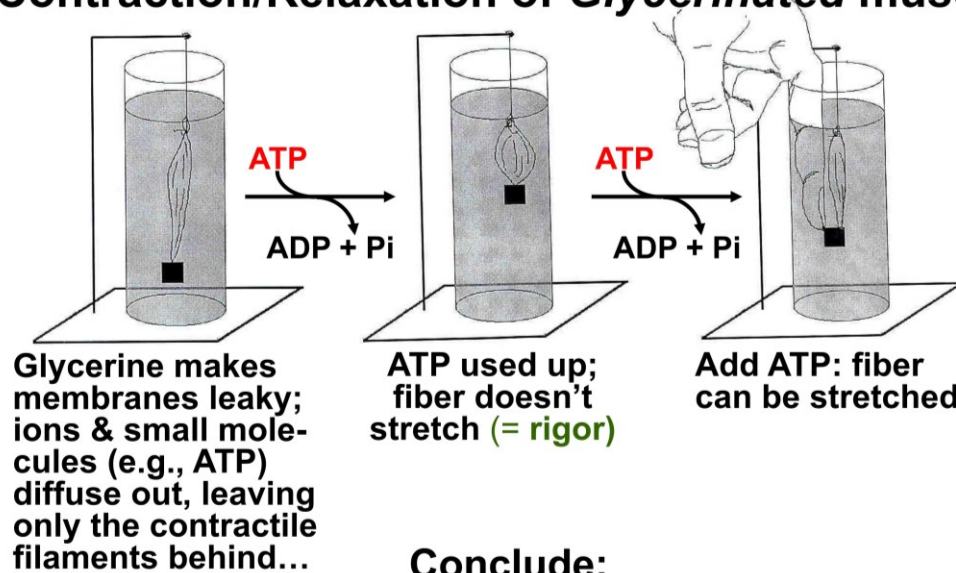
[330-2 The Sliding Filament Model of Skeletal-Muscle Contraction](#)

18.5.3 The Contraction Paradox: Contraction and Relaxation Require ATP

The role of ATP in fueling the movement of sliding filaments during skeletal muscle contraction was based in part on experiments with *glycerinated fibers*. These are muscle fibers that were soaked in glycerin, which permeabilizes the plasma membrane, allowing the soluble components to diffuse out of the muscle cells, but leaves intact most of the sarcomere-based contractile structure, as seen in electron micrographs.

Investigators found that if ATP and calcium were added back to glycerinated fibers, the ATP was hydrolyzed, and the fibers could still contract—and even lift a weight! Contraction of a glycerinated muscle fiber in the presence of ATP is illustrated below in Figure 18.16.

Contraction/Relaxation of *Glycerinated* muscle



Conclude:

- ATP hydrolysis is required for contraction
- ATP is required to stretch muscle (i.e., for relaxation)

Fig. 18.16: The skeletal muscle *contraction Paradox*: When given ATP, *glycerinated* muscle fibers contract and pull a weight. When all the ATP is hydrolyzed, the fiber can't stretch unless more ATP is added.

Assays showed that when all the added ATP had been hydrolyzed, the muscle remained contracted. It would not relax, even with the weight it had lifted still attached. Attempts to manually force the muscle back into its relaxed position didn't work. But the fiber could be stretched when fresh ATP was added to the preparation. Moreover, if the experimenter let go immediately after stretching the fiber, it would again contract and lift the weight. A cycle of forced stretching and contraction could be repeated until all of the added ATP was hydrolyzed. At that point, the fiber would again no longer contract—or if contracted, it could no longer be stretched.

The contraction paradox, then, was this: ATP hydrolysis is required for muscle contraction *as well as* for relaxation (stretching). The paradox was resolved when the functions of the molecular actors in contraction were finally understood. Here we review some of the classic experiments that led to this understanding.



[331 The Contraction Paradox](#)



18.6 Actin-Myosin Interactions In Vitro: Dissections and Reconstitutions

Several experiments hinted at the interaction of actin and myosin in contraction. For example, *actomyosin* was first observed as the main component of viscous skeletal muscle homogenates. Under appropriate conditions, adding ATP to such *actomyosin* preparations caused a decrease in viscosity. However, after the added ATP was hydrolyzed, the mixture became viscous again. Extraction of the nonviscous preparation (before it recondensed and the ATP was consumed)

led to the biochemical separation of the two main substances we now recognize as the **actin** and **myosin** (*thin* and *thick*) filaments of contraction. What's more, adding these components back together reconstituted the viscous *actomyosin* (now renamed **actinomyosin**). Adding ATP once again to the reconstituted solution eliminated its viscosity. The ATP-dependent viscosity changes of actinomyosin solutions were consistent with an ATP-dependent separation of thick and thin filaments. Do actin and myosin also separate in glycerinated muscles exposed to ATP, allowing them to stretch and relax? This question was answered with the advent of electron microscopy. The purification of skeletal muscle myosin from actin (still attached to Z-lines) is cartooned in Figure 18.17, showing what the separated components looked like in the electron microscope.

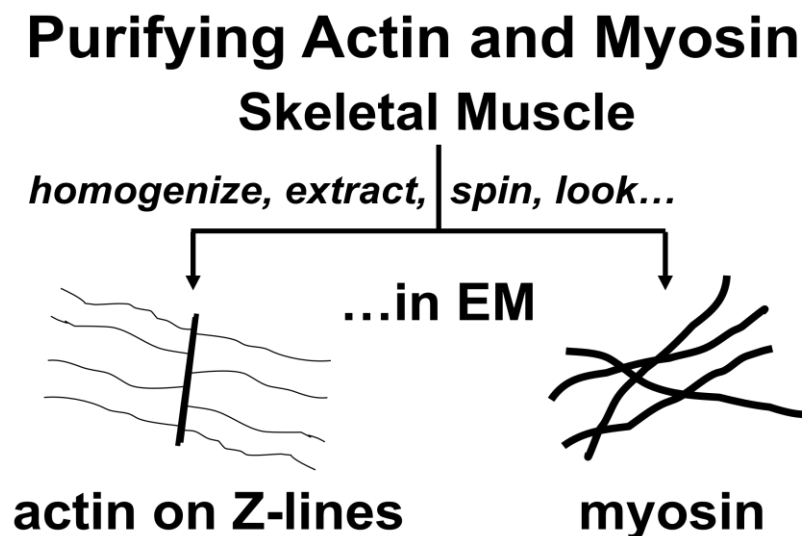
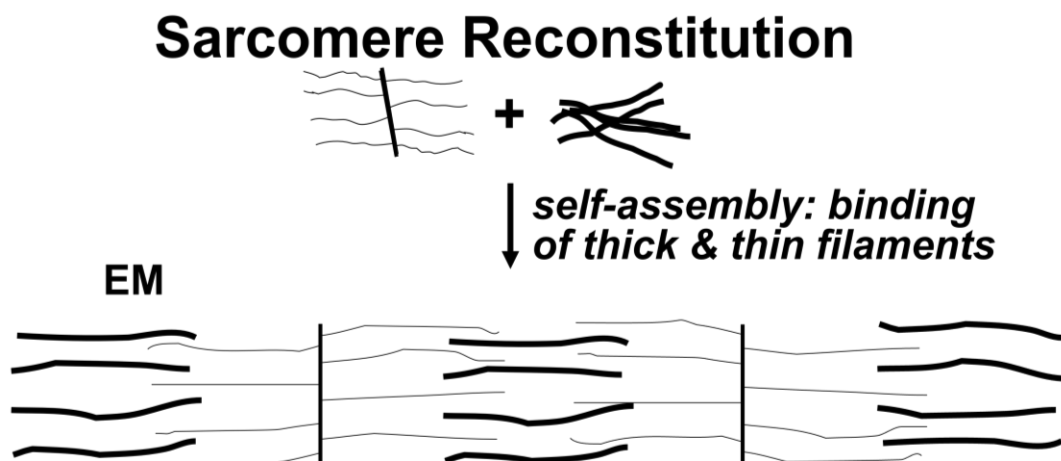


Fig. 18.17: Overview of an isolation of actin (thin) filaments (still on Z-lines) from myosin (thick) filaments. Next, after mixing actin Z-line and myosin fractions, electron microscopy of the resulting viscous material revealed thin filaments interdigitating with thick filaments (Figure 18.18).



These isolated sarcomere components bind... and behave in many ways as they would in intact muscle!

Fig. 18.18: Reconstitution of actin filaments (on Z-lines) with myosin filaments.

As expected, when ATP was added to these extracts, the solution viscosity dropped, and electron microscopy revealed that the myosin and actin filaments had again separated. The two components could again be isolated and separated by centrifugation.

In yet further experiments, actinomyosin preparations could be spread on over an aqueous surface, producing a film on the surface of the water. When ATP was added to the water, the film visibly “contracted,” pulling away from the edges of the vessel, reducing its surface area! Electron microscopy of the film revealed shortened, sarcomere-like structures, with closely spaced Z-lines and short I-bands—further confirming the sliding-filament model of muscle contraction.



[332 In Vitro & Electron Microscope Evidence for a Sliding-Filament Model](#)

When actin and myosin were further purified from isolated actinomyosin, the thick myosin rods could be dissociated into large myosin monomers. In fact, at ~599 Kd, myosin monomers are among the largest known proteins. Thus, thick filaments are massive polymers of huge myosin monomers! The molecular structure of myosin (thick) filaments is shown in Figure 18.19.

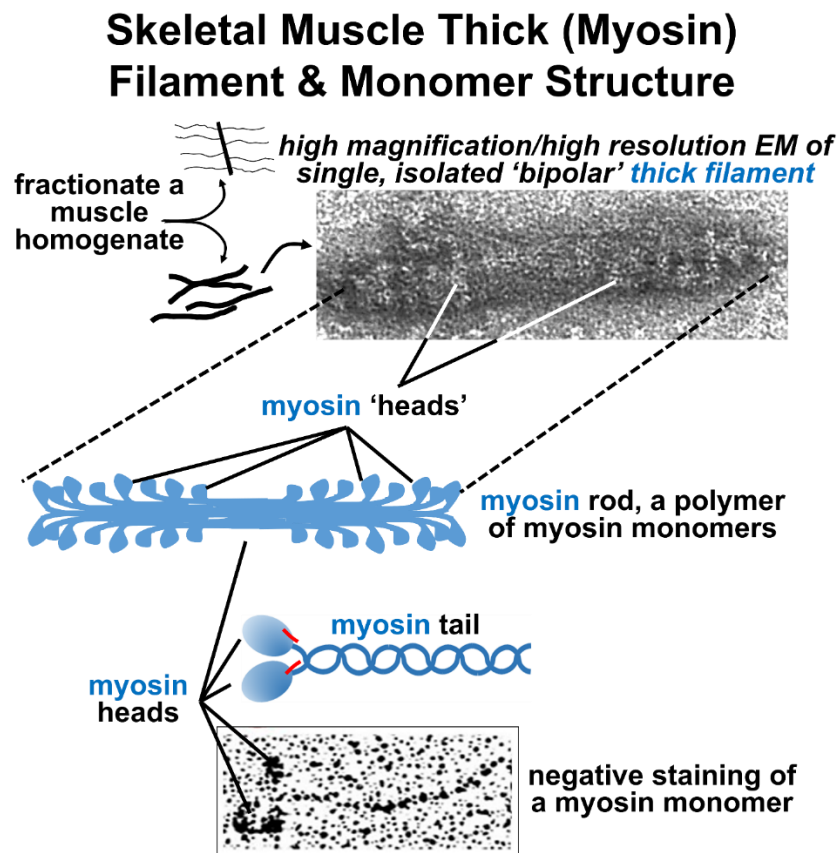


Fig. 18.19: Structure of a skeletal-muscle myosin filament and the myosin monomer. Shown is **myosin II**, the thick filament that spans both sides of the H zone in a sarcomere (upper). The head-and-tail structure of a myosin monomer is shown in the high-magnification electron micrograph and is illustrated in the cartoon (lower). The myosin monomer is itself a polymer of four polypeptides.

An early observation of isolated mammalian actin filaments was that they had no ATPase activity. We've seen isolated myosin preparations do have ATPase activity, but they would catalyze ATP hydrolysis only very slowly compared to intact muscle fibers. Faster ATP hydrolysis occurred only if myosin filaments were mixed with microfilaments (either on or detached from Z-lines). In the electron microscope, isolated myosin protein monomers each appeared to have a double-head and a single-tail region. Biochemical analysis showed that the monomers themselves were composed of the two heavy-chain polypeptides and two pairs of light-chain polypeptides, as shown in the Figure 18.19 illustration. In Figure 18.20, a high-magnification, high-resolution electron-micrograph simulation and corresponding drawings illustrate a myosin monomer and its component structures.

Separating Myosin Monomer Heads and Tails

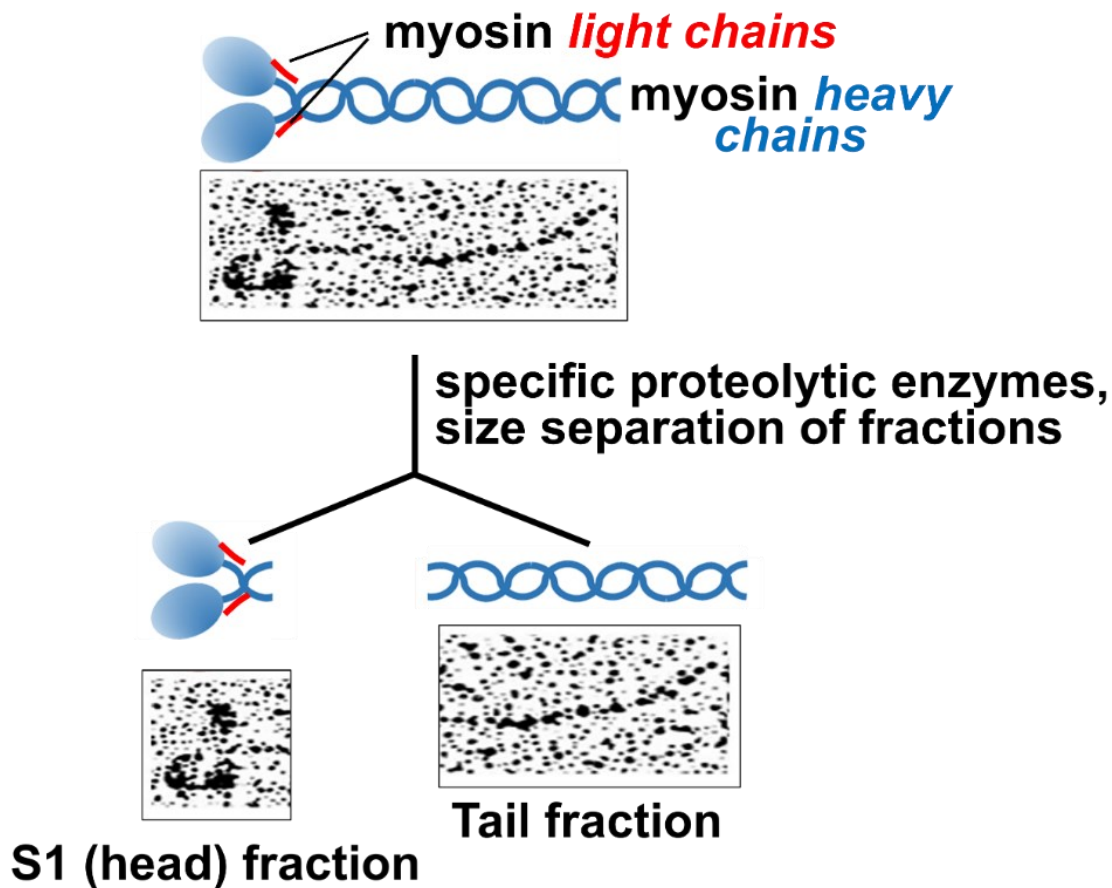


Fig. 18.20: Purified myosin monomers are digested with enzymes that hydrolyze peptide bonds between specific amino acids. This produces an *S1* (*head*) fragment and a *tail* fragment with different properties.

Proteolytic enzymes that only hydrolyze peptide linkages—and only between specific amino acids—“cut” the myosin monomers into *S1* (*head*) and *tail* fragments. Shown in Figure 18.20 are electron micrographs of enzymatic digest fractions separated by ultracentrifugation. The tail fragments are parts of the two-myosin heavy-chain polypeptides. The S1 fragments consist of a pair of light chains and the rest of the heavy chains.

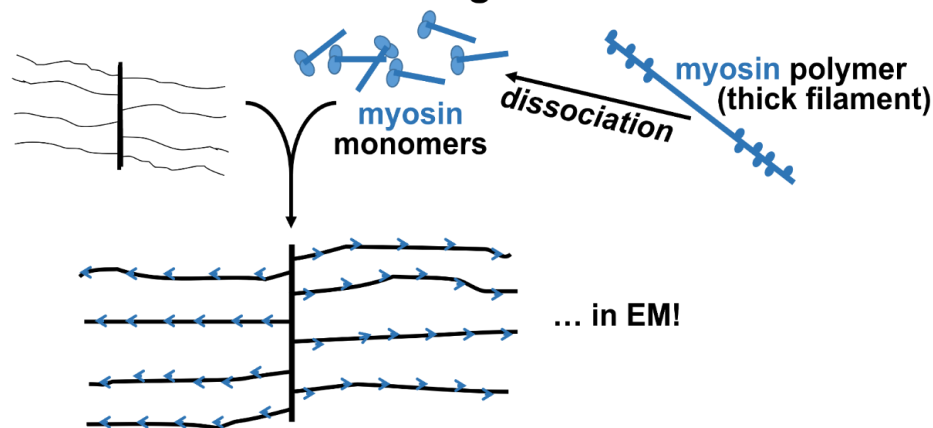
On further analysis, the S1 myosin head fraction had a slow ATPase activity, while the tails had none. The slow activity was not an artifact of isolation; mixing the S1 fraction with isolated actin filaments resulted in a higher rate of ATP hydrolysis. Thus, the myosin heads must be ATPases that bind and interact with actin microfilaments.



[333-2 Thick Filament & Myosin Monomer Structure](#)

In fact, S1 myosin heads bind directly to actin, *decorating* the actin with “**arrowheads**” visible in the electron microscope (see ^{18.4}[S1 Arrowheads On Muscle Actin](#)). Even intact myosin monomers could decorate muscle actin. These results are consistent with the requirement that myosin must bind to actin to achieve a maximum rate of ATPase activity during contraction. The arrowheads on *decorated actin* still attached to Z-lines are illustrated in Figure 18.21.

Myosin Monomers or S1 fragments “Decorate” Actin



Conclude:


- myosin monomers or S1 () fragments) bind to (i.e., ‘decorate’) actin on Z-lines.
- actin filaments ‘face away’ on opposite sides of Z-line; microfilaments show directionality (*polarity*)...

Fig. 18.21: Actin decoration by myosin-monomer S1 fragments in a kind of reconstitution experiment shows an opposing polarity of actin filaments on opposite sides of the Z-line.

Note that the “arrowheads” always face in opposite directions on either side of the Z-line. These opposing arrowheads suggest that the actin filaments attached to the two Z-lines of a sarcomere are drawn toward each other along the opposite sides of bipolar myosin rods. This is consistent with sliding filaments that draw Z-lines closer together during skeletal muscle contraction, shortening the sarcomeres. For another look at “arrowheads” and other aspects of muscle structure, check out the slide show at ^{18.5}[Muscle Structure & Physiology-J. Rosenbluth](#).



[334-2 Myosin Monomers & S1 Heads Decorate Actin](#)

18.7 Allosteric Change and the Microcontraction Cycle

Whereas dynein and kinesin are *motor proteins* that “walk” along microtubules, myosins are motor proteins that walk along microfilaments. All these motor proteins are ATPases that use the free energy of ATP hydrolysis to effect conformational changes that result in the walking (i.e., motility). In skeletal muscle, allosteric changes in myosin heads enable the myosin rods to do the walking along the sarcomere thin (F-actin) filaments. When placed in sequence, the different myosin head conformations illustrated in Figure 18.22 are the likely changes that would occur during a *microcontraction cycle*.

ATP & the Microcontraction Cycle

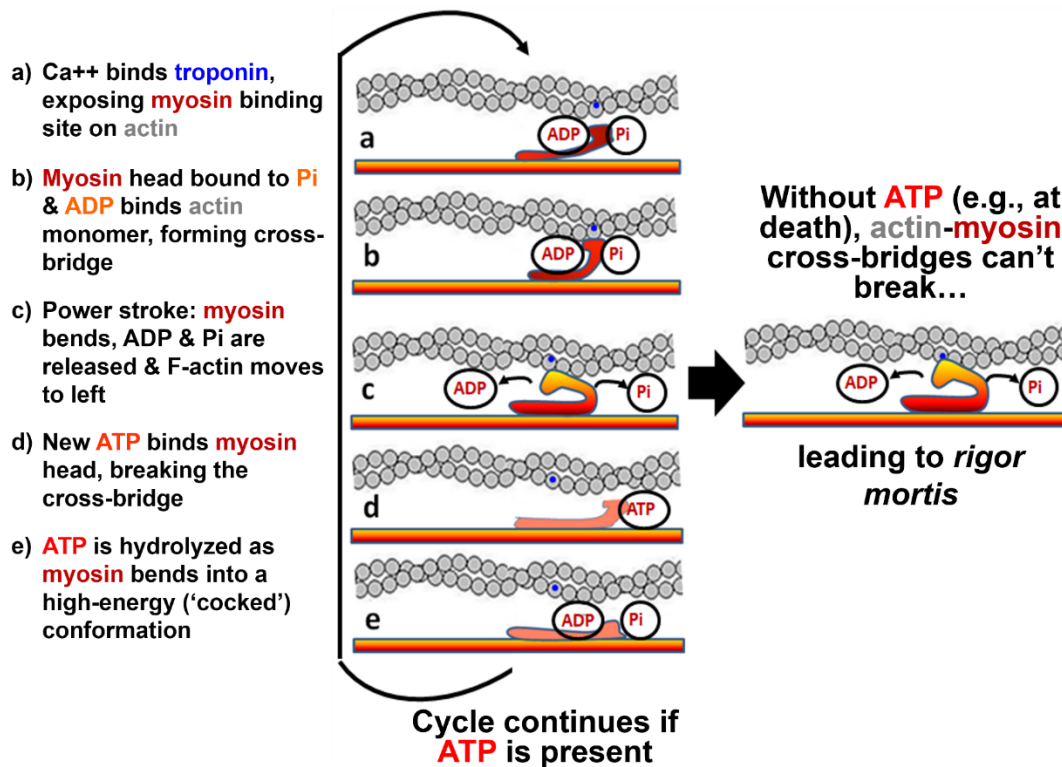


Fig. 18.22: The steps in the *microcontraction cycle* explain the muscle contraction paradox.

To help you follow the sequence find and follow the small dot on a single monomer in the actin filament. Here are the steps:

- a. Calcium is required for contraction; in the presence of Ca^{++} ions, myosin-binding sites on actin are open. (Ca^{++} -regulation of muscle contraction is discussed in detail shortly.)
- b. Myosin heads with attached ADP and Pi bind to open sites on actin filaments.
- c. Actin-myosin binding causes the *hinge region* of the myosin head to bend, *micro-sliding* (pulling) the attached microfilament; this is the *power stroke* (the dot on the actin has moved to the left. The myosin head remains bound to an actin monomer in the F-actin.

- d. ATP displaces ADP and Pi on myosin heads in this bent conformation. The resulting allosteric change in the head breaks the crossbridge between it and the actin.
- e. Once dissociated from actin, myosin heads catalyze ATP hydrolysis, resulting in yet another conformational change. The head, still bound to ADP and Pi, has bent at its hinge, taking on a high-energy conformation that stores the energy of ATP hydrolysis.

Microcontraction cycles of actin sliding along myosin continue if ATP and Ca^{++} are available. During repetitive *microcontraction cycles*, myosin heads on the thick filaments pull actin filaments attached to Z-lines, as stored free energy is released during *power strokes*, bringing the Z-lines closer together. The result is the shortening of the sarcomere, the muscle cells, and ultimately of the entire muscle. When the muscle no longer needs to contract, the Ca^{++} required for contraction is withdrawn; microcontraction cycles cease; the myosin heads remain in the high-energy conformation of step *e*, and the muscle can relax (stretch). Neural stimulation can cause another release of Ca^{++} that will again signal contraction.

The *microcontraction cycle* also stops when ATP is gone, the permanent state of affairs brought about by *rigor mortis* after death! At this time, myosin heads remain attached to the actin filaments in the state of muscle contraction that existed at the time of death. This is *rigor mortis* at the molecular level (shown in the illustration above). At the level of whole muscle, *rigor mortis* results in the inability to stretch or otherwise move body parts when ATP has once-and-for-all departed. You will have encountered this phenomenon if you watch any police or detective story that includes a coroner!

18.8 The Microcontraction Cycle Resolves the Contraction Paradox

The microcontraction cycle resolves the contraction paradox:

- **ATP is necessary for muscle contraction:** In step *e* in the illustration above, as ATP on myosin heads is hydrolyzed, the heads change from a low energy to a high-energy conformation. The myosin heads can now bind to actin monomers (step *b* in the microcontraction cycle). This results in the *power stroke* (step *c*). Free energy that is released by an allosteric change in myosin will pull the actin along the myosin, in effect causing a micro-shortening of the sarcomere—in other words, a contraction!
- **ATP is necessary for muscle relaxation:** At the end of step *c*, myosin remains bound to actin until ATP can again bind to the myosin head. Binding of ATP in step *d* displaces ADP and inorganic phosphate (Pi)...and breaks actin-myosin crossbridges. A removal of Ca^{++} from sarcomeres at the end of a contraction event blocks myosin-binding sites on actin, while the rapid breakage of actin-myosin crossbridges by ATP-myosin binding allows muscle relaxation and the sliding apart of the actin and myosin filaments (i.e., stretching). This leaves the myosin heads in the “cocked” (high-energy) conformation, ready for the next round of contraction.

No more paradox! The displacement of ADP and Pi by **ATP binding** to myosin heads breaks actin-myosin crossbridges, allowing for relaxation (stretching) of the sarcomeres, and hence

muscles. The **hydrolysis of the ATP** bound to myosin cocks the myosin head in a high free energy conformation whose energy is released during the microcontraction power stroke.

Electron-microscopic examination of myosin-monomer heads at different ionic strengths provides visual evidence that myosin heads are flexible and can take on alternate stable conformations, as would be expected during the microcontraction cycle (Figure 18.23).

Conformational States of Myosin Monomers

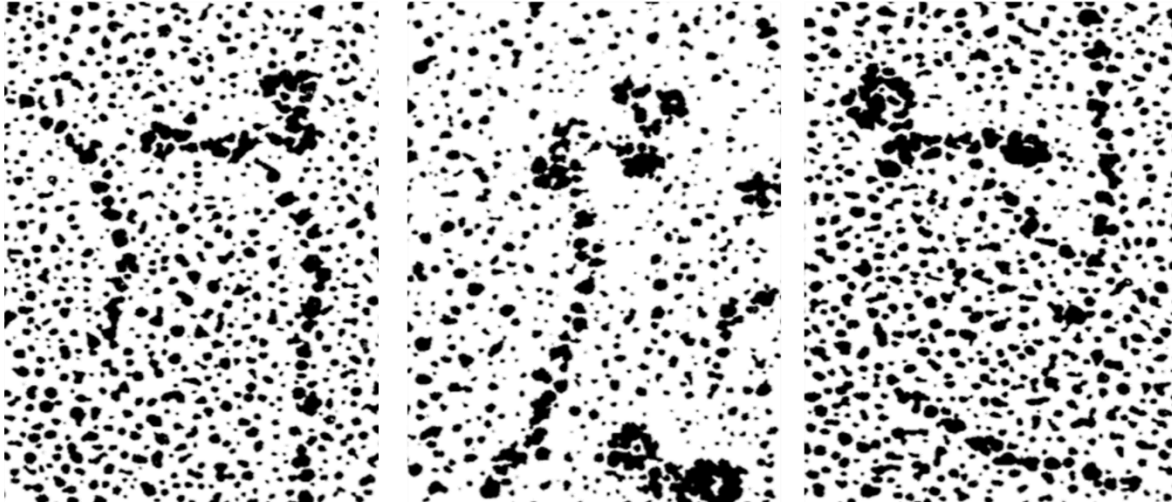


Fig. 18.23: The ability of flexible myosin heads to bend and to change conformation is consistent with their proposed activity during the microcontraction cycle.

The arrows point to myosin heads in different conformations. This visualization was made possible by treating the preparations with antimyosin antibodies (immunoglobulins). See a cool video of conformational change in myosin monomers at ^{18.6}[Myosin Heads in Action](#).



[335-2 Actin-Myosin Contraction Cycle Resolves the Contraction Paradox](#)



[336-2 ATP Binding & Hydrolysis Changes Myosin-Head Conformation](#)



18.9 Ca^{++} Ions Regulate Skeletal Muscle Contraction

Typically, the neurotransmitter acetylcholine released by a motor neuron binds to receptors on muscle cells to initiate contraction. Early experiments had already revealed that Ca^{++} was required, along with ATP, to get glycerinated skeletal muscle to contract. It was later shown that Ca^{++} ions were stored in the **sarcoplasmic reticulum** (the smooth endoplasmic reticulum) of intact muscle cells. As we have seen, an action potential generated in the cell body of a neuron propagates along an axon to the nerve terminal, or synapse.

The action potential at a **neuromuscular junction** that initiates contraction is summarized in the illustration in Figure 18.24 (below).

Ca⁺⁺ Regulation of Skeletal Muscle Contraction

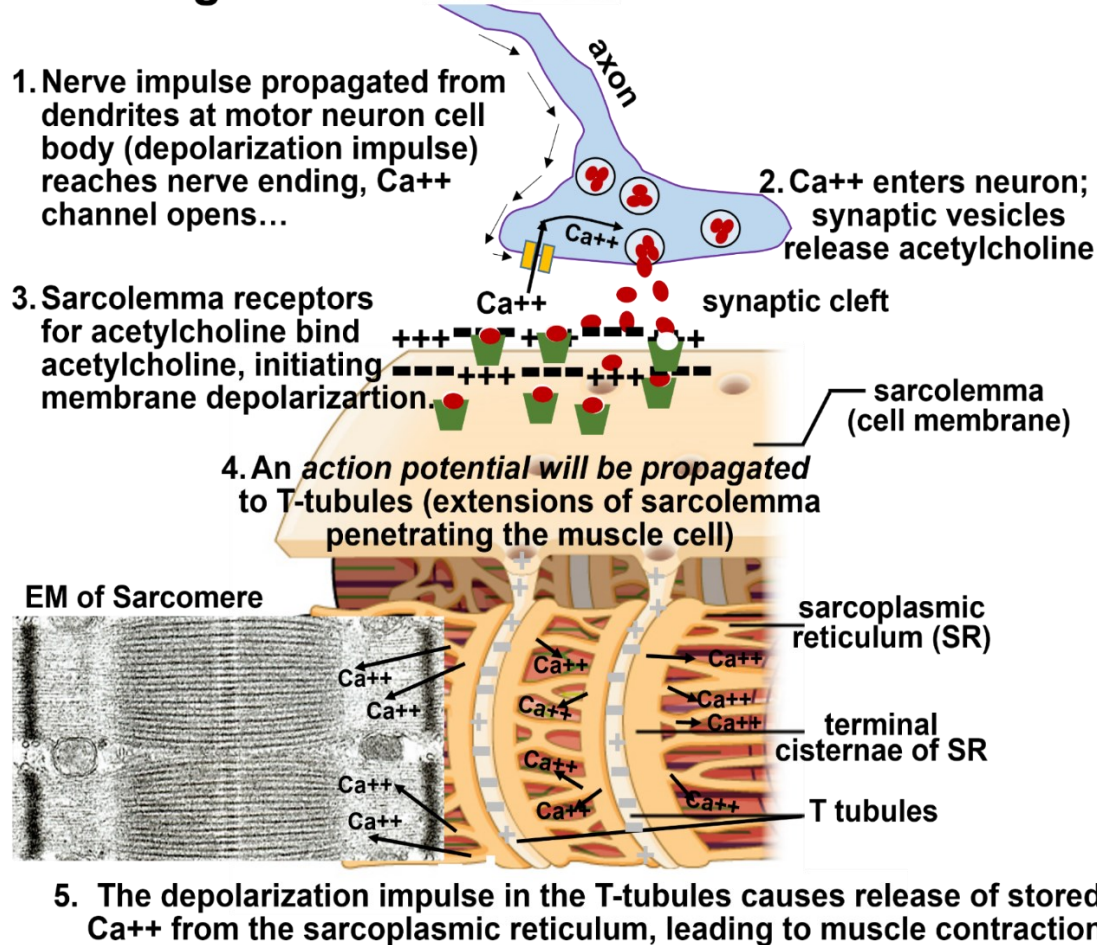
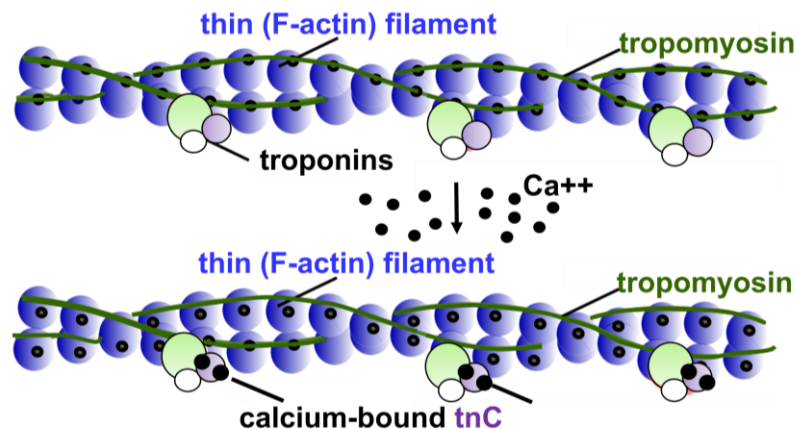


Fig. 18.24: A nerve impulse reaching a *neuromuscular junction* causes uptake of Ca⁺⁺ from the *synaptic cleft* (the space between a nerve ending and the muscle) into the nerve ending. This causes release of a neurotransmitter (acetylcholine) from synaptic vesicles that will trigger an action potential in the myocyte, leading to muscle contraction.

A neural impulse reaches the nerve ending (1), signaling a flow of Ca⁺⁺ ions into the nerve ending that cause synaptic vesicles to fuse with the cell membrane and release acetylcholine (2). The neurotransmitter crosses the synaptic cleft and binds to receptors (3), initiating an action potential in the muscle cell 4) that travels along the *sarcolemma* (the muscle plasma membrane) to points where it is continuous with *transverse tubules* (*T-tubules*). The action potential moves along the T-tubules and then along the membranes of the sarcoplasmic reticulum (4). This propagation of an action potential opens Ca⁺⁺ channels in the sarcoplasmic reticulum and T-tubules(5), releasing Ca⁺⁺ to bathe the myofibril sarcomeres where they bind to one of three *troponin* molecules; this will allow skeletal muscle contraction (i.e., to allow filaments to slide). The *three troponins* and a *tropomyosin* molecule are bound to actin filaments.

Experiments using *antitroponin* and *antitropomyosin* antibodies localized the three proteins in electron micrographs at regularly spaced intervals along actin filaments, as modeled in Figure 18.25 below.

Skeletal Muscle Thin (F-actin) Filament Structure



Ca⁺⁺ binds troponin C (TnC) subunits, exposes myosin-binding sites

Fig. 18.25: Thin skeletal muscle filaments consist of actin associated with troponins and tropomyosin, which play a role in regulating actin-myosin interactions during contraction.

In resting muscle, tropomyosin (a fibrous protein) lies regularly spaced along the actin filament, where it covers up the myosin-binding sites of seven G-actin monomer subunits in the microfilament. The *negatively stained* electron micrograph in ^{18.7}[Actin Bound to Tropomyosin](#) is consistent with the model illustrated in Figure 18.25. The cross section drawn in Figure 18.26 illustrates how **troponin T** (tropomyosin-binding troponin) and **troponin I** (*inhibitory troponin*) hold the tropomyosin in place, and how the binding of Ca⁺⁺ ions to troponin C regulates contraction.

Ca⁺⁺ ions regulate contraction by binding to troponin and causing conformational changes in the troponins and tropomyosin...

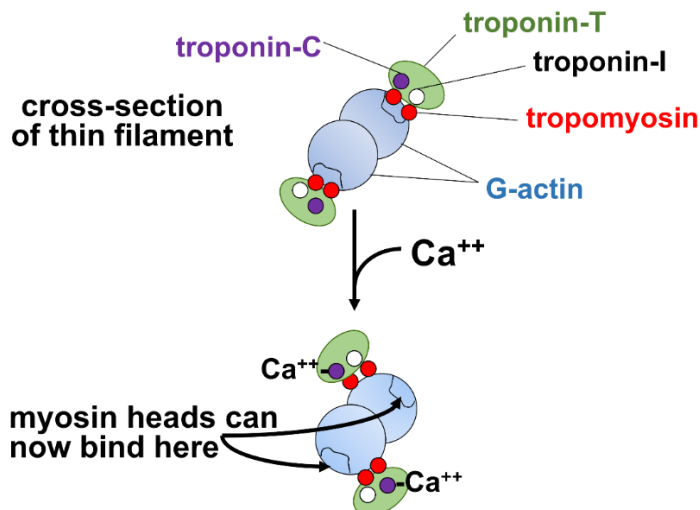


Fig. 18.26: Ca⁺⁺ ions bind troponin C to initiate the uncovering of myosin-binding sites on G-actin monomers of F-actin, resulting in allosteric changes in the troponins, and ultimately tropomyosin.

A chain reaction of conformational changes begins when Ca⁺⁺ ions bind to **troponin C**. The result is that the three-subunit troponin complexes that are bound to tropomyosin will shift

position along the filament to expose the myosin-binding sites on the G-actin subunits. Only after this shift can ATP-bound myosin in turn bind to actin and initiate the *microcontraction* cycle discussed earlier. The regulation of contraction by Ca^{++} is animated in the following link.



[337 Regulation of Skeletal Muscle Contraction by Calcium](#)

18.9.1 Muscle Contraction Generates Force

Contraction by ATP-powered sliding of thin along thick filaments generates force on the *Z-lines*. In three dimensions, the Z-lines are actually ***Z-discs***, to which the actin filaments are attached. The protein ***α -actinin*** in the Z-discs anchors the ends of the actin filaments to the discs. That way when the filaments slide, *force transduction* draws the Z-discs closer, shortening the sarcomeres. Another Z-disc protein, ***desmin***, is an intermediate filament organized around the periphery of the Z-discs. Desmin connects multiple Z-discs in a myofibril. With the Z-discs being kept in register, the muscle cell—and ultimately whole muscle contraction—is coordinated. Finally, actin filaments at the ends of the muscle cell must be connected to the plasma membrane so that *force transduction* will cause the muscle cell to shorten during myofibril contraction.

Several proteins, including ***syntrophins*** and ***dystrophin*** (an intermediate filament protein) anchor the free ends of microfilaments coming from Z-discs to the cell membrane. Still other proteins anchor the cell membrane in this region to the ECM (extracellular matrix) ***tendons***, which are in turn attached to bones! The force generated by myosin hydrolysis of ATP and the sliding of filaments in individual sarcomeres is thus transmitted to the ends of muscles to effect movement. If the name *dystrophin* sounds familiar, it should! The gene and its protein were named for a mutation that causes muscular dystrophy, resulting in a progressive muscle weakening.



[338 Contraction Generates Force Against Z-Discs and Cell Membranes](#)

18.9.2 The Elastic Sarcomere: Do Myosin Rods Just Float in the Sarcomere?

In fact, myosin rods do not “float” in sarcomeres but are instead anchored to proteins in the ***Z-discs*** and ***M-lines***. In 1954, R. Natori realized that when contracted muscle relaxes, it lengthens beyond its resting state, then shortens again to its resting length. Natori proposed that this elasticity must be due to a fiber in the sarcomere. Twenty-five years later, the elastic structure was identified as ***titin***, a protein that holds several molecular records! The gene for titin contains the largest number of exons (363) of known proteins. After actin and myosin, titin is also the most abundant protein in muscle cells. At almost 4×10^6 Da, the aptly named *titin* is also the largest known polypeptide, much larger than even myosin! As it extends from the Z-discs to the M-line of sarcomeres, titin coils around thick filaments along the way. The *Titin* is anchored at Z-discs by ***α -actinin*** and ***telethonin*** proteins. At the *M-line*, titin binds to ***myosin-binding protein C (MYBPC3)*** and to ***calmodulin***, among others (*myomesin*, *obscurin*,

skelamin...). Some if not all of these proteins must participate in keeping the myosin (thick) filaments positioned and in register in the sarcomere. This is like how desmin binds Z-discs to each other to keep sarcomeres in register. The location of titin and several other sarcomere proteins is illustrated in Figure 18.27.

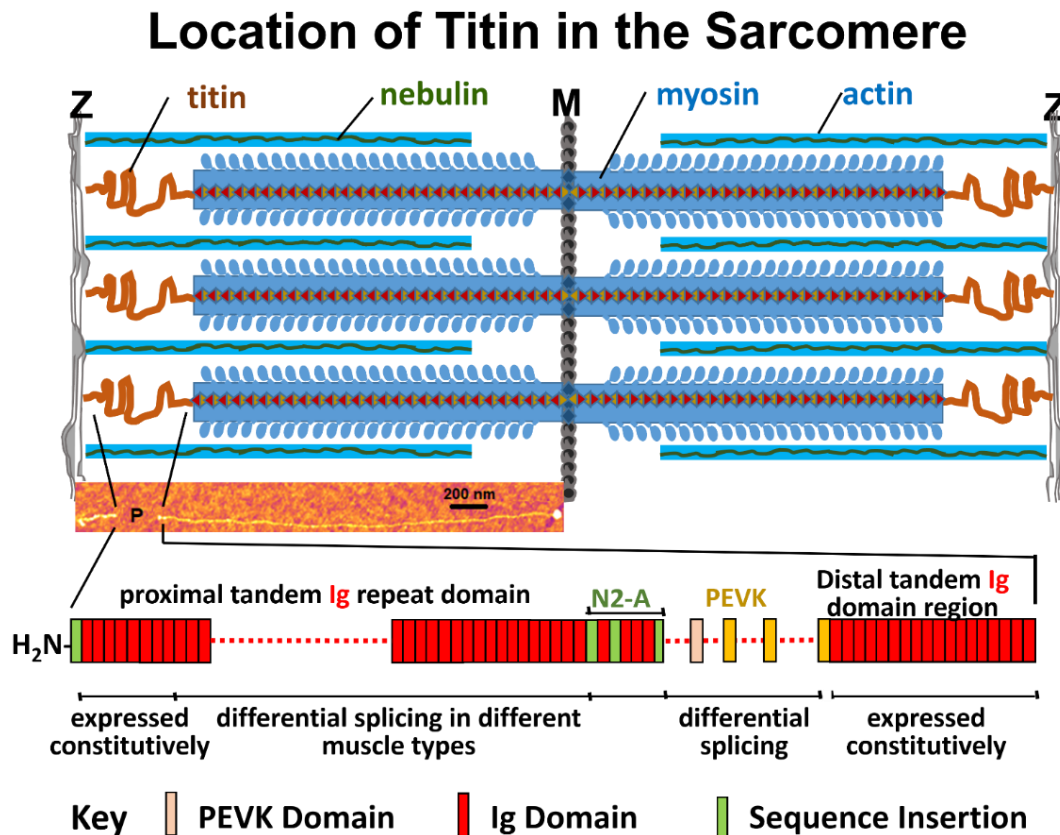


Fig. 18.27: Structure of *titin* and its location in the sarcomere.

Coiled titin molecules (red in the illustration) extend from the Z to M lines. The colorized electron micrograph of one extended titin molecule in the middle of the illustration should convince you of the length (35,213 amino acids!) of this huge polypeptide. Titin's elastic features are largely in a region labeled *P* in the electron micrograph—between the Z-discs and the ends of the myosin rods. The many domains of this *P* region are shown expanded at the bottom of Figure 18.27. With all the binding (and other) functions, you might expect that titin has many domains. It does! They include **Ig** (immunoglobulin) domains, **fibronectin** domains (not shown here), and **PEVK** and **N2A** domains, some of which help bind titin to α -actinin in Z-discs. Which ones and how many of the Ig and/or PEVK domains are present in a particular muscle depends on which alternative splicing pathway is used to form a titin mRNA. Over a micron long, titin functions as a molecular spring, as Natori predicted. Its coiled domains compress during contraction, passively storing some of the energy.

When a skeletal muscle relaxes, Ca^{++} is withdrawn from the sarcomere. ATP can still displace ADP from myosin heads, breaking actin-myosin crossbridges. When actin and myosin heads dissociate in the absence of Ca^{++} , the troponins and tropomyosin reverse their allosteric changes, once more covering myosin-head binding sites on F-actin. The muscle then stretches,

typically under the influence of gravity or an opposing set of muscles. But during contraction, 244 individually folded titin domains had been compressed, so during relaxation, these domains decompress. The stored energy of compression thus also helps to power muscle relaxation. At the same time, titin connections limit the stretch so that a potentially overstretched muscle can “bounce” back to its normal relaxed length.

In a particularly elegant experiment, Wolfgang A. Linke (et al.) provided a visual demonstration of myofiber *elasticity* consistent with the coiled-spring model of titin structure. They made antibodies to peptide domains on either side of the *PEVK* domain of *titin* (*N2A* and *I20-I22*) and attached them to *nanogold particles*. The gold particles were electron-dense and would appear as black granules in transmission electron microscopy. In the experiment, individual *myofibers* were stretched to different lengths, fixed for electron microscopy, and treated with the *nanogold*-linked antibodies. The *antibodies* localize to and define the boundaries of the titin *PEVK* domains in myofibers, stretched to different lengths (Figure 18.28, showing electron micrographs with simulated localization of nanogold particles, which reflect the actual results).

Nanogold Immunostaining Demonstrates Titin Elasticity in Glycerinated Myofibers

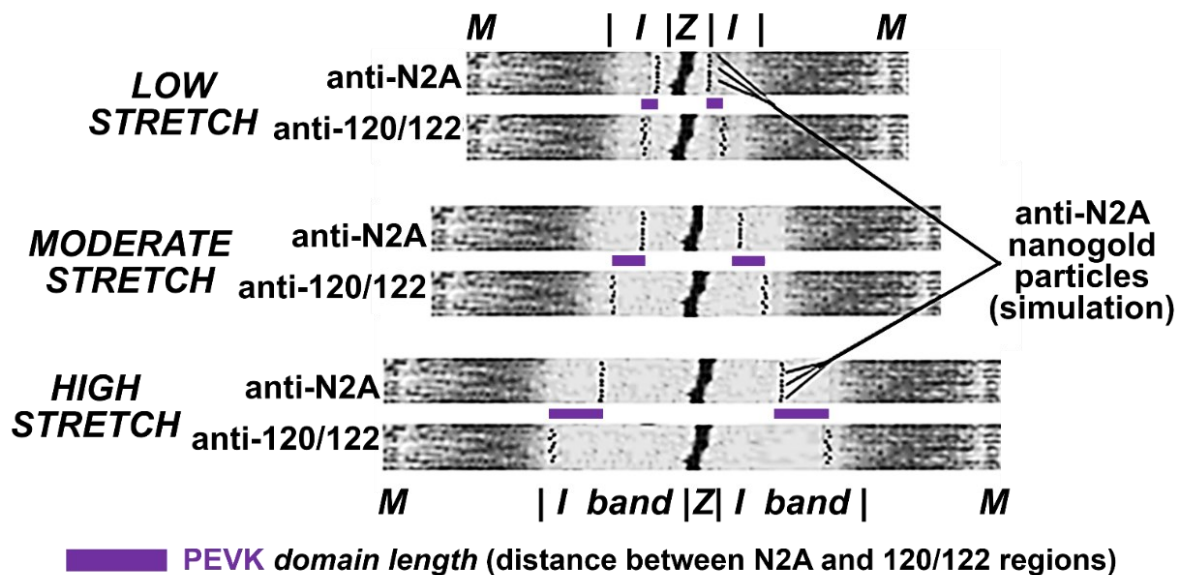


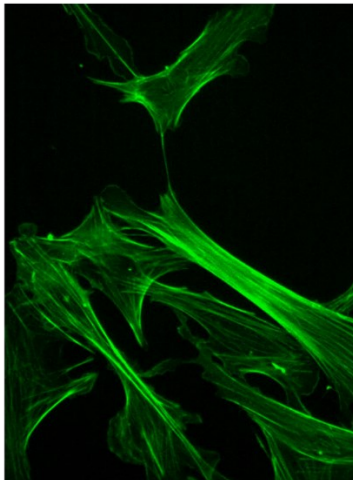
Fig. 18.28: Skeletal-muscle elasticity shown by the changes in location of anti-N2A- and anti-120/122-linked nanogold particles in sarcomeres. Note the increased separation of the PEVK domains targeted by the antibodies as a muscle fiber is increasingly stretched. Titin elasticity facilitates muscle relaxation.

In the experiment, increased stretch lengthened the I-bands on either side of the Z-lines of sarcomeres (blue bars under each pair of micrographs). Likewise, titin PEVK domains have also lengthened, as seen in the increased distance between the nanogold-linked N2A and the 120/122 antibody localization flanking the PEVK domains. This demonstration of titin (and thus, sarcomere) elasticity is consistent with a storage of some of the free energy of contraction when the titin is compressed, and a passive release of that energy during relaxation. Titin tethers thick filaments to Z-discs and M-lines and thus also limits sarcomere stretch during relaxation. Titin elasticity is animated at ^{18.8}[Elasticity of Titin Domain](#).

18.10 Actin Microfilaments in Nonmuscle Cells

Microfilaments are found throughout the Eukarya category. Electron microscopy revealed that thin (~10 nm) filaments permeate the cytoskeleton of eukaryotic cells (see Figure 18.3). They typically lie in the cortex of cells, just under the plasma membrane, where they support cell shape. These same microfilaments can also reorganize dynamically, allowing cells to change shape. A dramatic example of this occurs during cytokinesis in dividing cells, when the dividing (meiotic or mitotic) cell forms a *cleavage furrow* in the middle of the cell (discussed further in another chapter). The cortical microfilaments slide past each other with the help of *nonmuscle myosin*, progressively pinching the cell until it divides into two new cells. To test whether these 10 nm “microfilaments” were in fact actin, scientists placed myosin monomers or S1 myosin head fragments atop actins isolated from many different cell types. When seen in the electron microscope, such preparations always revealed that the 10 nm actin microfilaments were decorated with *arrowheads*, just like S1 fragments decorated muscle cell actin or Z-line-bound actin. See S1 arrowheads on cortical actin and on microvillar actin at ^{18.10}[S1 Fragment Decorates Cortical Actin](#) and ^{18.11}[S1 Fragment Decorates Ciliary Actin Bundle](#) (respectively) Clearly the cytoplasmic microfilaments are a form of F-actin. The role of cortical filaments in cell division is animated at ^{18.12}[Cortical Actin Filament Action in Cytokinesis](#). We now know that actin microfilaments are involved in all manner of cell motility, in addition to their role in cell division, enabling cell movement and cytoplasmic streaming within cells. They give intestinal microvilli strength; they even enable them to move independently of the passive pressures of *peristalsis*. Other examples of microfilaments in cell motility include the ability of *amoeba* and other *phagocytic cells* to extend *pseudopodia* to engulf food or foreign particles (e.g., bacteria). A well-studied example of microfilament-powered cell movement is the spread of fibroblast cells along surfaces, shown in Figure 18.29.

Immunofluorescing Fibroblast Stress Fibers



Role of Actin in Fibroblast Migration

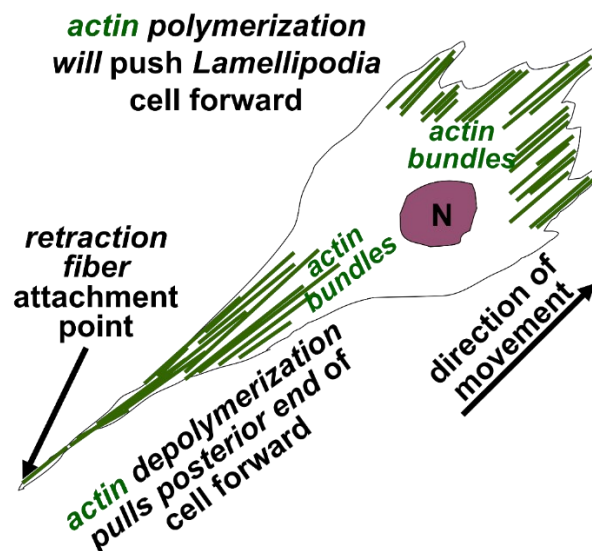


Fig. 18.29: In the anti-actin immunofluorescence micrograph of fibroblasts, actin localizes with stress fibers, which help maintain cell shape (left). Actin also localizes in lamellipodia and retraction fibers in migrating fibroblasts (right), orienting in the direction of movement.

Migrating fibroblasts move forward by extending lamellipodia (and thin filipodia beyond them), by assembling actin bundles along the axis of movement. In the immunofluorescence micrograph (left, Figure 18.29), the actin *stress fibers* that maintain cell shape fluoresce green; the dual roles of actin in fibroblast shape and movement are also illustrated at the right.

The extension of filipodia at the moving front of a fibroblast is mainly based on actin assembly and disassembly (not unlike motility based on microtubules). A **retraction fiber** forms at the hind end of the cell as the fibroblast moves forward. The retraction fiber remains attached to the surface (*substratum*) on which it is migrating until actin-myosin interactions (in fact, sliding) cause *retraction* of most of this "fiber" back into the body of the cell.

Studies of nonmuscle cell motility suggest the structure and interacting molecular components of stress fibers modeled in Figure 18.30.

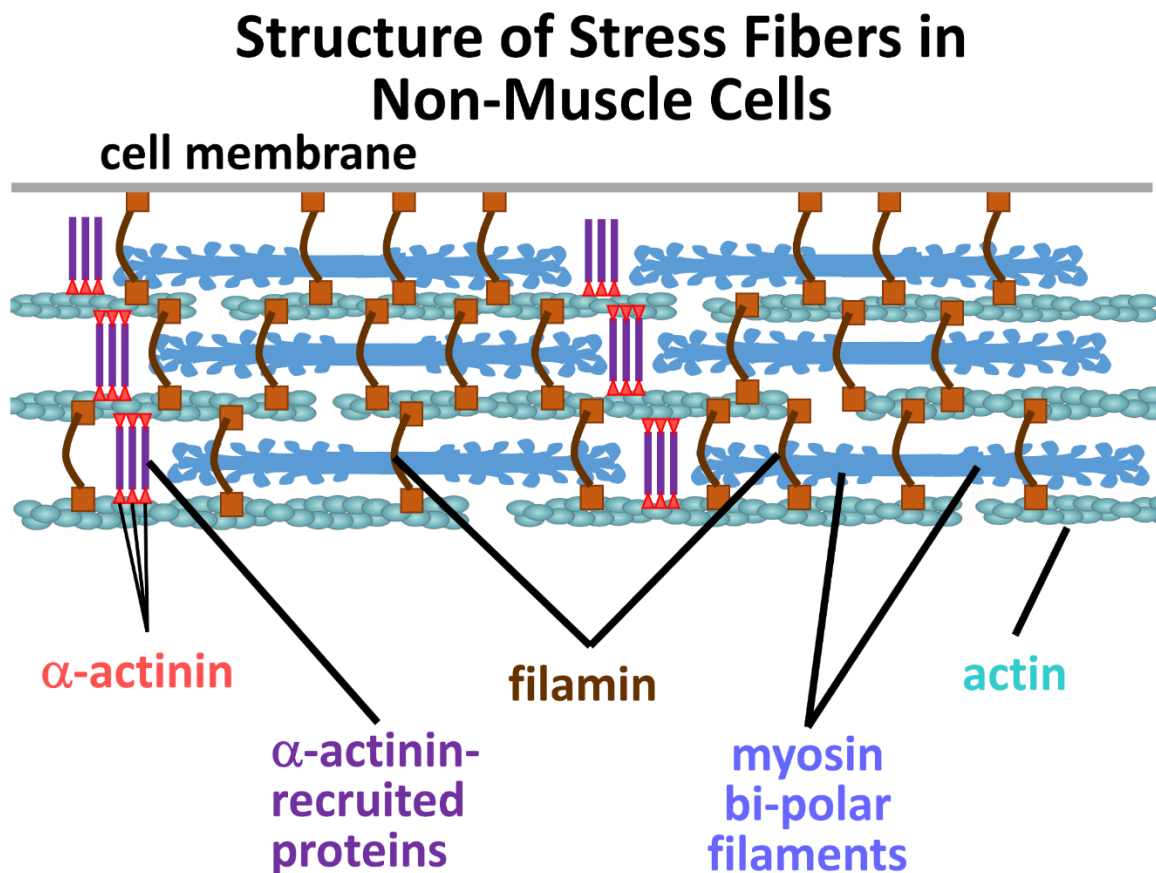


Fig. 18.30: The molecular structure of stress fibers: Myosin as well as other actin-binding proteins interact with actin in nonmuscle cell motility.

The illustration suggests roles for actin-binding proteins in the sliding of overlapping myosin and actin filaments during movement. This model may also explain the cytoplasmic streaming that distributes cellular components and nutrients throughout a cell. In fact, both movements involve actin-myosin interactions. **Filamin** in this drawing is shown holding actin filaments together at an angle, while **α -actinin** also helps to bundle the actin filaments. **Titin** (not shown) also seems to be associated with stress fibers.

Unlike the highly organized skeletal muscle sarcomeres, the proteins and filaments in stress fibers are not part of Z- or M-line superstructures. Were these less-organized nonmuscle stress-fiber actin bundles the evolutionary ancestors of muscle cell sarcomeres? Or vice-versa?

18.11 Both Actins and Myosins are Encoded by Large Gene Families

Actin may be the most abundant protein in cells! At least six different actin *isoforms* encoded by a large actin *gene family* have nearly identical amino acid sequences, all of which are involved in cytoskeletal function. While the β -actin isoform predominates, genes for some isoforms are expressed in a cell-specific manner. In mice, the loss of a γ -actin gene has little effect on the organism, while loss of the β -actin gene in mice is lethal at embryonic stages.

So what is the function of γ -actin? Does it even have one? Or is γ -actin a *vestigial* protein that, like our appendix, once served a now long-lost function? And what about all those other isoforms? To date, the amino acid sequence differences among actins don't predict dramatically different protein function. Despite yet undefined functional differences between them, virtually all actin isoforms are decorated by myosin monomers (or S1 heads), so any function should involve interaction with myosin. These observations make one wonder if most actin isoforms are adaptations, however subtle, such that the absence of one isoform would pose a significant threat to the survival of an organism. So, could isoform differences underlie some as-yet-unidentified physiological advantage to different cells? Or are they simply a bank of mutations lying in wait for natural selection to reveal a new function?

Studies in humans show that a nonlethal β -actin mutation correlates with delayed development and later neurological problems (e.g., epilepsy) and kidney and heart abnormalities. Still, people with such mutations can lead nearly normal, healthy lives, suggesting that a maintaining a bank of isoforms offer some protection against such nonlethal mutations.

Finally, genes encoding variant myosin isoforms comprise a large eukaryotic gene family. All isoforms have ATPase activity, and some (if not all) are clearly involved in cell motility. Unique functions are not yet known for the isoforms, but different myosin monomers can decorate actin, and myosins from one species can decorate actin filaments of other species, even across wide phylogenetic distances. With so many unknowns, one can still make a career studying why there are so many isoforms of actin and myosin!

18.12 Intermediate Filaments

The 10 nm *intermediate filaments* are proteins with extended secondary structure that don't readily fold into tertiary structures, and they have no known enzymatic activity. The extracellular *keratins* that make up fur, hair, fingernails, and toenails are proteins related to intermediate filaments. Keratins are bundles of rigid, insoluble extracellular proteins that align and bind to form stable, unchanging secondary structures. Intermediate filaments also permeate other cells, where they participate in regulating and maintaining cell shape. Recall their intercellular location in desmosomes, where they firmly bind cells together to confer tensile strength to tissues. Also recall the role of intermediate filaments in anchoring actin to

either Z-discs or plasma membrane *plaques* in muscle cells. There they transmit the forces of contraction to the coordinated shortening of the sarcomeres and ultimately to the actual shortening of a muscle. Finally, intracellular intermediate filaments are not confined to the cytoplasm; ***lamins*** are intermediate filaments that make up structural elements of the *nuclear lamina*, a kind of ***nucleoskeleton***. The structural features of intermediate filaments are illustrated in the cartoon in Figure 18.31.

Intermediate Filament Structure: Elasticity Results from Secondary Coil Structures

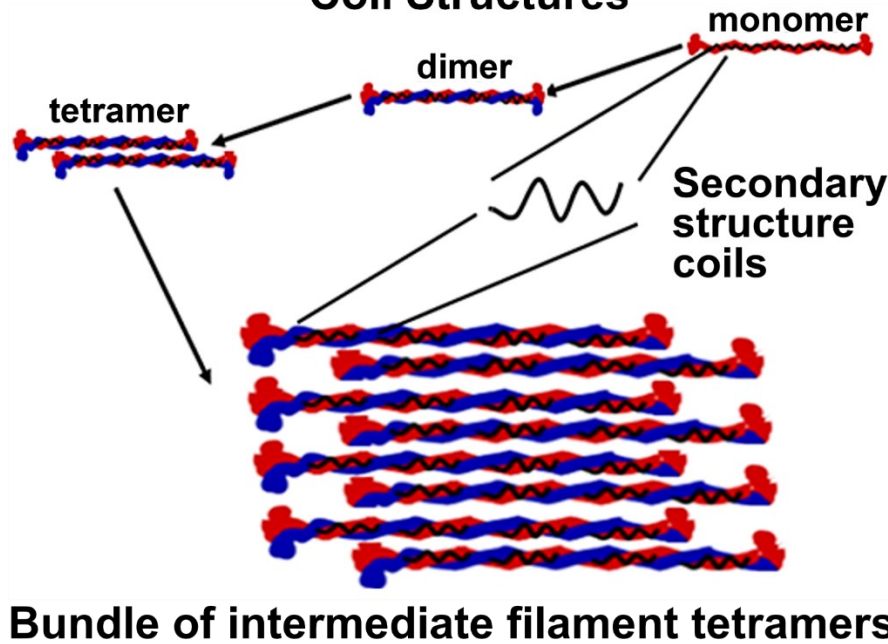


Fig. 18.31: The coiled secondary structure of the proteins permits the elasticity of intermediate-filament bundles that contributes to the viscosity of cytoplasm.

As we saw earlier, intermediate filament subunits have a common structure, consisting of a pair of monomers. The monomers are nonpolar; unlike microtubules and actin filaments, they do not have “plus” and “minus” ends. Globular domains at the C- and N-terminal ends of monomer filaments are separated by coiled-rod regions. The basic unit of intermediate filament structure is a dimer of monomers. Dimers further aggregate to form tetramers and larger filament bundles. Like microtubules and actin filaments, intermediate filament bundles can disassemble and reassemble as needed when cells change shape.

Unlike microtubules and actin, intermediate filaments can stretch—a property conferred by the coiled-rod regions of the filaments. This should be reminiscent of titin molecules. In the bundled intermediate filaments that permeate the cytoplasm of cells, the ability to stretch contributes to the viscosity of cytoplasm and is called ***viscoelasticity***. This elastic property is thought to allow actins and microtubules some freedom of movement of and within cells. Like the actins, there are many isoforms of intermediate filaments, but in this case with some known different functions.

Some iText & VOP Key Words and Terms

"9 + 2"	F-actin	myofiber
---------	---------	----------

α tubulin	F-actin polarity	myofibril
A-band	flagella	myosin
acetylcholine	fluorescence microscopy	myosin ATPase
acidic keratin	force transduction	myosin "heads"
actin	G-actin	neuromuscular junction
actin-binding proteins	hair, horn	nuclear lamina
actin-myosin interactions	I-band	plus, and minus ends
actin-myosin paradox	intermediate filaments	pseudopodia
action potential	intestinal microvilli	sarcomere
amoeboid movement	keratin	sarcoplasmic reticulum
ATPase	keratin isoforms	sarcolemma
axoneme	lamins	Secretion vesicle transport
β tubulin	membrane depolarization	skeletal muscle contraction
basal body	microfilaments	skeletal muscle relaxation
Ca^{++} regulation of contraction	microtubule assembly end	sliding-filament model
Ca^{++} release vs active transport	microtubule disassembly end	syncytium
cell motility	microtubule doublets	thick and thin filaments
centriole	Microtubule-organizing center	titin
cilia	microtubule polarity	transverse (T) tubules
contraction regulation	microtubule-associated proteins (MAPs)	tropomyosin
cortical cellular microfilaments	microtubules	troponin C
crossbridges	mitotic & meiotic spindle fibers	troponin I
cytoplasmic streaming	M-line	troponin T
cytoskeleton	motor proteins	tubulin heterodimer
desmosomes	MTOC	tubulins
dynein	muscle cell	viscoelasticity
evolution of actin genes	muscle fiber	Z-discs
evolution of myosin genes	myocyte	Z-line

CHAPTER 18 WEB LINKS





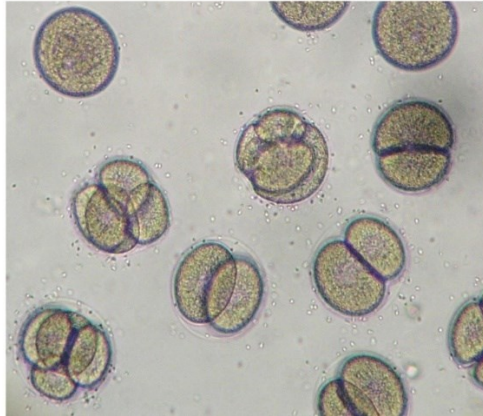
Chapter 19

Cell Division and the Cell Cycle

Separation of replication from cell division in eukaryotes; cell-cycle checkpoints, cyclins, and MPF; apoptosis, cancer, radiation, chemotherapy, and immunotherapy

Reminder: For inactive links, google key words/terms for alternative resources.

**Cleavage: the first mitoses
of fertilized sea urchin eggs
forms an embryo**



**Use this QR code to
watch real-time fruit
fly embryogenesis,**



or use the link below

[Real-Time Drosophila
Embryogenesis](#)

19.1 Introduction

Bacteria divide by binary fission, duplicating their nearly “naked” circular chromosomes as the cell enlarges and divides into two daughter cells. Under favorable conditions, bacteria (e.g., *E. coli*) can divide every twenty to forty minutes. Many eukaryotic cells stop dividing when they reach maturity (their *terminally differentiated state*), while those that do divide may have varying life spans between divisions, ranging from fifteen to twenty-four hours. Cell division begins with **mitosis**, the condensation of chromosomes from chromatin. **Cytokinesis** is the process near the end of mitosis that physically divides one cell into two new cells after the duplicated chromosomes are safely on opposite sides of the cell. Although mitosis lasts only about an hour in the life of the typical eukaryotic cell, it has been parsed into four or five phases (depending on whose text you are reading!); it is the last of these phases that overlaps cytokinesis. Usually, mitosis and cytokinesis together take about 1.5 hours. By far the longest period between successive cell divisions or the **cell cycle** is **interphase**, so-called because nineteenth-century microscopists who described mitosis saw nothing happening in cells when they were not actually dividing. But by the 1970s, experiments had shown that interphase itself was divisible into discrete phases of cellular activity called G_1 , S , and G_2 (occurring in that order). It turns out that **kinases** regulate progress through the cell cycle, catalyzing timely *protein phosphorylations*. The first of these kinases to be discovered was *mitosis-promoting factor* (**MPF**). Kinase-regulated events are **checkpoints** a cell must pass through to enter the next phase of the cell cycle. As you might guess, the failure of a checkpoint can have serious consequences. One consequence is **carcinogenesis**, a runaway proliferation of cancer cells. We begin this chapter with a brief description of binary fission in *E. coli*, followed by a closer look at the eukaryotic cell cycle. We close with a look at alternate fates of eukaryotic cells and at details of cellular end-of-life events, including **apoptosis**, or **programed cell death**.

Learning Objectives

When you have mastered the information in this chapter, you should be able to do the following:

1. Distinguish between mitotic and cell-cycle *phases* and what occurs in each of the phases.
2. Interpret the experiments leading to our understanding of the *separation* of chromosomal events from the duplication of the DNA contained in those chromosomes.
3. Describe and compare the roles of different cyclins and cdks in regulating the cell cycle.
4. Define cell-cycle checkpoints and how/why they monitor progress through the cell cycle, including their roles in detecting DNA damage and pathways to *apoptosis*.
5. State a hypothesis for how cell-cycling errors could make normal cells become *cancer* cells.
6. Compare cell-cycle disruption in somatic and germline cells that affect cancers.
7. List some examples of apoptosis in humans and other organisms.
8. Compare and contrast examples of apoptosis and *necrosis*.
9. Formulate a hypothesis to account for the degradation of cyclin after mitosis.
10. Explain the biochemical/molecular basis of several chemotherapies and their side effects.

19.2 Cell Division in a Prokaryote

The life of actively growing bacteria is not separated into discrete times for duplicating genes (i.e., DNA synthesis) followed by **binary fission** (dividing and partitioning the duplicated DNA into new cells). Instead, the single circular chromosome of a typical bacterium starts replicating even before fission is complete, so that the new daughter cells already contained partially duplicated chromosomes. Cell growth, replication, and fission as it might be seen in *E. coli* are illustrated in Figure 19.1.

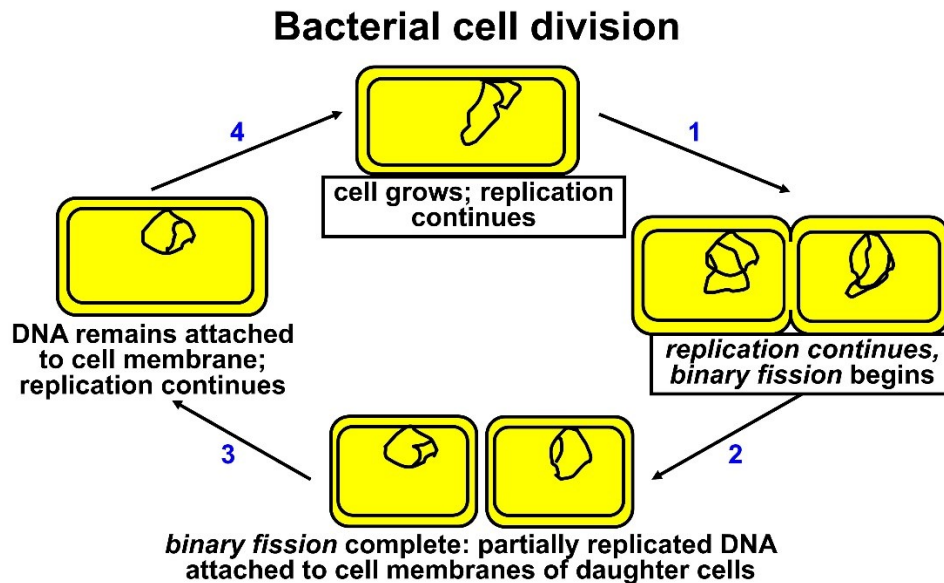


Fig. 19.1: Bacteria divide by binary fission. When growing in culture, they divide continually, partially replicating their circular DNA molecules during one division in preparation for the next.



In marked contrast to bacteria, the life cycle of a typical eukaryotic cell is divided into separate events. Next we will see how the timing of separated events of the cell cycle was discovered.

19.3 Cell Division in Eukaryotes

Depending on cell type, typical eukaryotic cells have a cell cycle of sixteen to twenty-four hours, which is divided into **phases**. One of these phases, **mitosis**, was discovered in the late 1800s with light microscopy (you may have seen mitosing onion root-tip cells in a biology class laboratory). In any given cluster of cells, some were seen to lose their nuclei and to form **chromosomes** (from *chroma*, colored; *soma*, bodies). In **mitosis**, paired chromosomes (**chromatids**) are attached at their centromeres. The chromatids were seen to separate and to be pulled apart by **spindle fibers**. Once separated and on their way to opposite poles of dividing cells, they are once again called **chromosomes**.

Thus, homologous chromosomes were equally partitioned to the daughter cells at the end of cell division. Because the same chromosomal behavior was observed in mitosis in diverse organisms, chromosomes were soon recognized as the stuff of inheritance, the carrier of genes! The events of mitosis, lasting about an hour, were seen as occurring in four phases over that short time (Figure 19.2).

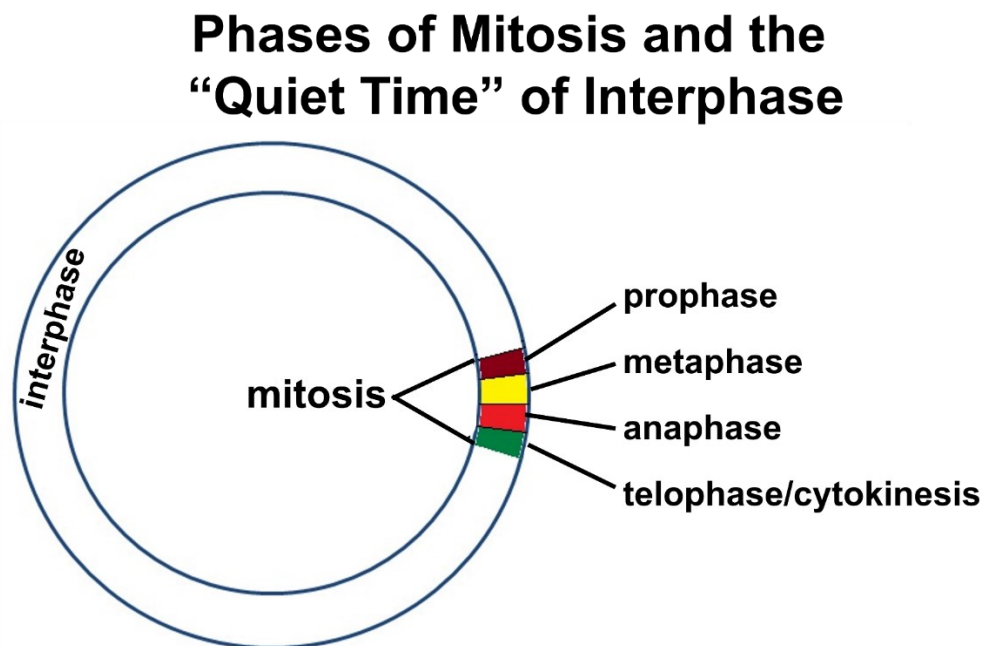


Fig. 19.2: Mitosis and cytokinesis in eukaryotic cells are separated in time, and mitosis is further divisible into five phases.

To the early microscopists, this short period of intense mitotic activity was in stark contrast to a much longer “quiet” time in the life of the cell. For lack of a more descriptive term, they called this period **interphase**. Also, depending on whom you ask, **cytokinesis** (the cell movements that divide a cell in two) might *not* be considered a part of mitosis. In that sense, we can think of three stages in the life of a cell: interphase, mitosis, and cytokinesis. Of course, it turned out that interphase is not cellular “quiet time” at all!

19.3.1 Defining the Phases of the Eukaryotic Cell Cycle

Correlation of the inheritance of specific traits with the inheritance of chromosomes was shown early in the twentieth century, most elegantly in genetic studies of the fruit fly, *Drosophila melanogaster*. At that time, chromosomes were assumed to contain the genetic material, and it was assumed that both were duplicated during mitosis. The first clue that this was not so came only after the discovery that DNA was in fact the chemical *stuff of genes*. The experiment outlined in Figure 19.3 was the first step in working out details of the cell cycle, distinguishing the time of chromosome formation from the time of DNA duplication.

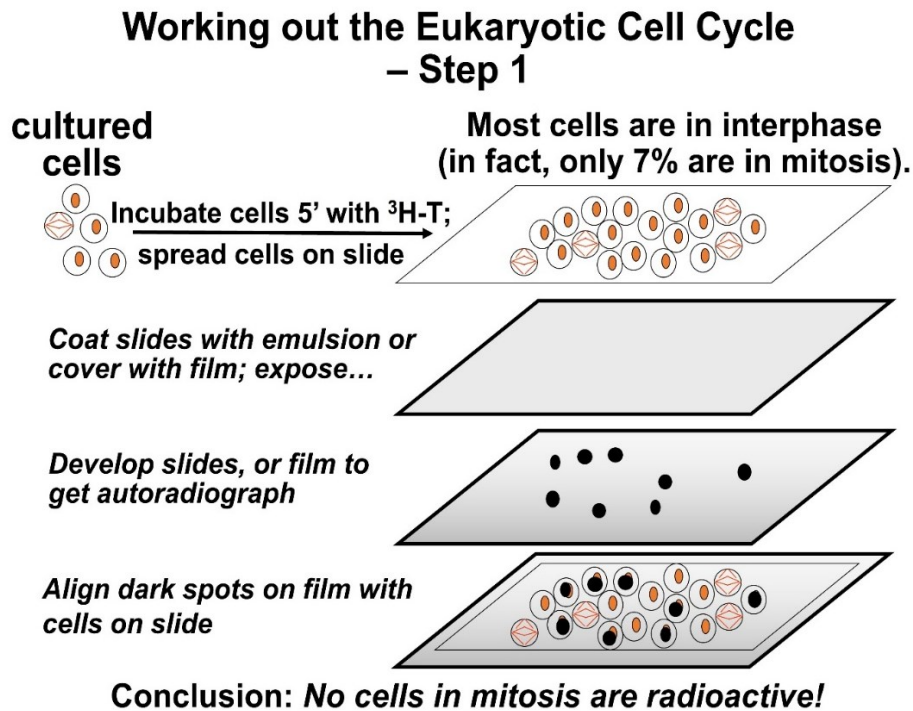


Fig. 19.3: This experiment asked how many cells in a random culture are in mitosis and how many were actively replicating their DNA. The results showed that mitosing cells were not synthesizing DNA.

Here are the details of the experiment:

1. Cultured cells were incubated with ^3H -thymine, the radioactive base that cells will incorporate into deoxythymidine triphosphate (dTTP, one of the four nucleotide precursors to DNA synthesis).
2. After a short period of culture, unincorporated ^3H -thymine was washed away, and the cells were fixed and spread on a glass slide.
3. Slides were dipped into a light-sensitive emulsion containing chemicals like the light-sensitive chemical coat on the emulsion-side of film. (Do you remember photographic film?)
4. After sufficient time to allow the radioactivity on the slide to "expose" the emulsion, the slides were developed (in much the same way as developing photographic film).
5. The resulting **autoradiographs** in the microscope revealed images in the form of dark spots created by exposure to **hot** (i.e., radioactive) DNA.

If DNA replicates in chromosomes undergoing mitosis, then when the developed film is placed back over the slide, any dark spots should lie over the cells in mitosis, and not over cells that

are not actively dividing. But observation of the autoradiographs showed that none of the cells in mitosis were radioactively labeled. On the other hand, some of the cells in interphase were! Therefore, DNA synthesis must take place sometime in interphase, before mitosis and cytokinesis, as illustrated in Figure 19.4.

Replication Must Precede Mitosis

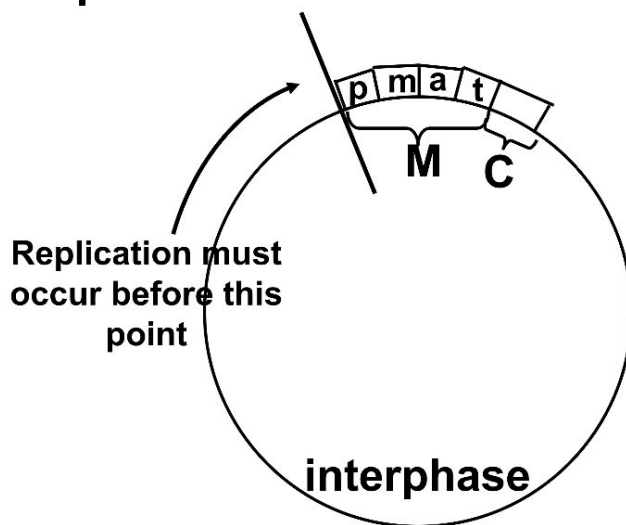


Fig. 19.4: The autoradiographic data from the experiment outlined in Fig. 19.3 demonstrates that DNA synthesis begins and ends some time before the beginning of mitosis.



[340-2 Experiments Revealing Replication in Cell-Cycle Interphase](#)



As we will see, the synthesis of nuclear DNA in eukaryotic cells (replication) occurs many hours before the start of mitosis, and in fact it lasts a good chunk of the time of interphase.

Next a series of **pulse-chase** experiments were done to determine when in the cell cycle DNA synthesis takes place. Cultured cells given a short **pulse** (exposure) to ^3H -thymine and then allowed to grow in a nonradioactive medium for different times (the **chase**). At the end of each chase time, cells were spread on a glass slide and again prepared for autoradiography. Analysis of the autoradiographs identified distinct periods of activity in interphase: **G₁** (*Gap 1*), **S** (a time of DNA *synthesis*), and **G₂** (*Gap 2*). Described here are the details of these very creative experiments, performed before it became possible to synchronize cells in culture so that they would all be growing and dividing at the same time:

1. Cells were exposed to ^3H -thymine for just five minutes (the **pulse**) and then centrifuged. The radioactive supernatant was then discarded.
2. The cells were rinsed and spun again to remove as much labeled precursor as possible.
3. The cells were resuspended in fresh medium containing unlabeled (i.e., nonradioactive) thymine and were further incubated for different times (the **chase** periods).
4. At each chase time, cells were washed free of radioactive precursor, then spread on glass slides.

After dipping the slides in light-sensitive emulsion and exposing and developing the film, the autoradiographs were examined, with the results shown in Figure 19.5.

Working out the Eukaryotic Cell Cycle – Step 2

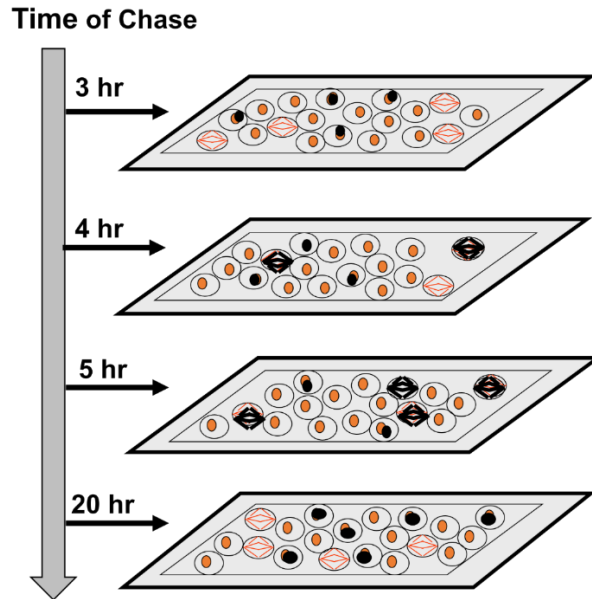


Fig. 19.5: These autoradiographs from a *pulse-chase* labeling experiment led to the identification of the phases of the eukaryotic cell cycle.

Here is a description of the results:

- After a three-hour (or shorter) chase period, the slides looked just as they would immediately after the pulse: that is, while 7% of the cells were in mitosis, none of those were radioactively labeled. In contrast, many *interphase* cells showed labeled nuclei.
- After four hours of chase, a few of the 7% of the cells that were in mitosis were labeled, along with others in interphase.
- After a five-hour chase, most cells in mitosis (still about 7% of cells on the slide) were labeled; many fewer cells in interphase were labeled.
- After a twenty-hour chase, none of the 7% of cells that were in mitosis were labeled. Instead, all the labeled cells were in interphase.

The graph in Figure 19.6 plots a count of *radiolabeled mitotic cells* against *chase times*.

Plot the # of labeled mitotic cells vs chase time:

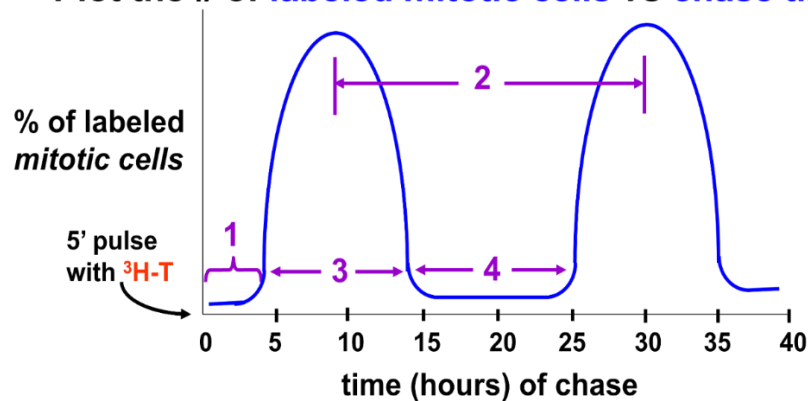
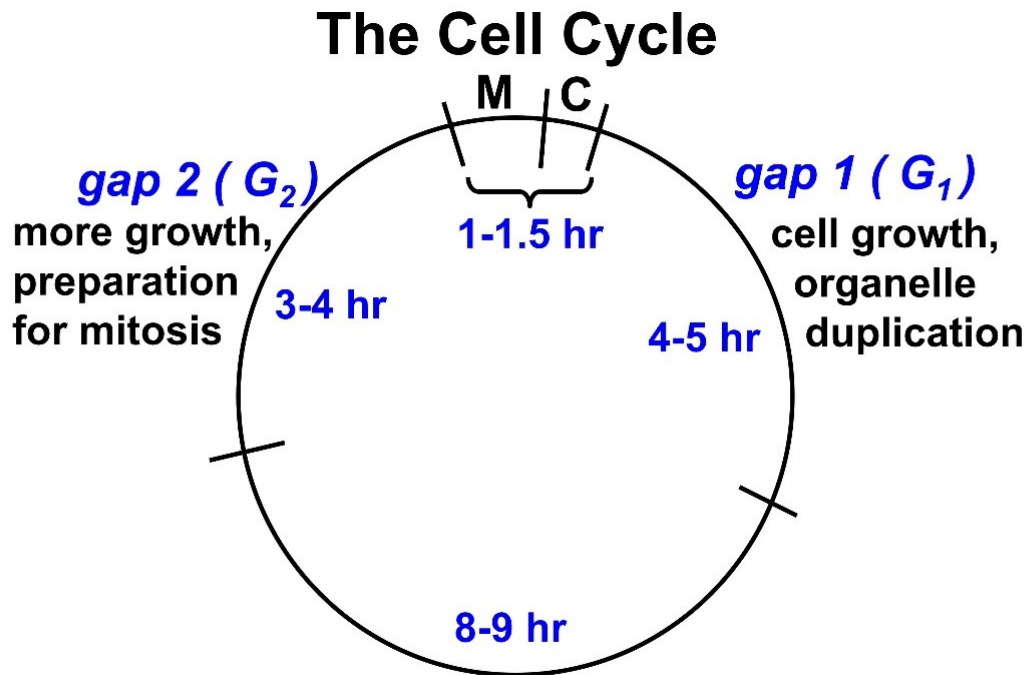


Fig. 19.6: A graph plotting the number of radioactive cells in mitosis over the time of chase in the pulse-chase experiment described in Figure 19.5 and in the text.

The plot defines the duration of events (phases) of the cell cycle as follows:

1. The first phase (interval **#1** on the graph) must be the time between the end of DNA synthesis and the start of mitosis, defined as **Gap 2 (G_2)**.
2. Cell doubling times are easily measured. Assume that the cells in this experiment doubled every twenty hours. This would be consistent with the time interval of twenty hours between successive peaks in the number of radiolabeled mitotic cells after the pulse (interval **#2**).
3. Interval **#3** is easy enough to define. It is the time when DNA is synthesized, from start to finish; this is the **synthesis**, or **S** phase.
4. One period of the cell cycle remains to be defined, but it is not on the graph! That would be the time between the end cell division (i.e., mitosis and cytokinesis) and the beginning of DNA synthesis (**replication**). That interval can be calculated from the graph as the time of the cell cycle (about twenty hours) minus the sum of the other defined periods of the cycle. This phase is defined as the **Gap 1 (G_1)** phase of the cycle.

Events in each phase of a typical eukaryotic cell cycle are summarized in Figure 19.7.



S - DNA synthesis, chromosome duplication, growth

Fig. 19.7: In a typical eukaryotic cell cycle, phases G_1 , S, and G_2 follow cytokinesis and G_2 immediately precedes prophase of mitosis. The length of time for each phase differs for different cell types.

During interphase (G_1 , S, G_2) cells grow in size, preparing for the next cell division. As you might guess, G_1 includes synthesis of enzymes and other proteins needed next for replication.

DNA replicates in the S phase, along with the synthesis of new *histone* and other proteins that will be needed to assemble new *chromatin*. G_2 is the shortest time of interphase and is largely devoted to preparing the cell for the next round of mitosis and cytokinesis. Among the proteins whose synthesis increases in this time are the *tubulins* and proteins responsible for condensing chromatin into paired chromatids that represent the duplicated

chromosomes. *Cohesin* is a more-recently discovered protein made in the run-up to mitosis. It holds centromeres of chromatids together until they are ready to separate.



[341-2 Events in the Phases of the Cell Cycle](#)

Recall that generation times of dividing cells range from sixteen to twenty-four hours. Atypical cells, like newly fertilized eggs, might divide every hour or so! In these cells, events that normally take many hours must be completed in just fractions of an hour.

19.3.3 When Cells Stop Dividing...

Terminally differentiated cells are those that spend the rest of their lives performing a specific function. These cells no longer cycle. Instead, shortly after entering G_1 , they are diverted into a phase called G_0 , as shown in Figure 19.8.

G_0 : When Terminally-Differentiated Cells Stop Dividing

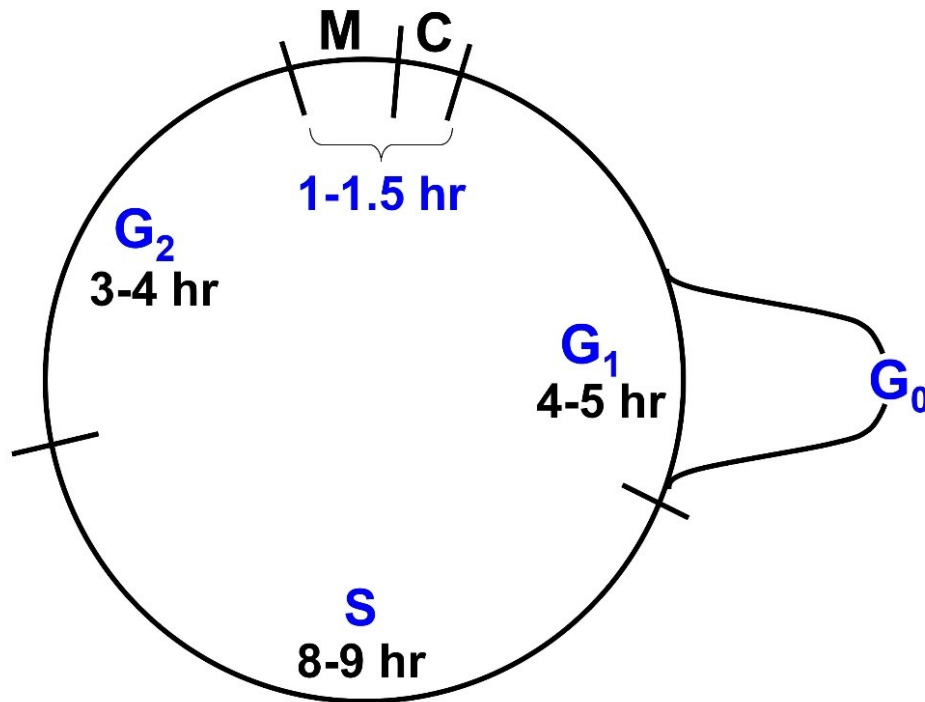


Fig. 19.8: Terminally differentiated cells no longer divide but instead enter the G_0 state. While they can sometimes resume cycling, such cells more typically experience cell death and replacement by stem cells.

Referred to as *terminally differentiated*, these cells normally never divide again. With a few exceptions (e.g., many neurons), most terminally differentiated cells have a finite lifespan, and must be replaced by stem cells. A well-known example is the erythrocyte, or red blood cell. With a half-life of about sixty days, these cells are regularly replaced by precursor reticulocytes produced in bone marrow.

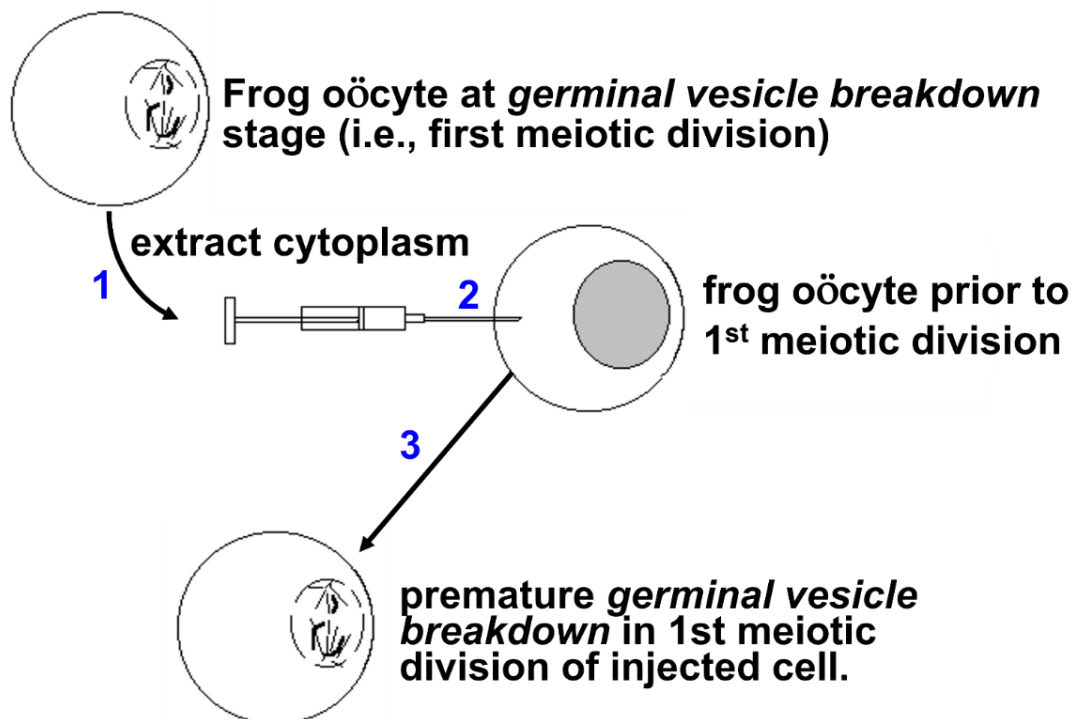
19.4 Regulation of the Cell Cycle

Progress through the cell cycle is regulated. The cycle can be controlled or put on “pause” at any one of several phase transitions. Such **checkpoints** monitor whether the cell is on track to complete a successful cell-division event. Superimposed on these controls are signals that promote cell differentiation. Embryonic cells *differentiate* as the embryo develops. Even after *terminal differentiation* of cells that form all adult tissues and organs, *adult stem cells* will divide and differentiate to replace worn out cells. Once differentiated, cells are typically signaled in G_1 to enter G_0 and to stop cycling. In some circumstances cells in G_0 are recruited to resume cycling. However, if this resumption should occur by mistake, the cells may be *transformed* to cancer cells. Here we consider how the normal transition between phases of the cell cycle is controlled.

19.4.1 Discovery and Characterization of *Maturation Promoting Factor (MPF)*

To monitor their progress through the cell cycle, cells produce internal chemical signals that tell them when it's time to begin replication or mitosis, or even when to enter into G_0 when they reach their terminally differentiated state. The experiment that first demonstrated a chemical regulator of the cell cycle involved fusing very large frog's eggs! The experiment is described in Figure 19.9.

First isolation of a cytoplasmic factor from *Xenopus* eggs that control the cell cycle:



Conclude: a cytoplasmic factor induces *Maturation*

Fig. 19.9: Experiment leading to the discovery of *meiosis-promoting factor (MPF)*, the first known chemical regulator of cell division.

The hypothesis tested here was that frog oocyte cytoplasm from *germinal vesicle-stage* oocytes (i.e., oocytes in mid meiosis) contain a chemical that causes the cell to lose its nuclear membrane, to condense its chromatin into chromosomes, and to enter meiosis. The test was to withdraw cytoplasm from one of these large, mid meiotic oocytes with a fine hypodermic needle and to inject it into a pre meiotic oocyte. The result of the experiment was that the mid meiotic oocyte cytoplasm induced premature meiosis in the immature oocyte.

Subsequently, a *Maturation Promoting Factor (MPF)* was isolated from the mid meiotic cells—a factor that when injected into pre meiotic cells, would cause them to enter meiosis. *MPF* turns out to be a protein kinase made up of two polypeptide subunits (Figure 19.10, below). *MPF* was subsequently shown to stimulate *somatic cells* in *G₂* to enter premature mitosis. So (conveniently) *MPF* can also be *Mitosis Promoting Factor*! Hereafter we will discuss the effects of an *MPF* as being equivalent in mitosis and meiosis. When active, *MPF* targets many cellular proteins.

Maturation Promoting Factor (MPF) could be purified from meiotic frogs' eggs...

MPF is a protein kinase

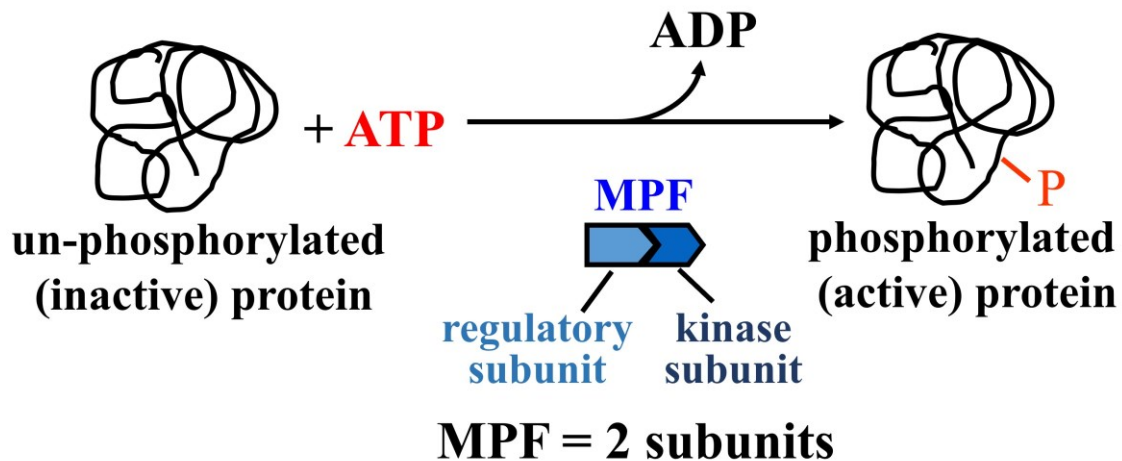


Fig. 19.10: MPF was shown to be a two-subunit *protein kinase* that transfers phosphates from ATP to several different proteins.



[342 Discovery of MPF Kinase and Its Role in Meiosis and Mitosis](#)



One subunit of *MPF* is *cyclin*. Cyclin was so named because its levels rise gradually after cytokinesis, peak at the next mitosis, and then fall. The other subunit, *cyclin-dependent kinase (cdk)*, contains the kinase-enzyme *active site*. Both subunits must be bound to make an active kinase. Assays of *MPF* activity as well as the actual levels of the two subunits over time during the cell cycle are graphed in Figure 19.11.

Regulation of MPF Activity

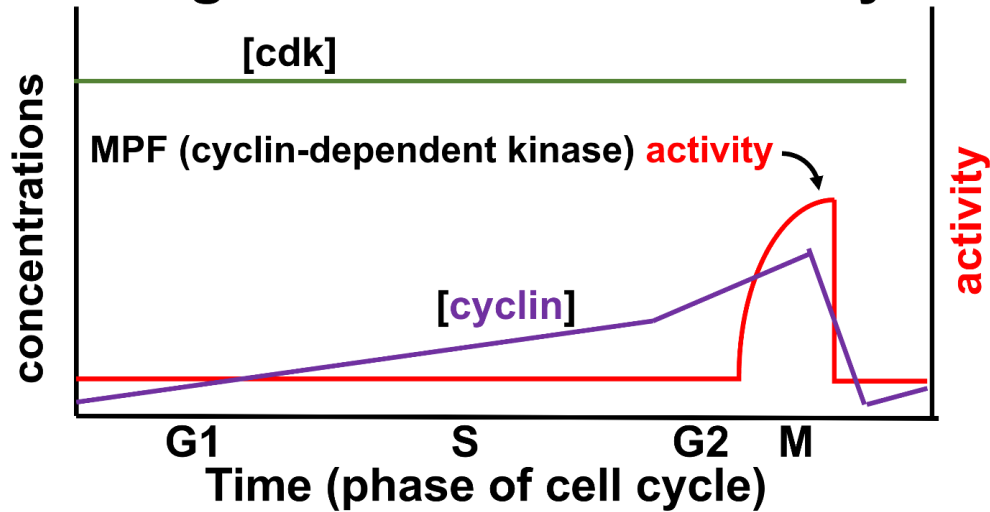


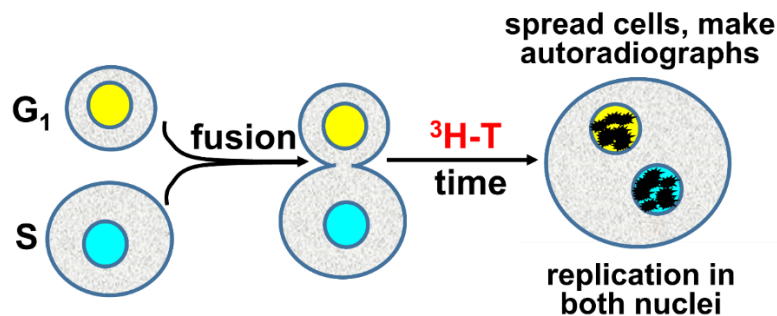
Fig. 19.11: Graph comparing plots of cellular *cdk* and *cyclin* levels over time with one of *MPF* activity.

Levels of the cdk subunit do not change significantly during the life of the cell. Because the kinase activity of MPF requires *cyclin*, it tracks the rise in cyclin near the end of the G₂, and its fall after mitosis. Cyclin begins to accumulate in G₁, rising gradually and binding to more *cdk* subunits. *MPF* reaches a threshold concentration in G₂ that triggers entry into mitosis. For their discovery of these central molecules, L. H. Hartwell, R. T. Hunt, and P. M. Nurse won the 2001 Nobel Prize in Physiology or Medicine.

19.4.2 Other Cyclins, CDKs, and Cell-Cycle Checkpoints

Other chemical signals accumulate at different points in the cell cycle. For example, when cells in S are fused with cells in G₁, the G₁ cells begin synthesizing DNA (visualized as ³H-thymine incorporation). Figure 19.12 describes an experiment showing control of progress to different phases of the cell cycle.

Cell Fusion Experiments Reveal Control at Other Phases of the Cell Cycle



Conclusion: S phase 'factors' induce G₁ cell replication (i.e., to begin the S phase).

Fig. 19.12: Cell-fusion experiments revealed additional chemical regulators of the cell-cycle activity.

An *S* phase factor could be isolated from the *S* phase cells. This factor also turns out to be a two-subunit protein kinase, albeit a *different* one from MPF. Just as MPF signals cells in *G*₂ to begin mitosis, the *S*-phase kinase signals cells in *G*₁ to enter the *S* phase of the cell cycle and to start replicating DNA. *MPF* and the *S*-phase kinase govern activities at two of several cell-cycle **checkpoints**. In each case, the activity of the kinases is governed by prior progress through the cell cycle. In other words, if the cell is not ready to begin mitosis, active *MPF* production is delayed until it is. Likewise, the *S*-phase kinase will not be activated until the cell is ready to begin DNA synthesis.



[343 Cell-Cycle Control at Check Points and the *G*₀ "Phase"](#)

Control of progress through the cell cycle is more intricate than we currently know, but the best-described checkpoints separate *G*₁, *G*₂, and *M* (Figure 19.13).

Cell Cycle Checkpoints

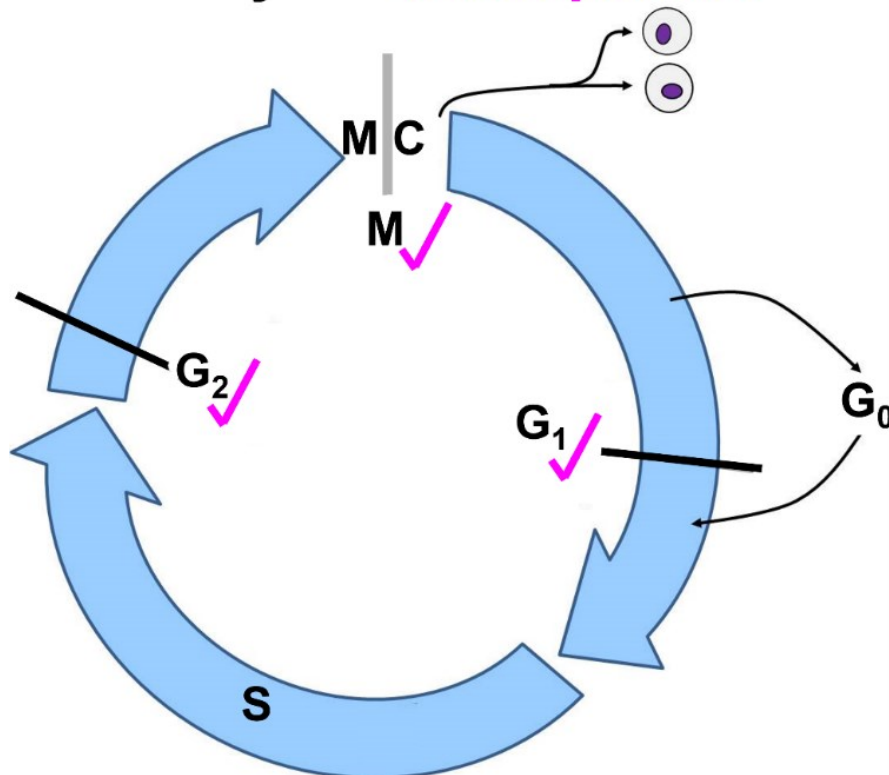


Fig. 19.13: Simplified diagram of cell-cycle checkpoints at which progress through the cycle is assessed. If progress through a phase is incomplete, cell-cycle regulators (kinases) delay onset of the next phase.

We generally envision checkpoints as monitoring and blocking progress until essential events of a current phase of the cell-cycle phase are completed. These kinases are part of molecular sensing mechanisms that act by phosphorylating cytoplasmic and/or nuclear proteins required by upcoming phases of the cycle. Let's look at some events that are monitored at these *checkpoints* in more detail.

19.4.2.a The G₁ Checkpoint

The **G₁ checkpoint** controls the transition from the G₁ to the S phase of the cell cycle. If actively dividing cells (e.g., stem cells) in G₁ fail to complete their preparation for replication, the S-phase kinase won't be produced, and the cells won't proceed to the S phase until the preparatory biochemistry catches up with the rest of the cycle. To enter S, a cell must be ready to make proteins of replication, like DNA polymerases, helicases, and primases, among others. Only when these molecules have accumulated to (or become active at) appropriate levels is it "safe" to enter S and begin replicating DNA.

Now, what about those cells that are fully differentiated? *Terminally differentiated cells* stop producing the active **G₁-checkpoint** kinase and stop dividing. Thus, they are arrested in **G₀** (see section 19.4.2.d).

19.4.2.b The G₂ Checkpoint

Passage through the **G₂ checkpoint** is only possible if DNA made in the prior **S** phase is not damaged. Or if it was, then that damage has been (or can be) repaired. (Review the proofreading functions of DNA polymerase and the various DNA-repair pathways.) Cells that do successfully complete replication and pass the **G₂ checkpoint** must prepare to make the proteins necessary for the upcoming mitotic phase. These include nuclear proteins necessary to *condense* chromatin into chromosomes, *tubulins* for making *microtubules*, and more. Only when levels of these and other required proteins reach a threshold can the cell begin mitosis.

Consider the following two tasks required of the **G₂ checkpoint** (in fact, any checkpoint):

- Sensing whether prior phase activities have been successfully completed
- Delaying transition to the next phase if those activities are unfinished

But what if sensing is imperfect and a checkpoint is leaky? A recent study suggests that either the **G₂ checkpoint** *is* leaky, or at least, that incomplete activities in the **S** phase are tolerated, and that some DNA repair is not resolved until mitosis is underway in M! See more about this at ¹⁹⁻¹ [DNA repair at mitosis](#).

19.4.2.c The M Checkpoint

The **M** checkpoint is monitored by the original MPF-catalyzed phosphorylation of proteins, which do the following:

Bind to chromatin, causing it to condense and to form chromatids

- a) Lead to the breakdown of the nuclear envelope
- b) Enable spindle-fiber formation and their attachment to chromatids
- c) Lead to the onset of mitosis

We have seen that the tension in the spindle apparatus at metaphase tugs at the kinetochores holding the duplicated chromatids together. As this tension reaches a threshold, MPF peaks and an activated **separase** enzyme catalyzes the separation of the chromatids at their centromeres. Starting in *anaphase*, continuing tension in the spindle apparatus draws the new chromosomes (mostly by their centromeres) to opposite poles of the cell. Near the end of

mitosis and cytokinesis, proteins phosphorylated by MPF initiate the breakdown of cyclin in the cell. Passing the **M** checkpoint means that the cell will complete mitosis and cytokinesis, and that each daughter cell will enter a new **G₁** phase.

During cell division, yeast cells seem to have the three checkpoints discussed here. More-complex eukaryotes use more *cyclins* and *cdks* to control the cell cycle at more checkpoints. Different *cyclins* show cyclic patterns of synthesis, while their *cdks* remain at constant levels throughout the cell cycle (as in *MPF*). Although the different *cdk* and *cyclin* gene families are evolutionarily conserved, each *cyclin/cdk* pair has been co-opted in evolution to monitor different cell-cycle events and to catalyze phosphorylation of phase-specific proteins. To learn more, see Elledge S. J. (1996) *Cell Cycle Checkpoints: Preventing an Identity Crisis*. Science 274:1664-1672.



[344-2 Cyclin/cdk Checkpoints for Cell-Cycle Phases](#)

19.4.2.d The **G₀** State

G₀ is not really a phase of the cell cycle, since the cells in **G₀** have reached a terminally differentiated state and have stopped dividing. For example, the terminally differentiated cells that make up mature tissues and organs no longer divide. But these cells do have a finite half-life. Recall that our red blood cells lack a nucleus. In circulation, they have a half-life of about 60 days and old erythrocytes are removed from circulation at 100-120 days. It is stem cells in bone marrow that replace aging red cells.

Most of our cells retain their nuclei, but when terminally differentiated cells in **G₀** age, they too are replaced by *adult stem cells*, which can still divide and differentiate. Life-time length in **G₀** can vary. Some cells live so long in **G₀** that they are nearly never replaced (muscle cells, neurons). Others live short lives in **G₀** (e.g., stem cells and some embryonic cells).

Lymphocytes are differentiated *immune-system* cells that can reenter **G₁**. Exposure of lymphocytes to foreign chemicals or pathogens activates **mitogens**, causing the cells to exit **G₀** to start dividing, and to produce the legions of cells that make antibodies that neutralize toxins and fight off pathogens. The **retinoblastoma** (**Rb**) protein is *mitogen*, a transcription factor that turns on genes leading to cell proliferation.

But what if cells resume cycling when they shouldn't? What if they are inappropriately signaled to exit **G₀**? Such cells are in trouble! Having escaped normal controls on cell division, they can become a focal point of cancer-cell growth. You can guess from its name that the retinoblastoma gene was discovered as a mutation that causes retinal cancer. For more about normal Rb protein function, and its interaction with a **G₁** *cdk*, check out the following link.



[345-2 Rb Gene Encodes the regulatory subunit of a Transcription Factor](#)

19.5 When Cells Die

As noted, few cell types live forever; most are destined to turn over (another euphemism for dying), mediated by **programmed cell death**, or **apoptosis**. Apoptosis occurs in normal development when cells are only temporarily required for a maturation process (e.g., embryonic development and metamorphosis). When older cells are no longer necessary or when cells are genetically damaged, they too are detected and signaled to undergo **apoptosis**. Frog metamorphosis is a well-known example of apoptosis. Signaled by **thyroid hormone**, tadpoles digest their tail cells, reabsorbing and recycling the digestion products. These in turn serve as nutrients to grow adult frog structures. Normal and apoptotic cells are compared in micrographs and drawings in Figure 19.14.

Normal (A) vs. Apoptotic (B, C) Cells

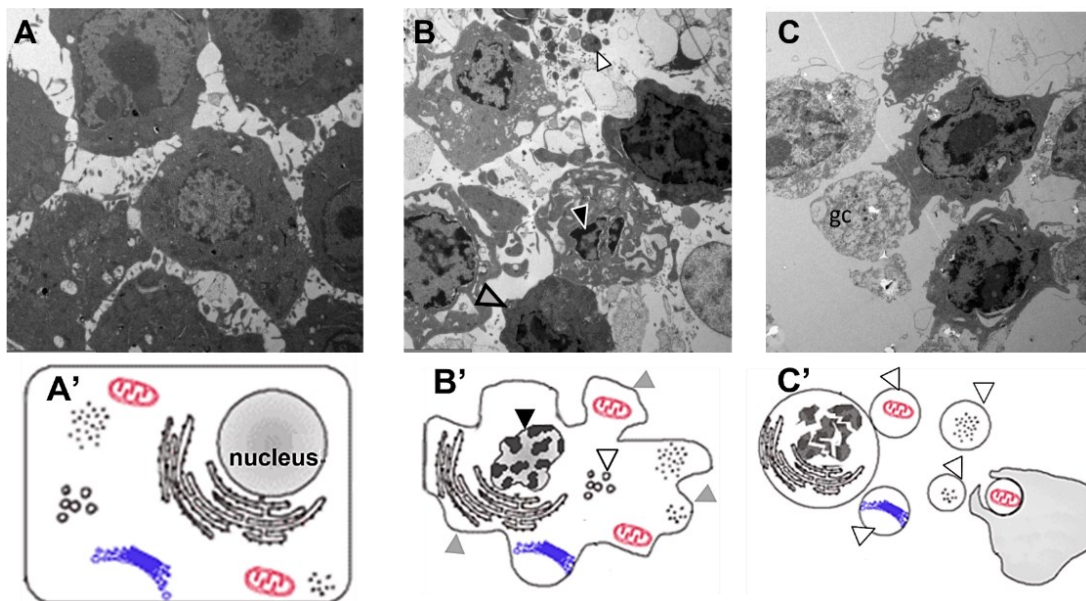


Fig. 19.14: Electron micrographs and corresponding drawings compare normal (A,A') with apoptotic cells (B,B' and C,C'). In (B,B'), a black arrowhead indicates *pyknosis* (nuclear condensation), gray arrowheads indicate membrane *blebs*, and a white arrowhead indicates apoptotic bodies. Later in apoptosis, nuclei break down (*karyorrhexis*, lower right in C; upper left in C'), leaving ghost cells (gc in C), and apoptotic bodies (arrowheads, C'). A phagocyte is engulfing an apoptotic body containing a mitochondrion in C'.

Apoptosis is first associated with nuclear fragmentations and cell shrinkage ([B, B'] and [C, C'] in Figure 19.14). Apoptotic bodies in the cell leave the cell in or as "blebs" (evaginations) where they may be engulfed by phagocytes. Look at ¹⁹⁻²[Cancer Cell Apoptosis](#) to see a time-lapse movie of live prostate cancer cells that were chemically induced to undergo apoptosis. Look for the formation of numerous apoptotic bodies; you might also see some nonapoptotic cells undergoing cytokinesis!

We now understand the basic metabolic pathway leading to apoptosis. During development, apoptotic cell death starts with an external signal programmed to appear at a specific time in development. Similar signals appear in response to cell obsolescence or genetic damage. The biochemical pathway to apoptosis is illustrated below in Figure 19.15.

***Bak* & *Bax* (*Bcl2*) make up outer mitochondrial membrane channels that when signaled, release **cyt C** into the cytosol:**

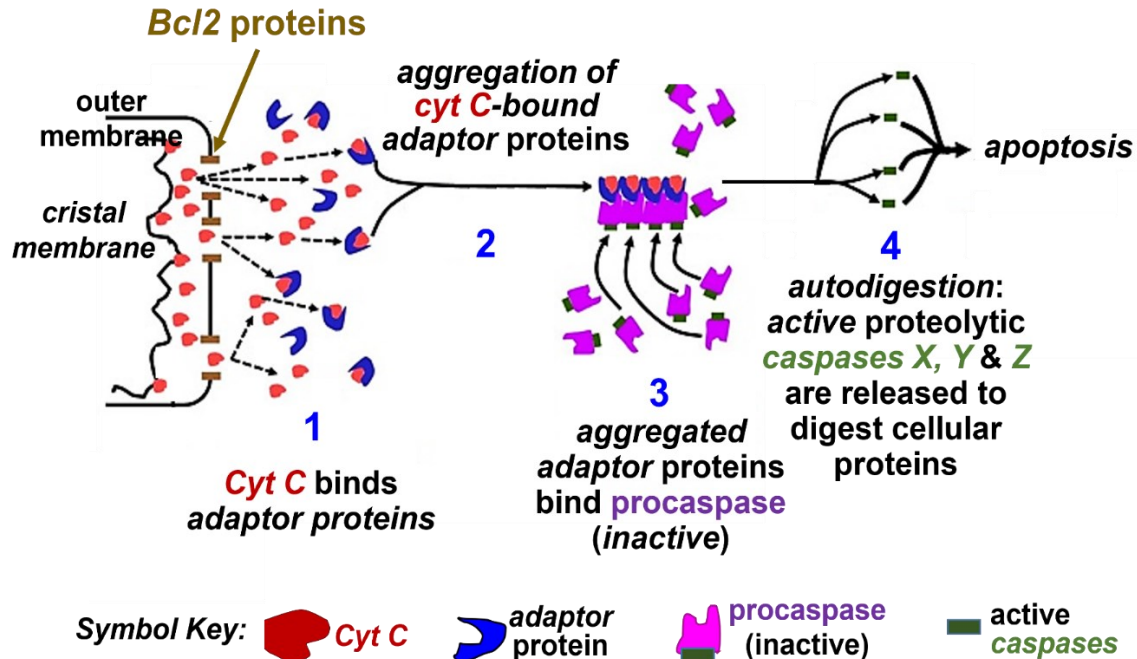


Fig. 19.15: Mitochondrial proteins play a role in the biochemistry of apoptosis (*programmed cell death*).

The signal molecule acts on target cells to induce transcription of ***Bcl2*** genes. *Bcl2* proteins ***Bak*** and ***Bax*** are outer-mitochondrial-membrane channel components that allow the release of *cytochrome C* into the cytoplasm. This sets off molecular events leading to *apoptosis*.

Mitochondrial exit of cytochrome C is possible because it is a *peripheral* membrane protein, only loosely bound to the cristal membrane. It exists in equilibrium between membrane-bound and unbound states. As some cytochrome C molecules exit the intermembrane space, others detach from the cristal membrane and follow.

In the cytosol, cytochrome C binds to **adaptor** proteins, which then aggregate. The cytochrome C-adaptor complex has a high affinity for a biologically **inactive procaspase**. Binding of *procaspase* to the *cytochrome C*-adaptor complex causes an allosteric change in the *procaspase*, releasing an active **caspase**.

Caspases are *proteolytic enzymes* that start the autodigestion of the cell. For their work in identifying apoptosis genes, Sydney Brenner, H. Robert Horvitz, and John E. Sulston shared the 2002 Nobel Prize in Physiology or Medicine.

In contrast to apoptosis, cells that die unexpectedly from external injury undergo ***necrosis***, an *accidental* rather than a programmed death. Key differences between the two cell deaths are summarized below in the illustration below (Figure 19.16).

Comparison of *Necrosis* (the dying of old, worn-out cells) and *Programmed cell Death (Apoptosis)*

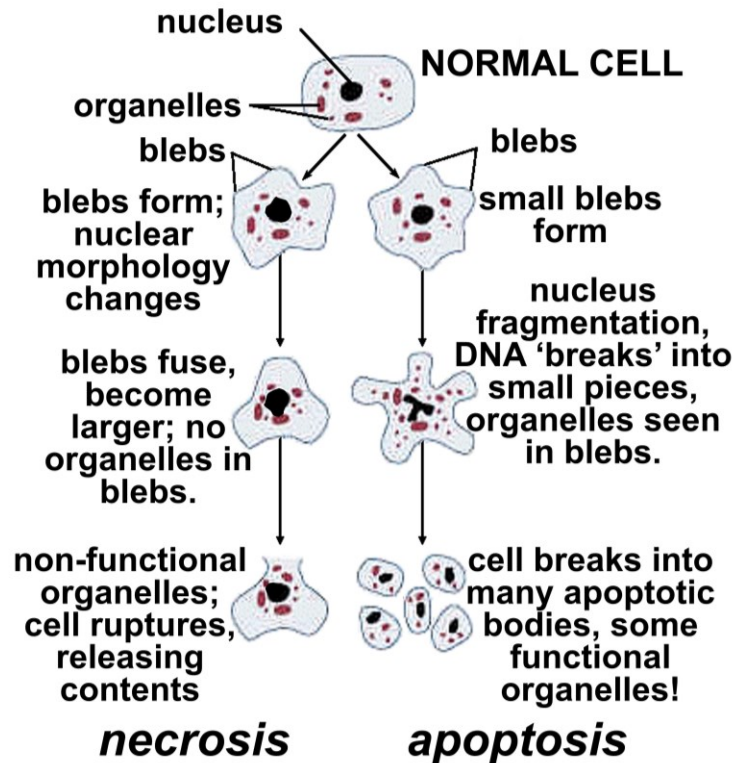


Fig. 19.16: Comparison of the cellular events of apoptosis and necrosis.

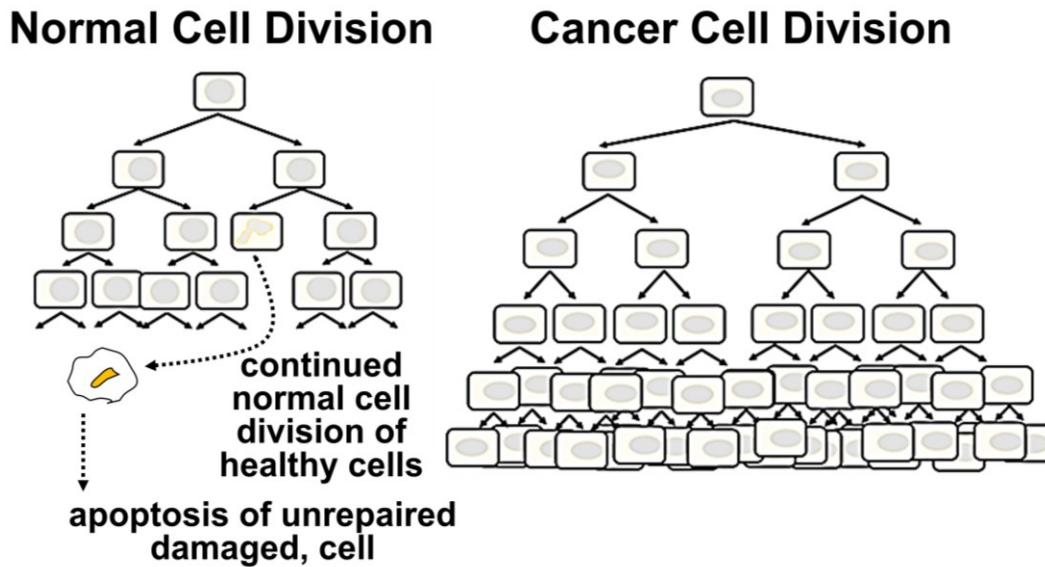
From the illustration,

- The nuclei of necrotic cells do not shrink or fragment while plasma membranes extend processes to form **blebs** and eventually burst, releasing the cell contents.
- The nuclei of apoptotic cells shrink and fragment while the plasma membrane **blebs** also form, enclosing organelles and eventually fusing. The organelles can remain functional for a time.

19.6 Disruption of the Cell-Cycle Checkpoints Can Cause Cancer

As we've noted, cycling cells continue to divide until they attain G_0 in the terminally differentiated state. Most terminally differentiated cells are cleared by *apoptosis* when they reach the end of their effective lives, to be replaced by stem cells. We also noted that accidental signaling can bring cells out of G_0 , leading to renewed cell proliferation. Even though such cells are abnormal, their defects may not be detected by apoptotic defense mechanisms. As a result, they undergo uncontrolled cell divisions, becoming cancer cells. Likewise, if a checkpoint fails or if a cell suffers physical damage to chromosomes during cell division, or if it suffers a debilitating somatic mutation in a prior S phase, it may self-destruct in response to a consequent biochemical anomaly. This is another example of *apoptosis*. But if a physically damaged or mutated cell should escape apoptotic clearance, it too may begin replication and

proliferation, producing cancer cells. Apoptotic clearance and uncontrolled cancer-cell proliferation are compared in Figure 19.17 (below).



19.17: Results of noncancer-cell divisions, in which abnormal cells are targeted for apoptosis (left), and cancer-cell division, in which damaged or altered cells escape apoptotic controls (right).



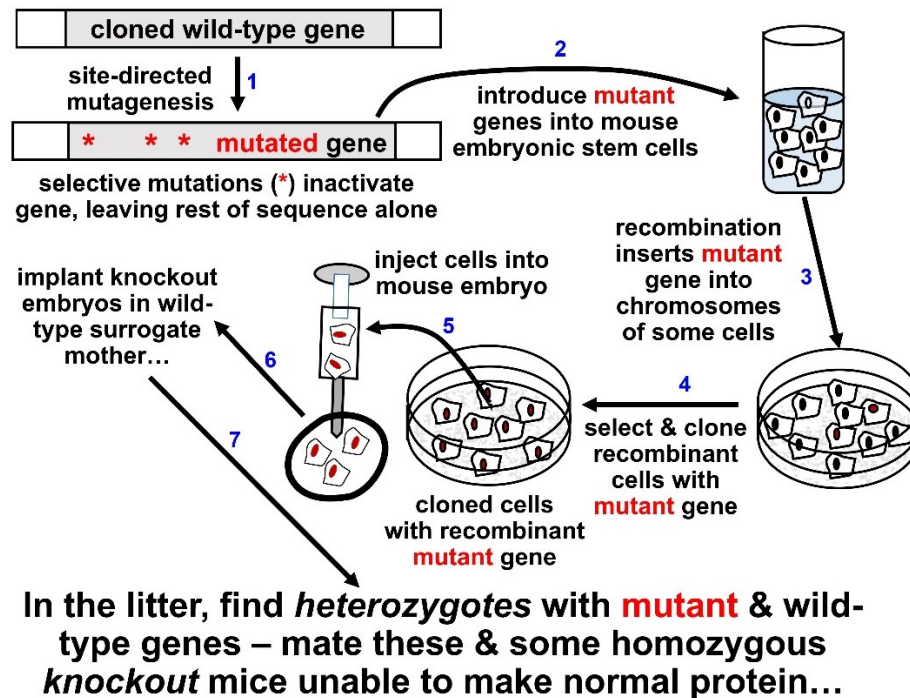
[346-2 Regulating Cell Death—Apoptosis vs Necrosis](#)

19.7 p53 Protein Mediates Normal Cell-Cycle Control

Cancerous growth can result if a normal dividing cell suffers a somatic mutation that disrupts normal cell-cycle control. The overexpression of *cdk*, for example, or higher *cyclin* levels in daughter cells that never drop; such cells would never stop cycling.

Other possibilities include a cell in G_0 that is stimulated to begin cycling again by an inappropriate encounter with a hormone or other signal. If undetected, these anomalies can transform normal cells to cancer cells. This is apparently the case for a mutation in the human *p53* gene that causes *Li-Fraumeni syndrome*. The symptoms include a high incidence of early-age bone, blood, and breast cancers (see ¹⁹⁻³[Li-Fraumeni Syndrome](#) for more). How the *p53* protein in regulating cancer cell formation was shown using *knock-out mice*. We noted earlier the use of *knock-out mutations* in studies of specific gene function. A protocol for making knock-out mutations in mice is shown below in Figure 19.18.

Making a *Knockout* Mouse: an Overview



Fig, 19.18: How to make knock-out mutant mice: Specific deoxynucleotides in a cloned gene of interest are altered by **site-directed mutagenesis**. The altered gene is inserted into embryonic stem cells, where it recombines with and replaces the homologous gene already in the chromosome. Recombinant clones are selected and reinserted into embryos, which are then incubated in the uteri of “foster mother” mice. Knockout mice in the newborn litter can be selected and studied for the effects of removing a gene.

For demonstrating how knock-out organisms can be genetically engineered, M. R. Capecchi, M. J. Evans, and O. Smithies shared the 2007 Nobel Prize in Physiology or Medicine. (Read all about it at ¹⁹⁻⁴[The Knockout Nobel Prize.](#))

The role of **p53** protein in regulating cancer-cell formation in mice was shown in a knock-out experiment, when the normal mouse *p53* genes were replaced with a pair of mutant *p53* genes that were unable to make functional **p53 protein**. The resulting **knock-out mice** developed tumors and cancers at a much higher rate than normal mice. In general, many **knock-out** organisms have been created to study the effects of **loss-of-function** mutations in vivo (i.e., in whole animals). Using similar strategies, **knock-in** protocols can insert normal genes into an organism lacking the genes, allowing study of the effects of **gain-of-function** mutations.

19.7.1 p53 is a DNA-Binding Protein

If a **G₀** cell is stimulated to start cycling again by an inappropriate encounter with a hormone or other signal, it may be transformed into a cancer cell. The job of the **p53 protein** is to detect such anomalies and to enable dividing cells to repair the damage before they proceed through the next cell-cycle checkpoint. Failing that, the p53 protein signals apoptosis of the cell.

p53 is a DNA-binding, gene-regulatory protein. Figure 19.19 (below) shows the structure of the p53 protein and how it binds to the DNA double helix.

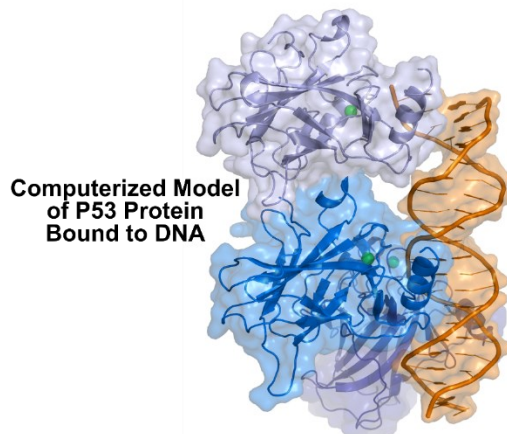


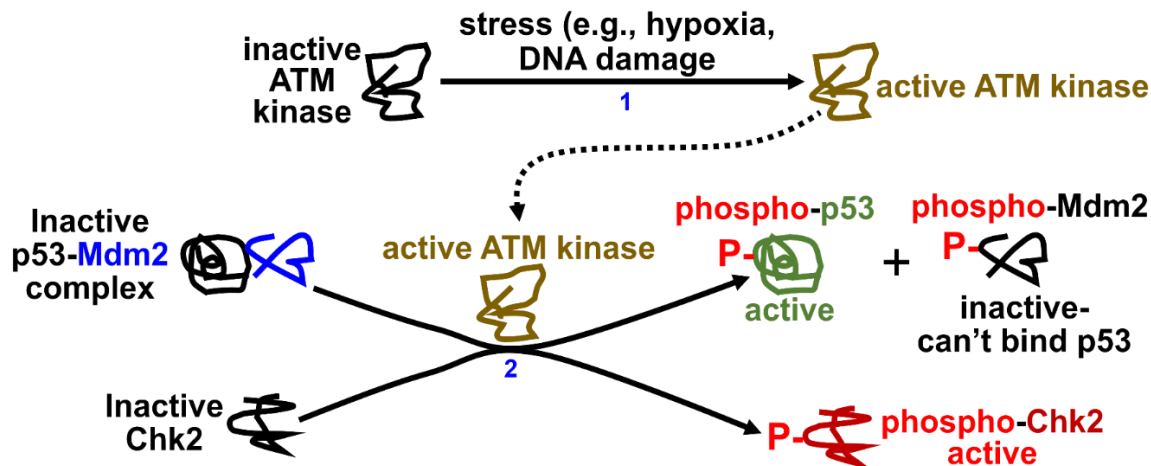
Fig. 19.19: Structure of the *p53* gene-regulatory protein bound to DNA. p53 was originally called a *tumor suppressor protein* because when mutated, tumors arose.

Mutations in the gene for the *p53* protein in humans (where it is called ***TP53***) are associated with many human cancers. Cultured cells of mutagenized *p53* genes also exhibit key characteristics of cancer cells, including uncontrolled replication and cell proliferation and the suppression of apoptosis.

19.7.2 How p53 Works to Salvage Cells

p53 protein is normally bound to an active ***Mdm2*** protein. The *p53* must separate from *Mdm2* for *p53* to enable cell progress through checkpoints. Physical or chemical stress on dividing cells, such as DNA damage during cell growth, can activate an ***ATM* kinase** that will phosphorylate *p53* and *Mdm2*, causing their dissociation *p53*. The same kinase also phosphorylates ***Chk2***. ATM-kinase-initiated events are detailed in Figure 19.20.

Stress-Activated **ATM** Kinase Phosphorylates *p53*, *Mdm2* and *Chk2*



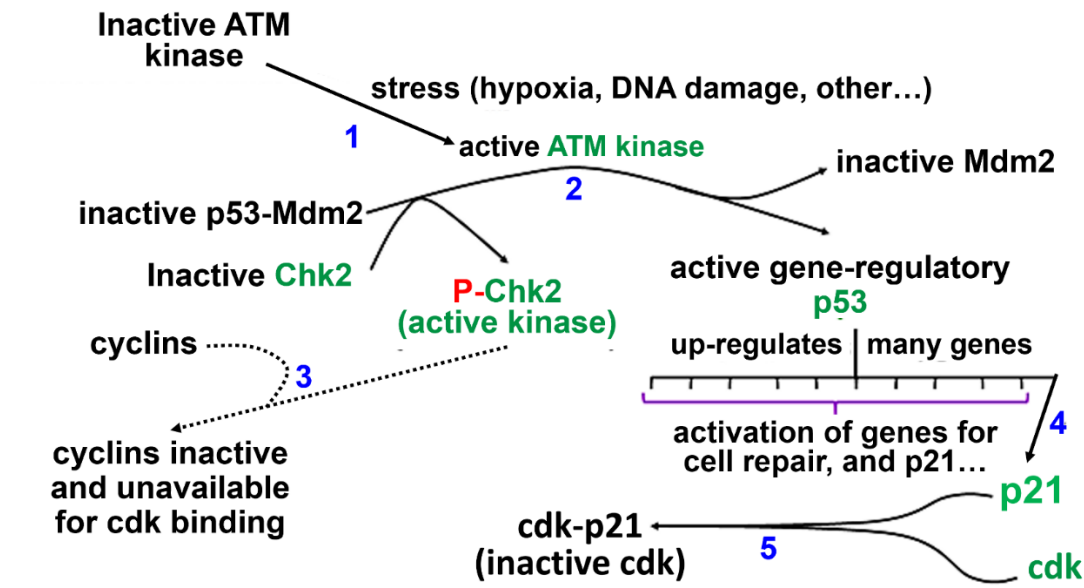
19.20: Normal functions of the *p53* gene-regulatory protein

Each of the proteins and enzymes phosphorylated by the ATM kinase has a role in cell-cycle checkpoint function and cell-cycle arrest while errors are corrected:

1. Now separated from Mdm2, active **phospho-p53** upregulates several genes, including the **p21** gene.
2. The **p21** protein binds to **cdks**; **cyclins** can't bind **p21-cdks**.
3. Active **phospho-Chk2** catalyzes cyclin phosphorylation; **phospho-cyclins** can't bind to **p21-cdks**.
4. The inability of **cyclins** to bind **cdks** specifically blocks the cell cycle between the **G₁** and **S** phases and the **G₂-to-M** phases.

Figure 19.21 illustrates the kinase-mediated events at cell-cycle checkpoints.

P53 and Chk2 Arrest Progress Through the Cell Cycle at a Checkpoint



...with no active **cyclin-cdk** formation, cell cycling stops

19.21: Role of the p53 in decision-making at cell-cycle checkpoints.

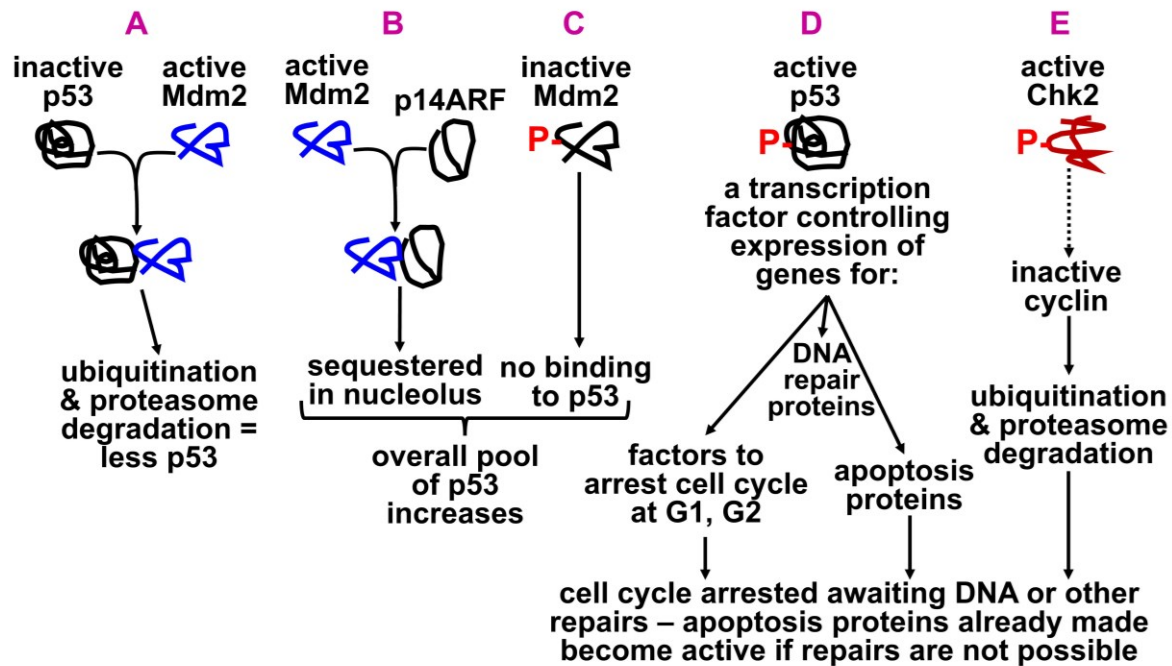
The cell cycle will remain arrested while the cell attempts to finish the essential biochemical activities necessary to correct stress-induced or other physical or chemical aberrations, before moving on to the next phase of the cycle. If DNA repairs or other corrections are successful, the cell can progress to the next phase. But what happens if repairs are unsuccessful?

19.7.3 How p53 Works When Cells Can't Be Saved

If DNA repair or other corrections fail, the damaged cell is routed to apoptosis. In this case, **proteasomes** target the **Chk2-cyclin** complex for degradation. Likewise, any **p53** remaining bound to unphosphorylated **Mdm2** is also targeted for proteasome destruction. The result is that the cell, unable to correct effects of stress or chemical damage or to repair DNA damage,

is now a target for **apoptosis**. Levels and activity of **p53** (and the other proteins) control the amount of **p53** protein available to respond to cell-cycling anomalies. When phosphorylated, active **p53** rapidly arrests the cell cycle and turns on genes for proteins needed for DNA repair *and* for apoptosis (just in case repair efforts fail!). Figure 19.22 summarizes these and related **p53** interactions and their consequences.

Activities and the Consequences of p53, Mdm2 & Chk2 Regulation by Phosphorylation and de-Phosphorylation



19.22: **A:** Inactive p53 & active Mdm2 bind, head for proteasome, & [p53] decreases; **B:** Active Mdm2 & p14ARF bind, both are sequestered in the nucleolus, increasing available [p53]; **C:** Inactive Mdm2 can't bind p53, increasing available p53. **D:** If ATM kinase is active, available p53 is phosphorylated & active p53 turns on genes for *DNA repair*, *apoptosis* & *protein factors* that arrest cell cycling at G1 or G2; **E:** Phosphorylated Chk2 activity causes cyclin inactivation, sending it to the proteasome, blocking cell cycling.

To sum up, **p53** suppresses malignant tumor growth by either of two methods:

- by allowing DNA or other cellular repair, after which **p53** and other proteins are inactivated and/or destroyed, and the cell cycle can resume.
- by upregulating apoptosis genes in case repairs aren't successful leading to apoptosis, thereby killing off damaged cells and blocking tumorigenesis.

This is why a mutant **p53** that reduces or eliminates **p21**-protein production, blocking essential DNA-repair, causes damaged cells to enter **S**, transforming them into cancer cells.

In a new twist, few whales or elephants die from cancer compared to humans, even though they have thousands of times more cells. At least for elephants, one reason may be that they have as many as twenty copies (40 alleles) of their p53 genes! Thus, mutation of one p53 allele may have little effect if the tumor-repressing effects of the remaining p53 genes prevail (¹⁹⁻⁵[Whales and Elephants Don't Get Cancer!](#)). Now it turns out that duplication of more than a few tumor suppressor genes not only protect large mammals from cancer, but originated in small animals, perhaps facilitating the evolution of large ones (¹⁹⁻⁶[Elephantine Evolution](#)).

19.8 The Centrality of p53 Action in Cell-Cycle Regulation

Because of its multiple roles in regulating and promoting DNA repair, and in controlling cell-cycle checkpoints, p53 has been called "*the Guardian of the Genome*"! Here is further evidence of this central role.

19.8.1 "Oncogenic Viruses"

Cancer-causing viruses include Human Papilloma Virus (HPV), Epstein-Barr Virus (EBV), human immunodeficiency virus (HIV), Hepatitis B and C viruses (HBV, HCV), Human Herpes Virus-8 (HHV-8), and Simian Virus 40 (SV40). The oncogenic SV40 virus was discovered because it was a contaminant of polio vaccines which were used in the 1960s.

While an association of SV40 and cancer in humans is yet unproven, there is a demonstrated link between SV40, p53, and cancer in other mammals. After an infection of cells by SV40, viral DNA enters the nucleus, where it can integrate into the host cell genome. Such SV40 infections are usually *latent*, meaning that they cause no harm. But when activated, they can lead to cellular transformation and the growth of malignant sarcomas in muscles, as well as tumors in other organs. If activated in an infected cell, the host-cell RNA polymerase II transcribes the SV40 genes to produce enzymes that replicate viral DNA and the viral structural proteins that encapsulate the DNA in a membrane envelope to make new viral particles. Since the relatively small SV40 genome does not encode all of the enzymes and factors needed for viral DNA replication, the infected cells also provide these factors, producing them only during the S phase. At that time, the SV40 large T antigen (already made soon after infection) enters the host-cell nucleus, where it regulates transcription of the genes essential to viral replication and viral particle formation.

All these strategies and activities may be OK for the SV40 virus, but the kicker is that the large T antigen also binds to p53, interfering with transcription of proteins whose genes are regulated by p53. Unable to exercise its normal checkpoint functions, the host cell starts dividing uncontrollably, forming cancerous tumors. Deregulation of the cell cycle by large T antigen ensures progress to the S phase and unregulated co-replication of viral and host-cell DNA.

19.8.2 p53 and Signal Transduction

Stress can activate signal transduction pathways. For example, mutations affecting the **MAP-K** (MAP kinase) signaling pathway can lead to tumorigenesis. This can be explained by the observation that when activated, the MAPK pathway leads to amplified production of a kinase that phosphorylates **p53**. Active **phospho-p53** in turn augments activation of the MAPK signal transduction pathway, and you may recall that MAPK signal transduction typically ends with a mitogenic response.

Another example of p53 interaction is with **FAK** (**focal adhesion kinase**) proteins. **FAK** activity is increased by *integrin*-mediated signal transduction. Recall that membrane integrins bind *fibronectin*, contributing to formation of the extracellular matrix. Elevated **FAK** activity participates in the regulation of cell-cell and cell-ECM adhesion at **focal adhesion points**.

Another role for FAK is to bind directly to inactive p53 and to increase p53-Mdm2 binding. As we have just seen, persistent p53-Mdm2 is targeted for ubiquitination—and ultimate destruction! In fact, abnormally high levels of FAK are associated with many different tumor cell lines—for example, ovarian, colon, breast, thyroid, skin (melanoma), and muscle (sarcoma) cell lines. These result when p53 is unable to properly activate cell-cycle checkpoints.

While these complex interactions under active study, they certainly confirm the central role of *p53* as both *guardian of the genome* and as *guardian of cell division*.

19.9 Cancer-Cell Growth and Behavior; Cancer Treatment Strategies

Different cancer cell-types have different growth and other behavioral properties. You may have heard of *slow-growing* and *fast-growing* cancers. *Colon* cancers are typically slow growing. Periodic *colonoscopies* that detect and remove colorectal tumors in middle-aged or older people can prevent the disease (although the risks of disease and the procedure itself must be balanced). *Pancreatic* cancers are fast growing and usually go undetected until they reach an advanced stage.

The twin goals of medical research are to detect the different cancers early enough for successful intervention and, of course, to find effective treatments.

19.9.1 Cancer-Cell Origins, Growth, and Behavior

A single mutated cell in a tissue can become the growth point of a *tumor*, essentially a mass of cells cloned from the original mutated one. *Benign tumors* or growths (e.g., breast and uterine *fibroids* in women or common moles in any of us) usually stop growing and are not life threatening. They are often surgically removed for the comfort of the patient (or because cells in some otherwise-benign tumors may have a potential to become cancerous).

Malignant tumors (also called *malignant neoplasms*) are cancerous and can grow beyond the boundaries of the tumor itself. When tumor cells are shed, they may enter the bloodstream and travel to other parts of the body, a phenomenon called *metastasis*.

Cancer cells that metastasize can become the focal points of new tumor formation in many different tissues. Because cancer cells continue to cycle and to replicate their DNA, they can undergo yet more somatic mutations. These further changes can facilitate metastasis and cancer-cell growth in different locations in the body.

19.9.2 Cancer Treatment Strategies

Many kinds of cancers originate in different tissues of the body (and most are capable of metastasis). All cancers share the property of uncontrolled cell division, albeit for different molecular (and not always well-understood) reasons. In addition to surgical removal of tumors, the two major cancer treatment strategies, developed in the twentieth century, all aim at disrupting replication in some way.

- **Radiation therapy** relies on the fact that most cells in our bodies do not divide and are thus less prone to mutational damage. They aim mutagenic radiation at tumors in the hope that the replicating DNA will be mutated at so many sites (i.e., genes) that the tumor cells can no longer survive or replicate properly.
- **Chemotherapy** is used to attack tumors that do not respond well to radiation or that are not easily reached by radiation technologies and to fight cancers that do not even form focused tumors (such as lymphomas and leukemias, involving lymph and blood cells). These *chemotherapies* also aim to derange replication or mitotic activities. For example, recall **cordycepin** (dideoxyadenosine triphosphate, or ddATP). When present during replication, ddATP is incorporated into a growing DNA chain, after which no additional nucleotides can be added to the DNA strand. That makes ddATP a potent chemotherapeutic disruptor of replication. **Taxol** is another chemo drug. In this case, it prevents the depolymerization of spindle-fiber microtubules, thus blocking mitotic anaphase and telophase in the latter part of the M and C phases of the cycle. **Colchicine** (a plant alkaloid) attacks cancer (and other dividing) cells by blocking microtubule polymerization in the first place, thus preventing spindle-fiber formation in mitotic prophase.

These therapies are not effective against all cancers, and of course, they don't target specific kinds of cancer cells. Their success simply relies on the fact that cancer cells proliferate rapidly and constantly in contrast to other, normal cell. Despite the differences in proliferation rates between cancer and normal cells, many if not all the side effects of radiation and chemotherapies result from the damage done to normal dividing cells (e.g., hair-follicle cells, accounting for hair loss among many cancer patients; or the damage and resulting depletion of blood cells that fail to be replaced by stem cells in bone marrow). For the last twenty years or so, much research has been focused on mobilizing the body's own immune system to create more specific, targeted cancer treatments that might avoid such side effects.

In fact, over a hundred years ago, Dr. William B. Coley read a paper about a cancer patient who had become infected with streptococcal bacteria but had survived infection...and emerged cancer-free. He searched for and found more anecdotal reports of patients who had suffered coinfection and survived cancer-free, some of whom remained in remission for as many as seven years! In an experiment that we would have frowned upon today, Coley, then at the Memorial Sloan Kettering Cancer Center, intentionally injected a terminal cancer patient with the bacteria! Remarkably, this patient (and later, many others similarly treated) emerged tumor-free upon recovery from the coinfection. For more details, check out ¹⁹⁻⁸[The Earliest Cancer Immunotherapy Trials](#)). Coley eventually started to inject heat-killed bacteria into his cancer patients, to reduce the danger of actual infection. These injections of what came to be known as *Dr. Coley's Toxins* also proved effective for some of his patients.

Dr. Coley's son, Bradley, became a doctor and succeeded his father as head of the bone-tumor service at the Sloan Kettering. Bradley Coley successfully treated many cancer patients with his father's toxins through the 1950s. William Coley is still thought of today as *the father of cancer immunotherapy*.

The effects of Coley's "toxins" were at first thought to be a direct antitumor action. But without a known mechanism for the effect and due to remaining concerns over treating

patients with infectious microbes, bacterial coinfection therapy was superseded by more predictable (somewhat) radiation and chemotherapy treatments.

Dr. Lloyd Old (also at Sloan Kettering) attempted coinfection therapy once again in the mid-1970s. In this case, he injected a vaccine made from attenuated *Mycobacterium bovis* (the cause of tuberculosis) directly into the bladders of patients with bladder cancer. Called the BCG (*Bacillus Calmette-Guérin*) vaccine, this immunotherapy was highly effective, approved by the FDA in 1990, and is still used today!

As an understanding of immunology grew, the idea of a direct effect of the bacteria on tumors faded. Already by 1948, Coley's phenomenon had become widely attributed to an immune response activated by the infection. In 1957 Frank Macfarlane Burnet realized that cancer cells arise from mutations in normal cells, causing their unregulated proliferation. He further suggested that cancer cells multiply by evading our immune system's surveillance mechanism. This *immunosurveillance* is the job of circulating T lymphocytes (T cells) to distinguish *self* from *non-self* (i.e., "foreign" molecules and cells different from our own). For their recognition of *acquired immunological tolerance* (the failure of *immunosurveillance*) F. M. Burnet and P. B. Medawar earned the 1960 Nobel Prize in Physiology or Medicine.

Perhaps an explanation of Coley's phenomenon and the success of BCG therapy were at hand. Were infectious agents boosting immune systems in some cancer patients to restore immunosurveillance and to allow T cells to recognize cancer cells as foreign...and kill them.

In the 1990s, scientists revisited the immune response to cancer, and by the turn of the twenty-first century, studies of cancer immunotherapy picked up steam (and more substantial research funding!). As a case-in-point, consider James Allison's studies that identified the T-lymphocyte surface receptor that detects foreign cells and allows their destruction. He discovered that another T-cell membrane protein, *CTLA-4* (*cytotoxic T lymphocyte antigen 4*) would bind to potential target cells and block their destruction. Allison reasoned that *CTLA-4* is an *immune checkpoint inhibitor* that evolved to slow the destructive capacity of the cellular immune response, leading to as false immune response against that the body's own cells. His next insight was that antibodies against *CTLA-4* could relieve the checkpoint, allowing T cells to resume immunosurveillance... and if so, that it could be a strategy for attacking cancer cells! Allison eventually demonstrated that *anti-CTLA-4* antibodies could kill melanoma tumors in mice and then that they were also effective against several cancers in humans. These findings would save hundreds of thousands of lives. He and Tasuku Honjo were recognized with the 2018 Nobel Prize in Physiology or Medicine *for their discovery of cancer therapy by inhibition of negative immune regulation*. New animal and human immunotherapy trials are promising. A few (including Allison's) were approved in the US by the FDA (Food and Drug Administration). As somatic mutants, cancer cells make proteins that are just "foreign" enough to elicit an immune response, however slight. This understanding leads to multiple and overlapping approaches to cancer immunotherapy aimed at improving immunosurveillance to get a strong immune response to cancer-cell antigens. Some strategies, like anti-CTLA-4 immunotherapy, seek to boost that innate immune response. Others seek to isolate or to synthesize unique cancer-cell antigens in vitro, which upon injection into a patient will generate an immune response strong enough to kill the cancer cells. Different immunotherapies are summarized in the following table.

Table 19.1

Summary of Cancer Immunotherapies in Use or Under Study					
Type	Monoclonal antibodies	Immune checkpoint inhibitors	Cancer vaccines	Non-Specific immunotherapies	Chimeric Antigen Receptor CAR T-Cell Therapies
Brief Description	Monoclonal antibodies (mAbs) are typically prepared from isolated, cloned immune cells B-lymphocytes (or B-cell) that secrete blood-borne antibodies, or IgGs. IgGs from such cloned B-cells, chosen because they produce IgGs to cancer cell surface proteins (antigens) can thus attack a very specific part of the cancer cell, leading to its destruction.	Immune checkpoints protect normal cells from immune attack. The PD-1 protein on T-cell surfaces normally binds to PD-L1 receptors on normal cell, cloaking them against T-cell attack. PD-L1 proteins on cancer cells also bind PD-1. Thus protected from immune response, they proliferate. The disruption the interaction of PD-1 and PD-L1 (and others), checkpoint inhibitors 'un-cloaks' cancer cells, allowing T-cells to attack and destroy them.	These are made using whole tumor cells or purified proteins (or bits there) encoded by mutant genes active in cancer cells. They are often injected with adjuvants (chemicals that boost an immune response). When recognized as 'foreign' by a patient's immune system, cancer cells are attacked. Some vaccines are made sourcing the patient's own cancer cells for antigens. Some effective vaccines can also be prepared against a cancer from any patient's tumor.	Tumor shrinkage in response to bacterial infection can be mimicked by signal molecules such as Cytokines that can enhance an otherwise weak immune cell response to cancer cells. Anti-cancer cytokines include interferons, interleukins, and even thalidomide! Examples of this activity include the isolation of tumor infiltrating lymphocytes (TILs) found in tumors. Their anti-cancer activity is low, but can be boosted by cytokine treatment to levels where they may shrink the tumor.	This is similar to cytokine-stimulation of tumor-infiltrating cells proliferation to attack a tumor. But in this case, T-lymphocytes isolated from a cancer patient's blood are genetically engineered to contain and express genes for receptors specific for antigens (proteins) on the surface of cancer cells (the CARs). Once multiplied by growth in culture, these cells with their engineered chimeric antigen receptors, are injected back into the patient, where they can target and kill the cancer cells.
Possible Susceptible Cancers	pancreatic cancer, brain tumors, breast cancer, cervical cancer, prostate cancer, lymphoma, colorectal cancer, kidney cancer, lung cancer, melanoma	melanoma, non-small cell lung cancer, Hodgkin lymphoma, bladder cancer..., and potentially, many others!	pancreatic cancer, brain tumors, breast cancer, cervical cancer, metastatic prostate cancer, lymphoma, colorectal cancer, kidney cancer, lung cancer, melanoma	metastatic melanomas, cervical squamous cell carcinoma, cholangio carcinoma	pancreatic cancer, brain tumors, breast cancer, neuroblastoma, acute myeloid leukemia, multiple myeloma, non-Hodgkin's lymphoma

As you can see from the table, the immuno-targeting of cancer cells has already proven to be highly effective. In some cases, the therapy is an example of *personalized medicine*, in which treatments are uniquely tailored to you as a patient. Yet these immunotherapies have some known issues:

- They are time and labor intensive... and costly to produce.
- While they may "cure" *you*, they may not work on someone else.
- Like radiation and chemotherapy, some immunotherapies come with their own unpleasant and sometimes-severe side effects.

The promise of the checkpoint blockade demonstrated by Allison continues to be an active area of research.

NOTE: The term *checkpoint inhibitor* in the context of immunotherapies differs from the *checkpoints* monitoring progress through the eukaryotic cell cycle. A more detailed discussion of cancer immunotherapies is at [Cancer.gov](#) (¹⁹⁻¹¹[Cancer Immunotherapy](#)).

Some iText & VOP Key Words and Terms

anaphase	G ₂ phase	necrosis
apoptosis	<i>Guardian of the Genome</i>	oncogenic viruses
ATM kinase	immunotherapy	p21
benign tumors	integrin	p53
cancer cells	interphase	PD-1 checkpoint protein
CDKs	invasive tumors	PD-L1
cell cycle	Li-Fraumeni Syndrome	programmed cell death
Cell-cycle checkpoints	M checkpoint	prophase
chemotherapy	M phase of the cell cycle	protein phosphorylation
Chk2	malignant tumors	proteasome
colchicine	MAPK	radiation therapy
cyclin	maturation	S phase
cyclin level in cell cycle	Maturation promoting factor	signal transduction
cytokinesis	Mdm2	SV40
dideoxyNTP chemotherapy	metaphase	T antigens
elephant p53 genes	metastasis	taxol
FAK	mitosis	telophase
G ₀ of the cell cycle	mitosis promoting factor	tumor suppressor protein
G ₁ checkpoint	mitotic phases	ubiquitination
G ₁ phase	MPF	
G ₂ checkpoint	mTOR signaling	

CHAPTER 19 WEB LINKS





19-7



19-8



19-9



19.10



19-11

Chapter 20

Origins of Life

A short history; Origins under reductive vs non-reductive conditions; Prebiotic chemical and metabolic evolution; Origins of communication (catalysis, autocatalysis, co-catalysis, biochemical pathways; Transition from the RNA World

Reminder: For inactive *links*, google key words/terms for alternative resources.

Primitive Ugly Homunculus in a Fossil Sperm...



20.1 Introduction

It is nearly universally accepted that there was a time, however brief or long, when the Earth was lifeless. Given that the cell is the basic unit of life, and that to be alive is to possess all the **properties of life**, any cell biology textbook would be remiss without addressing at some length, the questions of **when** and **how** the first cells appeared on our planet. **Abiogenesis** is the origin of life from non-living matter. Of course, any observation of abiogenesis in nature is no longer possible! But a combination of experiment and educated guesswork makes it possible to construct reasonable (if sometimes conflicting) scenarios to explain the **origins of life**, and hence our very existence.

In this chapter, we will see that different scenarios require consistent assumptions about climatic, geologic, thermodynamic, and chemical conditions that favored abiogenesis. The "right" conditions would lead to a prebiotic accumulation of organic molecules, chemical reactions and proto-structures that could support the formation of a cell. One might reasonably speculate such a **prebiotic laboratory** would have led to spontaneous experiments in **chemical evolution**, which chemical combinations best survived and interacted in a hostile pre-biotic environment, well before the origin of the first cell... or cells. Hence the chapter title "Origins of Life"!

While many consider the creation of the first cell to be a singular, one-off event, it has been argued that multiple independent origins were not only possible under these conditions, but also probable! According to Jeremy England, of MIT, the laws of thermodynamics dictate that "...when a group of atoms is driven by an external source of energy (like the sun or

chemical fuel) and surrounded by (heat) like the ocean or atmosphere), matter inexorably acquires the key physical attribute(s) associated with life". Check out *A New Physics Theory of Life* at ^{20.1}[Thermodynamic Laws Predict the Origins of life!](#) for more information. Here is a reminder of those key attributes, or properties of life:

Evolution:	long-term <i>adaptation/speciation</i>
Cell-based:	the cell is the fundamental <i>unit of life</i>
Complexity:	<i>dynamic order</i>; allows physical/biochemical change
Homoeostasis:	living things maintain <i>balance</i> between change and order
Requires Energy:	needed to do <i>work</i>, i.e., all cellular functions
Irritability:	immediate sensitivity and <i>response to stimuli</i>
Reproduction:	sort of self-explanatory, yes?!
Development:	<i>programmed change</i>; most obvious in multicellular organisms but found in some form in all organisms.

Remember that to be alive is to possess not just some, but all these properties! If entities with the required properties of life (i.e., cells) did originate independently, they would have reproduced to form separate populations of cells, each of which would embark on an independent evolutionary pathway. In this scenario, less successful populations would go extinct as successful ones become dominant. Successful organisms would have spread, spawning populations, evolving, and generating new species. This could go on until all other independent lineages of life were extinct.

The ***take-home message*** here is that if conditions on a prebiotic earth favored the formation of a single progenitor 'first cell,' then why not the formation of two or dozens or even hundreds of 'first cells? We may never know because there are no survivors of such lineages. If they ever existed, they are now extinct! We will see the evidence that there is only one successful population of cells from which common ancestor came all known living things. As to the question of ***when*** life began, we will look at geological and geochemical evidence suggesting the presence of life on Earth as early as 4.1 billion years ago. As for ***how*** life began, this remains the subject of ongoing speculation.

After a brief history of thought about life origins, *all* the scenarios to be described below try to understand the physical, chemical, and energetic conditions that might have been the ideal laboratory for prebiotic "*chemistry experiments*." We will examine two of these scenarios in some detail. What *all* Origins of Life scenarios *must* account for include the following:

- ***prebiotic synthesis* of organic molecules, monomers & polymers**
- **origins of *catalysis***
- ***sources of free energy* to sustain prebiotic chemistry**
- **beginnings of *metabolism* sufficient for life**
- **origins of *molecular communication* (information storage and retrieval)**
- ***enclosure* of life's chemistry behind a semipermeable boundary**
- ***the legacy* in all organisms of common *life and genetic chemistries***

Now let's consider some tricky definitions. If one believes (as is still common), that the origin of earthly life was so unlikely that it could only have happened once, then the first cell,

the **progenote**, is indeed the progenitor of us all, the common ancestor of all living things. The evolutionary tree in Figure 20.1 (below) puts the progenote at the root of the tree.

Our Descent from a Progenote

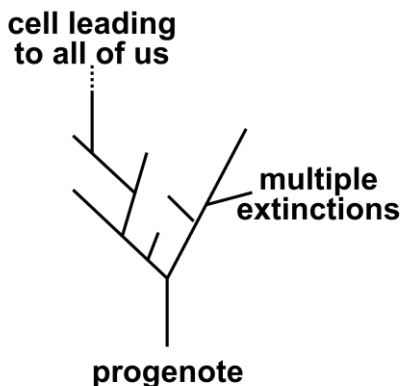


Fig. 20.1: Descent (evolution and speciation) from a single progenitor cell formed by sheer luck on a prebiotic Earth.

The progenote, that first and only cell, would have reproduced and diverged into multiple, separate, giving rise to multiple lineages. In other words, the descendants of the progenote would have evolved from the get-go. Like any populations of cells and organisms, whether from of a one or many "progenotes," descendant populations would have competed for resources (and ultimately survival) on an early Earth. Thus, one of these lineages (the dashed line in the illustration) would have survived as others fell to extinction, a process that continues today as the fossil record attests. Now let's imagine multiple independent 'first cells.' How would that change our view of life's descent? Figure 20.2 illustrates a hypothetical scenario.

Did a favorable environment generate a one-&-only single progenote?

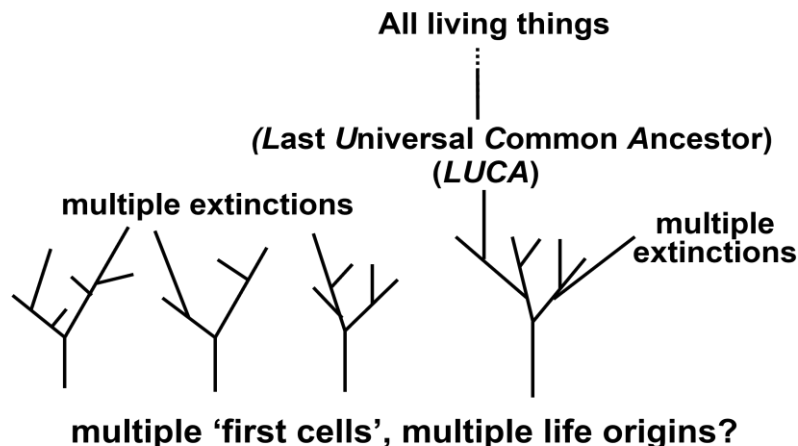


Fig. 20.2: Descent (evolution and speciation) from multiple "first" cells" assumes **no progenote** and the extinction of all but one lineage, whose descendants include our 'last universal common ancestor' (LUCA).

If we assume multiple life origins ("first cells"), there was **no progenote**. But without evidence (we have none!), we must surmise that the evolution of all but one of these "first cells" led to

extinction of their lineage, leaving only one path to the evolution of life on Earth, the one on the right. Descendants of the 'first cell' in that path survived to evolve their own lineages, among which many themselves went extinct. The surviving lineage that produced all living things traces back to the Last Universal Common Ancestor (LUCA).

In other words, whether life had one or many origins in a permissive prebiotic environment, we can anchor our assumptions on the evolution of life from a single origin. We will see clear evidence to support that all organisms alive today are descended from a single cell. In a scenario where only one cell population survives the competition with other populations, its evolved cells would have been the source of our ***Last Universal Common Ancestor***, or ***LUCA***. The ***LUCA*** must be that highly evolved cell whose genome, biochemistry, and basic metabolic infrastructure is shared among all things alive today (Figure 20.3).

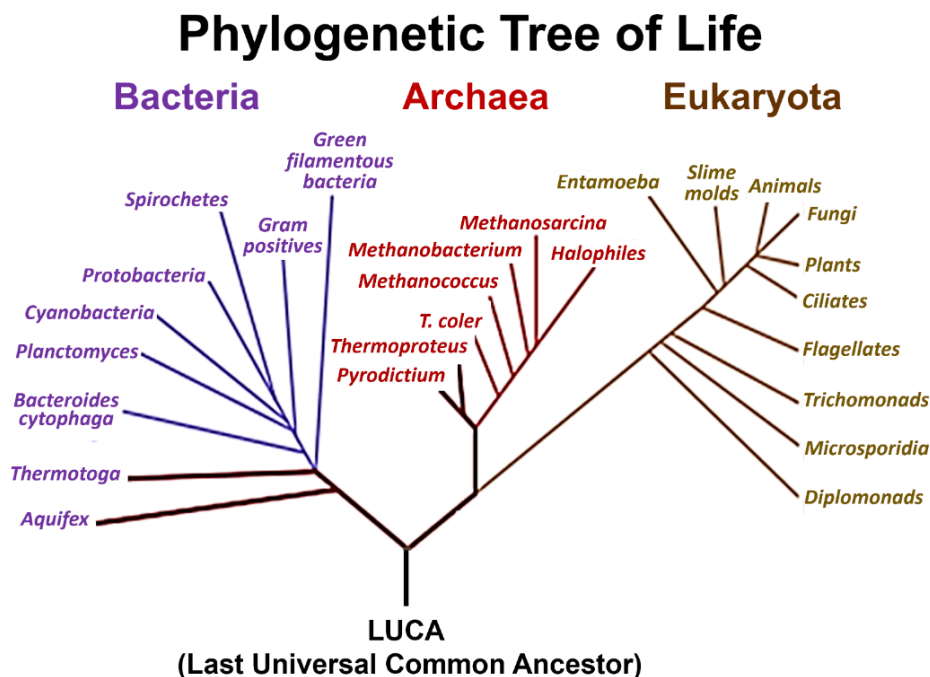


Fig. 20.3: Evolutionary (i.e., phylogenetic) tree of all living organisms showing descent of the 3 domains of life (Bacteria, Archea and Eukaryota) from a ***LUCA***. The node separating archaea from eukaryota is the LECA, (Last Eukaryotic Common Ancestor), but could just as well be called the LACA!

This and similar evolutionary trees for known living things on Earth are based on studies of nucleotide sequences of many genes in existing organisms. Whatever the prebiotic pathway (or pathways) to the first living cells on Earth, computers can be used to trace differences in gene nucleotide sequences back to changes that would have begun in that single common ancestor that we have defined as our ***LUCA***.

What is the basis for our conclusion that the progeny of multiple 'first cells' are no longer with us? Even if there were multiple life origins, each of the "first cells" might survive long enough to produce its own lineage of evolved cell populations. But the cellular biochemistries and structures of these independently lineages would likely have been quite different as they evolved their own solutions to the tasks of survival. Perhaps it will be easiest to imagine that they evolved very different genetic codes, even if they used RNA and DNA as their genetic material, and even if the best solution to storing and retrieving genetic

information was the three-base codon for their amino acids. If life originated multiple times and any of their descendants survived to this day, we would expect them to have a different genetic code, if not a totally different way to store genetic information. With only evidence of a shared genetic code among today's living things, we can conclude that but one of many early cell lineages went extinct, or that they never happened. Either way, among the descendants of *our* progenote is a lineage, a population that includes the first cell with a DNA genome with the same genetic code used by virtually all organisms alive today. This, *our LUCA*, and its descendants display the high level of conservation of gene and protein sequences that we find today. We'll look at more supporting evidence that this is so as we move through this chapter. We'll learn that this evidence is perhaps the strongest support for realized expectations. Let's begin by looking at the following basic requirements of any life-origins scenario:

- reduction of inorganic molecules to form organic molecules
- a source of free energy to fuel the formation of organic molecules
- a mechanism for anabolic metabolism, the synthesis of macromolecular polymers from organic monomers.
- a pathway to the catalytic acceleration of biochemical reactions
- separation of early biochemical experiments by a semipermeable boundary.

Then, we'll consider some proposed scenarios for the creation of organic molecules:

- import of organic molecules (or even life itself) from *extraterrestrial* sources.
- organic molecule synthesis on an Earth with a *reducing atmosphere*.
- organic molecule synthesis on an Earth with a *non-reducing atmosphere*.

We'll explore alternate free energy sources and pathways to the essential chemistry of life dictated by these alternatives. Then we look at specific scenarios of prebiotic chemical evolution. Finally, we consider how primitive (read "simpler") biochemistries could evolve into the metabolisms shared by all existing life forms today..., including a common genetic code!



347 What any Life Origins Scenario Must Explain

Learning Objectives

When you have mastered the information in this chapter, you should be able to do the following:

1. Explain how organic molecules would capture chemical energy on a *prebiotic earth*.
2. List the essential chemistries required for life and why they might have been *selected* during chemical evolution.
3. Discuss the different fates of prebiotically synthesized organic *monomers* and *polymers*, and how these fates would influence the origins of the first cells on Earth.
4. Compare and contrast two scenarios for *extraterrestrial origins* of organic molecules.
5. Summarize the arguments for and against Oparin's *primordial soup* hypothesis.
6. Summarize the evidence for and against life origins in a *non-reducing* Earth atmosphere.
7. Explain the difference between a *progenote* and the *LUCA*.

8. Discuss the evidence suggesting an origin of cellular life in the *late Hadean eon*.
9. Describe how life might have begun in deep ocean vents – compare the possibilities of life beginning in *black smokers* vs *white smokers*.
10. Argue for and against an *autotroph-first* scenario for cellular origins.
11. Explain why some investigators place significance on the early origins of free energy storage in *inorganic proton gradients*.
12. Speculate on what selective forces drove the evolution of electron transport.
13. Define *autocatalysis*, *co-catalysis*, and *co-catalytic sets*; provide examples.
14. Describe the significance and necessity of *coevolution* before life. In what ways is coevolution a feature of living things? Explain.

20.2 Thinking about Life's Origins: A Short Summary of a Long History

By all accounts, the Earth must have been a very unpleasant place soon after its formation! For that reason, the period from 4.8 to 4.0 billion years ago is called the ***Hadean Eon***, after *Hades* ('hell' to the ancient Greeks, but a great free energy source!). Until recently, geological, geochemical, and fossil evidence suggested that life arose between 3.8 and 4.1 billion years ago. More recently, the 2017 discovery of evidence for life in 3.95 billion year-old sedimentary rocks in Labrador points to a 4 billion year old earlier origin of life (see ^{20.2}[Oldest Evidence for Life on Earth-from Canada](#)).

In fact, questions about life's origins are probably "as old as the hills..." or at least as old as the ancient Greeks! We only have records of human notions of life's origins dating from biblical accounts and, just a bit later, from Aristotle's musings. Aristotle did not suggest that life began in 'hell.' Instead, he and other ancient Greeks speculated about life's origins by spontaneous generation, in the sense of abiogenesis, or life originating from non-life. Aristotle further speculated that the origins of life were gradual. Later, the dominant theological accounts of creation in Europe in the Middle Ages muted all notions of natural origins and evolution. While a few mediaeval voices ran counter to strict biblical readings of the creation stories, it was not until the Renaissance (in the fourteenth-seventeenth centuries) that an appreciation of ancient Greek humanism was reawakened and with it, scientific curiosity, and the ability to engage in rational questioning and research.

You may recall that Louis Pasteur in the mid-nineteenth century put to rest any lingering notions of life forming from dead (e.g., rotten, or fecal) matter. He showed that life would not form in sterilized nutrient solutions unless the broth was exposed to the air. Fewer know that much earlier, Anton Van Leeuwenhoek (the seventeenth century sheriff of Delft, amateur lens grinder and microscopist who first described pond water bacteria and protozoan animalcules) had already tested the notion of spontaneous generation. By observing open and sealed containers of meat over time, he became convinced that 'large' animals like fleas and frogs do not arise *de novo* from putrid meat or slime. He also declared that insects come from other insects, and not from the flowers that they visited. In 1859, no lesser a light than ^{20.3}[Charles Darwin](#) himself favored the idea that life might have begun in a "...warm little pond, with all sorts of ammonia and phosphoric salts, light, heat, electricity, &c., present, that a proteine compound was chemically formed ready to undergo still more complex changes." He even realized that these chemical constituents would not have survived in the atmosphere and waters of his day but must have been doing so in a prebiotic world.

In *On the Origin of Species*, Darwin referred to life having been 'created.' However, Darwin makes it clear that he was not referring to a biblical basis of creation. Rather, he meant that life originated "by some wholly unknown process" at a time before which there was no life.

Finally, Pasteur's 1861 contribution was the irrefutable, definitive proof that 'invisible' microbial life likewise did not arise by spontaneous generation. Thus, the creatures already on Earth could only arise by biogenesis (life-from-life), the opposite of abiogenesis, a term that now applies to only the first origins of life! Among Darwin's contemporaries and friends were **Charles Lyell** and **Roderick Murchison**, both geologists who understood much about the slow geological changes that shaped the Earth. Darwin was therefore familiar with the concept of extended periods of geological time, amounts of time he believed were necessary for the natural selection of traits leading to species divergence.

OK, *spontaneous generation* is disproven. But..., appropriate conditions on earth must have made life's origins spontaneous! Let's fast-forward now to the 1920s when J. H. B. S. Haldane and A. Oparin offered a hypothesis about the life's origins based on notions of the chemistry and physical conditions they believed might have existed on a prebiotic earth. Their proposal assumed that the Earth's atmosphere was hot, hellish, and reducing (i.e., filled with inorganic molecules able to give up electrons and hydrogens). There are more than a few hypotheses for which kinds of chemicals were already present on Earth, or that formed along with the planet about 4.8 billion years ago, but all posit a source of free energy for their formation. We'll start our exploration with Oparin and Haldane's reducing atmosphere. Then we look at possibility that life began under non-reducing conditions, with passing reference to a few other ideas!



[348 Early Ideas to Explain the Origins of Life](#)

20.3 Formation of Organic Molecules in an Earthly *Reducing Atmosphere*

A prerequisite to prebiotic chemical experimentation is a source of organic molecules. Just as life requires energy (to do anything and everything), converting inorganic molecules into organic molecules requires an input of *free energy*. As we have seen, most living things today get free energy by oxidizing nutrients or directly from the sun by photosynthesis. Recall that *all* the chemical energy sustaining life today ultimately comes from the sun. But before there were cells, how did organic molecules form from inorganic precursors? Oparin and Haldane hypothesized a reducing atmosphere on the prebiotic earth, rich in inorganic molecules with *reducing power* (like H_2 , NH_3 , CH_4 , and H_2S) as well as CO_2 to serve as a carbon source. They predicted that the *physical conditions* on a prebiotic earth would include lots of water (oceans); high heat (no free O_2); lots of ionizing (e.g., X^- , γ^-) radiation from space, (no protective ozone layer); frequent ionizing (electrical) storms generated in an unstable atmosphere; and lots of volcanic and thermal vent activity.

20.3.1 Origins of Organic Molecules and a Primordial Soup

Oparin suggested that abundant sources of free energy fueled the reductive synthesis of the first organic molecules to create what he called a "primeval soup." No doubt, he called this

primeval concoction a “soup” because it would have been rich in chemical (nutrient) free energy. **Harold Urey**, who had already won the 1934 Nobel Prize in Chemistry for discovering deuterium, and **Stanley Miller** tested the prediction that, under Haldane’s and Oparin’s prebiotic earth conditions, inorganic molecules can produce the organic molecules in what became known as the *primordial soup*. In their classic experiment, a mix of inorganic molecules provided with an energy source was reduced to very familiar organic molecules, supporting the Oparin/Haldane proposal. Their experiment is shown below in Figure 20.4.

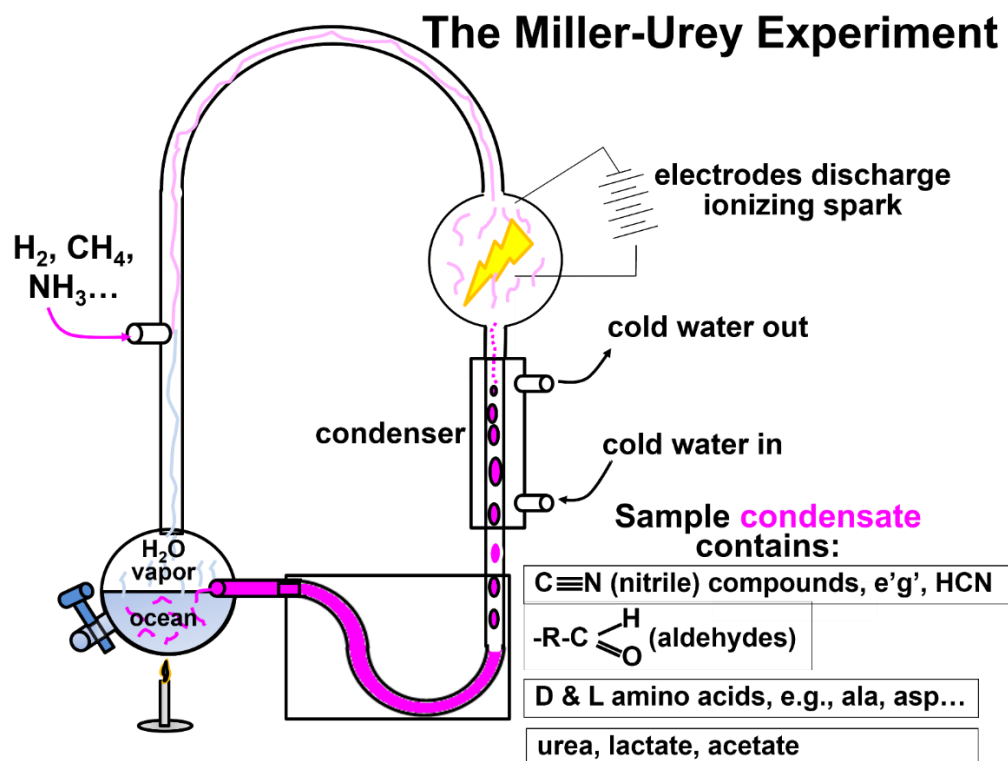


Fig. 20.4: The Miller & Urey experiment: organic molecules seen in living things today form spontaneously in the lab under *reducing* conditions like those expected in a prebiotic earth atmosphere.

Miller’s earliest published data indicated the presence of several organic molecules in their *ocean* flask, including a few familiar metabolic organic acids (lactate, acetate, several amino acids...) as well as several highly reactive *aldehydes* and *nitriles*. The latter can interact in spontaneous chemical reactions to form organic compounds. Later analyses further revealed purines, carbohydrates, and fatty acids in the flask. Some 50 years after Miller’s experiments (and a few years after his death), some un-analyzed samples collected from those early experiments were discovered. When their contents were re-analyzed with newer, more sensitive detection techniques, they were shown to contain organic molecules not originally reported, including 23 amino acids (to read more, see ^{20.4}[Surprise Goodies in the Soup!](#)).

Clearly, the thermodynamic and chemical conditions proposed by Oparin and Haldane would support *reductive synthesis* of organic molecules. But these evolving chemistries would have to be captured inside semipermeable aggregates (or boundaries) destined to be a cell. Proposals for such structures are discussed below. Such events in a nutrient-rich primordial soup would likely have led to the genesis of *heterotrophic* cells that could use environmental nutrients for energy and growth. This implies an early evolution of fermentative pathways like

glycolysis. But here's the rub: these first cells would have quickly consumed the nutrients in the soup, quickly ending the Earth's new vitality!

To stop life from becoming a dead-end experiment, one must propose the early evolution of **autotrophs**, cells that could capture free energy from inorganic molecules (**chemoautotrophs**) or even sunlight (**photoautotrophs**). As energy-rich organic nutrients in the 'soup' declined, autotrophs would be selected, for example photoautotrophs that could *fix* (i.e., reduce) CO₂, with H⁺ ions split from water. In this way, **photosynthesis** would replenish carbohydrates and other nutrients in the oceans and splitting water would add O₂ to the atmosphere. Oxygen would have been toxic to most cells at the time, but some of those cells already had the ability to survive oxygen. These survived, evolving into cells that could *respire*, i.e., use oxygen to *burn* environmental nutrients. In fact, respiratory metabolism must have followed hard on the heels of the spread of photosynthesis. After photosynthesis emerged, sometime between 3.5 and 2.5 billion years ago (the Archaean Eon), photosynthetic and aerobic cells and organisms achieved a natural balance to become the dominant species in our oxygen-rich world.

20.3.2 The Tidal Pool Scenario for an Origin of Polymers and Replicating Chemistries

In this scenario, prebiotic organic monomers concentrate in tidal pools and polymerize in the heat of a primordial day by dehydration synthesis. This uphill reaction requires free energy. Very high temperatures (**heat of baking**) can link monomers by dehydration synthesis in the lab and may have formed random polymers in tidal pool sediments, as shown in Figure 20.5, which further assumes dispersal of these polymers with the ebb and flow of the tides.

Prebiotic polymer synthesis... one scenario

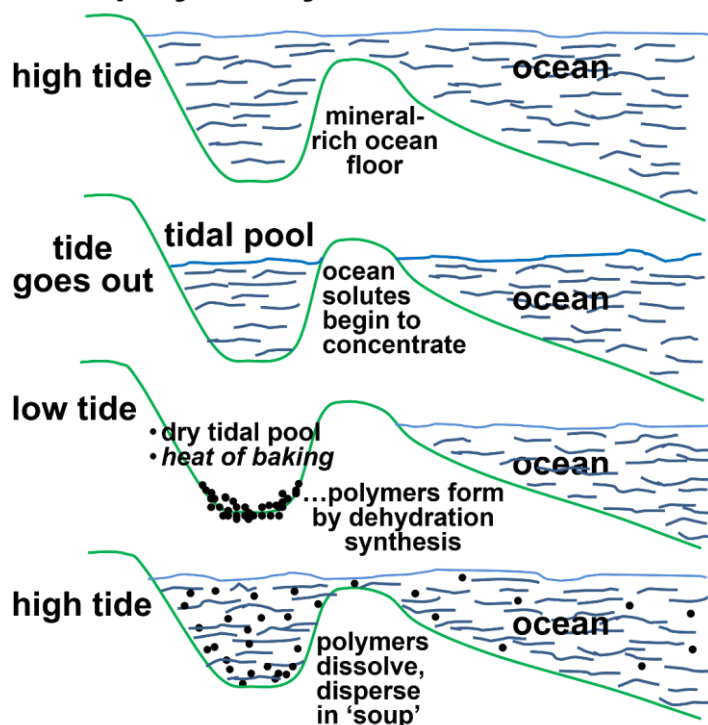


Fig. 20.5: Scenario for the synthesis of prebiotic polymers in tidal pools.

The catalytic properties of metals (e.g., nickel, platinum, silver, magnesium, and manganese) are exploited in the laboratories to speed up chemical reactions. Such metals were likely present in primordial ocean sediments in the early Earth's crust, including, just as they are there today. And their catalytic properties in soil and clay aggregates have also been demonstrated. In fact, metals like magnesium and manganese are now an integral part of many enzymes, consistent with an origin of biological catalysts as prebiotic benthic minerals. Finally, some metals like magnesium (Mg^{++}) readily associate with nucleotides. Before life, the microsurfaces of ocean sediments, if undisturbed, may have catalyzed the same or at least similar reactions repeatedly, leading to related sets of polymers. Consider the possibilities for RNA monomers and polymers, based on the assumption that life began in an "RNA world" (Figure 20.6). The concentration of putative organic monomers at the bottom of mineral-rich tidal pools may have accelerated polymerization reactions.

Making an RNA World: Replicating RNA Polymers In Tidal Pools

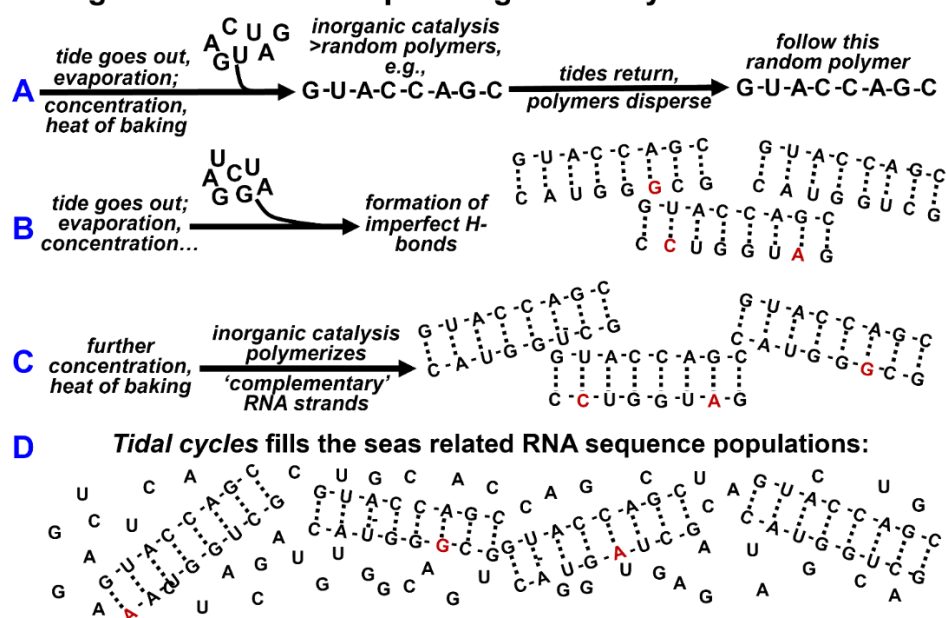


Fig. 20.6: Scenario for the synthesis of redundant prebiotic nucleic acid polymers in tidal pools.

The result predicted by this scenario is the formation not only of RNA-like polymers (perhaps only short ones at first), but of H-bonded double-stranded molecules that might effectively replicate at each cycle of concentration, polymerization, and dispersal. High heat could have supported polymerization, while catalysis enhanced the fidelity of template-based replication. Of course, repeated high heat or other physical or chemical attack might degrade newly formed polymers, unless some double strands were resistant to destruction. These would accumulate at the expense of the weaker, more susceptible ones. In this way, the *fittest* replicated molecules would be stable in the environment. The environmental accumulation of structurally related, replicable, and stable polymers reflects a prebiotic chemical *homeostasis* (one of those properties of life!). For this scenario to work, prebiotic nucleosides must have been available and stable. We'll revisit this issue shortly



Overall, the tidal pool scenario for life origins in a reducing environment hung together nicely for many decades. But there are now challenging questions about the premise of a prebiotic reducing environment. Newer evidence suggests an Earth atmosphere that was not at all reducing, casting doubt on the idea that heterotrophs were the first cells on the planet. Recent proposals posit alternative sources of prebiotic free energy and organic molecules that look quite different from those assumed by Oparin, Haldane, Urey, and Miller.

20.4 Origins of Organic Molecules in a *NON-Reducing Atmosphere*

A prebiotic non-reducing atmosphere is based on several assumptions: (1) The early Earth would have had insufficient gravity to hold H_2 and other light gasses; (2) as a result, "outgassing" would have resulted in a loss of H_2 and other reducing agents from the atmosphere; (3) The geological evidence suggests that the Earth's oceans and crust formed early in the Hadean Eon, just a few hundred million years after formation of the planet; (4) studies of 4.4-billion-year-old (early Hadean Eon) Australian *zircon* crystals suggest that their oxidation state is the same as modern day rocks, meaning that the early Hadean atmosphere was largely N_2 and CO_2 , a distinctly *non-reducing* one! Solid geological evidence of cellular life dates to 3.5-3.95 billion years ago (i.e., the *Archaean Eon*). Softer evidence of microbial life exists in the form of graphite and other possible footprints of life as old as 4.1 billion years ago, near the end of the Hadean Eon. If this is true, the discovery of an oxidizing Hadean atmosphere at least 3 billion years earlier (4.4 billion years ago) argues for the origins of life in a non-reducing atmosphere. Therefore, regardless of whether life began 3.5 or even 4.1 billion years ago, the evidence suggests that life's beginnings had to contend with a non-reducing environment. A colorized image of the Australian zircon is shown in Figure 20.7.

4.4 BYO Zircon Crystal Predicts Life Origins in an O_2 atmosphere

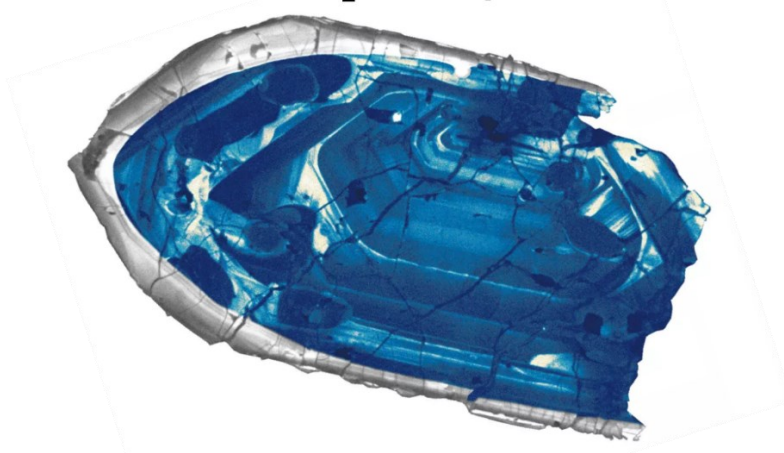


Fig. 20.7: The *oxidation state* of this Australian *zircon* crystal supports an oxidizing atmosphere on Earth as far back as 4.4 billion years ago. Photo Courtesy of J. W. Valley et al.

Before we look more closely at other evidence of life origins under non-reducing conditions, let's consider the *Panspermia*, the possibility that life came to Earth from extraterrestrial sources and a related hypothesis that prebiotic organic molecules came from *extraterrestrial* sources. Then we'll return to the question of how cells might have formed in localized, favorable *terrestrial* environments.

20.4.1 Panspermia – an Extraterrestrial Origin of Earthly Life

Panspermia posits that life itself arrived on our planet by hitchhiking on comets or meteorites. Since these are unlikely to have sustained life in space, they must have been a kind of interstellar 'mailbox' into which dormant life forms were deposited. The cells in the mailboxes must have been *cryptobiotic*. Examples of cryptobiosis exist today (e.g., bacterial spores, brine shrimp!). Once they were delivered to Earth's life-friendly environment, such organisms would emerge from dormancy, eventually populating the planet. But there is no evidence of dormant or cryptobiotic life on comets or meteorites, and no hard evidence to support *Panspermia*.

On the other hand, there is evidence at least consistent with an extraterrestrial source of organic molecules, and even more to support more terrestrial origins of life. In any case, notions of *Panspermia* (and even extraterrestrial sources of organic molecules) just beg the question of the conditions that would have led to the origins of life elsewhere!

Not a favored scenario, panspermia is still intriguing in the sense that it is in line with a likelihood that organic molecules formed soon after the *Big Bang*. Moreover, if ready-made organic molecules and water were available, we can expect (and many do!) that there is life on other planets. This possibility has stimulated discussion and serious funding of programs looking for signs of extraterrestrial life. It supported the earlier *Search for Extraterrestrial Intelligence* (SETI) program based on the assumption that life not only exists elsewhere, but that it evolved high level communication skills (and why not?)! More recently NASA funded *Rover's* search for signs of water on Mars and *Cassini's* recent discovery of water on *Enceladus*, one of Saturn's moons. For a fascinating story about meteorites from Mars that contain water and that are worth more than gold, check out ^{20.5}[Martian Obsession](#). For one about Titan (a moon of Saturn's), see ^{20.6}[Life on Titan?](#).

And then there is this! *Maybe, just maybe* we have **exported** earthly life to comets when they bumped into us and skittered away instead of crashing and burning. And *maybe, just maybe* they scraped up a few upper atmospheric microbes along their way to other outer-space encounters (upper-atmosphere microbial colonies). Will we discover earthlike forms of life on exoplanets as a result? Read about the possibilities and realities at. ^{20.7}[Maybe, just maybe...](#) Or did all this sharing of our gift of life just explain the presence of organic molecules found on some space objects that have visited us from time to time!

20.4.2 Extraterrestrial Origins of Organic molecules

Even if life did not come to us ready-made, could organic molecules already in space have arrived safely on Earth? They are abundant, for example in *interstellar clouds*, and could have become part of Earth as the planet and our solar system formed around 4.8 billion years ago. This suggests that there was no need to create them *de novo*. Meteorites, comets, and asteroids are known to contain organic molecules and could have brought them here during fiery impacts on our planet. Such bombardments would have been common 3.8 or more billion years ago. In this scenario the question of how (*not on Earth!*) free energy and inorganic molecular precursors reacted to form organic molecules..., is moot! A related hypothesis suggests that those fiery hits themselves provided the free energy necessary to synthesize organic molecules from inorganic ones... a *synthesis-on-arrival* scenario. With this hypothesis on the one hand,

we are back to an organic oceanic primordial soup. On the other, some have suggested that organic molecules produced in this way (not to mention any primordial life forms) would likely have been destroyed by the same ongoing impacts by extraterrestrial bodies; witness the relatively recent dinosaur extinction by an asteroid impact off the coast of Mexico some 65.5 million years ago. If organic molecules could not have come from space, then where from?



[350 Life Origins in a Non-Reducing Atmosphere?](#)

20.5 Organic Molecular Origins of Life Closer to Home

Deep in the oceans, far from violent meteoric encounters and rampant free energy of an ozone-less sky, deep-sea hydrothermal vents would be spewing reducing molecules (e.g., H_2S , H_2 , NH_4 , CH_4), much as they do today. Some vents are also high in metals such as lead, iron, nickel, zinc, and copper. Combined with their clay or crustal substrata, some of *these* minerals could have provided catalytic surfaces to enhance organic molecule synthesis. Could such localized conditions have been the focus of prebiotic chemical experimentation leading to the origins of life? Let's look at two kinds of deep-sea hydrothermal vents: *volcanic* and *alkaline*.

20.5.1 Origins in a High-Heat Hydrothermal Vent (*Black Smoker*)

The free energy available from a volcanic hydrothermal vent would come from the high heat temperatures ranging to 350°C) and the minerals and chemicals expelled from the Earth's mantle. Figure 20.8 (below) shows a volcanic hydrothermal vent.

A Black Smoker

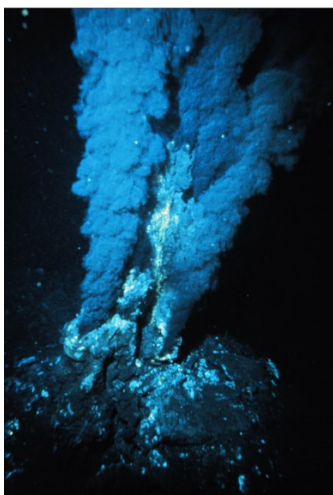


Fig. 20.8: An oceanic volcanic hydrothermal vent, or *black smoker*.

Conditions assumed for prebiotic volcanic hydrothermal vents could have supported the catalytic synthesis of organic molecules from inorganic precursors (see ^{20.8}[Volcanic Vents and Organic Molecule Formation](#)) or ^{20.9}[Volcanic Vents & Organic Molecule Formation-full paper](#)). Available catalysts would have been inorganic, for example metals like nickel and iron. chemical reactions tested include some that are reminiscent of biochemical reactions in chemoautotrophic cells alive today.

Günter Wächtershäuser proposed the ^{20,10}[Iron-Sulfur World Theory](#) of life's origins in these vents, also called "black smokers". These vents now spew large amounts of CH₄ and NH₄ and experiments favor the idea that iron-sulfur aggregates in and around *black smokers* could provide catalytic surfaces for the prebiotic formation of organic molecules like methanol and formic acid from dissolved CO₂ and the CH₄ and NH₄ coming from the vents. A variety of extremophiles (e.g., thermophilic archaea) now living in and around *black smokers* would seem to be testimony supporting a black smoker hypothesis for life origins. Wächtershäuser also realized that prebiotic selection acted not so much on isolated chemical reactions, but on aggregates of *metabolic reactions*. We might think of such metabolic *aggregates* as biochemical pathways or multiple integrated biochemical pathways. Wächtershäuser also proposed the selection of cyclic chemical reactions that released free energy that could then be used by other reactions. If so, some version of a prebiotic scenario of *metabolic evolution*, or selection of metabolic chemistries, something more complex than a simple chemical evolution, would have been essential to the origins of life.

While the idea of selecting metabolic pathways has great merit, there are problems with a life-origins scenario in volcanic hydrothermal vents. For one thing, their very high temperatures would likely have destroyed as many organic molecules as were created. Also, the extremophilic archaea now found around these volcanic vents cannot be the direct descendants of any cells that might have originated there. Woese's phylogeny clearly shows that archaea share a lineage with eukaryotes (not eubacteria). Thus, extremophilic cellular life originating in the vents must have given rise to a more moderate LUCA first, before then going extinct. Then, extremophiles we find in the black smokers would once again have had to evolve independently to re-colonize the vents! Such twists and turns militate against an extremophiles-first origins scenario. Given these concerns, recent proposals focus on life origins in less extreme *alkaline hydrothermal vents*.

20.5.2 Origins in an Alkaline Deep-Sea Vent (*White Smoker*)

Of the several scenarios discussed here, an origin of autotrophic life in *alkaline vents* is one of the more satisfying alternatives to a soupy origin of heterotrophic cells. For starters, at temperatures closer to 100°C-150°C, alkaline vents (*white smokers*) are not nearly as hot as are black smokers. An *alkaline hydrothermal vent* is shown below in Figure 20.9.

A White Smoker

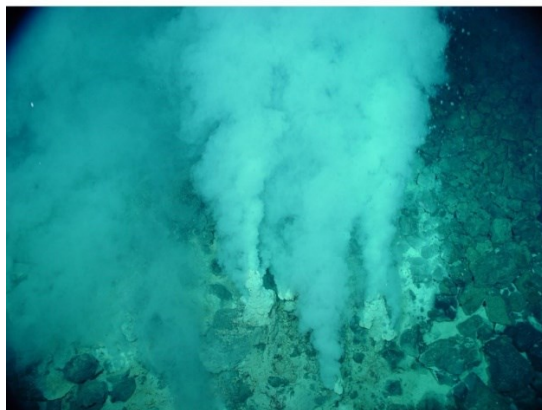


Fig. 20.9: An alkaline deep-sea hydrothermal ocean vent, or *White smoker*

Other chemical and physical conditions of alkaline vents are also consistent with an origins-of-life scenario dependent on *metabolic evolution*. For one thing, the interface of alkaline vents with acidic ocean waters has the theoretic potential to generate many different organic molecules [Shock E, Canovas P. (2010) *The potential for abiotic organic synthesis and biosynthesis at seafloor hydrothermal systems*. *Geofluids* 10 (1-2):161-92)].

In laboratory simulations of alkaline vent conditions, the presence of dissolved CO_2 favors ***serpentinization***, a reaction of water and heat with *serpentine*, an iron-containing mineral found on land and in the oceanic crust. Figure 20.10 is a sample of serpentine.

Serpentine



Fig. 20.10: *Serpentine* from Deer Lake in upper Michigan is a mineral also found in the oceanic crust that under conditions found in an alkaline vent, can form methane from CO_2 .

Experimental serpentinization produces hydrocarbons, and a warm aqueous oxidation of iron produces H_2 that could account for abundant H_2 in today's *white smoker* emissions. Also, during serpentinization, a mineral called *olivine* [$(\text{Mg}^{+2}, \text{Fe}^{+2})_2\text{SiO}_4$] reacts with dissolved CO_2 to form methane (CH_4). So, the first precondition of life's earthly origins, the energetically favorable creation of organic molecules, is possible in alkaline vents.

Proponents of cellular origins in a late-Hadean non-reducing ocean also realized that organic molecules formed in an alkaline (or *any*) vent would disperse and be rapidly neutralized in the wider acidic ocean waters. Somehow, origins on a non-reducing planet had to include a way to contain newly formed organic molecules from the start, and a way to power further biochemical evolution.

What then, were the conditions inside an alkaline vent that could have sequestered organic molecules and led to metabolic evolution and ultimately, life? Let's consider an intriguing proposal that gets at an answer!

The porous rock structure of today's alkaline deep-sea vents provides micro-spaces or micro-compartments that might, in a prebiotic past, have captured some of the alkaline liquids they emitted. In fact, conditions in today's *white smokers* also support the formation of hydrocarbon ***biofilms***, primitive organic membranes that could have been lined those porous rocky micro-compartments. Alkaline vent emissions would then be trapped within primitive biofilm membrane vesicles enclosed by rocky "cell walls."

An interesting consequence of this scenario would be the existence of a natural *proton gradient* between the *alkaline* contents of the micro-compartments and the surrounding *acidic* ocean waters. Did all this happen? Perhaps! Without a nutrient-rich environment, *heterotrophs-first* is not an option. That leaves only the alternate option: an *autotrophs-first* scenario for the origins of life. Nick Lane and his coworkers proposed that the *natural proton gradients* suggested above were the selective force behind the evolution of early metabolic chemistries in alkaline vents (^{20,11} [Prebiotic Proton Gradient Fuels Origin of Life](#)). Organized around biofilm compartments, prebiotic structures and chemistries would have harnessed the free energy of these proton gradients. In Lane's view, the first protocells, and therefore the first cells, may have been *chemoautotrophs*. If so, ask yourself if early autotrophy could have evolved without an electron transport system!

Finally, how might chemoautotrophic chemistries on a non-reducing planet have supported polymer formation, as well as polymer replication? Today we see storage and replication of information in nucleic acids as separate from enzymatic catalysis of biochemical reactions. But are they all that *separate*? If replication is the faithful reproduction of the information needed by a cell, then enzymatic catalysis ensures the redundant production of all molecules essential to make the cell! Put another way, if catalyzed *polymer* synthesis is the replication of the workhorse molecules that accomplish cellular tasks, then what we call 'replication' is simply the replication of nucleic acid *information* needed to faithfully reproduce these workhorse molecules. Was there an early, coordinated, concurrent selection of mechanisms for the catalyzed metabolism as well as catalyzed polymer synthesis and replication? We'll return to these questions when we consider origins of life in an *RNA world*.

An origin of life scenario in a non-reducing (and oxygen-free) atmosphere raises additional questions. Would proton gradients provide enough free energy to fuel and organize life's origins? If so, how did cells arising from prebiotic chemiosmotic metabolism harness the energy of a proton gradient? Before life, were protocells already able to transduce gradient free energy into chemical free energy? Was ATP selected to hold chemical free energy from the start? Or was it selected only concurrent with metabolic selection? Alternatively, was the relief of the gradient coupled at first to the synthesis of other high-energy intermediate compounds, with for example, thioester linkages? Later, how did cells formed in alkaline vents escape the vents to colonize the rest of the planet? By whatever series of primordial events the energy of a natural proton gradient would be initially captured, the *chemoautotrophic LUCA* must already have been using membrane-bound proton pumps coupled to electron transport, and a membrane ATP synthase to make a proton gradient and then harness its free energy to make ATP, since *all* of its descendants do so.

Finally, when did photoautotrophy (specifically *oxygenic photoautotrophy*) evolve? Was it a late evolutionary event? Is it possible that photosynthetic cells evolved quite early among some of the chemoautotrophic denizens of the white smokers, biding their time before exploding on the scene to create our oxygenic environment?



20.6 Heterotrophs-First vs. Autotrophs-First: Evolutionary Considerations

In the alkaline vent scenario, chemiosmotic metabolism predated life. Therefore, the first chemoautotrophic cells did not need the fermentative reactions required by cells in a heterotrophs-first origin scenario. Even though all cells alive today incorporate a form of glycolytic metabolism, *glycolysis may in fact, not be the oldest known biochemical pathway*, as we have thought for so long. In support of a late evolution of glycolytic enzymes, those of the archaea show little structural resemblance to those of bacteria. If fermentative heterotrophy was a late evolutionary development, then LUCA and its early descendants would lack a well-developed glycolytic pathway. Instead, the LUCA must have been one of many 'experimental' autotrophic cells, most likely a chemoautotroph deriving free energy from inorganic chemicals in the environment. To account for heterotrophy in the three domains of life, it must have evolved separately in the two antecedent branches descending from the last universal common ancestor of bacterial, archaeal, and eukaryotic organisms. The selection of similar traits (fermentative biochemistry in this case) in unrelated organisms is called **convergent evolution**.

The phylogenetic tree in Figure 20.11 below features an autotrophic LUCA in an *autotrophs-first* scenario. The tree traces a separate (convergent) evolution of heterotrophy in two branches of the descendants of the LUCA by tracing the spread of fermentative pathways in all living things on the familiar three domain phylogeny.

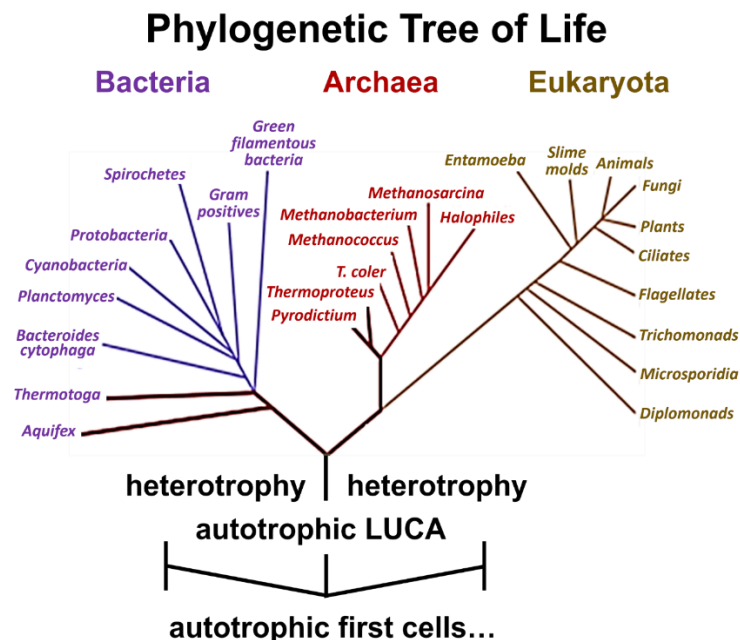


Fig. 20.11: This evolutionary (phylogenetic) tree of living organisms assumes autotrophic first cells and an autotrophic LUCA, with heterotrophy and heterotrophic metabolism evolving independently in ancestors of Bacteria on the one hand (left) and Archaea and Eukaryota on the other (right)

20.7 Life's Origins, A Summing Up

Speculation about life's origins begins by trying to identify a source of free energy with which to make organic molecules. The first cells might have been heterotrophs formed in a reducing Earth environment, from which autotrophs later evolved. On the other hand, the earliest cells

may have been autotrophs formed under non-reducing conditions in the absence of a primordial soup. Then, only after these autotrophs had produced enough nutrient free energy to sustain themselves, did heterotrophs belatedly emerge. Evidence suggesting that the Earth's atmosphere was a non-reducing one more than 4 billion years ago (soon after the formation of the planet) and suggesting life on Earth 3.95 billion years ago favor metabolic origins of autotrophic life in a thermal vent, likely an alkaline one. Questions nevertheless remain about life-origins under non-reducing conditions. Even the composition of the prebiotic atmosphere is still in contention. For now, let us put these concerns aside and turn to events that get us from the LUCA and its early descendants to the still more elaborated chemistries common to all cells today. The descriptions that follow are educated guesses about early pathways on the road to the familiar cellularity we see now on Earth. They mainly speculate on the selection of catalytic mechanisms, replicative metabolism, the web of intersecting biochemical pathways, and even more intricate chemical communications that organize cell function and complexity.



[352 Phylogenetic Support for Autotrophs-First Origins of Life](#)

20.8 Origins of Life Chemistries in an RNA World

Following the origins of organic monomers in a reducing environment in the tidal pool scenario, the energy for polymer formation came from cycling temperatures on an overheated Earth. In that scenario, we considered how chains of nucleotides could be synthesized and even replicated to form populations of nucleic acids with similar sequences. But if the prebiotic environment was non-reducing, where would the energy have come from to make any polymers, let alone ones that could replicate themselves? If you guessed that the energy was provided by a proton gradient between biofilm-enclosed acidic proto-cells and an alkaline ocean..., you would have been right! In this case, polymers would have been synthesized in enclosed spaces, and not in tidal pools only to be dispersed and diluted in the wider oceans. But then, how would replicative, informational, and catalytic chemistries have arisen from these organic monomers and polymers? Polypeptides would have formed, but they have no inherent chemical or structural basis for self-replication. Unlike polypeptides, we saw in describing the tidal pool scenario that polynucleotides (nucleic acids) do! In fact, as we have already seen, evidence is accumulating to support the hypothesis that life (and its first genome) originated in a ***RNA world***:

- Glycoside bond synthesis between prebiotic bases and sugars may have been possible.
- Today's RNAs include ***ribozymes*** that catalyze their own replication self-splicing introns).
- Some RNAs are part of ***ribonucleoproteins*** that have at least ***co-catalytic activities***, like ribosomes, the small ribonucleoproteins (snRNPs) in spliceosomes, and the secretory signal recognition particle.
- ***Retroviruses*** (e.g., HIV) store their genetic information in *RNA genomes* that may have been integral to the emergence of cellular life.

Ribozymes, ribonucleoprotein structures, and retroviruses may be legacies of a prebiotic RNA world. In fact, in an in vitro evolution study, self-replicating ribozyme polymerases in a test tube become more efficient at replicating a variety of increasingly longer

and more complex RNAs over time. For more about these autocatalysts that support an RNA world, check out ^{20,12}[Real & Artificial Ribozymes Catalyze RNA synthesis](#).

There are hypothetical *RNA world* scenarios for the origins of replicating, catalytic polymers. There is even an organic chemical, an **autocatalyst** that can catalyze its own synthesis. So, which may have come first? A self-replicating RNA or other self-replicating organic molecule? Arguably, chemical evolution of an autocatalytic RNA is a stretch, but at least one organic molecule, Amino Adenosine Triacid Ester (AATE), is a present-day self-replicating *autocatalyst*. Could a molecule like AATE have been a prebiotic prelude to the RNA world? Figure 20.12 (below) shows the structure and replication of AATE.

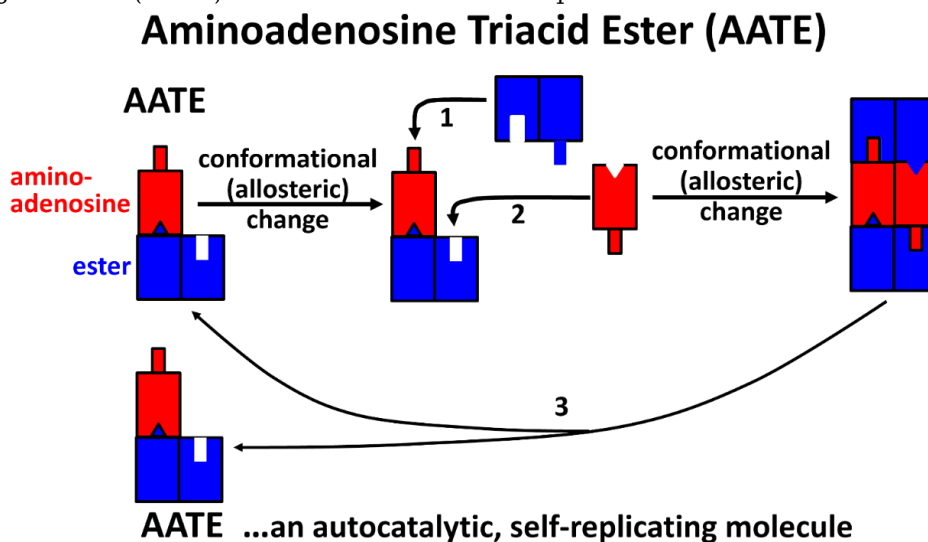


Fig. 20.12: Aminoadenosine triacid ester (AATE) catalyzes its own replication by the mechanism suggested here.

The replicative reaction proceeds in the following steps:

1. An aminoadenosine triacid ester molecule binds a second aminoadenosine.
2. A second triacid ester molecule binds to "di-aminoadenosine triacid ester." The two aminoadenosines, now in opposite orientations, can attract and bind a second triacid ester.
3. After bond-rearrangements, the "di-aminoadenosine, di-triacid ester" molecule separates into two molecules of AATE.

This reaction is catalytic because the stereochemistry of the reacting molecules creates an affinity of the AATE molecule first for another *free* aminoadenosine molecule. The structure formed allows (i.e., catalyzes) linkage of a second *free* triacid ester, leading to the separation of two AATE molecules. Subtle, sequential changes in the molecular conformation of the molecules result in the changes in affinities of the molecules for each other. In the replicative reaction, the AATE, *free* ester and *free* aminoadenosine concentrations would drive the reaction. Could AATE-like molecules have been progenitors of autocatalyzed polymer replication? And before that, could something like reduced NAD (NADH) have supplied free energy to fuel AATE synthesis in the first place? After all, it too contains a nucleotide! *Could replication of a prebiotic AATE-like molecule then have led to an RNA world?* Could primitive RNAs have been stabilized by binding to short prebiotic peptides, becoming forerunners of ribozymes? The possibility of a prebiotic AATE-like precursor to an RNA world is intriguing because the 'triacid' includes the purine adenosine! On the other hand, the possibility of

prebiotic replicating RNA-peptide complexes implies the origins of life in an **RNA-Protein world** (rather than exclusively RNA-world)! Whether life began in an RNA world or an RNA-protein world, catalyzed replication is of course another property of life.



[353 AATE: An Autocatalytic, Self-Replicating Organic Molecule](#)

20.9 Experimental Evidence for an RNA World

There is some circumstantial evidence for at least one organic autocatalyst as a candidate precursor to an RNA world. There is also evidence that the prebiotic chemistries could have supported the linkage of bases to sugars to make nucleosides (see ^{20.13}[Prebiotic Ribonucleotide Synthesis](#)). Also, cyclic nucleotides (e.g., cAMP) were shown to be stable in hot water (^{20.14}[Stable Prebiotic cAMP](#)). Could cyclic nucleotides have served as substrates for prebiotic RNA synthesis? In fact, investigators probing an *RNA world* hypothesis reported the requirements for RNA replication in a test tube. Adding salt, magnesium ions, a bit of RNA primer and A, U, G and C (precursor bases to RNA) to a buffered basic solution resulted in RNA synthesis... without an RNA polymerase! The reaction was slow and prone to error. But RNA was synthesized against the primer in a simple solution, the kind that might have been found on Earth at life's beginnings. The study further found that adding the base inosine to the reactions increased the rate as well as the accuracy of RNA replication under what were otherwise the same conditions. Of course, this is not what happens today! First, the bases are part of nucleosides and nucleotides that are the actual precursors of replication. And, while inosine is in fact found in some tRNAs, it does not result from the use of inosine nucleotide precursors, but from the chemical modification (deamination) of adenine bases already in the transcript as seen in Figure 20.13 (below).

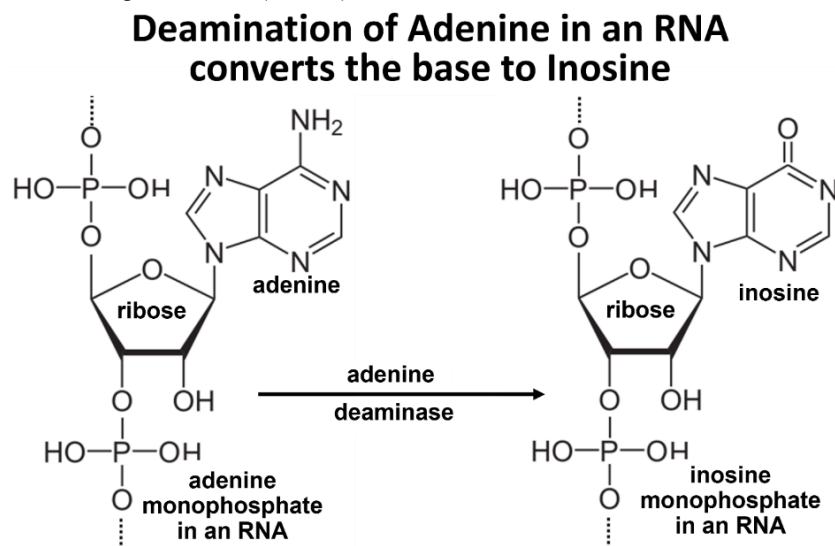


Fig. 20.13: The deamination of adenine in a ribonucleotide chain to inosine in a tRNA by the **adenine deaminase** enzyme.

OK, calling inosine the “missing ingredient” in life’s origins might be hyperbole. Still, if cell-free RNA synthesis using nucleobases is more efficient with inosine than without it,

could inosine have participated in an early RNA world at life's origins, to be replaced later by our DNA world? Read more about possible roles for inosine or other unusual bases in a prebiotic RNA world at ^{20,15}[Inosine-a Missing Ingredient in the Origin-of Life?](#).

20.10 Molecules Talk: Molecular Communication and Complexity

In our complex human society, we define **communication** by its **specificity**. Without a careful choice of words, our speech would at best, be a source of magnificent misunderstanding..., or just plain babel! What does this mean for prebiotic chemistries? In any prebiotic chemical evolution, selection would have favored the protective accumulation of longer-lived molecular aggregates. Over time, the same selective imperatives would create webs of such aggregates, increasing the *range and specificity* of molecular interactions in a very challenging environment. If this were to have occurred in an enclosed proto-cellular space, it would have resulted in a primitive *molecular communication* and the potential for a growing **complexity** (another property of life!). In fact, properties of life must have arisen before life itself, as they underlie the achievement of increasingly complex intermolecular communication. Simply put, a prebiotic (or for that matter a cellular) genetic change that alters the rate of one catalytic reaction (if not destructive) will drive the selection of changes in components of other, interconnected metabolic chemistries. Finally, if we agree that *molecular communication* required the evolution of catalytic specificity, then we must agree that the elaboration of complexity and order in fact, a property of life, required the selection of mechanisms of **regulation** and **coordination**.

20.10.1 Intermolecular Communication: Establishment of Essential Interconnected Chemistries

Earlier, we suggested that inorganic catalyst precursors to biological enzymes could be minerals embedded in clay or other substrata, providing surfaces that would naturally aggregate organic molecules and catalyze repetitive reactions. The initial objects of prebiotic selection must have included stable monomers and polymers outside, or as seems more likely, inside proto-cells. Later, chemical selection would have favored polymers that enhanced growth and reproduction of successful aggregates. These polymers were likely those that catalyzed their own synthesis, perhaps collaborating with inorganic catalytic minerals. The result would be the elaboration of a web of *interconnected chemical reactions* between molecules with high affinity for each other, thereby increasing the specificity of those reactions. In the context of life origins and evolution, **co-catalysis** describes the activities of these interconnected metabolic reactions.

As noted, high-affinity interactions are inherently **protective**. During prebiotic chemical and/or metabolic evolution, protected stable **molecular assemblies** would be targets of selection. Continuing co-evolution of catalysts, substrates, and co-catalytic reaction sets would lead to more sophisticated *molecular communication*. Once established, efficient biochemical reaction sets would be constrained against significant evolutionary change. Any change (mutation) that threatened this efficiency would mean the end of a prebiotic chemical (or for that matter, cell) lineage! This explains why we find common pathways for energy-generation (e.g., autotrophic and fermentative), reproduction (i.e., replication), and information storage and retrieval (DNA, RNA, protein synthesis) in all of LUCA's descendants.

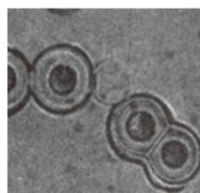
In other words, an organizing principle must have been selected to enable communication between molecules and their reactions. Such complex and effective communication requires **coordination**. In fact, effective communication is *defined* by coordination, the capacity to make *chemical decisions*. Selection of molecular aggregates that sequestered metabolic reactions behind a semipermeable membrane ensures that only certain molecules communicate with each other within a protocell (or cell). This sequestration is likely to have occurred repeatedly during chemical evolution, beginning with the synthesis of larger, polymeric molecules and possibly, an aggregation of primitive lipoidal molecules. We can think of increasingly effective catalysis in an enclosed, sheltered environment as **a conversation mediated by good speakers!** Thus, *coordination* is a property that started with prebiotic chemistry and likely co-evolved with life itself.

20.10.2 Origins of Coordination

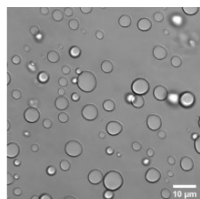
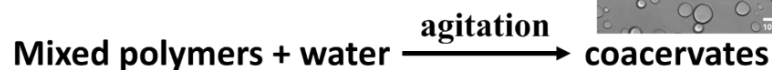
Let's look some possible structures churning around in the prebiotic chemistry set that might have self-assembled and sequestered compatible chemistries of life. Along with the alkaline vent *biofilm compartment*, **proteinoid microspheres**, **coacervates**, and **liposomes** have been considered as possible progenitors of biological membranes. Each can be made in the laboratory, shown to be semipermeable and in some cases can even replicate! Production of coacervates, proteinoid microspheres, and liposomes is summarized in Figure 20.14.

Candidates for Early Compartments (the general stickiness of things)

A. Proteinoid microspheres



B. Coacervates



C. Liposomes

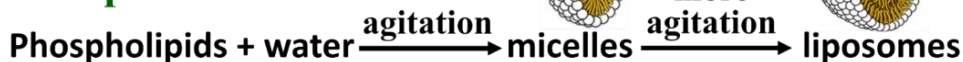


Fig. 20.14: Proteinoid microspheres, coacervates and liposomes can all be made in a laboratory and are candidates for boundary structures that could protect early prebiotic organic molecules and reactions. See more micrographs of these synthetic compartments at ^{20.16}[Candidates for Early Compartments](#).

Oparin had proposed that the action of sunlight in the absence of oxygen could cause ionized, oppositely charged organic molecules (e.g., amino acids, in a primordial soup to form "colloids," or "droplets" which were immiscible with water. These **coacervates** were actually produced in 1932, visualized by microscopy and shown to be a semi-permeable compartment. They even behaved as if they were able to grow and reproduce (also as Oparin originally suggested they might). **Sidney Fox** produced **proteinoid microspheres** from short peptides In the 1950s; they formed spontaneously from aqueous amino acid solutions heated to dryness, not unlike what happens in the tidal pool scenario of polymer formation from organic monomers. These can be seen by light and electron microscopy.

Liposomes are also easily made in a laboratory, but it isn't clear that they could have formed spontaneously on a pre-biotic earth. Nevertheless, cell membranes must have had acquired their phospholipid bilayer structure by the time of LUCA since we all have them! Prior to LUCA (perhaps in or soon after formation of *our* progenote), chemical rearrangements must have occurred to enable incorporation of a phospholipid bilayer into whatever semipermeable boundary life started with.

We have already considered the **biofilm** proposed for cellular origins in an alkaline vent, and that the formation of such biofilms in alkaline vents would have separated acidic ocean protons from the interior of such protocells, creating a proton gradient. Such a gradient could have driven the early evolution of chemiosmosis as a means of capturing chemical energy, complete with the eventual selection of ATP synthases and the enzymes of proton transport, again because all cells descendent from LUCA possess these biochemistries.

Of course, proteinoid microspheres, coacervates, biofilm-based membranes, and liposomes are not alive, and are therefore not cells. But one or another of them must have been where the enhanced **coordination of molecular communication** required for life began their elaboration.



[354-2 Protected Molecular Communication-Semipermeable Membranes](#)

An important **take-home message** here is that whatever the original structure of the first cells was, they arose soon after the organic chemical prerequisites of life began to acquire familiar metabolic functions. We need to see chemical and structural progress to cellularity as concurrent metabolic evolutionary events. At some point, selection of sequestered biochemistries led to **protocells**, then to the first cell or cells, each with all the properties of life. Finally, selection of highly specific pathways of communication between cellular molecules allowed cells themselves to talk to one another, engage in group activities, and eventually join to form multicellular organisms.

Multicellularity is of course a characteristic of many if not most eukaryotes. But watch an excellent TED Talk on bacterial intercellular communication by Dr. Bonnie Bassler at ^{20.17}[Intercellular Communication in Bacteria](#).

20.10.3 An RNA World: Origins of Information Storage and Retrieval

Let us accept for now that molecular communication began concurrently with the packaging of interconnected co-catalytic sets into semipermeable structures. Then the most 'fit' of these structures were selected for their efficient coordination of meaningful, timely chemical messages. Ultimately, coordination requires *information processing, storage and retrieval*, something we recognize in Francis Crick's *Central Dogma* of information flow from DNA to RNA to protein. Cells and organisms do coordination quite well, but what do its beginnings look like? The answer may lie in the pre-biotic RNA world. Figure 20.15 below is a statement of *The Central Dogma*, modified to account for the role of reverse transcription in the behavior of retroviruses and retrotransposons, both sources of lateral gene transfer (the exchange of genes between cells and organisms).

Information Storage & Retrieval Today

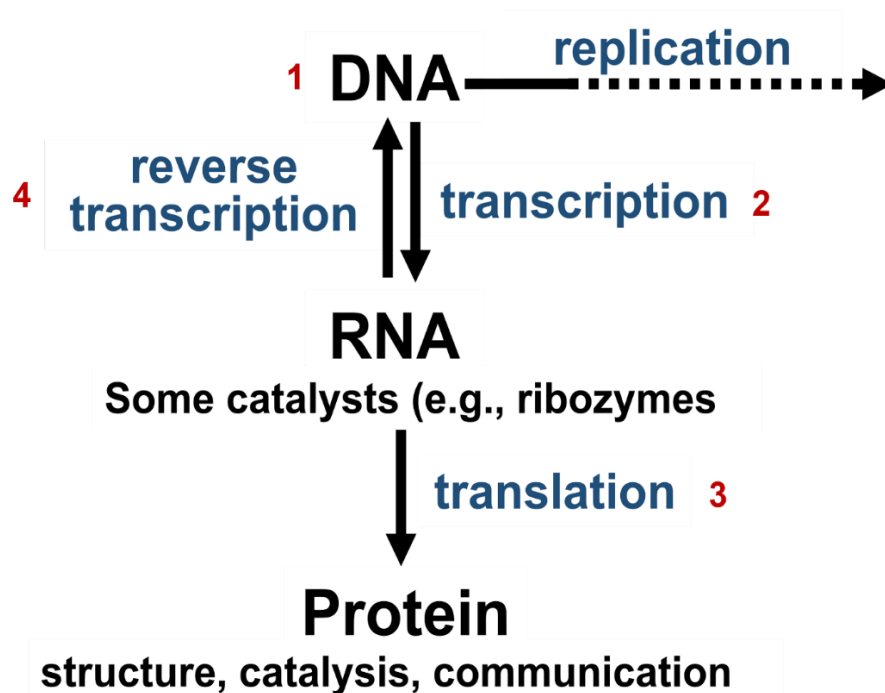


Fig. 20.15: Genetic information flows from DNA to RNA to protein (the *Central Dogma*) but can also flow from RNA to DNA by reverse transcription.

We do not really know how cells came to rely on DNA to store, pass on and mobilize genetic information, but we have presented reasons to believe that the first replicating nucleic acid was RNA, creating an *RNA world*. Here is the evidence that leads us to this conclusion.

- Based on the stem-and-loop and other structures that form when RNA molecules undergo internal H-bonding, we know that RNAs can take on varied and intricate shapes.
- Diverse conformations are consistent with the evolution of specificity in the interaction of RNAs with themselves and/or with other molecules in the prebiotic environment.
- RNAs, either alone as *autocatalysts* (for example, self-splicing mRNAs) or in catalytic ribonucleoprotein complexes (e.g., ribosomes, snRNPs) that exist in cells today.
- Some of these RNAs (specifically rRNAs), have a long phylogenetic heritage, shared by cells in all three domains of life.

The propensity of single stranded RNA molecules to fold based on internal H-bonding can lead to those diverse 3D shapes (tertiary structure). These structures could have interacted with other molecules in a prebiotic environment. Because they could be replicated according to different prebiotic scenarios, the same RNAs could also pass on simple genetic information contained in their base sequences. The combination of informational and catalytic properties in a single molecule is illustrated in Figure 20.16 below.

RNA: the first information storage molecule

Shape, Specificity, Catalysis

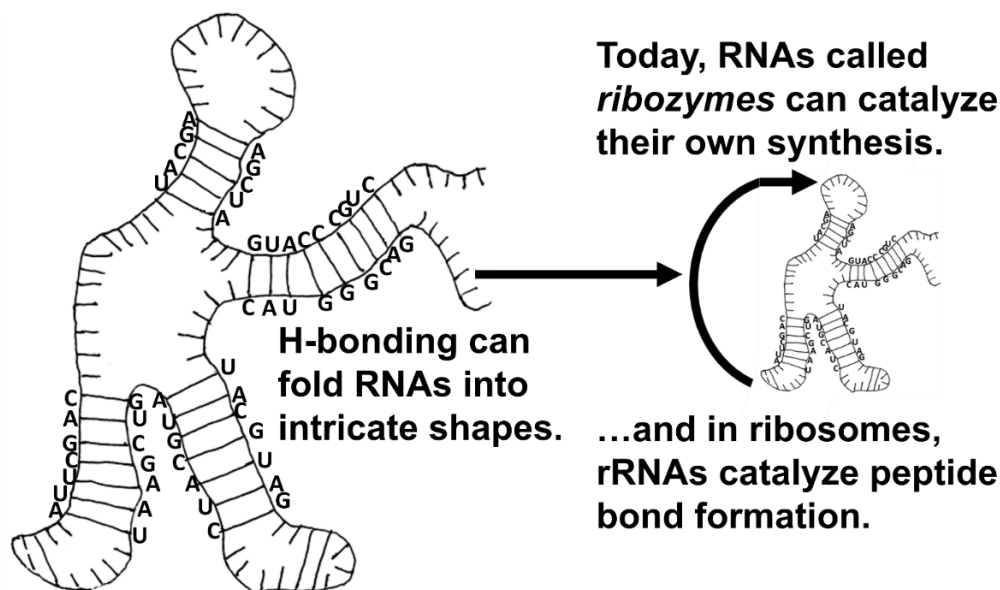


Fig. 20.16: RNA sequences contain genetic information. In retroviruses, they are molecules of inheritance. But single stranded RNAs made in cells can fold into 3-dimensions, creating specific shapes that can act as catalysts (e.g., ribozymes) combining information storage with catalytic activities.

The capacity of RNAs to be catalysts and warehouses of genetic information at the same time speaks to an efficient candidate for the first *dual* or *multi-purpose* polymer, a property that is not known and cannot be demonstrated for DNA. To read more about the proposed 'RNA worlds' in which life may have begun, see Cech T. R. (2012) [*The RNA Worlds in Context*. In *Cold Spring Harbor Perspectives in Biology* (Cold Spring Harbor, NY: Cold Spring Harbor press) 4(7):a006742e].



[355 Self-Replication: Information, Communication & Coordination](#)



20.10.4 From Self-Replicating RNAs to Ribozymes to Enzymes; From RNA to DNA

What might RNA catalysis beyond self-replication have looked like in simpler times? One can envision a pair of different RNA molecules, each uniquely folded, with affinities for two different amino acids (Fig. 20.17, below). After they bind to 'their' amino acids, the RNAs are attracted to each other, presumably by refolding to permit complementary H-bonding between

the RNAs (not shown). The resulting ribozyme catalyzes formation of a peptide linkage between the amino acids (i.e., a dehydration synthesis), after which the dipeptide and the two RNAs would separate.

Imagining Ribozyme Catalysis of Dipeptide Synthesis in an RNA World

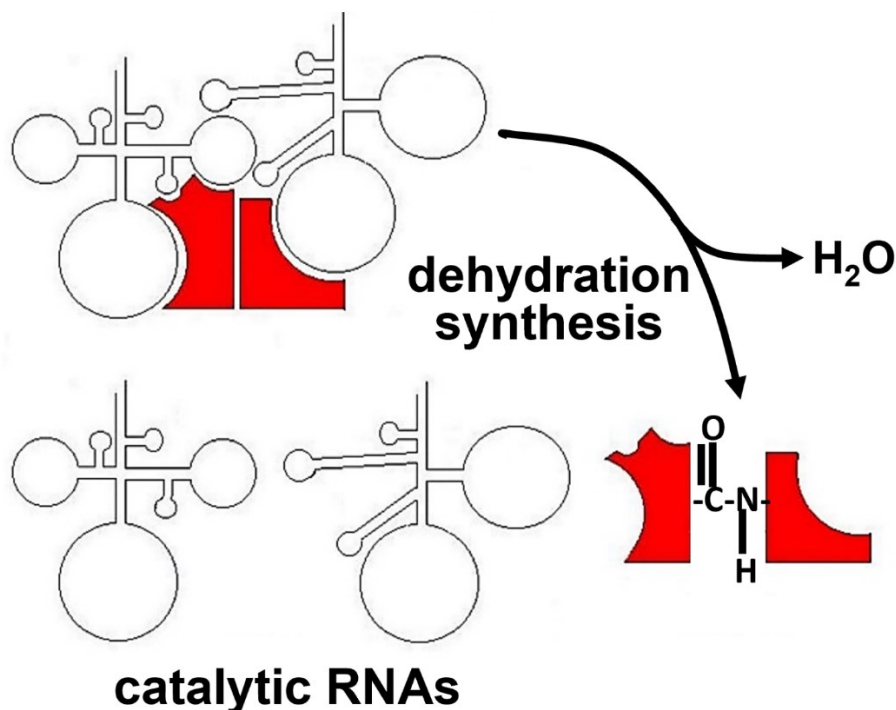


Fig. 20.17: Hypothetical origin of RNA catalysis, suggesting how some early (prebiotic) RNAs may have evolved to catalyze peptide bond formation between amino acids.

As we know, the formation of a peptide bond requires an input of free energy (recall that this is one of the most energy intensive reactions in cells). For now, assume a chemical energy source and let us focus on the specificities required for RNA catalytic activity.

We know now that tRNAs are the intermediaries between nucleic acids and polypeptide synthesis. So, it's fair to ask how the kind of activity illustrated above could have led to the tRNA-amino acid interactions we see today. There is no obvious spontaneous binding chemistry between today's amino acids and RNAs, but there may be a less obvious legacy of the proposed bindings. This has to do with the fact that the genetic code is universal, which means that any structural relationship between RNA and amino acids must have been selected early (at the start!) of cellular life on Earth. Here is the argument.

1. The code is indeed universal (or nearly so)
2. There is a correlation between the chemical properties of amino acids and their codons, for example:
 - Triplet codons for charged (polar) amino acids contain more G (guanine) bases
 - Triplet codons for uncharged amino acids more often contain a middle U (uracil) than any other base.

These correlations would mean that an early binding of amino acids to specifically folded RNAs was replaced in evolution by enzyme-catalyzed covalent attachment of an amino acid to a 'correct' tRNA, such as we see today.

What forces might have selected separation of the combined template and informational functions from most of the catalytic activities of RNAs? Perhaps it was the selection of the greater diversity of structure (i.e., shape) that folded polypeptides can achieve, compared to folded RNAs. After all, polypeptides are strings of 20 different amino acids compared to the four bases that make up nucleic acids. This potential for molecular diversity would in turn accelerate the pace of chemical (and ultimately cellular) evolution. A scenario for the transition from earlier self-replicating RNA events to the translation of proteins from mRNAs is suggested in Figure 20.18.

Possible First Steps to Separating RNA Template and Informational Functions from Most RNA Catalytic Activities

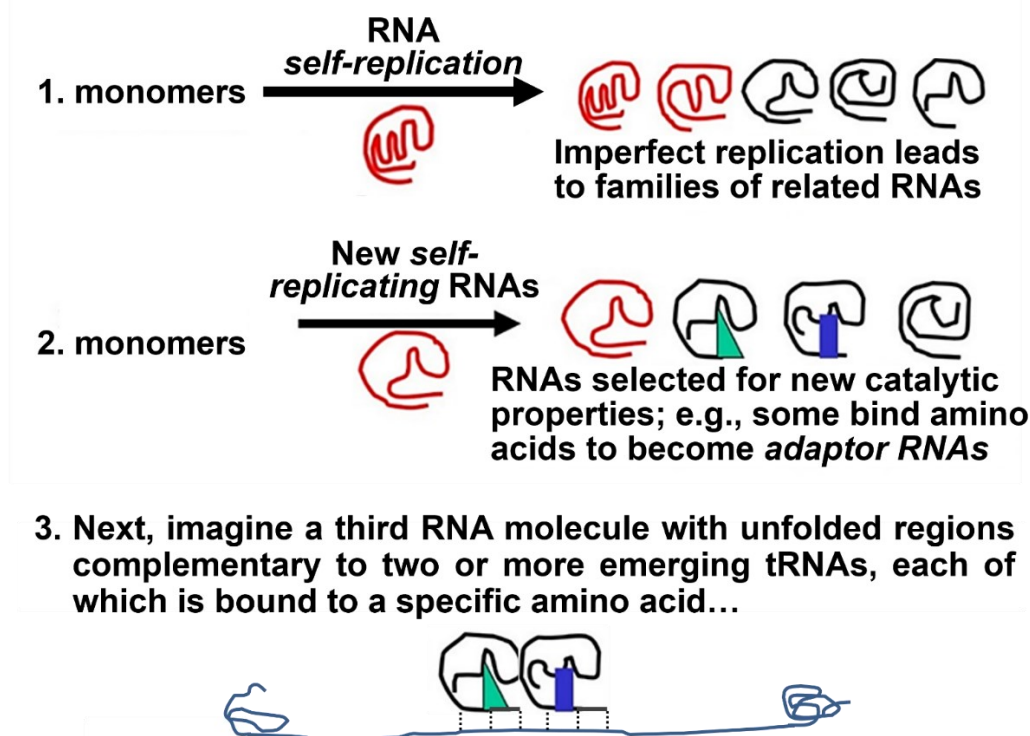


Fig. 20.18: Suggested steps in evolution from an RNA world to our DNA world. The first of the steps would be to divorce the self-replication activity of RNAs from their information storage function, as illustrated here.

Adaptor RNAs in the illustration will become tRNAs. The novel, relatively unfolded RNA is a presumptive mRNA (even though mRNAs can sometimes engage in intrastrand H-bonding). Thus, even before the intrusion of DNA into our RNA world, we can imagine selection of the defining features of the genetic code and mechanism of translation (protein synthesis) that characterizes all life on Earth. Next, we'll consider "best-speculations" for how RNA-based information storage and catalytic chemistries might have made the evolutionary transition to DNA-based information storage and largely protein-based enzyme catalysis.

20.10.4.a Ribozymes Branch Out: Replication, Transcription and Translation

The term *co-catalysis* describes biochemical reactions in which a catalyst accelerates a chemical reaction whose product feeds back in some way on its own synthesis. We saw this in action when we discussed allosteric enzyme regulation and the control of biochemical pathways. Catalytic feedback loops must have been significant events in the evolution of the *intermolecular communication* and the *metabolic coordination* required for life. Here we look at some scenarios for the transition from an RNA world to something more recognizable as today's nucleic acid information storage and protein-based catalytic metabolism.

If early RNAs catalyzed their own replication, they were functioning as primitive ribozymes. If some of these ribozymes bound and polymerized amino acids, they may have catalyzed the synthesis of short peptides. What if some of the polypeptides occasionally bound to their shapely catalytic RNAs and enhanced the catalytic properties of the aggregate? What if one of these aggregates enhanced the rate of synthesis of its own RNA, or evolved to catalyze other reactions useful to life? Such a structure might presage the ribosome, which is a ribonucleoprotein with catalytic properties. If later in evolution, a peptide changed just enough to accomplish the catalysis on its own, it might dissociate from its RNA, flying solo as a protein enzyme catalyst, as suggested in Figure 20.19.

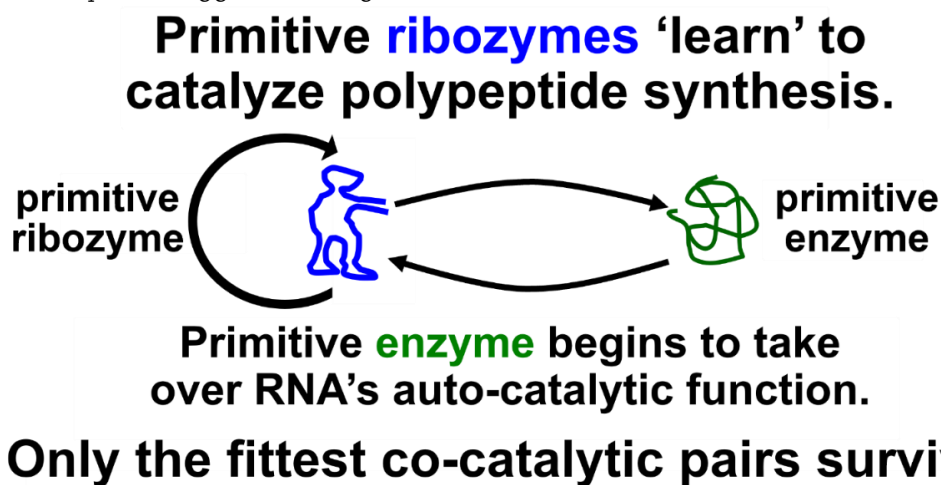


Fig. 20.19: If RNAs could bind peptides whose synthesis they catalyze, evolution could have selected ribozymes from some RNA-peptide complexes that did not separate. Later, the peptides themselves could have evolved to take over catalytic functions from ribozymes.

Selection favoring the synthesis of short oligopeptides and polypeptides is consistent with a catalytic diversification that led to the dominance of protein catalysts, i.e., enzymes. The primitive enzyme shown here must have been selected because at first, it assisted the autocatalytic replication of the RNA itself! Over time, the enzyme would evolve along with the RNA. This co-evolution then eventually replaced autocatalytic RNA replication with the enzyme-catalyzed RNA synthesis we recognize as transcription today. In this scenario, self-splicing pre-mRNAs and ribozymes are surviving remnants of an RNA world!



Let's turn now to some speculations about *how* an RNA world could make the transition to the DNA-RNA-protein world we have today.

20.10.4.b Transfer of Information Storage from RNA to DNA

The transfer of function from RNA to DNA is by no means a settled issue among students of life origins and early evolution. A best guess is that the elaboration of protein enzymes begun in the RNA world would lead to reverse transcriptase-like enzymes that copied RNA information into DNA molecules. The basic transfer scenario is illustrated below in Fig 20.20.

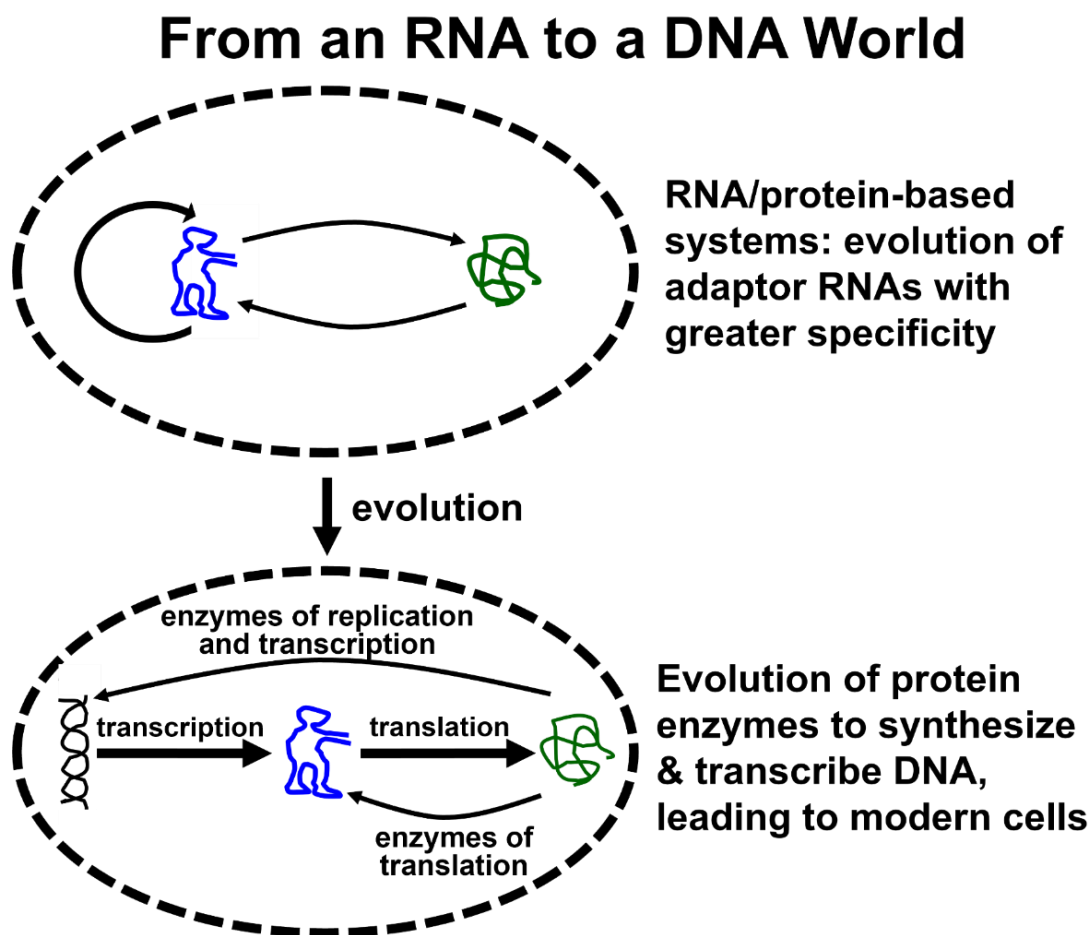


Fig. 20.20: A final evolutionary scenario gets us from the RNA world to a DNA world that governs most life on earth today.

DNA information may have been selected because DNA is chemically more stable than RNA. All cells alive today store information in DNA (only some viruses have an RNA genome). Therefore, transition to the use of DNA as an information molecule would have preceded the origin of life. At least, it must have occurred in the cells from which the LUCA arose. Details of this key change involve evolutionary steps yet to be worked out to everyone's satisfaction!



[357 The Transition from an RNA World to a DNA World](#)



20.11 The Evolution of Biochemical Pathways

The tale of the evolution of enzymes from ribozymes and of informational DNA from RNA, and the other metabolic chemistries behind prebiotic semipermeable boundaries is ongoing in cells today. Undoubtedly, early cellular metabolism involved only reactions crucial to life..., catalyzed by a limited number of enzymes. But, if evolution inexorably trends towards greater complexity of molecular communication and coordination—in other words, towards increasingly refined regulation of metabolism—how did the repertoire of enzymes get larger and how did biochemical pathways become more elaborate? We answered the first question elsewhere, when we discussed gene duplication (e.g., by unequal crossing over). Duplicated genes encoding the same enzyme could provide the raw material for new enzymes and new enzymatic functions.

Whether in cells or in prebiotic structures, we can hypothesize how a new chemical reaction could evolve. For example, assume that a cell usually gets molecule **D** required for an essential function, from an external, environmental source. What would happen if levels of **D** in the environment become *limiting*? Cells would die with insufficient **D**..., unless some cells in the population already have a duplicated, redundant gene that has mutated and now encodes an enzyme with the ability to make **D** in the cell. Such a cell might have coexisted with cells without the mutation, but a **D**-limited environment would select the mutant cell for survival and reproduction. Imagine the scenario illustrated in Fig 20.21.

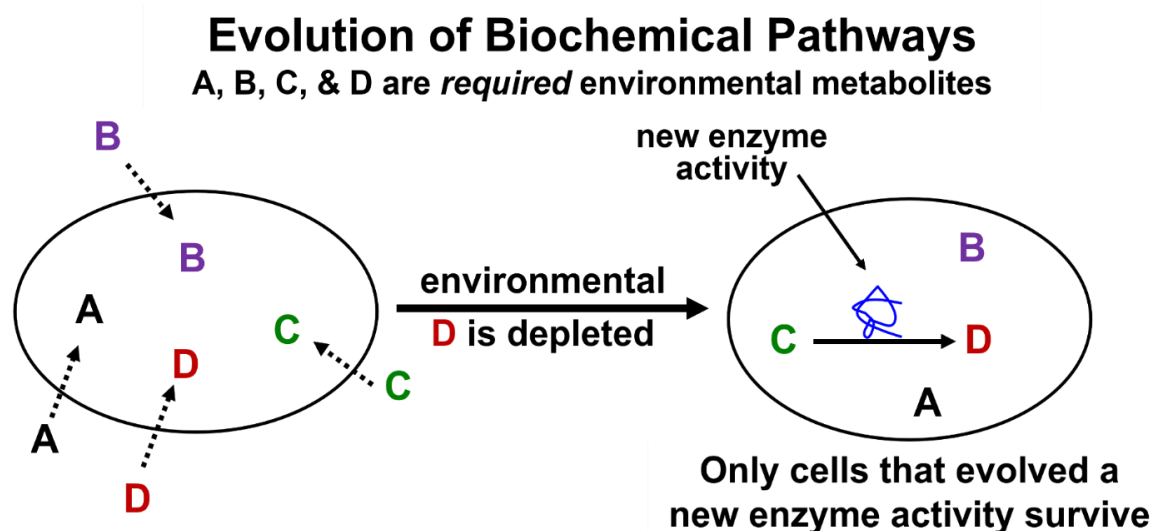


Fig. 20.21: One selective pressure that contributes to the complexity of biochemical pathways would be the depletion of a particular chemical resource, say molecule **D**, in the environment). The few cells in the population that happen to have an enzyme that can convert **C** to **D** will survive and proliferate.



[358 Origins and Evolution of Biochemical Pathways](#)

In a similar scenario, a mutation in a duplicated gene could result in a novel enzyme activity that can convert some molecule (e.g., **C** or **D**) in the cell into a new molecular product.

If the new enzyme and molecular product do not kill or debilitate the cell, the cell might survive to be selected by some future exigency.

20.12 A Grand Summary and Some Conclusions

Our consideration of how life began on Earth was intentionally placed at the end of this textbook, waiting for us to get a handle on how cells work. Clearly any understanding of life origins scenarios is very much a matter of informed if divergent speculations. Alternative notions for the origins of life entertained here all address events that presaged life under 'best-guess' hypothetical conditions. After trying to get a grip on prebiotic events, we asked how we got from what could have happened under a given set of prebiotic conditions to the cellular life we recognize today.

All proposals recognize that the first cells had all the properties of life (including evolution itself). Starting with that common understanding, all arguable scenarios try to navigate pathways from primitive, less controlled chemistries to more regulated and coordinated metabolisms, in other words from chemical simplicity to biochemical complexity. The chemical and metabolic evolution that began before life may have overlapped in time with cellular evolution, at least until the arrival of the LUCA.

While chemical evolution was mainly a series of selections by the physicality of a prebiotic world, the arrival of our LUCA, of life, contends with both that physical world, and with life itself. LUCA, the universal common ancestor, had already escaped the RNA world, replicating DNA, transcribing RNA, and translating mRNAs into polypeptides, all behind a semipermeable phospholipid bilayer.

Whether a heterotroph or (increasingly more likely) an autotroph, LUCA used the energy of ATP to power all of its cellular work, as do its descendants. Thus, evolution after the LUCA, is focused on continued selection of the complexities of metabolism that enables the spread and diversification of life from wherever it started. The selection of chemistries and characters encoded among by already, accumulated random neutral genetic changes is ongoing, and continues to increase the diversity of species and their spread to every conceivable ecological niche on the planet. The take-home message of this chapter should thus be that:

- Prebiotic chemicals interacted with each in a free energy rich environment.
- Protected chemistries that resisted environmental degradation were subject to a prebiotic selection (*chemical evolution*).
- Chemical evolution led to a diversified *chemistry set* with which to experiment with combinations that could produce an entity with all the properties of life..., the first cell.
- Life has been evolving ever since!
- An understanding of the molecular basis of evolution can help us understand how life may have begun, spread, diversified, and been sustained on Earth (or elsewhere!).

Some iText & VOP Key Words and Terms

AATE	Hadean eon	proteinoid microsphere
abiogenesis	heat of baking	protocell
adaptor RNA	heterotrophs-first	reducing atmosphere

alkaline hydrothermal vent	ionizing radiation	retroviruses
aminoadenosine triacid ester	Last universal common ancestor	ribonucleoproteins
Archean eon	liposome	ribozymes
autocatalysis	LUCA	progenote
autotrophs-first	metabolic evolution	RNA world
biofilm	molecular communication	serpentinite
biogenesis	non-reducing atmosphere	serpentinization
black smoker	ozone layer	spontaneous generation
chemoautotrophs	Panspermia	tidal pool scenario
coacervate	photoautotrophs	white smoker
co-catalysis	platypus	zircon
deep sea hydrothermal vent	primordial soup	

CHAPTER 20 WEB LINKS



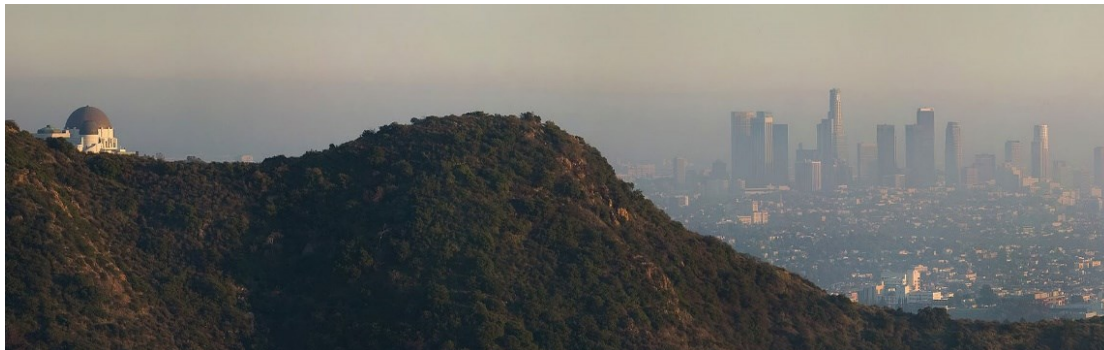


Epilogue

Reminder: If links have become inactive, google key words/terms for alternative sources.

We began this textbook with a testimony to the marvels of what science can accomplish. In Chapter 1 we defined the scientific enterprise as both intuitive and disciplined and learned that the goal of learning science and its methods is to prepare us to ask creative questions and to answer them. In Chapter 1 we learned that prebiotic chemical evolution set the stage for life's origins and that evolution is a key property of life itself, leading inexorably to the incredible diversity of life on which Darwin predicated his theory of natural selection. In Chapter 1 we also noted the natural role of extinction in favoring diversity... and how recent human activity accelerated the near extinction of the American bison and the final extinction of the passenger pigeon. The role of evolution in creating diversity is emphasized throughout the book. This is a moment to consider the threat of human activity to life on Earth that has only increased since the last known passenger pigeon died in 1914. Here are some more recent threats.

- **Smog Problems** in big cities (notably Los Angeles) date back to the 1940s. After much resistance to evidence that auto emissions were the cause, public acceptance came in the 1950s. Congress responded in 1963 by passing the first ^{E1}[Clean Air Act](#) and set some emission standards 2 years later. In 1975 the U.S. mandated catalytic converters on all new cars. California still has some of the worst air in the country. But ozone in Los Angeles has been at 40% of its mid-1970s level, despite twice the number of cars. Here is smog in L.A. in 2006:

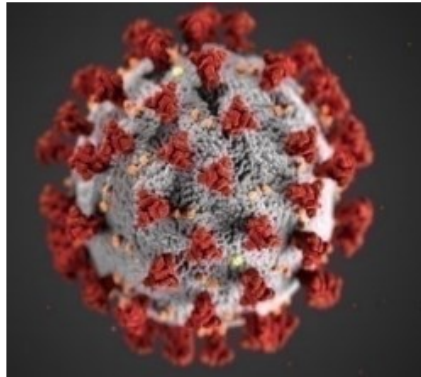


By Diliff-Own work; CC-BY-SA 3.0, ^{E2}[L.A. Smog in 2006](#)

And now, read about days of smog and ozone in June, 2020 at ^{E3}[Winning the Air Pollution Battle in L.A.?](#)

- **Plagues, epidemics, and pandemics** have been with us since ancient times. More recent ones include an 1889 influenza (flu) pandemic, thought to have started in Russia, which killed an estimated million people worldwide by 1890. World War I began 24 years later, ending in 1918 with 20 million (mostly men) dead. Then the flu pandemic of 1918 wasted no time killing an estimated 50,000,000+ people worldwide, proving once again, that nature always gets the upper hand. The 2009 swine flu (*H1N1pdm09*) pandemic killed an estimated 100,000-500,000+ people worldwide, coming after other recent gentle reminders that nature can still plague us (e.g., SARS in 2002-03, the 'bird flu in 2003-07). As if to

drive home this message, in 2020 we have *COVID-19*, a respiratory disease caused by *SARS-CoV-2*, a novel corona RNA virus (shown below).



COVID-19 Viral Particle: Public Domain; ^{E4}SARS-CoV-2 (CDC/Alissa Eckert, MS; Dan Higgins, MAMS) .

The red projections on the surface are composed of 'spike proteins' that enable the virus infect cells

COVID-19 surfaced in China in late 2019, spreading rapidly. By March 11, 2020, with >250,000 infections world-wide and >10,000 dead, the WHO (World Health Organization) labeled *COVID-19* a pandemic. With the *SARS-CoV-2* genome quickly sequenced and PCR-based and serological tests for the virus rapidly developed, several countries (e.g., China, Israel, Italy, South Korea...) seemed to have controlled the spread of *COVID-19* with the help of quarantines and shutdowns of schools, businesses and large entertainment and other events. *Pfizer*, *Moderna* and *Johnson & Johnson* became household names when these companies rapidly developed the first vaccines against *SARS-CoV-2*. The *Pfizer* vaccine first became available in the U.S. in December of 2020. The *Johnson & Johnson's* is a traditional vaccine, using a purified virus *spike protein* to stimulate the immune response. The *Pfizer* and *Moderna* vaccines are cDNA-based vaccines. The cDNAs encode the viral spike protein that, when injected, are taken up by cells that transcribe it, translate its spike protein mRNA and secrete it to elicit an immune response. You may be surprised to learn that Katalin Kariko's pursuit of the idea for mRNA vaccines long-predicate the *SARS-Cov-2* vaccine. Learn more at ^{E5}<https://www.youtube.com/watch?v=b3hWEC553sU>.

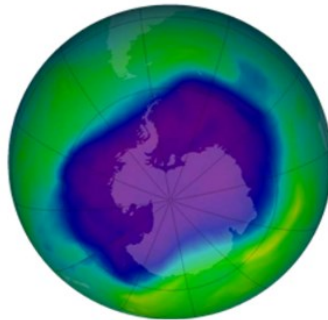
Our individual response to the pandemic has been to rally around our relatives, neighbors and even strangers to help them through tough times. We bring them food and medical supplies and keep up their spirits if they are in quarantine or physically isolated by their illness. But, despite the promise of testing and highly effective mRNA vaccines, the pandemic is still with us, assisted by the rapid spread a Greek alphabet-full of variants, against which the first generation of vaccines is less effective. The lethal ***Delta*** variant accounted for 99% of *COVID-19* infections between August and December 2021. Less deadly but highly infectious ***Omicron*** BA subvariants were next. As of this writing, ***Omicron BA.5*** accounted for 43% of infections worldwide and 54% of infections in the U.S. BA variants seem to account for more re-infections and infections of already vaccinated people and may also result in cases of "long *COVID*" a compendium of more or less varied post-infection symptoms. Check out the link at ^{E6}<https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html> for more details. Also as of this writing (July 2022), the world-wide *COVID-19* infection (case) count is >560,000,000 with >6,000,000 deaths. These figures include a massive spike in infections in early 2022, more than 3 times the

daily infection rate than any time since the pandemic began. Contributing to these figures are low vaccination rates (especially in developing nations), the spread of new variants, and premature re-openings and “business-as-usual” decisions virtually everywhere. Preventing the spread of COVID is now a race between vaccine development and distribution, and the emergence of new variants.

How soon will the pandemic end? While economies in developed countries recover from a subsiding pandemic, surging infections in parts of the developing world in Africa and South America reflect inequities of vaccine availability. These inequities even threaten even developed nations with infectious variants. Nevertheless, vaccines targeting more infectious COVID variants are being developed and in the developed world, the COVID-19 pandemic may already be on the way to becoming endemic and perhaps, like the flu, preventable with annual vaccinations. In addition, new antiviral medications in the pipeline (e.g., Paxlovid) even promise effective treatments for COVID-19. Unfortunately, these hopes come only after the damage already done by collaboration between the virus and us.

Still, after more than two years, we suffer a *plague* of ignorance and ill-advised behavior within the pandemic! There are about 35% of eligible Americans that are unvaccinated, many of whom distrust the science that produced the vaccines and believe the misinformation that spread virally on social media..., all of which led to the rise of the variants and surges of infection. The question is, do we now know how to cope with the next pandemic? This will depend on how we have learned from COVID-19, how we use available science, how our leaders respond, and how we all respond!

- **Polar ozone holes** were first discovered over the Antarctic in 1984; a smaller one now hovers over the Arctic.



The largest Ozone hole, over the Antarctic, slowly began to close after CFCs were banned in 1987. The blue and purple colors define this 10.6 million square mile zone of ozone depletion. Public Domain; ^{E7}[Ozone Over the Antarctic](#)

We worried as we watched the Antarctic ozone hole grow. Then we discovered that chlorofluorocarbons (CFCs) were the cause. CFCs are a propellant that used to be in spray cans filled with things like hair spray and paint. They are also a refrigerant that was part of every household refrigerator. Since the international 1987 Montreal Protocol banned CFCs, more countries agreed to the ban. While the expanse of the Antarctic ozone hole has *slowly* receded, it still reopens every year. The Arctic ozone hole opens periodically and varies in size. The one that opened in the winter of 2019-2020 was three times the size of Greenland, one of the largest (see the report in the figure above). Concern over upper atmospheric ozone depletion continues, and now, new information links it to climate change (for more information, see ^{E8}[Ozone Holes and Climate Change](#)).

- **Plastics** are produced from petrochemicals (i.e., oil), often by oil companies themselves. Most everyday disposable plastics are *not* biodegradable.



Public Domain: From ^{by} [Ocean Microplastics... not good](#)

Water bottles and milk jugs are typically recyclable. However, most other consumable plastics (straws, cups, plates, packing materials, particles in cosmetics and toothpaste, grocery bags, clamshells, tubs and other food wrappings...) are not, even when marked as recyclable with the 'chasing arrows.'



Public Domain, Wikipedia

They just cost too much to sort, clean up and recycle. Some ends up in landfills; most is shipped to the far east third world countries (e.g., Indonesia), intended for recycling. But it's too expensive to recycle even there. So, some is incinerated and much ends up dumped into the Pacific Ocean. This plastic debris now floats in *The Great Pacific Garbage Patch* (actually two patches), circulated by vortex current that distributes the plastic trash from the United States to Japan.

Either as whole pieces or as degraded microplastics, the trash accumulates in the bodies of marine animals feeding around the patches. Toxic polychlorinated biphenyls (PCBs) that leach from plastics made before 2001 (when PCBs were effectively banned worldwide) also spread through the marine food chain. The floating patches can also block light from penetrating the ocean water, killing off photoautotrophic marine phytoplankton and algae, the photoautotrophic primary producers reduce atmospheric CO₂ and feed marine life.

Microplastics and plastic debris has been found at 29,000 feet, atop Mt. Everest, and has settled more than 6000 feet to the bottom of the Pacific Ocean Marianna Trench. There, a new crustacean species was discovered in 2020, already having ingested a polyethylene-derived bit of plastic garbage. The new species was named *Eurythenes plasticus*!



CC-BY-SA; By Johanna Weston, Alan Jameson - Newcastle University, WWF Germany, Attribution, ^{E10}[The crustacean Eurythenes plasticus](#).

For more, check out the video at ^{E11}[Plastics, microplastics and Eurythenes plasticus video](#).

- **Climate change and global warming:** Are these an endpoint of human evolution? Recall our last chapter where we revisited in detail physical and chemical conditions that could support life chemistries and the origins of life itself, noting that evolution began the pathway to the generation of new species and the spectacular diversity on display to everyone in magazines like National Geographic, Smithsonian and the like.

By the 1960s, a few scientists saw a role of human activity in global warming. Now, 97% of scientists form the consensus that we caused a dramatic acceleration in climate change, raising global temperatures by 1° C. We are seeing stronger storms, more severe coastal flooding and inland droughts that most scientists view as the result. And, despite agreement on the human impact on climate change, international agreement has done little to slow increases in the greenhouse gas (90% CO₂ and CH₄) emissions that aggravate global warming. Tropical deforestation and the consumption of vehicle, residential and industrial fossil fuels continue with little restraint (check out the link at ^{E12}[EPA Overview of Greenhouse Gases](#) for some more details.

So, have we reached a temperature “tipping point” past which climate change is irreversible, ecological losses will drive extinctions and dramatically change the course of evolution, and above which, faith in solutions from science becomes a rush to cope? Is the recent spate of forest fires in California caused by climate change? Does an estimate of loss of a billion animals (not counting small critters like insects and worms...) in the southern Australia bush fires mean that we have reached or even surpassed that tipping point?



Firefighters move in to protect properties from an out-of-control bushfire in SW Sydney, Australia.

CC-BY-SA (From Helitak430, Dec. 2019, ^{E13}[2019 Australian Bush Fire](#))

...and Google for images of the 2022 deadly flooding in and around Sydney and elsewhere in Australia.

And should we also be concerned about the extensive and deadly flooding in Sydney (^{E14}<https://www.youtube.com/watch?v=xqldJfqMyWM>) and elsewhere in Australia?

With a measure of pessimism, young people are asking these questions and making demands of their elders. Click ^{E15}[Speaking Truth At the UN](#) to see Greta Thunberg's address to the United Nations 2019 Climate Action Summit. With a measure of optimism, I hope that the children will find the answers. As of this writing (2022) and a more recent climate action summit, there is more acceptance that climate change is upon us, and more international promises, but the criticism is that it's getting close enough to "too-late", and that the promises and good intentions may not go far enough to reverse climate change and its effects.

EPILOGUE WEB LINKS



Appendix I: List of Figures & Sources

Click or copy and paste the links below into your browser for access; some may no longer be active.

Chapter 1

Chapter Top-Art: Public Domain: left, R. Hooke, (1865) *Micrographia*; middle 19th century travelogue; right, Debbie Stinson photograph from <https://notesfromstillsong.blogspot.com/2012/02/eremitical-life.html>; and <https://www.pinterest.com/pin/138485757266449322/>

Fig. 1.1: Evolution of three domains showing a closer relationship between archaeobacteria and Eukaryotes. Illustration by G. Bergtrom

Fig. 1.2: Tobacco mosaic virus symptoms on a tobacco leaf.

USDA Forest Service, <http://www.forestryimages.org/browse/detail.cfm?imgnum=1402027> and https://en.wikipedia.org/wiki/Tobacco_mosaic_virus#/media/File:Tobacco_mosaic_virus_symptoms_tobacco.jpg

Fig. 1.3: Transmission electron micrographs of giant viruses, the AIDS (HIV) virus and the bacterium *E. coli*. *K. casanovai* is at least the same size and *M. horridgei* is twice the size of the bacterium. All the giant viruses, even the mimivirus, dwarf HIV, a typical eukaryotic virus. CC-BY, from Barthelmy, R.M., Faure, E. and Goto, <https://www.hilarispublisher.com/open-access/serendipitous-discovery-in-a-marine-invertebrate-phylum-chaetognatha-of-the-longest-giant-viruses-reported-to-date.pdf>

Fig. 1.4: Transmission and scanning electron micrographs of the gram-negative *E. coli* bacterium (left and middle), with its basic structure illustrated at the right.

- Left: from Peter Highton - Estate of Peter Highton a molecular biologist working at University of Edinburgh 1968-1990& Middle, CC-0 (Public Domain), https://en.wikipedia.org/wiki/Escherichia_coli
- Middle: CC 0, From photo by Eric Erbe, digital colorization by Christopher Pooley at USDA, EMU, ARS, <https://commons.wikimedia.org/w/index.php?curid=958857>
- Right: From Jeff Dahl - Own work, GFDL, <https://commons.wikimedia.org/w/index.php?curid=3647374>

Fig. 1.5: Localization of FtsZ proteins in Z rings in cross-sections in the middle of dividing *E. coli* cells. The immunofluorescence micrograph (left) is drawn at the right.

- CC SA. Fluorescence micrograph adapted from <https://commons.wikimedia.org/wiki/File:Zrings.png>
- Public domain, Illustrated cross section of a dividing bacterium showing location of FtsZ cortical ring protein https://commons.wikimedia.org/wiki/File:FtsZ_Filaments.svg
- Illustration by G. Bergtrom

Fig. 1.6: Electron micrographs of Carboxysomes in a cyanobacterium (left), carboxysomes after isolation. From https://en.wikipedia.org/wiki/Carboxysome#/media/File:Carboxysomes_EM; Tsai Y, Sawaya MR, Cannon GC, et al (June 2007). *Structural analysis of CsoS1A and the protein shell of the Halothiobacillus neapolitanus carboxysome*. PLoS Biol. 5 (6): e144

Fig. 1.7: Scanning electron micrograph of Halobacterium salinarium, a salt-loving bacterium. CC-BY; Adapted from: https://openi.nlm.nih.gov/imgs/512/185/3495301/PMC3495301_gbi0010-0424-f2.png

Fig. 1.8: Scanning electron micrograph of 'heat-loving' *Thermus aquaticus* bacteria. CC-0 (public domain) Adapted From: https://upload.wikimedia.org/wikipedia/commons/4/48/Thermus_aquaticus.JPG

Fig. 1.9: Illustration of the structural components of a typical animal cell. Source: Public Domain; From Mariana Ruiz, Image: Animal cell structure.svg <https://commons.wikimedia.org/w/index.php?curid=4266142>

Fig. 1.10: Illustration of the structural components of a typical plant cell. Source: Public Domain; From Mariana Ruiz, Image: Animal cell structure.svg, <https://commons.wikimedia.org/w/index.php?curid=4266142>

Fig. 1.11: **Fig. 1.11:** LEFT: Transmission electron micrograph of an insect cell nucleus showing the nuclear envelope (ne) and nucleolus (n); RIGHT: Drawing of a nucleus with chromatin (purple) and nuclear pores. Left; From Bergtrom et al., (1977) J. Ultrastr. Res. 60:395-406: Research by G. Bergtrom; Right; CC BY 3.0; Blausen.com staff (2014). "Medical gallery of Blausen Medical 2014". WikiJ. Medicine 1(12); Own work, <https://commons.wikimedia.org/w/index.php?curid=28223971>

Fig. 1.12: Simulated fluorescence micrograph showing the immunolocalization of antibodies against fibrillarin, coilin and ASF/SF2 protein to *nuclear bodies* (nucleolus, *Cajal Bodies* and nuclear speckles, respectively). Illustration by G. Bergtrom, adapted from a fluorescence micrograph in <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC138913/12>

Fig. 1.13: Drawing (LEFT) and fluorescence micrograph (RIGHT) of a cell in metaphase of mitosis: aligned chromosomes (chromatids) at the center of the cell (blue in the micrograph) are just about to be pulled apart by microtubules of the spindle apparatus (green) extending from the poles to the center of the cell. LEFT, Adapted from [Kelvinsonq; https://en.wikipedia.org/wiki/Metaphase](https://en.wikipedia.org/wiki/Metaphase); CC BY 3.0, unported license; RIGHT, Public Domain, from Afunguy-Transferred to Commons by Lije also using CommonsHelper; <https://commons.wikimedia.org/w/index.php?curid=5148470>

Fig. 1.14: To form a polysome, ribosomes (blue) assemble at the left on an mRNA molecule. As they move along the mRNA from left to right, they translate the message into a polypeptide (green), shown growing and emerging from ribosomes in the polysome. Illustration by G. Bergtrom.

Fig. 1.15: Transmission electron micrographs of cells from desiccated and fully hydrated *Selaginella lepidophylla* plants. Free ribosomes or ribosomal subunits in the desiccated cells (LEFT) appear to have organized to form polysomes in the hydrated plant cells (RIGHT). From Bergtrom et al. (1982) J. Ultrastr. Res 78:269-282. Research by G. Bergtrom

Fig. 1.16: An illustration of the differences between prokaryotic and eukaryotic ribosomes; Eukaryotic ribosomes and their subunits are larger and contain more proteins and larger ribosomal RNAs (rRNAs) than those of e.g., bacteria. The components were separated by sucrose density gradient centrifugation in which particles and macromolecules (like RNA) move through a sugar gradient at rates dependent on their mass (in effect, their size). Illustration by G. Bergtrom

Fig. 1.17: Transmission electron micrograph of a Golgi apparatus, or Golgi body. Adapted from OpenStax; CC BY 4.0; <https://cnx.org/contents/FPtK1znh@8.25:fEI3C8Ot@10/Preface;wikimedia.org/w/Golgi/index.php?curid=30131198>

Fig. 1.18: Illustration of 'packaged' protein traffic through a cell, from the RER to organelles (e.g., lysosomes) or to the plasma membrane for exocytosis (i.e., secretion). RER and Golgi vesicles are major sites for the modification (i.e., maturation) of packaged proteins. Illustration by G. Bergtrom

Fig. 1.19: Transmission electron micrograph of mitochondria (LEFT) and drawing of a mitochondrion (RIGHT).

- Electron micrograph from: Bergtrom et al. (1977) J. Ultrastr. Res. 60:395-405. Research by G. Bergtrom
- Public Domain; By Mariana Ruiz Villarreal LadyofHats, <https://commons.wikimedia.org/w/index.php?curid=6195050>

Fig. 1.20: Transmission electron micrograph of chloroplast that could have begun photosynthesizing (LEFT), and one that has photosynthesized long enough to accumulated starch granules (RIGHT). S, starch granule; T, thylakoids.

- Left: [GNU Free Documentation License](https://commons.wikimedia.org/w/index.php?curid=7153916); By and3k & caper437-Own work by uploaders, <https://commons.wikimedia.org/w/index.php?curid=7153916>
- Right: CC-BY; Transmission electron micrograph of chloroplast [T=thylakoids, S = starch] ; Adapted from PLOS one: <http://redoxbiologycenter.unl.edu/ee00b8d7-d7fc-43f8-bab9-5570b2dbd731.pdf>

Fig. 1.21: Transmission electron micrograph of a leucoplast, a chloroplast that has become filled with starch granules (S). From Bergtrom et al. J. Ultrastr. Res. 78:269-282. Research by G. Bergtrom

Fig. 1.22: Illustration of assembly, structure and polarity of microtubules, microfilaments and intermediate filaments alongside fluorescence micrographs made using fluorescent antibodies against isolated microtubule, microfilament and intermediate filament proteins (TOP, MIDDLE and BOTTOM, respectively).

- CC-BY 2.0; Adapted from: <http://bmcbiochem.biomedcentral.com/articles/10.1186/1471-2091-9-23> Open Access
- CC-BY-SA-2.5 and GNU.org; Adapted from https://en.wikipedia.org/wiki/media:MEF_microfilaments.jpg, By Y Tambe (Y tambe's file) via Wikimedia Commons
- CC-BY; Adapted from: www.plosone.org/article/info:doi/10.1371/journal.pone.0012115
- Illustrations by G. Bergtrom

Fig. 1.23: Transmission electron micrograph showing a flagellum (#1) emerging from a basal body (2). Number 3 is another basal body, this time in cross section. Public Domain, By Dartmouth Electron

Microscope Facility, Dartmouth College – <http://remf.dartmouth.edu/imagesindex.html> ,
<https://commons.wikimedia.org/w/index.php?curid=2787228>

Fig. 1.24: The characteristic “9+2” arrangement of microtubules seen in cross-sections of eukaryotic cilia and flagella is maintained the *axoneme*, a structure remaining after removing the plasma membrane from isolated cilia or flagella. Illustration by G. Bergtrom

Fig. 1.25: A cell fractionation involving disruption of cells and the isolation of sub-cellular components, including organelles, ribosomes and a soluble fraction, the cytosol. Separation is achieved by successive centrifugations at different speeds (centrifugal forces) that sediment subcellular structures on the basis of mass; the lowest mass structures require the highest G-forces (fastest spin), requiring an ultracentrifuge. Illustration by G. Bergtrom

Fig. 1.26: Organelles isolated by cell fractionation from eukaryotic cells.

- CC-BY-NC-ND-SA; Adapted from Hancock, 2009:
https://openi.nlm.nih.gov/detailedresult.php?img=PMC2762040_pone.0007560.g001&query=isolated+nuclei&it=xg&req=4&npos=23
- CC-BY-NC-ND-SA Adapted from Siekevitz and Palade , 1956:
https://openi.nlm.nih.gov/detailedresult.php?img=PMC2854378_JCB_201002147R_GS_Fig1&query=isolated+rough+endoplasmic+reticulum&it=xg&req=4&npos=5
- CC-BY-NC-ND-SA Adapted from Beznoussenko et al., 2015;
https://openi.nlm.nih.gov/detailedresult.php?img=PMC4394477_ijms-16-05299-g005b&query=isolated+golgi+vesicles&it=xg&req=4&npos=1
- CC-BY-NC-ND-SA; Adapted from Soubannier et al., 2012
https://openi.nlm.nih.gov/imgs/512/66/3530470/PMC3530470_pone.0052830.g001.png?keywords=abse nces
- CC-BY-SA: Adapted from S. Saleh, <https://commons.wikimedia.org/wiki/File:Chloroplast.jpeg>
- CC-BY-SA 4.0; Adapted from Boone et al. *Isolation of Plasma Membrane from HeLa Cell*,
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2107758/>
- CC-BY 2.0: From https://media.springernature.com/lw785/springer-static/image/art%3A10.1186%2F1471-2172-10-41/MediaObjects/12865_2009_Article_261_Fig3_HTML.jpg

Chapter 2

Chapter Top-Art: Images and videos are released under the [Pixabay License](#) and may be used freely for almost any purpose - even commercially. Attribution" mariolayaquerevalu MARIO OLAYA • Age 52 • TRUJILLO/PERU • Member since Oct. 5, 2017

Fig. 2.1: The shell (Bohr) and orbital models of atoms respectively emphasize the energy of electrons and space occupied by the electrons moving around the atomic nucleus. Illustration by G. Bergtrom

Fig. 2.2: The *Periodic Table of the Elements*, emphasizing those found in living things (blue, purple, green or brown). Illustration by G. Bergtrom

Fig. 2.3: *Excited* electrons gain their energy by absorbing light. Absorbance of one *photon* causes an electron to ‘jump’ to the next shell further from the atomic nucleus. Excited electrons have *potential energy* that is released as the electron returns to *ground state*, releasing heat or light (*fluorescence*). Illustration by G. Bergtrom

Fig. 2.4: Non-polar covalent bonds vs polar covalent bonds. Illustration by G. Bergtrom

Fig. 2.5: Electrons on the H atoms of water molecules are drawn close to the large, positively charged nucleus of the O atom. Thus, the H atoms ‘lose’ electrons and acquire a partial positive charge, while the oxygen atoms ‘gain’ those electrons and have a partial negative charge. The resulting polar covalent water molecules attract other water molecules. Illustration by G. Bergtrom

Fig. 2.6: Computer-generated space-filling (LEFT) and ‘ribbon’ models (RIGHT) of insulin structure. Source: CC-BY 2.5 <http://commons.wikimedia.org/wiki/File:InsulinMonomer.jpg>

Fig. 2.7: Ionic bonds in table salt (NaCl) crystals; Opposite charges on Na⁺ and Cl⁻ hold the ions together in a regular crystalline array. Illustration by G. Bergtrom

Fig. 2.8: Table salt (NaCl) dissolving in water. Water’s solvent properties result from its polar covalent structure, allowing interactions with Na⁺ and Cl⁻ in the crystal. Illustration by G. Bergtrom

Fig. 2.9: Charged groups on the macromolecule (e.g., protein) attract the partial charges on water molecules, hydrating the molecule. Illustration by G. Bergtrom

Fig. 2.10: Structural and geometric *isomers* of hydrocarbons create molecules with the same chemical formula but different shapes. Illustration by G. Bergtrom

Fig. 2.11: The two different generic molecules (*enantiomers*) are mirror images of one another. They form from alternate arrangements of the same molecular groups around a *chiral* carbon. Illustration by G. Bergtrom

Fig. 2.12: Two straight chain forms of glucose (*D*-glucose and *L*-glucose) are enantiomers (optical isomers), differing in the arrangement of the H atom and the OH group around C5. Illustration by G. Bergtrom

Fig. 2.13: A generic monomer is linked to a growing polymer by water removal (*dehydration synthesis*); Water addition across the linkage between monomers (*hydrolysis*) breaks the polymer back down to monomers. Illustration by G. Bergtrom

Fig. 2.14: *Dehydration synthesis* forms peptide linkage (circled); *hydrolysis* is the reverse reaction. Illustration by G. Bergtrom

Fig. 2.15: When straight chain glucose forms a cyclic molecule in solution in water, C1 becomes optically active (chiral), creating a racemic mixture of α (D)glucose and β (D)glucose enantiomers. Illustration by G. Bergtrom

Fig. 2.16: Polysaccharide sugar polymers form by dehydration synthesis: α (D)glucose monomers polymerize to form energy-storage molecules (e.g., starch, glycogen), while β -D-glucose monomers polymerize to form structural polysaccharides (starch, cellulose). Illustration by G. Bergtrom

Fig. 2.17: Two enantiomers (optical isomers) of Thalidomide (**R** and **S**) form in water. Public Domain; By Vaccinationist - (R)-(+)-Thalidomide & (S)-(-)-Thalidomide on PubChem, <https://commons.wikimedia.org/w/index.php?curid=50750574>

Chapter 3

Chapter Top-Art: Attribution: Released into the [public domain](#) by Attribution: Released into the Public Domain by Ysangkok: https://en.wikipedia.org/wiki/File:US_Beef_cuts.svg

Fig. 3.1: The four orders (levels) of protein structure. Primary, secondary and tertiary structures describe polypeptides

Fig. 3.2: Chemical characteristics of the 20 amino acids found in the proteins of cells. Illustrated by G. Bergtrom

Fig. 3.3: Partial polypeptide; the amino and carboxyl ends of a polypeptide define the polarity of the molecule, with positively charged amino and negatively charged carboxyl ends at physiological pH. Illustration by G. Bergtrom

Fig. 3.4: In the secondary structure of a polypeptide, more organized α -helical and β -pleated sheet structures are separated by less organized, random coil stretches of amino acids.

- Public Domain; From: Fvasconcellos - Own work. Created by uploader upon request by Opabinia regalis, <https://commons.wikimedia.org/w/index.php?curid=1587088>
- CC-0 (public domain); Adapted from: Jo, H et al. BMC Structural Biology (cc-by Open Access, <http://bmcsstructbiol.biomedcentral.com/articles/10.1186/s12900-014-0024-y>)

Fig. 3.5: Tertiary structure is created by non-covalent *hydrophobic* amino acid interactions as well as *H-bonding* in the interior of a polypeptide, leaving charged (hydrophilic) amino acid side chains to interact with water on the exterior of a typical "globular" protein. Stable covalent disulfide bonds between cysteine amino acids help stabilize tertiary structures. Illustration by G. Bergtrom

Fig. 3.6: Disulfide bridges form in oxidation reactions between SH (sulfhydryl) groups on cysteine amino acids ('residues') in a polypeptide. Illustration by G. Bergtrom

Fig. 3.7: Light micrograph of a sickled erythrocyte (red blood cell). CC BY 3.0; By OpenStax College - Anatomy & Physiology, Connexons Web site; <http://cnx.org/content/col11496/1.6/>

Fig. 3.8: Mutations in a normal gene for a prion protein (*PrP^c*) may produce some abnormally folded prion proteins (*PrP^{Sc}*). The misfolded *PrP^{Sc}* molecules interact with other (even normal) prions to mis-fold, precipitating *PrP^{Sc}* proteins into aggregated amyloid plaques. Adapted from: CC-BY-SA: M. R. Muntada (217) Prions: Special Proteins;

<https://allyouneedisbiology.wordpress.com/2017/11/12/prions-special-proteins/>

Fig. 3.9: Steps in the formation of β -amyloid plaques. Amyloid precursor proteins (APP) are embedded in cell membranes (upper left). Enzymes digest APP, releasing β -amyloid protein fragments (middle panel). Unregulated accumulation of β -amyloid protein results in the formation of extracellular amyloid plaques.

- Public Domain; By ADEAR: "Alzheimer's Disease Education and Referral Center, a service of the National Institute on Aging." - <http://www.nia.nih.gov/Alzheimers/Resources/HighRes.htm>, <https://commons.wikimedia.org/w/index.php?curid=4475153>
- Public Domain ; By National institute on aging - <http://www.nia.nih.gov/Alzheimers/Publications/UnravelingTheMystery/Part1/Hallmarks.htm> <https://commons.wikimedia.org/w/index.php?curid=3481613>
- Public Domain,; By National institute on aging - <http://www.nia.nih.gov/Alzheimers/Publications/UnravelingTheMystery/Part1/Hallmarks.htm> , <https://commons.wikimedia.org/w/index.php?curid=3481649>

Fig. 3.10: Non-fluorescent immunostaining of *tau* proteins of an Alzheimer's brain reveals *tau tangles*. CC-BY-SA 3.0; From Wikimedia Commons User Patho; [Neurofibrillary tangles in the Hippocampus of an old person with Alzheimer-related pathology, immunohistochemistry for tau protein.JPG](#)

Fig. 3.11: The formation of tau tangles occurs in diseased neurons that cannot maintain normal microtubule structure; in the absence of stabilizing tau molecules microtubules degenerate and tau proteins clump. Public Domain; Adapted from: https://commons.wikimedia.org/wiki/File:TANGLES_HIGH.jpg

Fig. 3.12: The vertebrate hemoglobin molecule, consisting of 4 globin subunits (two α and two β polypeptides). Each globin is associated with a heme group bound to iron. [GNU Free Documentation License](#); Adapted from: <https://en.wikipedia.org/wiki/Hemoglobin>

Fig. 3.13: The *Pleckstrin Homology* (PH) domains shown here are an example of common domains (maroon colored sequences) in two different proteins. Because they have a PH domain, both proteins can interact with cell signaling factors with roles in intercellular communication. CC BY-SA 3.0; By Fdardel - Own work, <https://commons.wikimedia.org/w/index.php?curid=17659815>

Fig. 3.14: Screenshot of NCBI portal to 3D protein structure database based on X-Ray crystallographs. At <https://www.ncbi.nlm.nih.gov/structure?db=Structure>

Fig. 3.15: Screenshot of NCBI protein structure database search leading to available macromolecular structures for human insulin.

Fig.3.16: Screenshot of the dynamic 3D X-Ray structure of a human insulin in the NCBI protein structure database.

Chapter 4

Chapter Top-Art: Public Domain; <https://pixabay.com/photos/space-universe-night-sky-sky-fog-11099/> (Wikiimages)

Fig. 4.1: The sum of all things contained in the universe (e.g., as mass and other, more familiar kinds of energy). In this simple "Venn" diagram, the universe is a closed system; nothing (mass, energy) can get into or exit the universe, consistent with the *First Law of Thermodynamics*. Illustration by G. Bergtrom

Fig. 4.2: Since even mass is a form of energy, the universe shown here consists of only two components, *entropy* and *free energy* (defined as available to do work), still consistent with the *First law of Thermodynamics*. Illustration by G. Bergtrom

Fig. 4.3: The reciprocal relationship between entropy and free energy graphed over time. If all activities in the universe proceed by increasing entropy, and if the universe consists *only* of free energy and entropy, then as entropy increases, free energy must be decreasing in the universe. Illustration by G. Bergtrom

Fig. 4.4: Basic design of a *bomb calorimeter*, with an inner chamber for conducting reactions, and an outer chamber containing water with a thermometer to measure temperature change (heat absorption or release, or ΔH) during the reaction. Illustration by G. Bergtrom

Chapter 5

Chapter Top-Art: Illustration by G, Bergtrom

Fig. 5.1: Illustration of the early *Lock-&-Key* mechanism for enzyme-substrate interaction, in which the substrate *key* fits an enzyme *lock*. Illustration by G. Bergtrom

Fig. 5.2: The *Induced Fit* mechanism of enzyme-substrate interaction. Initial interaction of substrate with enzyme based on mutual affinity causes an allosteric change in the enzyme. This results in the induced 'better' fit of substrate to enzyme, to be followed by catalysis. Illustration by G. Bergtrom

Fig. 5.3: Graph of change in free energy over time as chemicals A and B react. The high *Activation Energy* peak in the graph is the *free energy barrier* that A and B must overcome before much C and D can be made. This barrier is due mainly to thermal motion of the A and B molecules, with only rare encounters in relatively dilute solutions. Enzymes are even more efficient than inorganic catalysts in lowering the activation energy barrier to a reaction. Illustration by G. Bergtrom

Fig. 5.4: A generic biochemical pathway in which the final product of three enzymatic reactions accumulates. To prevent wasteful accumulation of the final product, the pathway is regulated, in this case, when some of the accumulating reaction product (E) binds to enzyme 1, blocking the reaction from A to form B. Illustration by G. Bergtrom

Fig. 5.5: *Allosteric effectors* are small metabolites whose amounts in a cell reflect the cell's metabolic status. They bind to enzymes at a regulatory site, causing an *allosteric* (conformational) change in the enzyme that affects the shape of the *Active Site*, either inhibiting catalysis (as suggested here) or stimulating it. Illustration by G. Bergtrom

Fig. 5.6: *Saturation kinetics* of an enzyme-catalyzed reaction. At high substrate concentrations, all the active sites of all enzymes are occupied. Under these conditions, the reaction occurs at its fastest rate. Illustration by G. Bergtrom

Fig. 5.7: Experimental protocol for determining the kinetics of an enzyme-catalyzed reaction. Illustration by G. Bergtrom

Fig. 5.8: This graph plots the rate of product (P) formation at 4 different substrate concentrations, while holding the enzyme concentration constant. Illustration by G. Bergtrom

Fig. 5.9: This graph shows the slopes of the initial reaction rates (red) taken from the curves in the previous graph (Fig. 5.8). Illustration by G. Bergtrom

Fig. 5.10: This graph plots the initial reaction rates (slopes, or v_0) for the reactions plotted in Fig. 5.9. The formula shown for this curve is that of a *rectangular hyperbola*. Illustration by G. Bergtrom

Chapter 6

Chapter Top-Art: From Evans Love - Own work, CC BY-SA 4.0,

https://en.wikipedia.org/wiki/File:Human_Metabolism_-_Pathways.jpg

Fig. 6.1: The *flow of free energy* through life, from visible light to chemical energy (e.g., photosynthesis of glucose) to the high-energy intermediates like ATP (fermentation or respiration) and finally to cellular work. Illustration by G. Bergtrom

Fig. 6.2: The energetics of ATP synthesis and hydrolysis. Illustration by G. Bergtrom

Fig. 6.3: Stoichiometry of glucose metabolism and ATP productions by glycolysis, the Krebs Cycle, electron transport and oxidative phosphorylation. Illustration by G. Bergtrom

Fig. 6.4: *Glycolysis*: Glucose (6-C) becomes 2 pyruvates (3-C); 2 ATPs are consumed in Stage 1 but 4 are made in Stage 2; The oxidation of glyceraldehyde-3-P marks the start of Stage 2. Illustration by G. Bergtrom

Fig. 6.5: Alternate Fates of Pyruvate: *Fermentation* (the anaerobic reduction to e.g., alcohol, lactate) or *Respiration* (aerobic oxidation of pyruvate using oxygen as a final electron acceptor, resulting in H_2O and CO_2 production). Illustration by G. Bergtrom

Fig. 6.6: In Reaction 1, phosphorylation of glucose to make glucose-6-P (G-6-P) consumes a molecule of ATP. If cellular energy needs are being met, G-6-P will be polymerized to make storage polysaccharides. The G-6-P can be retrieved by polysaccharide breakdown (hydrolysis) when the cells require nutrient energy; at that time, the G-6-P will resume glycolysis. Illustration by G. Bergtrom

Fig. 6.7: Free energy flow (exchange) for Reaction 1 of glycolysis. Illustration by G. Bergtrom

Fig. 6.8: Enzymatics of the Hexokinase reaction. This enzyme catalyzes a *biologically irreversible* reaction and is allosterically regulated. Illustration by G. Bergtrom

Fig. 6.9: Reaction 2 of glycolysis. G-6-P isomerase catalyzes the isomerization of G-6-P to F-6-P (i.e., the isomerization of glucose to fructose). The reaction is endergonic and reversible. Illustration by G. Bergtrom

Fig. 6.10: Reaction 3 of glycolysis. A kinase catalyzes phosphorylation of F-6-P to F1,6-di-P in a biologically irreversible reaction, consuming a molecule of ATP. Illustration by G. Bergtrom

Fig. 6.11: Reactions 4 and 5 of glycolysis. In reaction 4, F1,6-di-P is split into dihydroxyacetone phosphate (DHAP) and glyceraldehyde-3-P (G-3-P) by an aldolase enzyme. In reaction 5, DHAP is isomerized to G-3-P. Both reactions are endergonic, consuming free energy. These reactions mark the end of Stage 1 of glycolysis. Illustration by G. Bergtrom

Fig. 6.12: Reaction 6, a redox reaction catalyzed by G-3-P dehydrogenase, is freely reversible. G-3-P is oxidized to 1,3 diphosphoglyceric acid (1,3 diPG), transferring electrons as a hydride (H^-) ion to NAD^+ to make NADH. Illustration by G. Bergtrom

Fig. 6.13: Reaction 7, a reversible, *exergonic, coupled* reaction generates ATP using free energy released by the hydrolysis of one of the phosphates from 1,3diPG. The mechanism of ATP synthesis here is called *substrate-level phosphorylation*. Illustration by G. Bergtrom

Fig. 6.14: Reaction 8 is a reversible, *endergonic* reaction catalyzed by a mutase. Mutases catalyze transfer of a chemical group from one part of a molecule to another; here, 3-PG is converted to 2PG. Illustration by G. Bergtrom

Fig. 6.15: Reaction 9 is a reversible *endergonic* reaction. An *enolase* catalyzes the conversion of 2-PG to phosphoenol pyruvate (PEP). Illustration by G. Bergtrom

Fig. 6.16: Reaction 10 is catalyzed by a kinase that transfers the phosphate on PEP to ADP to make ATP and pyruvate. Pyruvate kinase catalyzes a coupled, highly exergonic and biologically irreversible reaction. Illustration by G. Bergtrom

Fig. 6.17: Free energy and ATP yields of complete glycolysis (a fermentation) and incomplete glycolysis (i.e., respiration). Percentages represent the (efficiency of ATP production. They are based on the ratios of the free energy captured as ATP to the free energy released by the different pathways for metabolizing glucose. From the data, incomplete glycolysis is a more efficient way to extract nutrient free energy. Illustration by G. Bergtrom

Fig. 6.18: Comparison of incomplete glycolysis to gluconeogenesis. The two pathways are essentially the reverse of one another, except for the *bypass enzymes* in gluconeogenesis (shown in green), required to get around biologically irreversible enzymes of glycolysis. Illustration by G. Bergtrom

Fig. 6.19: The *Cori cycle* reveals the relationship between glycolysis and gluconeogenesis. Lactic acid produced by complete glycolysis in active skeletal muscle goes to the liver where it could be converted to pyruvate, and then to glucose. CC-BY-SA 3.0; From

https://commons.wikimedia.org/wiki/File:Cori_Cycle.SVG and [GNU Free Documentation License](#)
User:Petaholmes, [user:PDH](#), [user:Eyal Bairev](#)

Fig. 6.20: Control of gluconeogenesis by hormones of the *hypothalamic-pituitary axis*. A need for glucose for energy stimulates hypothalamic hormones that, in turn, stimulates the release of adrenocorticotrophic hormone (ACTH) from the pituitary gland. ACTH then stimulates release of glucocorticoids from the adrenal glands. Glucocorticoid hormones stimulate many tissues to control gluconeogenic glucose production as well as the use alternate nutrient fuels by many cells in the body. Illustration by G. Bergtrom

Fig. 6.21: Entry of pyruvate into a mitochondrion, followed by its oxidation to acetyl-S-Coenzyme A (acetyl-S-CoA). Illustration by G. Bergtrom

Fig. 6.22: Pyruvate dehydrogenase catalyzes pyruvate oxidation to Ac-S-CoA, releasing a molecule of CO_2 and reducing NAD^+ to NADH. Illustration by G. Bergtrom

Fig. 6.23: Highlights of the Krebs Cycle. The first reaction is condensation of Acv-S-CoA and oxaloacetate (OAA) to form citric acid (citrate). Four of the reactions in the cycle are redox reactions that create reduced electron carriers (NADH, FADH_2), while one reaction is coupled to GTP synthesis by substrate level phosphorylation. Illustration by G. Bergtrom

Chapter 7

Chapter Top-Art: Illustration by G. Bergtrom

Fig. 7.1: Overview of electron transport and oxidative phosphorylation. NADH and FADH_2 oxidation feed electrons into electron transport, releasing free energy that powers proton (H^+ ions) pumps to force H^+ ions out of the mitochondrion. The resulting H^+ gradient fuels ATP synthesis as the protons flow into the mitochondrial matrix through a regulated *ATP synthase* in the cristal membrane (at the right in the drawing). Illustration by G. Bergtrom

Fig. 7.2: The flow of protons through the cristal membrane ATP synthase relieves the proton gradient, releasing free energy that fuels ATP synthesis in the mitochondrial matrix. Illustration by G. Bergtrom

Fig. 7.3: A bacterial electron transport system can pump protons and other cations (Na^+ , Ca^{++}) from the cytoplasm into the periplasmic space, at the left on the inner (cell) membrane. Relief of the proton gradient can power an ATP synthase or motility by fueling a spinning flagellum (at the right). In both cases, proton flow into the cell powers complex protein motors. Illustration by G. Bergtrom

Fig. 7.4: Summary of carbon flow through life. Organic carbon is oxidized to release free energy and make ATP. Sunlight provides free energy for photosynthesis to turn inorganic carbon (CO₂) into organic carbon (e.g., glucose). Illustration by G. Bergtrom

Fig. 7.5: The *light dependent* and *light independent* ('dark') reactions of photosynthesis. The light-dependent reactions 'split' water, releasing oxygen and protons. The light-independent reaction 'fix' CO₂ in organic molecules. Illustration by G. Bergtrom

Fig. 7.6: Experimental design to test the hypothesis that photosynthesis is supported by absorption of light by chlorophyll pigments. Illustration by G. Bergtrom

Fig. 7.7: Graph showing the *action spectrum* of photosynthesis, showing that photosynthesis is supported across a wide range of visible light wavelengths. Illustration by G. Bergtrom

Fig. 7.8: Graph plotting the absorbance spectrum of purified chlorophylls, showing two main peaks of absorbance. Illustration by G. Bergtrom

Fig. 7.9: Low (LEFT) and high (RIGHT) power transmission micrographs of typical chloroplasts, thylakoid membranes and grana (thylakoid stacks). From and3k and caper437 - Own work by uploaders, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=7153976>

Fig. 7.10: Graph superimposing the action spectrum of photosynthesis over the absorbance spectra of different plant pigments. The multiple absorbance peaks of different chloroplast pigments are consistent with the action spectrum. Illustration by G. Bergtrom

Fig. 7.11: In the "Z-Scheme" of photosynthesis, PSI and PSII pigments absorb light energy that excites electrons captured by electron acceptors. Electrons excited from PSI reduce NADPH, the starting point of the *light-independent* reactions. Electrons excited out of PSII come from splitting water (releasing oxygen). These electrons will flow down a photosynthetic electron transport chain to replace electrons lost from PSI. Along the way, free energy from those excited electrons fuels ATP synthesis. Illustration by G. Bergtrom

Fig. 7.12: *Cyclic Photophosphorylation* cycles electrons excited from PSI along part of the photosynthetic electron transport chain, bypassing PSII and returning the electrons to PSI. ATP made during Cyclic photophosphorylation is used for cellular work when there is no need or capacity to make more sugar. Illustration by G. Bergtrom

Fig. 7.13: Overview of the *Calvin Cycle* for carbon fixation in C₃ (most) plants. Three RuBP molecules *fix* 3 CO₂ molecules, catalyzed by ribulose biphosphate carboxylase –oxygenase (RUBISCO). The three 6C molecules produced split into six 3C carbohydrates, becoming six glyceraldehyde-3-P (GA3P in the illustration). Five of these go on to regenerate 3 RuBP molecules; the 6th G3P waits for the Calvin Cycle to repeat, producing another G3P. Two G3Ps from 2 turns of the cycle are substrates for glucose synthesis. Illustration by G. Bergtrom

Fig. 7.14: *Calvin Cycle Arithmetic*: Making one glucose and regenerating 5 RubP molecules. Illustration by G. Bergtrom

Fig. 7.15: Photorespiration is a C₃ plant strategy to survive a dry spell or a heat wave. Water-starved C₃ plants resort to *fixing* oxygen using RuBP. This reaction is catalyzed by RUBISCO, the same enzyme that fixes CO₂ to RuBP, but using its oxygenase activity. Illustration by G. Bergtrom

Fig. 7.16: Crassulacean Acid Metabolism in CAM plants. To survive in arid climates and minimize water loss, CAM plants open their stomata and fix CO₂ in oxaloacetate (OAA) that is converted to malic acid to be stored until daylight. Stomata close at daybreak to conserving water, while malic acid releases CO₂ to be *re-fixed* to make glucose. Illustration by G. Bergtrom

Fig. 7.17: *C₄ Photosynthesis* pathway reactions in tropical plants. C₄ plants separate their 'dark' reactions into 2 cell types. *Mesophyll* cells can keep stomata open part of the day, allowing CO₂ into the cells where it is fixed to PEP to make OAA and then malate. Unlike CAM plants, the malate is not stored, but transferred to *bundle sheath* cells where it can release CO₂ to be *re-fixed* by the Calvin Cycle. Illustration by G. Bergtrom

Chapter 8

Chapter Top-Art: Photos from left to right:

- Fair use, <https://en.wikipedia.org/w/index.php?curid=22197113>
- http://www.nobelprize.org/nobel_prizes/medicine/laureates/1958/
- Public Domain, <https://commons.wikimedia.org/w/index.php?curid=3677045>
- CC-BY <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0030341>
- Public Domain, <https://commons.wikimedia.org/w/index.php?curid=15615954>

- Public Domain, <https://commons.wikimedia.org/w/index.php?curid=15616085>
 - Fair use, <https://en.wikipedia.org/w/index.php?curid=24959067>
- Fig. 8.1:** The experiments of F. Griffith demonstrating the existence of a chemical that could transfer a genetic trait (i.e., virulence) between bacteria. Illustration by G. Bergtrom
- Fig. 8.2:** The experiments of O. Avery et al. demonstrating for the first time that the chemical stuff of genes, of inheritance is DNA. Illustration by G. Bergtrom
- Fig. 8.3:** Life cycle of a bacteriophage: The phage *coat* remains attached to a cell after infection, but the chemicals inside the phage enter the infected cell. Illustration by G. Bergtrom
- Fig. 8.4:** The experiments of Alfred Hershey and Martha Chase demonstrating that the chemical stuff of viral genes is DNA. Illustration by G. Bergtrom
- Fig. 8.5:** *Photo 51*, the X-Ray crystallograph of DNA taken by Rosalind Franklin revealed 3 crucial molecular dimensions that ultimately led to the double helical DNA model. *Fair Use* (Ref. Wikipedia); By Source (WP:NFCC#4), <https://en.wikipedia.org/w/index.php?curid=38068629>; https://en.wikipedia.org/wiki/Photo_51
- Fig. 8.6:** The 3 molecular dimensions from Franklin's *Photo 51* was best explained by proposing a DNA double helix (Left). Model building by Watson and Crick revealed the antiparallel structure of the double helix (Right).
- Public domain; From: https://commons.wikimedia.org/wiki/File:DNA_double_helix_vertikal.PNG
 - CC BY-SA 3.0; By Madprime (talk · contribs) - Own work The source code of this SVG is valid. This vector image was created with Inkscape., <https://commons.wikimedia.org/w/index.php?curid=1848174>
- Fig. 8.7:** Three possible modes of DNA replication could be imagined. Illustration by G. Bergtrom
- Fig. 8.8:** Meselson and Stahl predict experimental outcomes based on 3 possible modes of DNA replication. Illustrated by G. Bergtrom
- Fig. 8.9:** Light micrograph of a human female chromosome stained for banding patterns, showing the *centromere* constriction and defining *telomeres* (chromosome ends). Public Domain; Adapted from National Human Genome Research Institute; Cropped from Human male karyotype.jpg, <https://commons.wikimedia.org/w/index.php?curid=40960106>
- Fig. 8.10:** Spread of human mitotic cell chromosomes. CC BY-SA 3.0; By Steffen Dietzel - Own work, <https://commons.wikimedia.org/w/index.php?curid=1369763>
- Fig. 8.11:** Computer-colored human *karyotype*. Public Domain; Adapted From <https://commons.wikimedia.org/w/index.php?curid=7853183C> Courtesy: National Human Genome Research Institute - Found on: National Human Genome Research (USA); copied from wikipedia:en.
- Fig. 8.12:** Transmission micrograph of *euchromatin* and *heterochromatin* in the nucleus. From Bergtrom et al., (1977) J. Ultrastr. Res. 60:395-406; Research by G. Bergtrom
- Fig. 8.13:** Low salt fractionation of interphase nuclei yields 10nm nucleosome *beads on a string*.
- Upper; From Bergtrom et al., (1977) J. Ultrastr. Res. 60:395-406: Research by G. Bergtrom;
 - Lower left; CC-BY-SA 3.0; Adapted from: https://commons.wikimedia.org/wiki/File:Chromatin_nucleofilaments_%28detail%29.png
- Fig. 8.14:** High salt chromatin extraction from nuclei or high salt treatment of 10 nm filaments yields 30 nm *solenoid* structures, essentially coils of 10 nm filaments.
- Electron micrograph of nucleus, From Bergtrom et al., (1977) J. Ultrastr. Res. 60:395-406: Research by G. Bergtrom
 - CC-BY-SA 3.0; Adapted from: https://commons.wikimedia.org/wiki/File:Chromatin_nucleofilaments_%28detail%29.png
 - CC BY-SA 4.0; Alt: Adapted from: Richard Wheeler <https://en.wikipedia.org/w/index.php?curid=53563761>
- Fig. 8.15:** Five different levels (orders) of chromatin structure. CC-BY-SA 3.0; Adapted From <https://en.wikipedia.org/wiki/Chromatin>.
- Fig. 8.16:** Short times of DNase I digestion of 10 nm filaments leaves behind shortened *beads on a string*. A long time of DNase I digestion releases single nucleosomes associated with some DNA. Illustration by G. Bergtrom
- Fig. 8.17:** Electrophoresis of nucleosomal proteins from digested 10 nm filaments reveals five different *histones* with the stoichiometry shown. Gel illustration by G. Bergtrom; nucleosome illustration adapted from CC-BY-SA drawing on <https://commons.wikimedia.org/wiki/File:Chromatosom.png>

Fig. 8.18: Acid extraction of chromatin to remove histones and leave more basic non-histone proteins behind. See original micrograph at

<https://www.flickr.com/photos/185778834@N06/49368658231/> Illustration by G. Bergtrom.

Fig. 8.19: Bacterial *conjugation* (sex in bacteria): The *F plasmid* (Fertility plasmid) can transfer bacterial chromosomal DNA from the F^+ mating strain of *E. coli* to an opposite, F^- mating strain, leading to a measure of genetic diversity. Illustration by G. Bergtrom

Fig. 8.20: Possible results of conjugation in *E. coli*. The *F* plasmid transferred from an F^+ donor cell may or may not integrate into the F^- (recipient) cell. If integration occurs, the recipient cell becomes *Hfr*, with a high frequency of conjugation. Illustration by G. Bergtrom

Fig. 8.21: During conjugation with *Hfr donor* strain cells, portions of donor chromosomal DNA follow the excised plasmid across the conjugation bridge into the F^- recipient cell. Illustration by G. Bergtrom

Fig. 8.22: Map of the *E. coli* chromosome based on conjugation of F^- recipient cells with different strains of *Hfr* donor cells indicated as *Hfr* 1, 2, ... The different *Hfr* donors are cells in which the *Hfr* plasmid integrated at different place on a recipient cell chromosome. When these different *Hfr* strains transfer DNA to new F^- cells, they bring along different regions of the donor cell chromosome (and thus different genes) into their conjugated F^- cells. Illustration by G. Bergtrom

Fig. 8.23: The lysogenic life cycle of phage. Illustration by G. Bergtrom

Chapter 9

Chapter Top-Art: From CC BY-SA 2.0; By Ion Chibzii from Chisinau, Moldova. "Hands of mother"

<https://commons.wikimedia.org/w/index.php?curid=12799400>

Fig. 9.1: Illustration of J. Cairns' autoradiographs of DNA extracted from *E. coli* cells allowed to grow on 3H -thymidine for more than one generation of cells so that all cellular DNA would become radioactive. Silver tracks represented in the figure are exposures of bacterial chromosomal DNA. Illustration by G. Bergtrom

Fig. 9.2: An ordering of Cairns' autoradiograph images to suggest the progress of replication of the *E. coli* circular chromosome. Illustration by G. Bergtrom

Fig. 9.3: Image of an *E. coli* circular chromosome with two *replication forks* (RFs) unwinding and replicating DNA in both directions from an *origin of replication* (OR); the *E. coli* OR is a defined sequence. Illustration by G. Bergtrom

Fig. 9.4: Eukaryotic replication enlarges replicons that eventually merge

Fig. 9.5: *DNA polymerase* activity grows a DNA strand from the 5' to 3' direction. CC BY-SA 3.0;

Adapted from: I. Madprime, <https://commons.Wikimedia.org/w/index.php?curid=2527732>

Fig. 9.6: Detection of replication errors and correction by DNA polymerase proofreading. CC BY-SA 3.0; Adapted from: I. Madprime, <https://commons.Wikimedia.org/w/index.php?curid=2527732>

Fig. 9.7: Initiation proteins recognize and begin unwinding the double helix by bending *E. coli* DNA at the OR. Helicase then continues to unwind the DNA. SSBs (single stranded binding proteins) stabilize unwound DNA as DNA polymerase and other enzymes begin replication in both directions away from the OR. Illustration by G. Bergtrom

Fig. 9.8: 5'-to-3' replication creates the problem at the replication fork shown here. Illustration by G. Bergtrom

Fig. 9.9: Hypothesis proposing that at least one DNA strand at a replication fork (the *lagging strand*) is synthesized in pieces, each starting with an RNA primer that must be later correctly stitched together. Illustration by G. Bergtrom

Fig. 9.10: Growth curves for wild type and a mutant T4 bacteriophage phage that synthesized a slow-acting DNA ligase enzyme. This suggested that slower growth might be due to inefficient ligation of lagging strand DNA fragments made during replication. the Illustration by G. Bergtrom

Fig. 9.11: Steps in the synthesis of DNA against the lagging template strand. Illustration by G. Bergtrom

Fig. 9.12: A pair of replication forks (RFs), each assembling proteins and enzymes required for replication, including DNA primases. Illustration by G. Bergtrom

Fig. 9.13: The problem with linear chromosomes is that the lagging strand at the end (telomere) of a double helix can't be primed and thus cannot be replicated, causing chromosome shortening at each cell division. Illustration by G. Bergtrom

Fig. 9.14: The *ribonucleoprotein* enzyme *telomerase* resolves the dangers of chromosome shortening by using its RNA sequence to generate repeats on the telomeric end of lagging strand DNAs. Illustration by G. Bergtrom

Fig. 9.15: The *Processive Replication* hypothesis unites all components needed to replicate both strands of DNA in a *replisome*, so both strands are replicated in the same direction. Illustration by G. Bergtrom

Fig. 9.16: Unwinding a circular DNA molecule (or any double helix that is rigidly associated with chromosomal proteins wherever it is not replicating) causes the DNA to twist and coil on itself, a phenomenon called *supercoiling*. Cairns' saw such supercoils in *E. coli* chromosomes. Illustration by G. Bergtrom

Fig. 9.17: Spontaneous depurination results in hydrolytic removal of a guanine or adenine from a nucleotide, resulting in a nucleotide deletion at that site in the DNA during replication. Illustration by G. Bergtrom

Fig. 9.18: Exposure of DNA to UV light can cause adjacent thymine to dimerize, resulting in deletion of two nucleotides at that site in the DNA during replication. Illustration by G. Bergtrom

Fig. 9.19: -NH₂ removal (deamination) of a base in one DNA strand results in a base substitution during replication. Illustration by G. Bergtrom

Fig. 9.20: Base Excision Repair mechanisms can detect and fix e.g., a deamination prior to replication. Illustration by G. Bergtrom

Fig. 9.21: Nucleotide Excision Repair mechanisms can detect and fix e.g., a pyrimidine dimer prior to replication. Illustration by G. Bergtrom

Fig. 9.22: DNA Mismatch Repair can detect and fix post-replication errors (i.e., mismatched base pairing).

- CC-BY 4.0; Adapted from Kenji Fukui,
https://commons.wikimedia.org/wiki/File:DNA_mismatch_repair_Ecoli.png
- Illustration by G. Bergtrom

Fig. 9.23: Non-Homologous End-Joining detects and repairs double-stranded breaks in DNA, but can leave nucleotide deletions. Illustration by G. Bergtrom

Fig. 9.24: Single-Stranded Break Repair by *Homologous Recombination* can fix a break in one DNA strand when the break is detected at a replication fork. The fix leaves an accurate repair with no deletion or base substitution. Illustration by G. Bergtrom

Fig. 9.25: Double-Stranded Break Repair by *Homologous Recombination* can detect and fix a break in double stranded breaks in DNA. Illustration by G. Bergtrom

Chapter 10

Chapter Top-Art: Photo by G. Bergtrom

Fig. 10.1: Ribosomal RNAs and proteins in a bacterial ribosome. Illustration by G. Bergtrom

Fig. 10.2: Transfer RNA (tRNA) associated with a ribosome, held in place by codon/anticodon complementarity as well as tRNA-ribosomal forces. Illustration by G. Bergtrom

Fig. 10.3: *Polysomes* form along an mRNA. Multiple ribosomes can assemble at the 5' end of an mRNA and then sequentially translate multiple polypeptides. Illustration by G. Bergtrom

Fig. 10.4: Transcription - The basic Steps: *Initiation* is the recognition of a promoter sequence near the transcription start site by *RNA polymerase*. *Elongation* is the successive addition of nucleotides to a growing RNA strand. *Termination* occurs at the end of the gene, releasing the newly made RNA. Illustration by G. Bergtrom

Fig. 10.5: An *Operon* is a contiguous group of 2 or more genes that are transcribed as a single messenger RNA that will be translated into two or more polypeptides. Illustration by G. Bergtrom

Fig. 10.6: Eukaryotic mRNAs, rRNAs and tRNAs are transcribed by different RNA polymerases and are processed by different mechanisms to yield usable, mature RNAs. Illustration by G. Bergtrom

Fig. 10.7: Bacterial transcription requires initiation factors (i.e., σ -factors) to help the RNA polymerase find and bind to a gene promoter to begin transcription. Illustration by G. Bergtrom

Fig. 10.8: After transcription has been initiated, σ factors soon fall away as elongation continues. Illustration by G. Bergtrom

Fig. 10.9: *rho-dependent termination* of bacterial transcription involves recognition of sequences at the end of the gene and transcript. Illustration by G. Bergtrom

Fig. 10.10: *rho*-independent termination of bacterial transcription involves complementary bases near the end of the transcript that form a stem/loop, or *hairpin loop* structure that serves as a termination signal. Illustration by G. Bergtrom

Fig. 10.11: Comparison of transcription in prokaryotes and eukaryotes; the roles of three RNA polymerases in Eukaryotes. Illustration by G. Bergtrom

Fig. 10.12: Eukaryotic mRNA transcription initiation begins with recognition of a gene promoter (including the *TATA box*) by initiation factor TFIID. The sequential association of a series of initiation factors and finally, RNA polymerase II with the mRNA. RNA polymerase phosphorylation starts transcription. Illustration by G. Bergtrom

Fig. 10.13: 5S rRNA transcription in eukaryotes is initiated from an *internal promoter* by RNA polymerase III. After binding to the promoter, the polymerase re-positions itself near the transcription start-site. Illustration by G. Bergtrom

Fig. 10.14: Transcription and processing of eukaryotic mRNA. Many mRNAs in eukaryotes are encoded by split genes containing coding regions (exons) and non-coding regions (introns). These genes are transcribed as precursor (pre-) mRNAs that must be processed by splicing before they can be used in translation. Illustration by G. Bergtrom

Fig. 10.15: Splicing of a eukaryotic pre-mRNA involves association of the primary transcript (*pre-mRNA*) with *snRNPs*, small ribonuclear proteins that catalyze cleavage of the pre-RNA at 5' and 3' splice sites followed by ligation of the 3' to 5' exon ends. Illustration by G. Bergtrom

Fig. 10.16: Formation of a spliceosome by the binding of *snRNPs* to an mRNA leads to hydrolysis of splice sites and the formation of a lariat structure from the intron remnants. Illustration by G. Bergtrom

Fig. 10.17: Immunofluorescence staining with antibodies to coilin and SMN protein show that *Cajal bodies* and *Gems* aggregate when undifferentiated cells (panels A and C) are stimulated to differentiate (panels B and D). Since mutations in *coilin* and *SMN* proteins are associated with splicing defects, co-localization of Cajal bodies and Gems suggests their co-involvement of *snRNP* function. Creative Commons Deed, NC, Attribution 2.5 Adapted from: <https://www.ncbi.nlm.nih.gov/pubmed/19735367>

Fig. 10.18: *Exon shuffling*, transfer/copying of an exon from one gene to another, can occur by unequal recombination (cross-over) between different genes based on regions of sequence similarity within introns. Illustration by G. Bergtrom

Fig. 10.19: mRNA capping results from a 5'-to-5' condensation reaction linking the 5' end of an mRNA to a methylated guanine triphosphate (CH₃-GTP). Illustration by G. Bergtrom

Fig. 10.20: mRNA polyadenylation; this is the addition of multiple adenine nucleotides to the 3' end of an mRNA, catalyzed by the enzyme *poly(A) polymerase*. Other than histone mRNAs, most eukaryotic mRNAs are polyadenylated. Illustration by G. Bergtrom

Fig. 10.21: Posttranscriptional processing of the 45S rRNA encoding three of the four eukaryotic rRNAs by hydrolytic cleavage. Illustration by G. Bergtrom

Fig. 10.22: Early (1906) low power (left) and high power (right) light micrographs of stained amphibian (salamander) oocyte nucleolar DNA in amphibian oocytes showing the "chromatic fibers" that (resemble) a bottle brush, now called *lampbrush chromosomes* because they were also reminiscent of brushes used to clean the chimneys of old-fashioned oil lamp chimneys. We now know that the bristles are nascent strands of rRNA. For a more recent, high resolution transmission electron micrograph, see <http://www.cellimagelibrary.org/images/11043>. Public Domain; O. Hertwig, <Image:O.Hertwig1906Fig2-6.jpg> & <https://commons.wikimedia.org/wiki/File:O.Hertwig1906Fig5.jpg>

Fig. 10.23: Characteristic clustering of eukaryotic tRNA genes. Illustration by G. Bergtrom

Fig. 10.24: Structure of a Yeast tRNA. CC-BY-SA 3.0; Own work; Yikrazuul; <https://commons.wikimedia.org/w/index.php?curid=10126790>

Fig. 10.25: The assembly of ribosomal subunits and their export to the cytoplasm is coordinated with the synthesis and association with the rRNAs and their ribosomal proteins. Illustration by G. Bergtrom

Fig. 10.26: Role of the 5' guanosine CAP and poly-A binding proteins and a nuclear transport factor in the export of an mRNA to the cytoplasm. Illustration by G. Bergtrom

Chapter 11

Chapter Top-Art: Both images are in the Public Domain; other source information not found

Fig. 11.1: The Universal RNA Genetic Code Dictionary. CC-BY-SA; From: https://en.wikipedia.org/wiki/Genetic_code

Fig. 11.2: Colinearity of genes and proteins (polypeptides) in bacteria. Illustration by G. Bergtrom

Fig. 11.3: A *single-base overlapping genetic code* would get more genetic information in less DNA! Illustration by G. Bergtrom

Fig. 11.4: Nirenberg et al.'s fractionation of bacterial cell RNAs. Illustration by G. Bergtrom

Fig. 11.5: Nirenberg et al.'s reconstitution of fractionated bacterial cell RNAs. Illustration by G. Bergtrom

Fig. 11.6: Nirenberg and Leder's experiment that led to breaking the entire genetic code. Illustration by G. Bergtrom

Fig. 11.7: Cloverleaf (LEFT) and computer-generated 3D (RIGHT) structures of a phenylalanyl-tRNA. Sources:

- CC BY-SA 3.0; By Yikrazuul - Own work, <https://commons.wikimedia.org/w/index.php?curid=10126790>
- CC-BY 3.0; Adapted from Qinhua Zhou at English Wikipedia, via Wikimedia Commons; [tRNA-Phe yeast 1ehz - Transfer RNA - Wikipedia](#)

Fig. 11.8: A, P and E sites, mainly on the large ribosomal subunit, involved in mRNA translation. Illustration by G. Bergtrom

Fig. 11.9: Three steps of amino acid *activation* (tRNA acylation) require ATP hydrolysis. Illustration by G. Bergtrom

Fig. 11.10: Discovery of *met-tRNA* and *formyl-met-tRNA*, despite only one codon for methionine. Illustration by G. Bergtrom

Fig. 11.11: Bacterial Translation Initiation: association stabilized of the small ribosomal subunit with an mRNA requires two initiation factors (IF1 and IF3). The association is by complementary H-bonding between a region of the 16S rRNA and the Shine-Delgarno sequence near the 5' end of the mRNA. Illustration by G. Bergtrom

Fig. 11.12: Bacterial Translation Initiation: *GTP-bound IF2* enables binding of the first aminoacyl tRNA (the initiator *fmet-tRNA_{fmet}*) to the 30S ribosomal subunit/mRNA complex to create a bacterial *initiation complex*. Illustration by G. Bergtrom

Fig. 11.13: Bacterial Translation Initiation: addition of the large ribosomal subunit to the *initiation complex*, *hydrolysis of GTP* and dissociation of initiation factors completes assembly of the ribosome on an mRNA. Illustration by G. Bergtrom

Fig. 11.14: Bacterial Translation Elongation: EF2-GTP and EF3 facilitate entry of the second aminoacyl tRNA into the A site of the ribosome to begin elongation; GTP is hydrolyzed. Illustration by G. Bergtrom

Fig. 11.15: Bacterial Translation Elongation: The *peptidyl transferase ribozyme* (part of the ribosome) catalyzes peptide bond formation between the *fmet* on its tRNA_f and the second amino acid (aa2). The *fmet* is transferred from its tRNA_f as it condenses with an aa2 still linked to its tRNA_{aa2}. Illustration by G. Bergtrom

Fig. 11.16: Bacterial Translation Elongation: Translocase (on the ribosome) catalyzes movement of the ribosome along the mRNA, exposing the next codon in the A site with its attached peptidyl tRNA now in the peptidyl (P) site. Illustration by G. Bergtrom

Fig. 11.17: Bacterial Translation Termination: translocation stalls when the ribosome reaches a termination codon for which there is no corresponding aminoacyl tRNA. Then, GTP is hydrolyzed as a protein *release factor* enters A site, causing release of the new polypeptide and dissociation of the ribosome from the mRNA. Illustration by G. Bergtrom

Chapter 12

Chapter Top-Art: Illustration by G. Bergtrom

Fig. 12.1: Transcription of the *E. coli lac* operon produces a single mRNA encoding three polypeptides (nicknamed 'z', 'y' and 'a') which is translated into 3 enzymes involved in lactose metabolism. Illustration by G. Bergtrom

Fig. 12.2: The *lacZ*, *lacY* and *lacA* genes of the *lac* operon are all controlled by the single promoter to the left of the *lacZ* gene. A regulatory gene (*lacI*) with its own promoter lies further to the left of the *lac* operon. Illustration by G. Bergtrom

Fig. 12.3: *Negative regulation* of the *lac* operon: The *I* gene transcribes a *repressor* protein. When bound to the operator sequence (between the promoter and beginning of the Z gene), RNA polymerase is blocked and the operon cannot be transcribed. Transcription will require derepression (i.e., removal of the repressor protein). Illustration by G. Bergtrom

Fig. 12.4: *Negative regulation of the lac operon:* When present, lactose enters cells and is converted to allolactose. Accumulating allolactose binds to the *lac* repressor protein. Illustration by G. Bergtrom

Fig. 12.5: *Negative regulation of the lac operon:* Allolactose binding causes allosteric change in the *lac* repressor that then dissociates from the operator sequence, allowing RNA polymerase to transcribe the operon. Illustration by G. Bergtrom

Fig. 12.6: *Positive regulation of the lac operon:* If lactose is present but glucose is unavailable, the operon is *derepressed AND induced*. Increased cAMP binds the *Catabolite Activator Protein (CAP)*, which binds to the *lac* operator, inducing *lac* gene transcription. The *lac* operon is maximally transcribed under these conditions. Illustration by G. Bergtrom

Fig. 12.7: *Positive regulation of the lac operon:* cAMP-CAP-bound operator bends the double helix. This loosens the H- bonds between the bases, making transcription factor access to template strands easier. Illustration by G. Bergtrom

Fig. 12.8: *lac operon regulation by inducer exclusion:* High glucose levels accelerate glycolysis, depleting cellular phosphate. Low phosphate leads to the dephosphorylation of EIIAGlc, among other proteins. Dephosphorylated EIIAGlc is an inhibitor of *lactose permease* (i.e., the *lacA* protein). Without active permease, lactose can't get into cells, allolactose cannot be made, and the *lac* gene cannot be transcribed. Illustration by G. Bergtrom

Fig. 12.9: Computer generated structure of the tetrameric *lac* repressor bound to DNA via *helix-turn-helix motifs*. CC-BY-SA; From: http://en.wikipedia.org/wiki/Lac_repressor

Fig. 12.10: Transcription of the 5-gene tryptophan (*trp*) operon: The 5 resulting enzymes are in the pathway for tryptophan synthesis. A tryptophan (*trp*) repressor gene to the left of the operator is always transcribed. Illustration by G. Bergtrom

Fig. 12.11: When there is sufficient tryptophan in the cells, excess tryptophan binds to and changes the conformation of the *trp* repressor protein, which then recognizes and binds the *trp* operon operator, blocking RNA polymerase and repressing operon transcription. Illustration by G. Bergtrom

Fig. 12.12: Replacement of an egg nucleus with that from an adult cell enabled cloning of embryos and eventually complete organisms, proving that adult cells contain all of the genes required to program development of, and to make the organism. Illustration by G. Bergtrom

Fig. 12.13: Potentials for gene regulation in eukaryotes: Control of transcript and/or protein abundance, as well as protein activity can regulate cellular metabolism. CC-BY-SA 3.0; Adapted From <https://en.wikipedia.org/wiki/Chromatin>; Illustration by G. Bergtrom

Fig. 12.14: Cis-acting sequence elements associated with a gene function as gene regulators when bound to regulatory proteins (trans-acting factors). Regulatory sequences may be near the gene promoter (proximal) or far from it (distal). Enhancers are examples of distal regulatory sequences. Illustration by G. Bergtrom

Fig. 12.15: Assembly of a eukaryotic transcription complex on a regulated gene. Distal regulatory protein/DNA interactions can cause DNA bending, recruiting transcription initiation factors and proximal regulatory factors to the transcription complex. CC BY 3.0; Adapted from Kelvinsong - Own work, <https://commons.wikimedia.org/w/index.php?curid=23272278> and https://en.wikipedia.org/wiki/Transcription_factor#/media/File:Transcription_Factors.svg

Fig. 12.16: Three DNA binding motifs commonly found in trans-acting (i.e. regulatory) protein factors.

- CC BY-SA 4.0; By Latacca - Own work, <https://commons.wikimedia.org/w/index.php?curid=37140265>
- CC BY-SA 3.0; By Zephyris at the English language Wikipedia, <https://commons.wikimedia.org/w/index.php?curid=2426895>
- Public Domain; From: Boghog at en.wikipedia - Own work Public Domain, <https://commons.wikimedia.org/w/index.php?curid=15966067>

Fig. 12.17: Steroid hormones regulate genes. Once in a target cell, the hormone binds a receptor either in the nucleus, or as in this example, in the cytoplasm. An active cytoplasmic hormone-receptor complex forms and translocates to the nucleus. Inside the nucleus, the steroid hormone receptor complex becomes transcription factor, binding to regulatory DNA sequences to turn genes on or off or simply modulate their transcription. Illustrated by G. Bergtrom

Fig. 12.18: Large or polar signal molecules bind to membrane receptors, resulting in *signal transduction*. Here, a *second messenger* molecule (cAMP) forms in the cytoplasm to deliver the hormonal message. A *phosphorylation cascade* leads to responses including gene regulation and/or control of existing protein activity. Illustration by G. Bergtrom

Fig. 12.19: *Enzyme-linked receptors* transduce hormonal signals directly, activating receptor kinases on the cytoplasmic surface of the cell, initiating phosphorylation cascades leading to any of several responses. Illustration by G. Bergtrom

Fig. 12.20: Chromatin Organization: different levels of chromatin structure result from differential association of DNA with chromosomal proteins. CC-BY-SA 3.0; Adapted From <https://en.wikipedia.org/wiki/Chromatin>

Fig. 12.21: Low or high salt nuclear chromatin extraction yields 10 nm filaments or 30 nm *solenoids* respectively.,

- From Bergtrom et al., (1977) J. Ultrastr. Res. 60:395-406: Research by G. Bergtrom; Research by G. Bergtrom
- Lower left; Public Domain; from Wikimedia Commons, Electron micrograph of decondensed chromatin from chicken erythrocytes; https://commons.wikimedia.org/wiki/File:Chromatin_nucleofilaments.png
- Lower right;; Adapted from OpenStax Biology 2nd Edition, licensed under a [Creative Commons Attribution 4.0 International License](#),

Fig. 12.22: Chemical modification of histones can *open* or *close* chromatin to transcription. CC BY-SA 1.0; By Annabelle L. Rodd, Katherine Ververis, and Tom C. Karagiannis. <http://www.hindawi.com/journals/lymph/2012/290685/>, <https://en.wikipedia.org/w/index.php?curid=42441420>

Fig. 12.23: Extracellular pressure (arrowheads) on a CED domain or activating spring-like arms by stretching the cell membrane (e.g., from inside the cell (arrow) opens a piezo-linked ion channel. Illustration by G. Bergtrom

Fig. 12.24: Summary of health, birth, death and demographic records by L/. O. Bygren for the town of Överkalix

Chapter 13

Chapter Top-Art: Illustration by G. Bergtrom, and recycling bin photos courtesy of Pixabay.com; Public Domain, <https://pixabay.com/vectors/recycle-bin-container-recycling-24543/> and <https://pixabay.com/vectors/recycling-container-bin-boxes-41078/>

Fig. 13.1: Bacterial guanine riboswitch: The 'switch' is an mRNA encoding an enzyme in the guanine synthesis pathway. Excess guanine binds and distorts the stem-loop structure of the mRNA, causing termination of further gene transcription. Illustration by G. Bergtrom

Fig. 13.2: The array of CRISPR/Cas9 genes, spacers and other components in the genome of *Streptococcus pyogenes*. Illustration by G. Bergtrom

Fig. 13.3: Phage infection triggers formation of *CRISPR/Cas9* array. Illustration by G. Bergtrom

Fig. 13.4: Phage spacers in a CRISPR/Cas9 array (derived from an earlier phage infection) protects against phage re-infection. Illustration by G. Bergtrom

Fig. 13.5: Anatomic illustration of the roundworm *Caenorhabditis elegans*. CC-BY 2.5 J.D. McGhee (2007) *The C. elegans intestine*, *Worm Book* -The Online Review of *C. elegans* Biology; http://www.wormbook.org/chapters/www_intestine/intestine.html

Fig. 13.6: Mechanism of action of siRNA (small interfering RNA) defense against foreign DNA. Illustration by G. Bergtrom

Fig. 13.7: Mechanism of action of miRNA (micro RNA) in degrading unwanted (e.g., old or damaged) RNA. Illustration by G. Bergtrom

Fig. 13.8: Synthesis of piRNA (thick blue arrows) and two pathways for blocking transposition in *D. melanogaster* testis (black arrows 1 and 2). Illustration by G. Bergtrom

Fig. 13.9: Review of the steps in translation initiation. Illustration by G. Bergtrom

Fig. 13.10: Regions and specific sequences known to be involved in regulating translation. Illustration by G. Bergtrom

Fig. 13.11: Specific inhibition of ferritin polypeptide synthesis by iron regulatory proteins. Illustration by G. Bergtrom

Fig. 13.12: Regulation of globin mRNA Translation initiation by hemin, a heme precursor. Illustration by G. Bergtrom

Fig. 13.13: Role of GCN2 kinase in regulating translation by phosphorylating initiation factor eIF2B, preventing formation of an initiation complex. Illustration by G. Bergtrom

Fig. 13.14: Multiple short *open reading frames* (uORFs) found in the 5' untranslated region (5'UTR) of GCN4 mRNA play a role in regulating its translation. Illustration by G. Bergtrom

Fig. 13.15: GCN4 translation is reduced when cellular amino acid levels are normal. Illustration by G. Bergtrom

Fig. 13.16: GCN4 translation is highest when cellular amino acid levels are reduced (amino acid starvation). GCN4 then regulates the expression of many genes, including those encoding enzymes needed for amino acid synthesis. Illustration by G. Bergtrom

Fig. 13.17: Mechanism of ubiquitin and proteasome action in the removal/destruction of old and damaged proteins in cells, and the recycling of their amino acids. Illustration by G. Bergtrom

Chapter 14

Chapter Top-Art: Public Domain

Fig. 14.1: Renaturation kinetics protocol: Double-stranded (ds)DNA is mechanically sheared to ~10Kb fragments, heated to denature the DNA, then cooled to let single strands find their complements and renature. Illustration by G. Bergtrom

Fig. 14.2: Plot of rat dsDNA formed over time during renaturation of denatured DNA. Illustration by G. Bergtrom

Fig. 14.3: The 3 'phases' of the curve in Fig. 14.2 highlighted to identify the 3 fractions of repeated and almost unique DNA sequences in the rat genome. Illustration by G. Bergtrom

Fig. 14.4: Plot of *E. coli* dsDNA formed over time during renaturation of denatured DNA. Illustration by G. Bergtrom

Fig. 14.5: DNA complexity revealed by plotting rat and *E. coli* DNA renaturation kinetics as the percent of re-associated dsDNA over CoT (Concentration of re-associated dsDNA X Time). Illustration by G. Bergtrom

Fig. 14.6: Experimental demonstration of *satellite DNA* in *low CoT* renaturation kinetic fractions. Illustration by G. Bergtrom

Fig. 14.7: Life-style of maize.

- Left, CC-BY-SA 3.0: From https://commons.wikimedia.org/wiki/File:Corntassel_7095.jpg ; By Spedona (en:Image:Corntassel_7095.jpg) via [Wikimedia Commons](#)
- Middle, CC-BY-SA 3.0; From: https://commons.wikimedia.org/wiki/File:Cornsilk_7091.jpg ; By The original uploader was Pollinator at English Wikipedia [GFDL (<http://www.gnu.org/copyleft/fdl.html>)] via [Wikimedia Commons](#)
- Right, CC-BY 3.0; From: Kiesselbach, T. A. https://commons.wikimedia.org/wiki/File:Double_Fertilization.jpg ,By The original uploader at English Wikipedia (Transferred from en.wikipedia to Commons, via [Wikimedia Commons](#)

Fig. 14.8: Mosaic corn cobs. [GNU Free Documentation License](https://en.wikipedia.org/wiki/Maize), V. 1.2 <https://en.wikipedia.org/wiki/Maize>

Fig. 14.9: Experimental cross of females with CCBzbz triploid genotype with males having a C'C'BzBzDsDs genotype with progeny genotypes and phenotypes based on Mendelian assumptions. Illustration by G. Bergtrom

Fig. 14.10: Triploid genotypes expected the cross in Fig. 14.9. Illustration by G. Bergtrom

Fig. 14.11: McClintock's interpretation of the results of the triploid cross shown in Fig. 14.9. Illustration by G. Bergtrom

Fig. 14.12: Structure of a bacterial *IS element*. Illustration by G. Bergtrom

Fig. 14.13: Structure of a bacterial *Tn element*. Illustration by G. Bergtrom

Fig. 14.14: Structure of a complex transposon (*Mu phage*). Illustration by G. Bergtrom

Fig. 14.15: Life cycle options for bacteriophage. Illustration by G. Bergtrom

Fig. 14.16: The transposon option for *Mu Phage*. Illustration by G. Bergtrom

Fig. 14.17: Structure of a eukaryotic *Class II DNA transposon*. Illustration by G. Bergtrom

Fig. 14.18: *Cut-&-paste transposition* of a DNA transposon. Illustration by G. Bergtrom

Fig. 14.19: *Replicative transposition* of a DNA transposon. Illustration by G. Bergtrom

Fig. 14.20: Structure of yeast *Ty*, an RNA *LTR retrotransposon*. Illustration by G. Bergtrom

Fig. 14.21: Structure of a *LINE*, a *non-LTR retrotransposon*. Illustration by G. Bergtrom

Fig. 14.22: Structure of the *Alu SINE*, a *non-LTR retrotransposon* or **retroposon**. Illustration by G. Bergtrom

Fig. 14.23: *Extrachromosomally-primed transposition* of a LINE. Illustration by G. Bergtrom

Fig. 14.24: *Target-primed transposition* of a LINE. Illustration by G. Bergtrom

Fig. 14.25: Alignment of *consensus* amino acid sequences of bacterial IS transposases with Mu phage and Tc1 mariner transposases reveals conservation of **D**, **D** and **E** amino acids (upper case) at key positions in the sequence. Other amino acids are shared between some but not all of the sequences (lower case). Slashes are variable gaps in the alignments. Adapted from

https://www.researchgate.net/publication/11063189_Presence_of_a_Characteristic_D-D-E_Motif_in_IS1_Transposase

Fig. 14.26: Comparison amino acid sequences of the *COPIA* retrotransposon and a retroviral (HIV) integrase with typical transposase sequences. The alignments reveal conservation of the **D**, **D** and **E** amino acids in the *DDE* domain of the enzymes. Other amino acids are shared between some but not all of the sequences (lower case). Slashes are variable gaps in the alignments. Adapted from

https://www.researchgate.net/publication/11063189_Presence_of_a_Characteristic_D-D-E_Motif_in_IS1_Transposase

Fig. 14.27: Retroviral and retrotransposon reverse transcriptases share a common evolutionary ancestor. Tree drawn by G. Bergtrom, based on data from [Nakamura et al. 1997](#) and [Eickbush 1994](#)

Fig. 14.28: Steps of paired DNA transposon-mediated exon shuffling. Illustration by G. Bergtrom

Chapter 15

Chapter Top-Art: Public Domain NASA .gov image;

<https://www.nasa.gov/sites/default/files/thumbnails/image/hand.jpg>

Fig. 15.1: Retroviral infection requires *reverse transcriptase* to make a copy (cDNA) the viral genome, which will then replicate and reproduce new viruses. Illustration by G. Bergtrom

Fig. 15.2: Most eukaryotic mRNAs have *poly(A)* tails; Poly(A) RNA is isolable by *oligo d(T)* chromatography. Illustration by G. Bergtrom

Fig. 15.3: Reverse transcriptase supplied with deoxynucleotides, mRNAs and an oligo d(T) primer will catalyze synthesis of DNA copies (cDNAs) of the mRNAs. Illustration by G. Bergtrom

Fig. 15.4: The first cDNA strand often forms a loop at its 3' end that can serve as a primer to synthesize a (ds)cDNA. Reverse transcriptase is also a DNA polymerase and can catalyze 2nd strand synthesis. Illustration by G. Bergtrom

Fig. 15.5: Bacterial restriction endonucleases (REs) recognize and hydrolyze 'foreign' DNA (e.g., phage DNA), blocking infection; most REs cut DNA at specific short DNA sequences; often leaving staggered ends. Illustration by G. Bergtrom

Fig. 15.6: Preparing cDNAs for insertion into plasmid vectors prior to cloning cDNAs sequences. Illustration by G. Bergtrom

Fig. 15.7: Making recombinant plasmids containing cDNA inserts. Illustration by G. Bergtrom

Fig. 15.8: Making the cDNA library. Illustration by G. Bergtrom

Fig. 15.9: Only cells containing plasmids will grow on agar containing *ampicillin* since the plasmids contain an ampicillin resistance gene. Illustration by G. Bergtrom

Fig. 15.10: *Replica plating* creates a *filter replica* of colonies from one agar plate to be grown on another plate, e.g., containing 'selective' media containing for example an antibiotic that screens for recombinant plasmids. Illustration by G. Bergtrom

Fig. 15.11: *Replica filters* are lysed in situ (in place), leaving partially denatured DNA (including recombinant plasmid DNA) from the colonies where the cells used to be. Filters can be probed for a sequence of interest. Illustration by G. Bergtrom

Fig. 15.12: A denatured, sequence-specific probe can be used to probe lysed replica filters for a sequence of interest, here detected as a radioactive colony (see Fig. 15.12 and text for details). Illustration by G. Bergtrom

Fig. 15.13: X-ray exposure of a replica filter incubated earlier with radioactive probe. The developed *autoradiograph* shows a dark spot where probe hybridized to colony DNA. Illustration by G. Bergtrom

Fig. 15.14: Chemical structure of a di-deoxynucleotide. Illustration by G. Bergtrom

Fig. 15.15: The four tubes in a manual DNA sequencing protocol contain the same ingredients, except that each contains a different di-deoxynucleotide. Illustration by G. Bergtrom

Fig. 15.16: Cartoon of results of the four different manual sequencing reactions. The dideoxy-terminated DNA fragments are electrophoresed to read the sequence of a DNA (see Fig 15.16). Illustration by G. Bergtrom

Fig. 15.17: Drawing of a manual dideoxy sequencing autoradiograph with a readable sequence. Illustration by G. Bergtrom

Fig. 15.18: Automated DNA sequence chromatograph readout. CC BY 3.0; Adapted from Intechopen book (Book chapters are distributed under the Creative Commons Attribution 3.0)

<https://www.intechopen.com/books/nucleic-acids-from-basic-aspects-to-laboratory-tools/nucleic-acid-isolation-and-downstream-applications>

Fig. 15.19: Overview of shotgun sequencing of DNA. CC-0 (Public domain, By InfoCan, from Wikimedia Commons); https://commons.wikimedia.org/wiki/File:DNA_Sequencing_gDNA_libraries_tr.jpg

Fig. 15.20: Basics of Nanopore DNA Sequencing, one of several 'next gen' sequencing technologies. Illustration by G. Bergtrom

Fig. 15.21: Overview of Southern blotting, a technique that transfers electrophoresed DNA to a filter for hybridization with a probe. Illustration by G. Bergtrom

Fig. 15.22: Genomic DNA restriction enzyme digest is mixed with phage vector with compatible 'sticky' ends and ligated to make recombinant DNAs. Illustration by G. Bergtrom

Fig. 15.23: Mixing recombinant phage DNA with phage coat proteins creates infectious recombinant phage particles. Illustration by G. Bergtrom

Fig. 15.24: Recombinant phage infection, the replication of phage DNA and production of new phage leads to host cell lysis, allowing the collection of new recombinant phage. Illustration by G. Bergtrom

Fig. 15.25: Phage plaques formed on a bacterial lawn; each plaque is a genomic clone. Illustration by G. Bergtrom

Fig. 15.26: Replica filters of plaques probed with a sequence of interest (in this case a globin DNA sequence) locates autoradiographic spots where probe hybridized to recombinant phage. Illustration by G. Bergtrom

Fig. 15.27: Progress of a PCR reaction: at the end of the second PCR cycle, the strands of DNA that will be geometrically amplified in succeeding PCR cycles have been synthesized. Illustration by G. Bergtrom

Fig. 15.28: Agarose gel electrophoresis of PCR amplified DNAs, stained with ethidium bromide to be detectable under fluorescent light: PCR products of the amplification of a globin gene clone from the insect *Chironomus thummi*. The sizes (lengths in base pairs) of the 4 PCR products in lanes 2-5 are close to those of the bands in the standard DNA size ladder in lane 1. Adapted from Kao et al. (1994), J. Mol. Evol. 38: 241-249. Research by G. Bergtrom

Fig. 15.29: Electrophoretic DNA fingerprint. Public domain; By James Tourtellotte, photo editor of CBP Today[1],

http://www.cbp.gov/xp/cgov/newsroom/multimedia/photo_gallery/afc/laboratories/13_5flab_5fhiresa.xml (file Cbp13_5flab_5fhires.jpg, part of America's Frontline Photography / CBP Laboratories Photography), <https://commons.wikimedia.org/w/index.php?curid=2875876>

Fig. 15.30: Simulated glass slide microarray showing multiple color fluorescent spots, indicating a hunt for more than one DNA sequence at the same time. Illustration by G. Bergtrom

Chapter 16

Chapter Top-Art: Public Domain; <https://pixabay.com/photos/lion-mosaic-art-museum-berlin-510159/>

Fig. 16.1: Structural and Space-filling models showing components of a phospholipid. CC-BY; Adapted from OpenStax Biology 2nd Edition, licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Fig. 16.2: Examples of common membrane phospholipids. Public Domain CC-0 1.0: Adapted from <https://commons.wikimedia.org/w/index.php?curid=63827468>

Fig. 16.3: Phospholipid bilayer membrane. Public Domain; Adapted from Mariana Ruiz Villarreal, LadyofHats - Own work, <https://commons.wikimedia.org/w/index.php?curid=3032610>

Fig. 16.4: Low and high magnification Transmission electron micrographs of *trilamellar* membrane structure. Public domain; Adapted from: http://en.wikipedia.org/wiki/History_of_cell_membrane_theory By Sandraamurray

Fig. 16.5: Model of the eukaryotic plasma membrane demonstrating asymmetry, with a sugarless cytoplasmic surface and a glycoprotein rich extracellular surface. Public domain, from: https://commons.wikimedia.org/wiki/File:Cell_membrane_detailed_diagram_en.svg or https://upload.wikimedia.org/wikipedia/commons/thumb/d/da/Cell_membrane_detailed_diagram_en.svg/877px-Cell_membrane_detailed_diagram_en.svg.png By LadyofHats.

Fig. 16.6: Illustration of a scanning electron micrograph of freeze-fractured plasma membrane with pits and mounds showing on opposing phospholipid layers of the membrane. Illustration by G. Bergtrom

Fig. 16.7: Fluorescent antibodies against human and mouse membrane proteins localize their cell surface antigens (i.e., proteins) on opposite poles of recently fused cells. Illustration by G. Bergtrom

Fig. 16.8: Immunofluorescent human (green) and mouse (red) membrane proteins diffuse and mix over time in fused cells. Illustration by G. Bergtrom

Fig. 16.9: Factors affecting membrane fluidity (i.e. rate of diffusion of membrane components). Illustration by G. Bergtrom

Fig. 16.10: Illustration of a micelle. Public domain; Adapted from:

https://upload.wikimedia.org/wikipedia/commons/c/c6/Phospholipids_aqueous_solution_structures.svg

Fig. 16.11: A liposome is a synthetic lipid bilayer that can be formed from micelles. Public domain; Adapted from:

https://upload.wikimedia.org/wikipedia/commons/c/c6/Phospholipids_aqueous_solution_structures.svg

Fig. 16.12: Properties of plasma membranes may differ in ways determined by cell-to-cell associations during tissue development. The lavender (short arrows), green (arrow heads), and purple (long arrows) regions indicate basal differentiation of a sheet of epithelial cells. Illustration by G. Bergtrom

Fig. 16.13: Integral membrane proteins may penetrate or span the membrane; Also shown are peripheral proteins and lipoproteins. Illustration by G. Bergtrom

Fig. 16.14: Integral transmembrane proteins cross the membrane one or more times. Illustration by G. Bergtrom

Fig. 16.15: Hydrophobic amino acids of glycophorin form a helix that spans the red blood cell plasma membrane. Illustration by G. Bergtrom

Fig. 16.16: Hydropathy plots correlate the degree of amino acid hydrophobicity with its position in a polypeptide. Long regions of hydrophobic amino acids suggest a possible membrane protein domain. Illustration by G. Bergtrom

Fig. 16.17: Examples of the many functions of membrane proteins. Illustration by G. Bergtrom

Fig. 16.18: The glycocalyx is the sugar-rich region on the extracellular surface of cells, the result of covalently bound sugars on glycoproteins and glycolipids. It is the basis of many cell functions and is associated with other macromolecules to form an extracellular matrix. CC-BY; Adapted from:

http://cnx.org/The_Glycocalyx_with Illustration by G. Bergtrom

Fig. 16.19: The extracellular matrix (ECM) forms when proteins (e.g., fibronectin, collagen) and other macromolecules (e.g., proteoglycans) associate non-covalently with membrane proteins and elements of the glycocalyx. CC-BY; From: http://cnx.org/Extracellular_Matrix

Fig. 16.20: Blood groups with membrane proteins and elements of the glycocalyx. Public domain; Adapted from InvictaHOG - Own work, Public Domain,

<https://commons.wikimedia.org/w/index.php?curid=1088507>

Chapter 17

Chapter Top-Art: Illustration by G. Bergtrom and Public Domain: <https://pixabay.com/illustrations/traffic-lights-traffic-light-phases-2147790/>

Fig. 17.1: Graph plotting rates of passive and facilitated transport show that facilitated diffusion is a saturable process. Illustration by G. Bergtrom

Fig. 17.2: Three kinds of facilitated diffusion. Illustration by G. Bergtrom

Fig. 17.3: Facilitated transport proteins (e.g., glucose transporter) undergo sequential allosteric changes as they recognize a solute (e.g., glucose) and then transfer and release it on the other side of the membrane. Illustration by G. Bergtrom

Fig. 17.4: Effects of differences in solute concentrations inside and outside *animal cells* on the movement of water into or out of the cells, i.e., *Osmosis*. Illustration by G. Bergtrom

Fig. 17.5: Osmotic effects of differences in solute concentrations inside and outside *plant cells* on the movement of water into or out of the cells, i.e., *Osmosis*. Correct cytoplasmic solute concentrations are achieved by accumulating excess water in the *tonoplast*. Illustration by G. Bergtrom

Fig. 17.6: Plasmodesmata are cell wall 'tunnels' connecting plant cells that allow direct movement of water between cells to transmit and balance changes in osmotic pressure in one part of a plant (e.g., roots) throughout the plant. Illustration by G. Bergtrom

Fig. 17.7: Light micrograph of Paramecium highlighting a contractile vacuole, an organelle that pumps out excess water from the cell, protecting it from osmotic shock. CC-BY-SA 3.0; From:

http://commons.wikimedia.org/wiki/File:Paramecium_caudatum_Ehrenberg_1833.jpg

Fig. 17.8: The *resting potential* of a cell results from an ionic concentration imbalance across the plasma membrane. Sodium and chloride ions concentrations are higher inside cells while potassium ions are at higher levels outside the cell. Thus, the cytoplasm is slightly negative compared to the extracellular fluid. Illustration by G. Bergtrom

Fig. 17.9: After a change in the resting potential (e.g., depolarization) of a cell, the ion balance across the cell membrane is restored by an ATP-powered sodium/potassium pump. Illustration by G. Bergtrom

Fig. 17.10: A patch-clamp device can measure ion flow through voltage-gated channels in a membrane during a depolarization event. Illustration by G. Bergtrom

Fig. 17.11: Voltage changes across e.g., a muscle cell membrane during an action potential correlated with the flow of specific ions, determined with the patch-clamp device. Illustration by G. Bergtrom

Fig. 17.12: Neurotransmission starts with an *action potential* in the cell body of a neuron. The resulting action potential is propagated down the axon to the nerve terminal *synapse*. The *depolarization pulse* initiates Ca^{++} flow into the neuron through *voltage-gated channels* that eventually cause neurotransmitter release into the synaptic cleft between the neuron and the responding cell. The neurotransmitter binds to a *ligand gated channel* on the responding cell, causing Na^{+} ions to flow into the responding cell, leading to an action potential. Illustration by G. Bergtrom

Fig. 17.13: Illustration of three main kinds of *endocytosis*, routes/mechanisms of import of extracellular materials into cells. Illustration by G. Bergtrom

Fig. 17.14: Illustration of the stages of receptor-mediated *endocytosis*. Drawn by G. Bergtrom based on http://www.zoology.ubc.ca/~berger/b200sample/unit_8_protein_processing/images_unit8/14_18.jpg

Fig. 17.15: Molecular details of receptor-mediated *endocytosis*. Illustration by G. Bergtrom

Fig. 17.16: A low-density lipoprotein coated with cholesterol. Adapted from <http://cdn.intechopen.com/pdfs-wm/39539.pdf> by G. Bergtrom

Fig. 17.17: Rough endoplasmic reticulum (RER), Golgi and other vesicles participate in the trafficking of proteins destined for secretion or vesicular packaging/storage in the cell. Illustration by G. Bergtrom

Fig. 17.18: Determining the role of RER in the synthesis of secretory and other packaged protein synthesis. Illustration by G. Bergtrom

Fig. 17.19: Molecular details (steps) of the *Signal Hypothesis*. Illustration by G. Bergtrom

Fig. 17.20: Integral membrane-spanning proteins have one or more very hydrophobic *stop transfer* signals in addition to their signal sequence. Such membrane proteins cross the membranes one or more times during their synthesis. Illustration by G. Bergtrom

Fig. 17.21: Sorting and directing the traffic of secreted, vesicular and other proteins in cells. Illustration by G. Bergtrom

Fig. 17.22: Nuclear proteins made in the cytoplasm contain a positively *charged nuclear localization signal*. This signal binds to the electronegative region of a *nuclear transport receptor* that then binds nuclear pore fibrils, guiding the protein into the nucleus. Illustration by G. Bergtrom

Fig. 17.23: Nuclear proteins destined for mitochondria are synthesized with an N-terminal *mitochondrial signal sequence* that is removed by a *mitochondrial signal peptidase*. Since polysomes do not attach to mitochondria, a different mechanism engages the N-terminal signal with membrane proteins required for transfer. Illustration by G. Bergtrom

Fig. 17.24: *Tight*, *Gap* and *Adherens* junctions between animal cells involve different membrane proteins. Public domain (all drawings); Top left, https://commons.wikimedia.org/wiki/File:Cellular_tight_junction-en.svg; Middle, https://commons.wikimedia.org/wiki/File:Gap_cell_junction-en.svg; Top right, https://en.wikipedia.org/wiki/Adherens_junction;

- CC BY-SA 2.5 Lower left micrograph adapted from Eric Yale Hayden, <https://commons.wikimedia.org/w/index.php?curid=6223080>;

- Public domain; Middle
https://en.wikipedia.org/wiki/Gap_junction#/media/File:Annular_Gap_Junction_Vesicle.jpg,
- CC-BY-SA; <http://www.cellimagelibrary.org/images/7620> (M. Farquhar & G. E. Palade (1963) J. Cell Biol. 17:375-412.)

Fig. 17.25: Glycocalyx protein interactions; roles in cell-cell recognition and attachment. Illustration by G. Bergtrom

Fig. 17.26: LEFT, Transmission electron micrograph of a Melanoma cell exosomes (bar, 100nm). Some of the larger vesicles show an irregular, cup-like shape (arrows). RIGHT: Transmission electron micrograph of a reticulocyte endosome containing exosomes about to fuse with the plasma membrane (bar, 200 nm).

- CC-BY-NC-SA 3.0 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3575529/>
- CC-BY-NC-SA 3.0 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3575527/>

Fig. 17.27: Migration and attachment of cells during development requires the synthesis of membrane proteins that will recognize and bind proteins on the membranes and glycocalyx of other cells, forming cell junctions and enabling *tissue formation*. At the same time, the specificity of these membrane proteins prevents incorrect cell-cell associations. Integrin is an early cell membrane protein 'receptor' that interacts with many proteins to form a glycocalyx and then respond to cell-surface proteins of other cells. Illustration by G. Bergtrom

Fig. 17.28: Density-dependent inhibition (aka. Contact Inhibition) occurs when cells multiplying and spreading on a surface cease dividing when there is no more room on the surface. Cancer cells have lost the property of contact inhibition and keep growing over one another in layers. Loss of contact inhibition is correlated with an absence of gap junctions in cancer cells. Illustration by G. Bergtrom

Fig. 17.29: Signal transduction by *effectors* (e.g., hormones) can lead to many different effects in the cytoplasm as well as in nuclei. Illustration by G. Bergtrom

Fig. 17.30: G-proteins cycle between bound GTP and bound GDP. They are intermediates in signal transduction between an active, effector-bound membrane receptor and adenylate cyclase (follow detailed steps in the text). Illustration by G. Bergtrom

Fig. 17.31: The *fight-or-flight response* to adrenalin (*adrenalin rush*): cAMP activates *protein kinase A*; a *phosphorylation cascade* leads to the release of glucose into the circulation. cAMP is a second messenger of signal transduction for many different cellular responses. Illustration by G. Bergtrom

Fig. 17.32 G-proteins also mediate signal transduction through different membrane receptors and different enzymes, as well as *protein kinase C*. These interactions generate different 2nd messenger molecules that lead to a phosphorylation cascade and many different responses in different target cells (follow detailed steps in the text). Illustration by G. Bergtrom

Fig. 17.33: The effects of *MAP kinase* include phosphorylation and activation of DNA-binding transcription factors and other nuclear proteins that lead to cell proliferation. Illustration by G. Bergtrom

Chapter 18

Chapter Top-Art: Public Domain; <https://pixabay.com/vectors/lifting-box-wood-heavy-crate-24401/>

Fig. 18.1: LEFT, drawing of metaphase spindle fibers RIGHT, fluorescence micrograph of a mitotic spindle showing chromosomes (blue) and spindle fiber (green).

- LEFT, adapted from [Kelvinsong](https://en.wikipedia.org/wiki/Metaphase); <https://en.wikipedia.org/wiki/Metaphase>; CCBY 3.0 Unported License;
- RIGHT, Public Domain, from Afunguy-Transferred to Commons by Lije also using CommonsHelper; <https://commons.wikimedia.org/w/index.php?curid=5148470>

Fig. 18.2: Illustration of assembly, structure and polarity of microtubules, microfilaments and intermediate filaments alongside fluorescence micrographs made using fluorescent antibodies against isolated microtubule, microfilament and intermediate filament proteins (TOP, MIDDLE and BOTTOM, respectively).

- CC-BY 2.0; Adapted from: <http://bmcbiochem.biomedcentral.com/articles/10.1186/1471-2091-9-23> Open Access
- CC-BY-SA-2.5 and GNU.org; Adapted from https://en.wikipedia.org/wiki/Cytoskeleton#/media/File:MEF_microfilaments.jpg Y Tambe (Y Tambe's file) via Wikimedia Commons

- CC-BY; Adapted from: www.plosone.org/article/info:doi/10.1371/journal.pone.0012115
 - Illustrations by G. Bergtrom
- Fig. 18.3:** Illustration and immunofluorescence microscopic localization of microtubules, microfilaments and intermediate filaments in cells.
- Public domain; From: https://en.wikipedia.org/wiki/Fluorescence_microscope
 - CC-BY-SA 3.0 From: <http://en.wikipedia.org/wiki/Vimentin#mediaviewer/File:VIMENTIN.JPG> By Simon Caulton
- Fig. 18.4.** Transmission electron micrograph of the characteristic 9 triplet microtubule array of centrioles and basal bodies. From: Chernov et al. (2008) BMC Biochemistry 2008; 9:23: Public Domain; http://commons.wikimedia.org/wiki/File:Spindle_centriole_embryonic_brain_mouse_TEM.jpg
- Fig. 18.5:** Starting mitotic anaphase, spindle fiber microtubules exert forces that separate and pull chromatids apart and also push the poles of the cell apart. Illustration by G. Bergtrom
- Fig. 18.6:** Disrupting a kinetochore spindle fiber demonstrates a strong force from microtubule disassembly that quickly pulls a pair of chromatids towards centrioles. Illustration by G. Bergtrom
- Fig. 18.7:** Transmission-electron micrographs of a basal body (left) and of cross sections of a pair of 9 + 2 microtubule arrays (right), each of which would emerge from a basal body (white line #1). The position of the 9-triplet microtubule array of microtubules seen in Fig. 18.4 is indicated by white line #2.
- Public domain; From: http://en.wikipedia.org/wiki/File:Chlamydomonas_TEM_09.jpg as per: <http://remf.dartmouth.edu/imagesindex.html>
 - Public domain; From: http://en.wikipedia.org/wiki/File:Chlamydomonas_TEM_17.jpg as per: <http://remf.dartmouth.edu/imagesindex.html>
- Fig. 18.8:** Transmission electron micrographs of axonemes show that microtubules are made up of a ring of 13 tubulins (LEFT). Microtubules in cilia or flagella are arranged in typical 9+2 arrays (MIDDLE). The enlarged cross-section illustrates specific microtubule-associated proteins (MAPs) that maintain the 9+2 structure, including dynein motors that powers motility.
- CC-BY 4.0; Adapted from: <http://bmcbiochem.biomedcentral.com/articles/10.1186/1471-2091-9-23>
 - Public domain; From: http://en.wikipedia.org/wiki/File:Chlamydomonas_TEM_17.jpg as per: <http://remf.dartmouth.edu/imagesindex.html>
 - Illustration of axoneme components by G. Bergtrom
- Fig. 18.9:** Roles of *dynein* and *kinesin* in retrograde (backwards) and anterograde (forward) movement of cargo vesicles along microtubules.
- CC-BY; Adapted from J. E. Duncan & L. S. B. Goldstein (2006) The Genetics of Axonal Transport and Axonal Transport Disorders, PLOS Cognitive Neuroscience Channel at <https://doi.org/10.1371/journal.pgen.0020124>
- Fig. 18.10:** Removing the membrane from isolated cilia or flagella leaves behind the axoneme; when provided ATP, both the isolated structures and demembranated axoneme will beat as the ATP hydrolyzed. Illustration by G. Bergtrom
- Fig. 18.11:** Detergents can dissociate axonemes into individual and doublet microtubules. Dialysis of separated microtubules to remove the detergent cause the microtubules to re-associate (*reconstitute*) into a sheet with connections resembling those seen in intact axonemes.. Illustration by G. Bergtrom
- Fig. 18.12:** Comparison of ciliary beat with propeller like generation of flagellar wave motion.
- CC-BY 3.0; Adapted From: <http://en.wikipedia.org/wiki/File:Flagellum-beating.svg>
- Fig. 18.13:** Light micrograph of skeletal muscle stained to showing characteristic striations. CC-BY: From OpenStax: Anatomy and Physiology; <https://openstax.org/books/anatomy-and-physiology/pages/10-1-overview-of-muscle-tissues>
- Fig. 18.14:** Skeletal muscle organization and the anatomy of a muscle cell sarcomere.
- LEFT, CC-BY 4.0; From: [https://commons.wikimedia.org/wiki/File:1007_Muscle_Fibes_\(large\).jpg](https://commons.wikimedia.org/wiki/File:1007_Muscle_Fibes_(large).jpg) By OpenStax via Wikimedia Commons; [Inset: From : <File:Skeletal muscle 横纹肌.1.JPG> - Wikimedia Commons]
 - RIGHT, CC-0 (Public Domain) From User:Sameerb (en:WP; Author User:Sameerb in English WP) , via Wikimedia Commons <http://upload.wikimedia.org/wikipedia/commons/a/a4/Sarcomere.gif>
- Fig. 18.15:** Transmission electron micrograph and a corresponding illustration showing sarcomere shortening, consistent with the sliding of filaments during skeletal muscle contraction.
- CC-BY-SA 4.0 (LEFT); From: <https://www.mrgscience.com/topic-112-movement.html>

- GNU FDL (RIGHT); Adapted from <https://upload.wikimedia.org/wikipedia/commons/6/6e/Sarcomere.svg> By Richfield, David (2014). "Medical gallery of David Richfield". WikiJournal of Medicine 1 (2). DOI:10.15347/wjm/2014.009. ISSN 2002-4436 (GFDL licensure: <http://www.gnu.org/copyleft/fdl.html>), from Wikimedia Commons
- Fig. 18.16:** The skeletal muscle *Contraction Paradox*: Given ATP, *glycerinated* muscle fibers contract and pull a weight. When all the ATP is hydrolyzed the fiber can't stretch unless more ATP is added. Illustration by G. Bergtrom
- Fig. 18.17:** Overview of the isolation of actin thin filaments (still on Z-Lines) from myosin thick filaments. Illustration by G. Bergtrom
- Fig. 18.18:** Reconstitution of actin thin filaments (on Z-Lines) with myosin filaments. Illustration by G. Bergtrom
- Fig. 18.19:** Structure of a skeletal muscle myosin filament and the myosin monomer. Shown is *myosin II*, the thick filament that spans both sides of the H zone in a sarcomere (upper). The head-&-tail structure of a myosin monomer is shown in the high magnification electron micrograph and is illustrated in the cartoon (lower). The myosin monomer is itself a polymer of four polypeptides.
- CC-BY 4.0; Adapted from <http://plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.0020356> By D. Hostetter et al. (2004) PLoS Biology
 - CC-BY 4.0; Adapted from: Adapted from <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0093272> By H. Sugi et al. (2014) PLOS One; PLoS ONE 9(6): e93272; doi:10.1371/journal.pone.0093272
 - Illustrations by G. Bergtrom
- Fig. 18.20:** Digestion of purified myosin monomers with enzymes that hydrolyze peptide bonds between specific amino acids produces an *S1 head* and a *tail* fraction with different properties.
- CC-BY 4.0; Adapted from <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0093272> , By H. Sugi et al. (2014) PLOS One; PLoS ONE 9(6): e93272; doi:10.1371/journal.pone.0093272
 - Illustrations by G. Bergtrom
- Fig. 18.21:** Illustration of myosin decoration by myosin monomer S1 fragments, showing opposing polarity of actin filaments on opposite sides of the Z-line. Illustration by G. Bergtrom
- Fig. 18.22:** Steps in the *Microcontraction Cycle* explain the muscle contraction paradox. Illustration by G. Bergtrom
- Fig. 18.23:** The ability of flexible myosin heads to bend and change conformation is consistent with their proposed activity during the microcontraction cycle. CC-BY 4.0; Adapted from <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0093272> By H. Sugi et al. (2014) PLoS ONE 9(6): e93272; doi:10.1371/journal.pone.0093272
- Fig. 18.24:** Innervation leads to Ca^{++} release from *sarcoplasmic reticulum* to regulate contraction.
- CC-BY 4.0; From S. Guarnieri et al. (2013) <https://doi.org/10.1371/journal.pone.0053267> and <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0053267#pone-0053267-g006>
 - CC-BY 4.0; <https://opentextbc.ca/anatomyandphysiology/chapter/10-2-skeletal-muscle/>
 - Overlay text and illustration by G. Bergtrom
- Fig. 18.25:** Skeletal muscle thin filaments consist of actin associated with troponins and tropomyosin. These actin-associated proteins participate in the response to Ca^{++} ions to regulate interactions with myosin. Public domain; Adapted from: <https://upload.wikimedia.org/wikipedia/commons/thumb/a/a3/Myofilament.svg/994px-Myofilament.svg.png>
- Fig. 18.26:** Ca^{++} ions initiate uncovering myosin binding sites on actin by binding to *troponin-C*, resulting in allosteric changes in the troponins and ultimately tropomyosin. Illustration by G. Bergtrom
- Fig. 18.27:** Structure of *titin* and its location in the sarcomere.
- CC-BY 4.0; Colorized micrograph of titin adapted from Zsolt Mártonfalvi and Miklós Kellermayer (2014): <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0085847>
 - Illustration by G. Bergtrom
- Fig. 18.28:** Elasticity of skeletal muscle shown by the location of anti-N2A-linked and anti-120/122-linked nanogold particles in sarcomeres. Note the increased separation of the PEVK domains targeted by the antibodies as a muscle fiber is increasingly stretched. This titin elasticity facilitates skeletal muscle relaxation; Adapted from <https://commons.wikimedia.org/w/index.php?curid=764619>, By User:Sameerb

(en:WP; Author User:Sameerb in English WP), Copyrighted free use; via Wikimedia Commons. Adaptations included adding simulated nanogold particles and elongating parts of the sarcomere to create an image replicating results of work from the laboratory of Wolfgang Linke.

Fig. 18.29: In the anti-actin immunofluorescence micrograph of fibroblasts, actin localizes with stress fibers which help maintain cell shape (LEFT). Actin also localizes in lamellipodia and retraction fibers in migrating fibroblasts (RIGHT), orienting in the direction of movement.

- Adapted from CC BY-SA 3.0 <https://commons.wikimedia.org/w/index.php?curid=378394> By Y Tambe - Y Tambe's file
- Illustration by G. Bergtrom

Fig. 18.30: The molecular structure of stress fibers; myosin as well as other actin-binding proteins interact with actin in nonmuscle cell motility. Illustration by G. Bergtrom

Fig. 18.31: Coiled secondary structure of the proteins permits elasticity of intermediate filament bundles that contribute to the viscosity of cytoplasm. Illustration by G. Bergtrom

Chapter 19

Chapter Top-Art: Public Domain: <https://pixabay.com/photos/microscope-sea-urchin-egg-splitting-1276131/>

Fig. 19.1: Bacteria divide by binary fission. When growing in culture, they divide continually, partially replicating their circular DNA molecules during one division in preparation for the next. Illustration by G. Bergtrom

Fig. 19.2: Mitosis and cytokinesis in eukaryotic cells are separated in time, and mitosis is further divisible into 5 phases. Illustration by G. Bergtrom

Fig. 19.3: First experiment demonstrating that replication and cell division in eukaryotes are separate events.. Illustration by G. Bergtrom

Fig. 19.4: The autoradiographic data from the experiment outlined in Fig. 19.3 demonstrates that DNA synthesis begins and ends some time before the beginning of mitosis. Illustration by G. Bergtrom

Fig. 19.5: These autoradiographs from a *pulse-chase* labeling experiment led to the identifying the phases of the eukaryotic cell cycle (see text for explanation). Illustration by G. Bergtrom

Fig. 19.6: Graph plotting the number of radioactive cells in mitosis over time of chase in the pulse-chase experiment described in Fig. 19.5 (see text for further explanation). Illustration by G. Bergtrom

Fig. 19.7: A typical eukaryotic cell cycle: phases G1, S and G2 follow cytokinesis, and G2 immediately precedes prophase of mitosis. The length of time for each phase differs for different cell types. Illustration by G. Bergtrom

Fig. 19.8: Terminally differentiated cells no longer divide, entering the Go state. While they can sometimes resume cycling, such cells more typically experience cell death and replacement by stem cells. Illustration by G. Bergtrom

Fig. 19.9: Experiment leading to the discovery of *meiosis-promoting factor* (MPF), the first known chemical regulator of cell division. Illustration by G. Bergtrom

Fig. 19.10: MPF was shown to be a two-subunit *protein kinase* that transfers phosphates from ATP to several different proteins. Illustration by G. Bergtrom

Fig. 19.11: Graph comparing plots of cellular *cdk* and *cyclin* levels over time with one of *MPF* activity. Illustration by G. Bergtrom

Fig. 19.12: Cell fusion experiments revealed additional chemical regulators of the cell cycle activity. Illustration by G. Bergtrom

Fig. 19.13: Simplified diagram of cell cycle checkpoints at which progress through the cycle is assessed. If progress through a phase is incomplete, cell cycle regulators (kinases) delay onset of the next phase. Illustration by G. Bergtrom

Fig. 19.14: Apoptotic Cells: In the 3 electron micrographs and corresponding illustrations, normal cells (A & A') are followed by apoptotic cells [B & B'] and [C & C']. In [B & B'], a black arrowhead indicates nuclear condensation (*pyknosis*), membrane blebbing (grey arrowheads) and apoptotic bodies (white arrowhead). In [C & C'] the nucleus has broken down (*karyorrhexis*) and the cell itself cell has fragmented into apoptotic bodies (white arrowheads). In C', a phagocyte is engulfing one of them.

- Electron micrographs, CC-BY; Adapted from: V. Wilhelmi et al. (2013) Zinc Oxide Nanoparticles Induce Necrosis and Apoptosis in Macrophages in a p47phox- and Nrf2-Independent Manner. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0065704>

- Illustrations: Public Domain; File adapted from a German Wikipedia entry:
<https://commons.wikimedia.org/wiki/File:Apoptosis.png> or <http://de.wikipedia.org/wiki/Bild:Apoptose-german.png>

Fig. 19.15: Biochemical steps of *apoptosis*, or *programmed cell death*. Illustration by G. Bergtrom

Fig. 19.16: Comparison of the cellular events of apoptosis and necrosis. Public Domain; Adapted from National Institute on Alcohol Abuse and Alcoholism (NIAAA) via Wikimedia Commons;

https://commons.wikimedia.org/wiki/File:Structural_changes_of_cells_undergoing_necrosis_or_apoptosis.png

Fig. 19.17: Results of non-cancer cell divisions in which abnormal cells are targeted for apoptosis, and cancer cell division, in which damaged or altered cells escape apoptotic controls. Illustration by G. Bergtrom

Fig. 19.18: Making knockout mutant mice. Specific nucleotides in a cloned gene of interest are altered by *site-directed mutagenesis*. The altered gene is inserted into embryonic stem cells where it recombines with and replaces the homologous gene already in the chromosome. Recombinant clones are selected re-injected into embryos that are then incubated in the uteri of 'foster mother' mice. Knockout mice in the newborn litter can be selected and studied for the effects of removing a gene. Illustration by G. Bergtrom

Fig. 19.19: Structure of the *P53* gene-regulatory protein bound to DNA. *P53* was originally called a *tumor suppressor protein* because when mutated, tumors arose. CC-BY-SA 3.0; From Thomas Splettstoesser, via Wikimedia Commons

<https://upload.wikimedia.org/wikipedia/commons/b/bb/P53.png>

Fig. 19.20: Normal function of the *P53* gene-regulatory protein. Illustration by G. Bergtrom

Fig. 19.21: Role of the *P53* in decision-making at cell cycle checkpoints. Illustration by G. Bergtrom

Fig. 19.22: Summary of different roles of *p53* protein in protein degradation and apoptosis, cell cycle progress and DNA repair. Illustration by G. Bergtrom

Chapter 20

Chapter Top-Art: Public Domain photos; <https://pixabay.com/illustrations/sperium-cum-sperm-dark-winner-326157/> and <https://pixabay.com/photos/halloween-pumpkin-scary-giant-309296/>

Fig. 20.1: Descent (evolution and speciation) from a single progenitor cell formed by sheer luck on a prebiotic Earth.. Illustration by G. Bergtrom

Fig. 20.2: Descent (evolution and speciation) from multiple progenotes, or "first" cells". Illustration by G. Bergtrom

Fig. 20.3: Evolutionary (i.e., phylogenetic) tree of all living organisms, showing descent from the *Last Universal Common Ancestor (LUCA)* of the 3 domains of life (Bacteria, Archaea and Eukaryota).

Public Domain; Adapted from: https://en.wikipedia.org/wiki/Phylogenetic_tree

Fig. 20.4: Classic Miller & Urey experiment showing that organic molecules found in living things today could be synthesized in a laboratory under *reducing* conditions similar to those expected for a prebiotic earth atmosphere. Illustration by G. Bergtrom

Fig. 20.5: Scenario for the synthesis of prebiotic polymers in tidal pools. Illustration by G. Bergtrom

Fig. 20.6: Scenario for the synthesis of redundant prebiotic nucleic acid polymers in tidal pools.

Illustration by G. Bergtrom

Fig. 20.7: Analysis of this Australian *zircon* supports an oxidizing atmosphere on a prebiotic earth.

Valley, J. W. et al. (2014) Hadean age for a post-magma-ocean zircon confirmed by atom-probe tomography, *Nature Geoscience* vol. 7, p 219-223; <http://dx.doi.org/10.1038/ngeo2075>; Used by permission of J. W. Valley, University of Wisconsin - Madison;

<http://geoscience.wisc.edu/geoscience/people/faculty/john-valley/john-valley-incle-on-zircons/>

Fig. 20.8: An oceanic volcanic hydrothermal vent, or black smoker. Public domain:

<https://commons.wikimedia.org/w/index.php?curid=262511>

Fig. 20.9: An oceanic volcanic hydrothermal vent, or *White smoker*. Public domain:

http://oceanexplorer.noaa.gov/explorations/04fire/logs/hirez/champagne_vent_hirez.jpg

Fig. 20.10: *Serpentinite* from Deer Lake in upper Michigan. It is a mineral also found in the oceanic crust that under conditions found in an alkaline vent, can form methane from CO₂. From: Gabriel HM, Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=48672474>

Fig. 20.11: Evolutionary (i.e., phylogenetic) tree of all living organisms reveals 3 domains of life: Bacteria (Prokarya), Archaeobacteria (archaea) and Eukaryotes (Eukaryota). Public Domain; Adapted from:

https://en.wikipedia.org/wiki/Phylogenetic_tree

Fig. 20.12: *Aminoadenosine triacid ester* (AATE) catalyzes its own replication by the mechanism suggested here. Illustration by G. Bergtrom

Fig. 20.13: The deamination of adenine in a ribonucleotide chain to inosine in a tRNA by the *adenine deaminase* enzyme. Public domain; adapted from

<https://commons.wikimedia.org/wiki/File:Adenosinmonophosphate>

Fig. 20.14: *Proteinoid microspheres*, *coacervates* and *liposomes* can all be made in a laboratory and are candidates for boundary structures that could have protected early prebiotic organic molecules and reactions.

- Liposome image adapted from: <https://upload.wikimedia.org/wikipedia/commons/c/c6/Micelles-Liposomes>;
- Proteinoid microsphere image adapted from CC-0 (public domain; http://ejournal8.com/journals_n/1450718865.pdf; I. Ignatov and O. Mosin , 2015. *S. Miller's Experiments in Modelling of Non-Equilibrium Conditions with Gas Electric Discharge Simulating Primary Atmosphere*. European Journal of Molecular Biotechnology 10, 197-209
- Coacervate image adapted from <https://commons.wikimedia.org/Coacervates/index.php?curid=92728896>; By Spruijtlab - Own work, CC BY-SA 4.0,
- Illustration by G. Bergtrom

Fig. 20.15: Genetic information flows from DNA to RNA to protein (the *Central Dogma*) but can also flow from RNA to DNA by reverse transcription. Illustration by G. Bergtrom

Fig. 20.16: RNA Sequences contain genetic information, and in retroviruses, they are molecules of inheritance. As single stranded molecules, RNAs can fold into 3-dimensional shapes, creating specific shapes that can act as catalysts (e.g., ribozymes). Thus RNAs can combine information storage with catalytic activities. Illustration by G. Bergtrom

Fig. 20.17: Hypothetical origin of RNA catalysis, suggesting how some early (prebiotic) RNAs may have evolved to catalyze peptide bond formation between amino acids. Illustrated by G. Bergtrom

Fig. 20.18: Suggested steps in evolution from an RNA world to our DNA world. The first of these steps would be to divorce the self-replication activity of RNAs from their information storage function, as illustrated here. Illustration by G. Bergtrom

Fig. 20.19: If RNAs could bind peptides whose synthesis they catalyze, evolution could have selected ribozymes from some RNA-peptide complexes that did not separate. Later, the peptides themselves could have evolved to take over catalytic functions from ribozymes. Illustration by G. Bergtrom

Fig. 20.20: A final evolutionary scenario gets us from the RNA world to a DNA world that governs most life on earth today. Illustration by G. Bergtrom

Fig. 20.21: One selective pressure that contributes to the complexity of biochemical pathways would be the depletion of a particular chemical resource, say molecule D, in the environment). The few cells in the population that happen to have an enzyme that can convert C to D will survive and proliferate. Illustration by G. Bergtrom

Appendix II: Context-Embedded YouTube Videos

The following links are inactive in the print version of this book (CMB5p). Please use QR codes in this book or download the digital *Annotated CMB5e* at







https://dc.uwm.edu/biosci_facbooks_bergtrom/





to access to the links.

Chapter 1: Cell Tour, Life's Properties and Evolution, Studying Cells







-  [100-2 The RER-Rough Endoplasmic Reticulum](#)
-  [101 Ribosomes and Polysomes](#)
-  [102 Golgi Vesicles and the Endomembrane System](#)
-  [103-2 Smooth Endoplasmic Reticulum](#)
-  [104-2 The Nucleus](#)
-  [105-2 Endosymbiosis-Mitochondria & Chloroplasts](#)
-  [106-2 Filaments & Tubules of the Cytoskeleton](#)
-  [107-2 Dissecting the Cell-a Cell Fractionation Scheme](#)
-  [108-2 Isolated Nuclei](#)
-  [109-2 Isolated RER](#)
-  [110-2 Isolated Golgi Vesicles](#)
-  [111-2 Lysosomes & Peroxisomes](#)
-  [112-2 Isolated Mitochondria](#)
-  [113-2 Isolated Chloroplasts](#)
-  [114-2 Isolated Membranes Form Vesicles](#)
-  [115 Properties of Life](#)
-  [116 The Universal Genetic Code](#)
-  [117 Origins of Life](#)
-  [118 Life Origins vs Evolution](#)
-  [119 The Random Basis of Evolution](#)
-  [120-2 Genomic Fossils-Molecular Evolution](#)
-  [121-2 Electron Microscopy](#)

Chapter 2: Basic Chemistry, Organic Chemistry and Biochemistry








-  [122-2 Atoms and Elements](#)<https://youtu.be/-mLaDeYV6E>
-  [123 Electron Energy and Fluorescence](#)
-  [124 Covalent Bonds](#)
-  [125-2 Water-Hydrogen & Ionic Bonds](#)
-  [126 Organic Molecules Monomers and Polymers](#)
-  [127-2 Carbohydrates-Sugars & Polysaccharides](#)

-  [128-2 Lipids-Triglycerides-Phospholipids](#)
-  [129-2 Proteins-Amino Acids & Polypeptides](#)
-  [130-2 DNA & RNA-Nucleotides & Nucleic Acids](#)
-  [131-2 Shape & the Specificity of Molecular interactions](#)







Chapter 3: Details of Protein Structure

-  [132-2 Amino Acid Sequence & Protein Primary Structure](#)
-  [133-2 Protein Secondary Structure](#)
-  [134-2 Protein Tertiary structure](#)
-  [135-2 Disulfide Bridges Stabilize 3° Structure](#)
-  [136-2 Protein Quaternary Structure & Prosthetic Groups](#)
-  [137-2 Protein Domain Structure & Function](#)







Chapter 4: Bioenergetics





-  [138 Different Kinds of Energy; Chemical Equilibrium](#)
-  [139 First Law of Thermodynamics](#)
-  [140-2 Second Law of Thermodynamics](#)
-  [141 Deriving Closed System Thermodynamics](#)
-  [142 Determining DH and DG in Closed Systems](#)
-  [143 Determining DS in Closed System](#)
-  [144 The Energetics of Open Systems](#)

Chapter 5: Enzyme Catalysis and Kinetics







-  [145 Enzymes vs Other Catalysts](#)
-  [146-2 Induced Fit mechanism of Enzyme Action](#)
-  [147 Enzyme Activation Energy](#)
-  [148-2 Allosteric Regulation of Enzyme Activity](#)
-  [149 Measuring Enzyme Kinetics](#)
-  [150 Graphing Enzyme Kinetic Data](#)

Chapter 6: Glycolysis, The Krebs Cycle and the Atkins Diet









-  [151 Overview of Glycolysis](#)
-  [152-2 Glycolysis Stage 1, Reaction 1](#)
-  [153 Glycolysis Stage 1; Reactions 2-5](#)
-  [154 Glycolysis Stage 2; Reaction 6](#)
-  [155-2 Glycolysis Stage 2; Reactions 7-10](#)
-  [156 Fermentation: Regulation of Pyruvate Reduction is NOT Allosteric!](#)

-  [157-2 Balance Sheet of Glycolysis](#)
-  [158-2 Gluconeogenesis & the Atkins Diet](#)
-  [159-2 Highlights of the Krebs Cycle](#)
-  [160 Discovery of the Krebs Cycle](#)












Chapter 7: Electron Transport, Oxidative Phosphorylation and Photosynthesis

-  [161 Mitochondrial Electron Transport Oxidizes Reduced Electron Carriers](#)
-  [162-2 Finding the Free Energy of Electron Transport](#)
-  [163-2 Separating Electron Transport from Oxidative Phosphorylation](#)
-  [164-2 Proton Pumps Store Free Energy of the ETC in Proton Gradients](#)
-  [165-2 Proton Gates Capture Proton Gradient Free Energy as ATP](#)
-  [166 Balance Sheet of Respiration](#)
















Chapter 8: DNA Structure, Chromosomes and Chromatin

-  [167-2 Transformation In & Out of Mice-Griffith, McCarty et al.](#)
-  [168 Hershey and Chase: Viral Genes are in Viral DNA](#)
-  [169-2 Unraveling the Structure of DNA](#)
-  [170 Replication is Semiconservative](#)
-  [171-2 DNA, Chromosomes, Karyotypes & Gene Maps](#)
-  [172-2 Nucleosomes-DNA and Protein](#)
-  [173-2 Chromatin Structure-Dissecting Chromatin](#)
-  [174-2 Histones and Non-Histone Proteins](#)
















Chapter 9: DNA Replication and Repair

-  [175 Seeing E. coli Chromosomes](#)
-  [176 Semiconservative Bidirectional Replication From Two RFs](#)
-  [177 Multiple Replicons in Eukaryotes](#)
-  [178 DNA Polymerases and Their Activities](#)
-  [179 Replication Initiation in E coli](#)
-  [180-2 Okazaki Experiments-Solving a Problem at an RF](#)
-  [181 Okazaki Fragments are Made Beginning with RNA Primers](#)
-  [182 Replication Elongation in E coli](#)
-  [183 Telomerase Replicates Telomeres to Prevent Chromosome Shortening](#)
-  [184 Processive Replication](#)
-  [185-2 Topoisomerases Relieve Supercoiling During Replication](#)

Chapter 10: Transcription and RNA Processing



-  [186 Transcription Overview: Ribosomes and Ribosomal RNAs](#)
-  [187 Transcription Overview: Demonstrating the Major RNAs](#)
-  [188-2 Transcription Overview- Basics of RNA Synthesis](#)
-  [189 RNA Polymerases in Prokaryotes and Eukaryotes](#)
-  [190 Post Transcriptional Processing Overview](#)
-  [191 Details of Prokaryotic Transcription](#)
-  [192-2 Eukaryotic mRNA Transcription](#)
-  [193 Recognition of Transcription factors at Promoters](#)
-  [194-2 The Discovery of Split Genes](#)
-  [195 mRNA Splicing](#)
-  [196-2 Origin of Introns](#)
-  [197 Intron Evolution: What was selected here?](#)
-  [198 mRNA 5' Capping and 3' Polyadenylation](#)
-  [199-2 rRNA Transcription & Processing](#)
-  [200-2 tRNA Transcription and Processing](#)

Chapter 11: The Genetic Code and Translation

-  [201 The Genetic Code Dictionary](#)
-  [202 Speculations About a Triplet Code](#)
-  [203 Deciphering the First Codon](#)
-  [204 Deciphering all 64 Triplet Codons](#)
-  [205-2 tRNA Structure and Base Modifications](#)
-  [206 Translation Initiation: mRNA Associates with 30S Ribosomal Subunit](#)
-  [207 Initiation Complex Formation](#)
-  [208 Adding the Large Ribosomal Subunit](#)
-  [209 Elongation: Elongation Factors and GTP](#)
-  [210 Elongation: A Ribozyme Catalyzes Peptide Linkage Formation](#)
-  [211 Elongation: Translocase Moves Ribosome Along mRNA](#)
-  [212 Adding the Third Amino Acid](#)
-  [213 Big Translation Energy Costs](#)
-  [214 The Fates of fMet and Met; Cases of Posttranslational Processing](#)
-  [215-2 Translation Termination](#)

Chapter 12: Regulation of Transcription and Epigenetic Inheritance


-  [216 Overview of Prokaryotic Gene Regulation](#)







-  [217 Regulation of the lac Operon](#)
-  [219 Repression of the Tryptophan \(TRP\) Operon](#)
-  [220 An Experiment: All of an Organism's Cells Have the Same Genome](#)
-  [221-2 Many Options for Regulating Eukaryotic Genes](#)
-  [222-2 Transcription Factors Bind DNA Near & Far](#)
-  [223-2 Assembling a Eukaryotic Transcription Initiation Complex](#)
-  [224-2 Transcription Factor Domains-Motifs Bind Specific DNA Sequences](#)
-  [225 Chemicals That Control Gene Expression](#)
-  [226 Steroid Hormones Regulate Gene Transcription](#)
-  [227 Signal Transduction Can Lead to Gene Regulation](#)
-  [228 Question: Is Euchromatic DNA Transcribed?](#)
-  [229 Experiment and Answer: Euchromatin *is* Transcribed](#)
-  [230 Epigenetic Inheritance: First Inkling](#)
-  [231-2 Experimental Demonstration of Germ-Line Epigenetic Inheritance](#)

Chapter 13: Posttranscriptional Regulation of Gene Expression

-  [232 Riboswitches Interrupt Bacterial Transcription](#)
-  [233 Small Metabolites Also Regulate Bacterial mRNA translation](#)
-  [234 siRNA Posttranscriptional Regulation](#)
-  [235 Did siRNA Coopt RISC from a Strategy to Trash Corrupt or worn out RNA?](#)
-  [236 miRNA Posttranscriptional Regulation](#)
-  [237 Translation Regulation of Globin Polypeptide Synthesis](#)

Chapter 14: Repetitive DNA; A Eukaryotic Phenomenon









-  [238 Discovery of Repetitive DNA](#)
-  [239 CoT Curves and DNA Complexity Explained!](#)
-  [240 Identifying Different Kinds of DNA Each CoT Fraction](#)
-  [241 Some Repetitive DNA Functions](#)
-  [242 What Interested McClintock About Maize](#)
-  [243 Variegated Maize Kernels result from “Loss” of the Ds Gene](#)
-  [244 Discovery of Mobile Elements and the Ac-Ds System](#)
-  [245 The Ac-Ds System Today](#)
-  [246 Bacterial Mobile Elements](#)
-  [247 Introduction to Eukaryotic Transposons](#)
-  [248 Eukaryotic Class II \(DNA\) Transposition](#)
-  [249 Introduction to Features of Retrotransposition](#)
-  [250 LTR Retrotransposons](#)










-  [251 LINEs \(Long Interspersed Nuclear Elements\): LTR Retrotransposons](#)
-  [252 SINEs \(Short Interspersed Nuclear Elements\): Non-LTR Retrotransposons](#)
-  [253 Extrachromosomally Primed Retrotransposition](#)
-  [254 Target Primed Retrotransposition](#)
-  [255 Transposon Evolution](#)
-  [256 Transposons-Junk or Not](#)

Chapter 15: DNA Technologies

-  [257 Overview of DNA Technologies](#)
-  [258 Isolate mRNA and Make cDNA](#)
-  [259 Reverse Transcriptase](#)
-  [260 Restriction Enzymes and Recombinant DNA](#)
-  [261 Recombine a cDNA Insert with a Plasmid Vector](#)
-  [262 Making the cDNA Library](#)
-  [263 Making a Replica Plate](#)
-  [264 Probing a Replica Plate Filter](#)
-  [265 Probe Replica Filters, Get a Clone and Play With It!](#)
-  [266 Treating Cancer with Dideoxynucleosides](#)
-  [267 Manual Dideoxy Sequencing](#)
-  [268 Automated Sequencing Leads to Large Genome Projects](#)
-  [269 Genomic Libraries: Make and Package Recombinant Phage DNA](#)
-  [270 Infect Host with Recombinant Phage to Make a Genomic Library](#)
-  [271 Screen a Genomic Library, Pick and Grow a Phage Clone](#)
-  [272 PCR: Design and Synthesize Opposing Oligonucleotide Primers](#)
-  [273 PCR: The Amplification Reaction](#)
-  [274 The Power of PCR: Some Examples](#)
-  [275-2 The Power of Microarrays](#)




Chapter 16: Membrane Structure

-  [276-2 Membrane Lipids & Phospholipid Bilayer](#)
-  [277-2 Experiments with & Uses of Liposomes](#)
-  [278-2 Properties of Proteins Embedded in a Phospholipid Bilayer](#)
-  [279 Different Membrane Compositions](#)
-  [280-2 Freeze Fracture Electron Microscopy of Cell Membranes](#)
-  [281-2 EM Cytochemical Demonstration of Membrane Asymmetry](#)
-  [282-2 Electrophoretic Demonstration of Membrane Asymmetry](#)
-  [283 Two Demonstrations of Membrane Fluidity: The Fluid Mosaic](#)

-  [284-2 Factors Influencing Membrane Fluidity](#)
-  [285-2 Membrane Domains-Regional Differentiation of a Plasma Membrane](#)
-  [286-2 Domains of Membrane Proteins](#)
-  [287-2 Hydropathy Predicts Hydrophobic Domains and Membrane Proteins](#)
-  [288-2 Diversity of Membrane Protein Structure & Function](#)
-  [289-2 Pore Proteins May Cross the Membrane Many Times](#)
-  [290-2 Red Blood Cell \(Erythrocyte\) Membrane Protein Functions](#)
-  [291 The Path to Sugar Coated Cells](#)
-  [292-2 The Extracellular Matrix](#)

Chapter 17: Membrane Function


-  [293-2 Passive & Facilitated Diffusion](#)
-  [294-2 Osmosis](#)
-  [295 Potassium Leakage Helps to Maintain Cellular Resting Potentials](#)
-  [296 Active Transport by the Sodium/Potassium Pump](#)
-  [297 A Patch-Clamp Device Can Record Membrane Potential and Ion Flow](#)
-  [298 Patch-Clamp Measurement of Membrane Resting Potential and Depolarization](#)
-  [299 Gated Ion Channels](#)
-  [300-2 Types of Ion Gated Channels](#)
-  [301 Gated Ion Channels Open and Close in Sequence During an Action Potential](#)
-  [302 Propagating an Action Potential Along an Axon](#)
-  [303-2 Gated Ion Channels Open & Close During an Action Potential](#)
-  [304 Formulating the Signal Hypothesis: Early Experiments](#)
-  [305 Testing the Signal Hypothesis](#)
-  [306-2 Details of the Signal Hypothesis](#)
-  [308-2 Integral Proteins Have Stop-Transfer Sequences](#)
-  [309 Protein Traffic to Nuclei and Mitochondria](#)
-  [310-2 Cell Junction Structure and Function](#)
-  [311-2 Glycocalyx-Sugars Covalently link to Plasma Membrane Proteins](#)
-  [312 Cell Adhesion Molecule Functions in the Glycocalyx](#)
-  [313 Formation of a Glycocalyx, Normal Development and Cancer](#)
-  [314 The Role of the Extracellular Matrix in Cell Migration and Development](#)
-  [315-2 Introduction to Signal Transduction](#)
-  [316-2 G-Protein Signal Transduction](#)
-  [317 G-Protein Activation of Protein Kinase A and the Fight-or-Flight Response](#)

-  [318-2 G-Protein Activation of Protein Kinase C and Phospholipase C](#)
-  [319 Receptor Kinase Signal Transduction](#)
-  [320-2 The RAS Oncogene-its Normal Mitogenic Effects and Cancer](#)



Chapter 18: The Cytoskeleton and Cell Motility











-  [321-2 Introduction to the Cytoskeleton](#)
-  [322-2 Microtubules, Microfilaments and Intermediate Filaments in Cells](#)
-  [323-2 Demonstration of the Polarity & Dynamics of Microtubules](#)
-  [324 Microtubules in Mitotic Spindle Fibers Generate Force on Chromatids](#)
-  [325-2 Microtubule Motor Proteins](#)
-  [326-2 9+2 Microtubule Array in Axonemes that Beat](#)
-  [327-2 Proof of Sliding Microtubules in Flagella and Cilia](#)
-  [328 Bacterial Flagella are Powered by a Proton Gradient](#)
-  [329 The Effects of Different Drugs on Microtubules... and Cancer](#)
-  [330-2 The Sliding Filament Model of Skeletal Muscle Contraction](#)
-  [331 The Contraction Paradox](#)
-  [332 In Vitro and Electron Microscope Evidence for a Sliding Filament Model](#)
-  [333-2 Thick Filament & Myosin Monomer Structure](#)
-  [334-2 Myosin Monomers & S1 Heads Decorate Actin](#)
-  [335-2 An Actin-Myosin Contraction Cycle Resolves the Contraction Paradox](#)
-  [336-2 ATP Binding & Hydrolysis Changes Myosin Head Conformation](#)
-  [337 Regulation of Skeletal Muscle Contraction by Calcium](#)
-  [338 Contraction Generates Force Against Z Lines and Cell Membranes](#)

Chapter 19: Cell Division and the Cell Cycle

-  [339 Binary Fission](#)
-  [340-2 Experiments Revealing Replication in Cell Cycle Interphase](#)
-  [341-2 Events in the Phases of the Cell Cycle](#)
-  [342 Discovery of MPF Kinase and Its Role in Meiosis and Mitosis](#)
-  [343 Cell Cycle Control at Check Points and the Go "Phase"](#)
-  [344-2 Cyclin-cdk Checkpoints for Cell Cycle Phases](#)
-  [345-2 Rb Gene Encodes Transcription Factor Regulatory Subunit](#)
-  [346-2 Regulating Cell Death - Apoptosis vs Necrosis](#)


Chapter 20: The Origins of Life


-  [347 What any Life Origins Scenario Must Explain](#)
-  [348 Early Ideas to Explain the Origins of Life](#)


-  [349 Life Origins in a Reducing Atmosphere?](#)
-  [350 Life Origins in a Non-Reducing Atmosphere?](#)
-  [351 Life Origins in a Thermal Vent](#)
-  [352 Phylogenetic Support for Autotrophs-First Origins of Life](#)
-  [353 AATE: An Autocatalytic, Self-Replicating Organic Molecule](#)
-  [354-2 Protected Molecular Communication-Semipermeable Membranes](#)
-  [355 Self-Replication: Information, Specificity, Communication, Coordination](#)
-  [356 Information Storage and Retrieval in an RNA World](#)
-  [357 The Transition from an RNA World to a DNA World](#)
-  [358 Origins and Evolution of Biochemical Pathways](#)


Reviews/Comments for, earlier editions/versions of
Cell and Molecular Biology
What We Know & How We Found Out


Reviews (Open Textbook Library, University of Minnesota)


2022: "No other text offers a broader understanding of this exciting science." Zhiming Liu, Professor of Biology, New Mexico University: 

2021: "Gerald Bergtrom's Basic Cell and Molecular Biology... textbook is a tour de force showcasing his passion for teaching... Bergtrom's periodic updates makes sure that the text stays accurate and relevant." Adriana LaGier, Assoc.Professor, Grand View University: 

2021: "This as comprehensive and up to date as any upper-level Cell Biology text I have used. I think is suitable for both higher-level college and introductory graduate school cell biology classes... It is excellent..." Philip Rock, Professor of Biology, Virginia Wesleyan University: 

2020: "The text is easy to read. It almost feels as if the instructor is talking to the reader. Topics are presented in a clear manner with the learning objectives in mind. Meltem Arikan, Adj. Faculty, Massachusetts Maritime Academy: 

2017: "...great introductory text that includes the basics of cell and molecular biology. Each chapter includes... video presentations which add to comprehension. There is a theme based on evolution utilized throughout each chapter with appropriate examples. Main points are reinforced with guided exercises for students." Kate Kenyon, Assoc. Professor, Umpqua Community College: 

2016: "(The book) is written in a narrative form with a somewhat casual tone which makes it easy to read and easy to follow... Learning objectives are outlined at the beginning of each chapter. Throughout the entire text evolution is a constant theme, providing context, rationale, and examples of its importance in the biological sciences". Brendan Mattingly, Acad. Prog. Associate, University of Kansas: 

Comments

2022: "Thank you for providing such a valuable open access text..." (Ali Azghani, Fulbright Scholar & Professor of Biology, Univ. of Texas at Tyler)

2020: "I appreciate the approach your textbook takes..., providing students with a relatable narrative on how things were discovered, rather than simply stating the facts." (Dr. Maria Vassileva, MVSc, PhD; Assoc. Prof. of Biological Science, Nagoya University)

2020: "I have used Campbell and Raven, Hillis, and am now considering the Openstax textbook and find yours so much superior. And then the additional resources like your short, recorded PowerPoints (are) amazing in their conciseness and clarity" (Crima Pogge, Instructor, City College of San Francisco)

2019: "I simply LOVE your book, your historical approach to the study of biology, your challenges for the students that stimulate critical thinking... (Daniela Fadda, Istituto di Istruzione Superiore Desantis Deledda [Linguistic H. S.], Sardinia Italy)

An OER hardcopy (print) version of the Annotated CMB5e iText (CMB5p) is available; search Bergtrom at <https://www.lulu.com/shop/>.