

December 2021

Protein-Ligand Binding as a Tool to Identify Antiviral Drugs.

Rajdeep Virdi
University of Wisconsin-Milwaukee

Follow this and additional works at: <https://dc.uwm.edu/etd>



Part of the [Biochemistry Commons](#)

Recommended Citation

Virdi, Rajdeep, "Protein-Ligand Binding as a Tool to Identify Antiviral Drugs." (2021). *Theses and Dissertations*. 2845.
<https://dc.uwm.edu/etd/2845>

This Dissertation is brought to you for free and open access by UWM Digital Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UWM Digital Commons. For more information, please contact scholarlycommunicationteam-group@uwm.edu.

PROTEIN-LIGAND BINDING AS A TOOL TO IDENTIFY ANTIVIRAL DRUGS

by

Rajdeep S. Viridi

A Dissertation Submitted in
Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy

in Chemistry

at

The University of Wisconsin-Milwaukee

December 2021

ABSTRACT

PROTEIN-LIGAND BINDING AS A TOOL TO IDENTIFY ANTIVIRAL DRUGS

by

Rajdeep S. Virdi

The University of Wisconsin- Milwaukee, 2021
Under the Supervision of Professor David N. Frick

Viruses are the most abundant biological entities on the planet. A virus is a collection of essential genetic material encapsulated in a protein coat that is incapable of replicating without a host. A virus must inject its genetic material into a suitable host cell to utilize host machinery to replicate. During the process of replication, a virus hijacks cellular functions, avoids or inhibits host antiviral defenses, and sometimes causes disease in the host organism. One effective way to fight viral infection is to identify molecules that inhibit the function of essential viral proteins. The studies described in this dissertation focus on identifying such molecules. Even if such molecules are not developed into drugs, they could be useful as molecular probes to study the biological role of these viral protein targets. The two targets studied here were helicases, which bind ATP and nucleic acids, and macrodomains, which bind ADP-ribose. The proteins were isolated from either the hepatitis C virus (HCV) or SARS-CoV-2, the virus that causes COVID-19.

To facilitate similar drug discovery screens with helicases, two assays were designed to monitor DNA and ATP binding to viral helicases. DNA binding was monitored with Förster resonance energy transfer (FRET), fluorescently labeled HCV helicase, and fluorescently labeled oligonucleotides. Factors effecting binding were examined, such as the length of the duplex, the

length of a single-stranded overhang, and whether the overhang had a 5' or 3' end. Less FRET was observed with longer length duplex, and longer length single-stranded overhang.

To monitor the interaction of ATP and a helicase, the steady state rates of ATP hydrolysis catalyzed by nsp13, a helicase from SARS-CoV-2, were measured. Based on the rates of helicase-catalyzed ATP hydrolysis, the catalytic rate constant, K_{cat} , was calculated. To monitor the binding of DNA (or RNA) to nsp13, rates of ATP hydrolysis were examined in the presence and absence of various oligonucleotides. Oligonucleotides with a length of 18–20 base pairs stimulated the helicase-catalyzed ATP hydrolysis by increasing rates of ATP hydrolysis, but poly(U) and other longer polynucleotides had no impact.

In the similar binding studies, Isothermal Titration Calorimetry (ITC) was used to show ADP-ribose binding to the macrodomain (Mac1) of SARS-CoV-2 was enthalpy (ΔH) driven. The interactions between Mac1 and nucleotides similar to ADP-ribose were also investigated with ITC. However, their interaction was not similar to that of ADP-ribose. The next step was to screen a library of 2,500 compounds in search of potential antiviral candidates or compounds that would bind to Mac1, blocking its interaction with ADP-ribose using Differential Scanning Fluorimetry (DSF). Compound effects were confirmed with DSF and ITC, and imatinib methane sulfonate was shown to bind Mac1 in a similar manner to that of ADP-ribose.

In this dissertation, multiple methods were employed to examine the interaction between protein and small molecule, protein-DNA, and protein-RNA. Even though not all methods and techniques employed resulted in identifying potential antivirals, these are the initial studies required to move forward before proceeding to identify potential antivirals.

Dedicated to

Dr. Zack Shana and all those who perceive potential in others, inspire and cultivate them so one day they can be their best selves.

TABLE OF CONTENTS

LIST OF FIGURES.....	VIII
LIST OF TABLES	VIII
LIST OF ABBREVIATIONS	X
ACKNOWLEDGEMENTS	XIII
1. INTRODUCTION.....	1
2. LITERATURE REVIEW: BIOLOGY	8
2.1. HEPATITIS C VIRUS (HCV)	8
2.1.1. <i>Hepatitis C Virus Genome</i>	10
2.1.2. <i>Hepatitis C Virus Structural Proteins</i>	11
2.1.3. <i>Hepatitis C Virus non-structural proteins</i>	12
2.1.4. <i>Hepatitis C Virus non-structural protein 3 helicase (NS3h)</i>	15
2.2. SEVERE ACUTE RESPIRATORY SYNDROMES CORONAVIRUS 2 (SARS-CoV-2).....	18
2.2.1. <i>Introduction to Coronavirus</i>	18
2.2.2. <i>Coronavirus infection and life cycle</i>	19
2.2.3. <i>Spike Protein initiating host infection</i>	19
2.2.4. <i>Translation of nonstructural proteins (NSP's) and structural Proteins (SP)</i>	20
2.2.5. <i>Replication and transcription complex (RTC)</i>	21
2.2.6. <i>Formation of DMVs and Assembly of Virion</i>	21
2.2.7. <i>SARS-CoV-2 non-structural proteins:</i>	22
2.3. MACRODOMAIN FROM MACROH2A.	34
2.3.1. <i>Structural differences between MacroH2A and canonical H2A</i>	34
2.3.2. <i>Structural changes in MacroH2A affecting protein-protein interaction</i>	35
2.3.3. <i>MacroH2A interaction with HDAC proteins</i>	36
2.3.4. <i>MacroH2A inhibits gene expression by masing regulatory region</i>	37
2.3.5. <i>MacroH2A interaction with Activating transcription factor 2 (AF2)</i>	39
2.3.6. <i>Macrodomains as standalone proteins</i>	40
3. LITERATURE REVIEW: CHEMISTRY.....	42
3.1. BACKGROUND THERMODYNAMICS	42
3.1.1. <i>Energy</i>	42
3.1.2. <i>The First Law of Thermodynamics:</i>	42
3.1.3. <i>The Second Law of Thermodynamics</i>	44
3.1.4. <i>Calorimetry</i>	47
3.2. ISOTHERMAL TITRATION CALORIMETRY (ITC)	48
3.2.1. <i>Example Analysis of data</i>	52
3.2.2. <i>The significance of Entropy and Enthalpy</i>	53
3.2.3. <i>The significance of Enthalpy in Drug Discovery</i>	56
3.3. COMPOUND OPTIMIZATION IN DRUG DISCOVERY	59
3.4. CHROMATOGRAPHY	64
3.4.1. <i>Introduction</i>	64
3.4.2. <i>Mobile and stationary phase</i>	64
3.4.3. <i>Separation of components</i>	66
3.4.4. <i>Calculations for Column efficiency</i>	68
3.4.5. <i>Factors affecting Column efficiency</i>	69
3.4.6. <i>Calculations for resolution from a chromatogram</i>	71

3.4.7. Ion Exchange Chromatography.....	72
3.4.8. Gel Filtration Chromatography	75
3.4.9. Gel Filtration Selection of stationary phase.	75
3.4.10. Gel Filtration Tips for better separation.....	76
3.5. FÖRSTER RESONANCE ENERGY TRANSFER (FRET).	77
3.5.1. FRET as a molecular ruler.....	77
3.5.2. FRET Efficiency.	79
3.5.3. Estimating complex concentration [AD].....	81
3.5.4. FRET chapter summary.	83
3.6. DIFFERENTIAL SCANNING FLUORIMETRY (DSF).....	85
3.6.1. Estimating melting temperature (T_m)	85
3.6.2. Estimation of EC_{50}	86
3.6.3. Estimating binding affinity (K_d)	87
3.7. SELECTED DRUGS FROM DSF HITS	89
3.7.1. Trifluoperazine (antipsychotic drug)	89
3.7.2. Irinotecan hydrochloride trihydrate (Topoisomerase 1 inhibitor)	90
3.7.3. Nifedipine (Calcium Channel Blocker)	92
3.7.4. Lactams: Cefaclor and Cefatrizine (Antibiotics)	93
3.7.5. Benzimidazoles: Omeprazole and Rabeprazole (Proton Pump Inhibitors).....	95
4. MATERIALS & METHODS.....	98
4.1. GENE SYNTHESIS.....	98
4.2. PROTEIN PURIFICATION (MAC1).....	98
4.3. PROTEIN PURIFICATION (Nsp13-HIS-TAG).....	99
4.4. ISOTHERMAL TITRATION CALORIMETRY (ITC)	100
4.5. DIFFERENTIAL SCANNING FLUORIMETRY (DSF)	100
4.6. MOLECULAR MODELING & VIRTUAL SCREENING.....	101
4.7. NUCLEOTIDE SEQUENCE OF PET28-COVID-Nsp13	102
4.8. AMINO ACID SEQUENCE OF Nsp13-HISTAG ENCODED BY PET28-COVID-Nsp13	103
4.9. NUCLEOTIDE SEQUENCE OF PET21-COVID-Nsp13	103
4.10. AMINO ACID SEQUENCE OF HISTAG-Nsp13 ENCODED BY PET21-COVID-Nsp13	104
4.11. MALACHITE GREEN ASSAY	104
4.11.1. Theory	104
4.11.2. Protocol.....	105
4.12. FRET OLIGONUCLEOTIDE SEQUENCE	106
5. USING FRET TO DETERMINE THE BINDING OF NS3H TO SSDNA AND DSDNA	107
5.1. INTRODUCTION.....	107
5.2. EXAMINING FACTORS THAT COULD INDIRECTLY AFFECT THE DONOR EMISSION.	108
5.3. ESTIMATING DISSOCIATION CONSTANT AND STOICHIOMETRY CONSTANT.....	113
5.4. EXAMINE NS3H BINDING WITH PARTIALLY DSDNA.	115
5.5. NS3H SEEMS TO BIND TO SSDNA IN A SPECIFIC ORIENTATION.	117
5.6. USING FRET TO EXAMINE THE EFFECT OF INCREASING THE OVERHANG ON PARTIAL DSDNA.....	120
5.7. USING FRET TO IDENTIFY THE EFFECT OF INCREASING THE DUPLEX SIZE	122
5.8. SUMMARY	124
6. VARIABLE NSP3 MAC1 DOMAIN OF SARS-COV-2 RETAINS THE ABILITY TO BIND ADP-RIBOSE	126
6.1. INTRODUCTION.....	126
6.2. PURIFICATION AND ISOLATION OF MAC1 DOMAIN	129
6.3. DETERMINE OPTIMAL CONDITIONS FOR ITC EXPERIMENTS.....	131
6.3.1. ATP binding to Mac1	133

6.3.2. <i>ATP-Glucose binding to Mac1</i>	134
6.3.3. <i>AMP binding to Mac1</i>	135
6.3.4. <i>Adenosine binding to Mac1</i>	136
6.3.5. <i>cAMP binding to Mac1</i>	137
6.3.6. <i>ADP binding to Mac1</i>	138
6.4. SUMMARY OF ITC DATA WITH NUCLEOTIDES.	139
6.5. OPTIMIZATION OF DSF ASSAY CONDITIONS.	141
6.6. DSF TO MONITOR THE BINDING OF ADP-RIBOSE WITH MAC1.	143
6.7. DSF TO ESTIMATE THE BINDING OF VARIOUS NUCLEOTIDES WITH MAC1.	145
6.8. SUMMARY	146
7. DISCOVERY OF DRUG-LIKE LIGANDS FOR THE MAC1 DOMAIN OF SARS-COV-2	148
7.1. INTRODUCTION	148
7.2. SCREENING OF DRUGS AND DRUGS-LIKE COMPOUNDS VIA DSF.....	149
7.3. ORTHOGONAL “VIRTUAL” SCREEN.	155
7.4. ITC WITH SELECTED COMPOUNDS.....	156
7.4.1. <i>*Nifedipine binding to Mac1</i>	157
7.4.2. <i>Trifluoperazine binding to Mac1</i>	158
7.4.3. <i>Irinotecan binding to Mac1</i>	159
7.4.4. <i>Imatinib methane sulfonate binding to Mac1</i>	160
7.4.5. <i>*Estradiol binding to Mac1</i>	161
7.4.6. <i>*Flunisolide binding to Mac1</i>	162
7.4.7. <i>*Telmisartan binding to Mac1</i>	163
7.4.8. <i>Rabeprazole binding to Mac1</i>	164
7.4.9. <i>Omeprazole binding to Mac1</i>	165
7.5. DISCUSSION	166
7.6. SUMMARY	168
8. SARS-COV-2 HELICASE.....	172
8.1. INTRODUCTION	172
8.2. PURIFICATION OF HIS-NSP13 AND NSP13-HIS.	173
8.3. ATPASE ASSAY TO EVALUATE HIS-NSP13 ACTIVITY.....	177
8.4. NSP13 IS STIMULATED IN THE PRESENCE OF 18-20 BASES LONG OLIGONUCLEOTIDES.....	179
8.5. NSP13 CHAPTER SUMMARY.....	180
9. REFERENCES.....	183
10. CURRICULUM VITAE.....	201





















LIST OF FIGURES






















FIGURE 3.2-1 EXAMPLE OF RAW DATA COLLECTED FROM ITC	49
FIGURE 3.2-2. EXAMPLE ANALYSIS OF ITC DATA.....	52
FIGURE 3.2-3. THE PROGRESSION OF HIV INHIBITORS' BINDING AFFINITY.	57
FIGURE 3.2-4 THE DEVELOPMENT OF HIV INHIBITORS IS MEASURED BY THERMODYNAMIC DATA.....	57
FIGURE 3.2-5 THE DEVELOPMENT OF STATINS IS MEASURED BY THERMODYNAMIC DATA.	58
FIGURE 3.3-1 A THERMODYNAMIC OPTIMIZATION PLOT.	59
FIGURE 3.3-2 AN EXAMPLE OF LIPE PLOT.....	63
FIGURE 3.4-1. A TYPICAL CHROMATOGRAM OF A MIXTURE CONTAINING TWO COMPONENTS.	66
FIGURE 3.7-1: PPI ACTIVATION MECHANISM: OMEPRAZOLE AND RABEPRAZOLE.	97
FIGURE 5.2-1 NS3H TITRATION WITH UNLABELED DNA, AND A TAMRA LABELED OLIGONUCLEOTIDE.	112
FIGURE 5.3-1. ESTIMATED DISSOCIATION CONSTANT AND STOICHIOMETRY COEFFICIENTS.....	114
FIGURE 5.4-1 TITRATION OF YFP LABELED NS3H WITH TAMRA-LABELED ON 3' OR 5' END OF PARTIAL DSDNA.....	117
FIGURE 5.5-1 YFP-NS3H TITRATION WITH 3' AND 5' TAM.....	119
FIGURE 5.6-1 TITRATION OF YFP-NS3H WITH VARIOUS LENGTH SSDNA OVERHANG ON PARTIAL DSDNA.....	121
FIGURE 5.7-1 TITRATION OF YFP-NS3H WITH VARIED DUPLEX SIZE ON PARTIAL DSDNA.....	123
FIGURE 6.1-1 SEQUENCE DIVERGENCE BETWEEN POTENTIAL DRUG TARGETS IN SARS-CoV AND SARS-CoV-2.	127
FIGURE 6.1-2 VARIATION IN THE MAC1 DOMAINS OF CORONAVIRUSES.	129
FIGURE 6.2-1 15% SDS PAGE OF SARS-CoV-2 MAC1 DOMAIN.	130
FIGURE 6.3-1 MAC1 TITRATION WITH ADP-RIBOSE.....	131
FIGURE 6.3-2 MAC1 TITRATION WITH ADP-RIBOSE.....	132
FIGURE 6.3-3 ATP TITRATION WITH MAC1 DOMAIN.....	133
FIGURE 6.3-4 MAC1 TITRATION WITH ATP-GLUCOSE.	134
FIGURE 6.3-5 MAC1 TITRATION WITH AMP.	135
FIGURE 6.3-6 MAC1 TITRATION WITH ADENOSINE.	136
FIGURE 6.3-7 MAC1 TITRATION WITH cAMP.....	137
FIGURE 6.3-8 MAC1 TITRATION WITH ADP.	138
FIGURE 6.4-1 THERMODYNAMIC DATA FOR ALL NUCLEOTIDES WERE OBTAINED USING ITC.	139
FIGURE 6.5-1 OPTIMIZATION OF DSF ASSAY CONDITIONS.	143
FIGURE 6.6-1 METHODS USED TO ESTIMATE THE BINDING AFFINITY OF MAC1 DOMAIN.	145
FIGURE 7.2-1 DSF SCREENS OF FDA-APPROVED DRUGS AND DRUG-LIKE COMPOUNDS FOR SARS-CoV-2 MAC1 LIGANDS.....	150
FIGURE 7.2-2 IDENTIFYING INTERFERING COMPOUNDS.	151
FIGURE 7.2-3 SELECTED HIT COMPOUNDS FROM THE NIH CLINICAL COLLECTION.	152
FIGURE 7.2-4 CONCENTRATION RESPONSE ASSAYS.....	154
FIGURE 7.3-1 COMPARISON OF DSF AND AUTODOCK SCREENING RESULTS.	155
FIGURE 7.4-1 MAC1 TITRATION WITH NIFEDIPINE.	157
FIGURE 7.4-2 MAC1 TITRATION WITH TRIFLUOPERAZINE.	158
FIGURE 7.4-3 MAC1 TITRATION WITH IRINOTECAN.	159
FIGURE 7.4-4 MAC1 TITRATION WITH IMATINIB METHANE SULFONATE.....	160
FIGURE 7.4-5 MAC1 TITRATION WITH ESTRADIOL.	161
FIGURE 7.4-6 MAC1 TITRATION WITH FLUNISOLIDE.	162
FIGURE 7.4-7 MAC1 TITRATION WITH TELMISARTAN.....	163
FIGURE 7.4-8 MAC1 TITRATION WITH RABEPRAZOLE.....	164
FIGURE 7.4-9 MAC1 TITRATION WITH OMEPRAZOLE.	165
FIGURE 7.5-1 REPRESENTATIVE STRUCTURES OBTAINED USING AUTODOCK VINA VIRTUAL SCREENING.	166
FIGURE 7.5-2 THERMODYNAMIC DATA OF SELECTED COMPOUNDS OBTAINED USING ITC.....	167
FIGURE 8.2-1 15% SDS GEL ON BL21(DE3) WHOLE CELL BEFORE AND AFTER INDUCTION.....	174
FIGURE 8.2-2. 15% SDS GEL SARS-CoV-2 NSP13.....	176
FIGURE 8.3-1 KINETIC ASSAYS WITH NSP13.....	178
FIGURE 8.4-1 NSP13 IS STIMULATED IN THE PRESENCE OF 18-20 BASES LONG OLIGONUCLEOTIDE.	180




















LIST OF TABLES

TABLE 3.1-1 ENTHALPY (ΔH), ENTROPY (ΔS) DICTATES REACTION SPONTANEITY.	46
TABLE 3.1-2 SUMMARY OF ΔG AND $\ln(K)$ BASED ON EQUATION 3.1-15.	47
TABLE 3.3-1 SUMMARY OF EACH QUADRANT OF THE THERMODYNAMIC OPTIMIZATION PLOT.....	61
TABLE 6.7-1 EFFECTS OF VARIOUS NUCLEOTIDES ON SARS-CoV-2 Mac1.....	146

LIST OF ABBREVIATIONS

	ORF	Open reading frames
	RTC	Replication and transcription complexes
	sgRNA	Subgenomic RNA
	TSRs	Transcription regulatory sequences
	DMV	Double membrane vesciles
	UTR	Untranslated region
	3CLpro	3C-like protease
	3Ecto	Nsp3 ectodomain
	ADPr	ADP-ribose
	ADRP	ADP-ribose-1''-phosphate phosphatase
	ARTD	ADP-ribosyltransferases diphtheria toxin-like
	β SM	betacoronavirus-specific marker
	CHIKV	Chikungunya virus
	CM	convoluted membrane
	DMV	double-membrane vesicle
	DPUP	Domain Preceding Ubl2 and PL2pro
	DUB	deubiquitinating
	ER	Endoplasmic Reticulum
	GST	glutathione S-transferase
	hISG15	human interferon-stimulated gene 15

 HCoV	human coronavirus
 HEV	hepatitis E virus
 HVR	hypervariable region
 IBV	infectious bronchitis virus
 IFN	interferon
 IL-6	Interleukin 6
 IRF	interferon regulatory factor
 mISG15	mouse interferon-stimulated gene 15
 Mpro	main protease
 Mac	macrodomain
 MARYlation	mono-ADP-ribosylation
 MDM2	mouse double minute 2 homolog
 MERS	Middle-East respiratory syndrome
 MHV	mouse hepatitis virus
 MKRN	makorin ring finger protein
 N	nucleocapsid
 NAB	nucleic-acid binding domain
 NF- κ B	nuclear factor kappa-light-chain-enhancer of activated B cells
 Nsp	non-structural protein
 NTD	N-terminal domain
 ORF	open reading frame

 PABP	poly(A)-binding protein
 PARP	poly(ADP-ribose) polymerase
 PARylation	poly- ADP-ribosylation
 RCHY1	RING finger and CHY zinc-finger domain-containing protein 1
 RID	Ras-interacting domain
 PLpro	papain-like protease
 R.M.S.D.	root-mean-square deviation
 RTC	replication/transcription complex
 SARS	severe acute respiratory syndrome
 SUD	SARS-unique domain
 TGEV	transmissible gastroenteritis virus
 TM	transmembrane
 TNF	tumor necrosis factor
 TRS	transcriptional regulatory sequence
 Ubl	ubiquitin-like
 Ub	ubiquitin
 USP	ubiquitin-specific protease
 UTR	untranslated region
 VEEV	Venezuelan equine encephalitis virus

ACKNOWLEDGEMENTS

I would like to express my gratitude to Prof. David Frick, my principal investigator, for allowing me to join his lab and for providing assistance and support during my degree program. Without his aid and support, this thesis and the published articles would not be possible. Working in this laboratory has honed my ability to think critically and independently, as well as equipped me to be a more effective scientist.

Second, I want to express my gratitude to Prof. Valerica Raicu for assisting me anytime I needed help with FRET studies, while also ensuring that I understood the theory. Similarly, I would like to thank Prof. Ionel Popa for his assistance with the magnetic tweezers assay. I really appreciate your time, materials, and knowledge.

Following that, I am appreciative of all the insightful input offered at committee sessions by all committee members. I believe that each member of my committee assisted me at some point throughout my degree program, and I would want to express my gratitude for the assistance. Listed alphabetically are Prof. Shama Mirza and Prof. Nicholas Silvaggi.

Lastly, I have not forgotten all the graduate students, post-docs, and research scientists who have provided their assistance. In alphabetical order: Dr. Gabriel Biener, Dr. Josie Corby, Dr. Narayan Dahal, Sabita Sharma, and Dr. Mike Stoneman. I am very grateful and want to express my sincere thanks to each person on this list, and to anybody else I may have forgotten. Thank you all. Furthermore, this research was supported in part by National Institutes of Health Grant R01 AI088001.

1. Introduction

The interaction of biological macromolecules with one another or with other small molecules to form a complex is the basis for all activities in living organisms. Proteins are an important class of biological macromolecules that perform their tasks by interacting with other molecules. These interactions may take place between identical proteins, i.e., homodimers, homotrimers, etc., between two separate proteins, i.e., heterodimers, or between proteins and a small molecule, referred to as a protein-ligand interaction. The interaction of protein with a variety of ligands such as DNA, RNA, and small molecules is investigated using Förster resonance energy transfer (FRET), isothermal titration calorimetry (ITC), and differential scanning fluorimetry (DSF), and the effect of this interaction on protein activity is also examined using an ATPase activity assay. The macrodomain from SARS-CoV-2 nsp3, the helicase from SARS-CoV-2 nsp13, and Hepatitis C Virus (HCV) NS3 were utilized to study their interactions with small molecules, DNA, and DNA/RNA, respectively.

The factors affecting the protein-DNA interaction between the helicase domain of hepatitis C virus (HCV) nonstructural protein 3 (NS3h) and DNA were examined. Such factors include the length of the duplex, the length of a single-stranded overhang, and whether the overhang had a 5' or 3' end. The study presented here is preliminary, but similar assays can be used to screen for compounds in search of antivirals that may prevent helicase function by blocking ATP binding site, helicase binding to the nucleic acid, or blocking conformational

changes required for helicase transition on nucleic acid, thereby disabling one of the crucial components required in HCV viral replication.

The HCV is a blood-borne pathogen that infects the liver. Its genome consists of 96,000 bases that encodes for 3000 amino acids (Moradpour & Penin, 2013). Upon infection, the whole genome is translated from a single open reading frame encoding for ten structural and non-structural proteins flanked by 5' and 3' untranslated regions. NS3 is the third non-structural protein co-translated with NS4A as part of a larger polyprotein but is then cleaved into two distinct mature proteins by the NS3 serine protease (Suzich et al., 1993a). NS3 is a 70 kDa, 630 aa multifunctional protein having an N-terminal serine protease and a C-terminal helicase. The protease found at NS3's N-terminal is 180 amino acids long, whereas the helicase located at NS3's C-terminal is 430 amino acids long (Suzich et al., 1993a). Despite being covalently bound as part of the same nonstructural protein, both helicase and serine protease can be expressed and isolated independently of each other. For the studies presented here, the helicase domain was expressed independently of the protease domain, and the yellow fluorescent protein was covalently attached to the N-terminal with an His-tag on the C-terminal (Corby et al., 2017).

The interaction between the helicase and DNA was monitored using FRET. FRET is a process in which energy is transferred from donor to acceptor. Since the non-radiative transfer of energy from donor to acceptor cannot be measured directly, FRET is estimated by comparing the magnitude of the donor and acceptor fluorescence. If there is an interaction, the donor's emission will decrease and the acceptor's emission will increase, producing a composite donor plus acceptor emission spectra. The extent of the interaction is estimated based on the value of

FRET efficiency. An experiment with a lower FRET efficiency conducts less FRET. Therefore, we assume that the helicase is bound farther from the acceptor, whereas in an experiment with the higher FRET efficiency value, we assume that the helicase is bound closer to the acceptor. The FRET efficiency is typically calculated by spectral unmixing of the composite donor and acceptor emission spectra. In the studies presented here, FRET efficiency is estimated based on donor quenching only. Therefore, approximations to the FRET theory were made to accommodate the absence of spectral unmixing. The details and explanation of the approximations are discussed in chapter 3.5.

Next, the interaction between a protein and a small molecule was investigated using Isothermal Titration Calorimetry (ITC). ITC is used to experimentally investigate the complete thermodynamic profiles consisting of the binding free energy, entropy, and enthalpy, and to accurately predict the binding affinity. In an ITC experiment, a complete thermodynamic profile is estimated by directly measuring the heat exchange during complex formation at a constant temperature, and it has established itself as the gold standard for identifying the forces involved in complex formation (Velázquez-Campoy et al., 2004).

The protein under investigation is a macrodomain from SARS-CoV-2's nsp3. SARS-CoV-2 is a virus that is responsible for the COVID-19 pandemic that has killed millions of patients worldwide. COVID-19 is a respiratory disease that affects the respiratory tract, including the lungs, causing symptoms such as breathing problems and loss of taste and smell. Upon infection, the SARS-CoV-2 viral genome encodes 16 non-structural proteins that are essential for viral survival and replication (Lal, 2010). Nsp3 is one of the 16 nonstructural proteins with three

macrodomains located towards the n-terminal, which are later processed into mature proteins by one of the viral proteases. To avoid confusion among three macrodomains, the first macrodomain located towards the n-terminal of nsp3 is Mac1.

Macrodomains in general are not specific to coronavirus, or even viruses; they are found in the majority of living organisms, including bacteria and humans. A macrodomain is a three-layered $\alpha/\beta/\alpha$ sandwich fold and it is not restricted to one function. Some macrodomains have been shown to have ADP-ribose-1''-phosphate phosphatase activity. It was also demonstrated that macrodomains from several positive sense ssRNA viruses are also capable of hydrolyzing single ADPr (de-MARylation) or multiple ADPr residues from proteins (de-PARylation). MARylation and PARylation are reversible posttranslational protein modifications usually utilized by host cells as a signal for various cellular processes, including the initiation of antiviral immune response by PARPs.

Previously, other labs have shown macrodomain from other coronaviruses bound to ADP-ribose (M.P. Egloff et al., 2006; Cho et al., 2016a). However, Mac1 from SARS-CoV-2 had the largest sequence variation compared to other coronaviruses. Our initial study on the Mac1 was decided based on large sequence variations with respect to its closest member in the coronavirus family. Despite the largest sequence variation, we confirmed ADP-ribose binding to Mac1 with a complete thermodynamic profile. To further investigate ADP-ribose binding, similar nucleotides were also investigated. According to our ITC data, only the binding of ADP-ribose was enthalpy-driven, whereas the binding of other nucleotides was significantly different. To support the thermodynamic differences in binding between ADP-ribose and other nucleotides. A crystal

structure of cAMP bound to mac1 was compared with a crystal structure of ADP-ribose bound to mac1 by Michalska et al. (2020). Upon comparison, the thermodynamic differences we measured with ITC reflected key differences in binding also seen in crystal structures.

After it was confirmed that the thermodynamic differences that we measured via ITC accurately reflect the binding differences, our next goal was to screen a library of 2500 compounds in search of drugs or drug-like ligands that can bind to Mac1, replacing ADP-ribose. While ITC is an excellent tool for examining the underlying interaction between a protein and a ligand, it cannot be used to screen a library of compounds since each test takes an inordinate amount of time. As a result, we employed differential scanning fluorimetry (DSF), on the premise that the protein is stabilized upon binding. The protein's stability is determined by assessing the temperature at which it unfolds, and the degree of unfolding was measured by the degree of fluorescence change in SYPRO orange dye (Boivin et al., 2013; Li & Zhang, 2021). The observed increase in fluorescence of SYPRO orange dye is directly correlated with the concentration of unfolded protein or hydrophobic groups exposed during protein unfolding.

After the binding of ADP-ribose to Mac1 was confirmed via ITC, we investigated if similar binding parameters could also be obtained via DSF. Fortunately, the binding parameters determined using DSF were equivalent to those determined via ITC. The next task was to screen small libraries of drugs and drug-like compounds using DSF. The screen makes use of the following libraries: the National Institutes of Health (NIH) clinical collection (726 compounds), the National Cancer Institute (NCI) mechanistic set (540 compounds) and Sigma-Aldrich's 1280 compound Library of Pharmacologically Active Compounds (LOPAC).

2546 compounds from three libraries were screened, with several compounds giving an impression of change in melting temperature. Other methods were also employed to eliminate false positives. These hits were carefully screened again and, if the results were verified, these compounds were bought fresh to be used with ITC. Imatinib ends up being one of the compounds with similar thermodynamic parameters as ADP-ribose, and it has been reported by other labs to have some antiviral efficacy against SARS-CoV-2 (Emadi et al., 2020.; Gasmi et al., 2020; Morales-Ortega et al., 2020). On the other hand, some laboratories have also concluded that Imatinib is not a very effective SARS-CoV-2 drug (Zhao et al., 2020).

Lastly, we evaluated the activity of the SARS-CoV-2 helicase in the presence and absence of DNA and RNA. To examine the helicase activity, the concentration of inorganic phosphate released upon ATP hydrolysis was measured against time using the malachite green assay. In this experiment, free inorganic phosphate reacts with ammonium molybdate to generate a phosphomolybdate complex. This complex subsequently interacts with three molecules of malachite green dye, resulting in a change in color at 630nm. The resulting change in absorbance at 630 nm is directly correlated to the concentration of inorganic phosphate in the solution.

Malachite green assay is an ATPase assay that was used to assess the activity of SARS-CoV-2 helicase. It is a 601 amino acid long multidomain protein. It is a helicase of superfamily 1 that unwinds RNA and DNA duplexes unidirectionally in the 5' to 3' direction, and the translocation is powered by dNTP hydrolysis (Mickolajczyk et al., 2020; Weber & McCullagh, 2021). After it was confirmed that nsp13 is active, DNA and RNA of various sizes were added to the sample to investigate if oligonucleotides can be used to stimulate helicase activity. Based on

the data, 18-20 base pair long oligonucleotides tend to stimulate the helicase by increasing its activity to hydrolyze ATP, whereas poly(U) and other polynucleotides longer in length had no effect.

In this dissertation, Förster resonance energy transfer (FRET), isothermal titration calorimetry (ITC), and differential scanning fluorimetry (DSF) were utilized to study protein-DNA, protein-RNA, and protein-ligand interactions with macrodomain and helicase from SARS-CoV-2, and helicase from Hepatitis C Virus (HCV), and the effect of this interaction on protein activity was examined using an ATPase activity assay. Even though not all methods and techniques employed resulted in identifying potential antivirals, these are the initial studies required to move forward before proceeding to identify potential antivirals.

2. Literature Review: Biology

2.1. Hepatitis C Virus (HCV)

The hepatitis C virus is a blood-borne pathogen that infects the liver. The viral particle's surface is composed of viral envelope glycoproteins, and within glycoprotein is a nucleocapsid composed of core proteins that interacts with the viral RNA genome. It is transmitted via blood and it is transported to the liver via the bloodstream. The virus circulates in the bloodstream as a component of very-low-density lipoproteins and low-density lipoproteins. Once the viral particle reaches the liver, its entry into the hepatocytes requires binding of viral glycoproteins E1 and E2 with host scavenger receptor SCARB1 and the tetraspanin protein CD81. After the viral particle is bound to the entry factors, it further interacts with claudin 1 (CLDN1) and occludin before gaining entry into the hepatocytes via clathrin-mediated endocytosis (Miao et al., 2017).

Upon entry, a clathrin-coated endosomal membrane vesicle containing a viral particle is transported near the endoplasmic reticulum with the assistance of the motor protein called dynein. Acidification of the clathrin-coated endosomal membrane causes a conformational change leading to the fusion of viral nucleocapsid glycoprotein and endosomal membranes, resulting in uncoating and release of the viral genome into the cytosol. The positive sense viral RNA in the cytosol serves as a template for RNA replication. The translation of the viral genome is initiated in a cap independent fashion upon the assembly of the host ribosome on the viral genome internal ribosomal entry site located at the 5'untranslated region (J.-C. Cheng et al., 1999) (Yi & Lemon, 2003). A signal sequence at the start of the translated polyprotein enables

the ribosome to recognize and attach to the translocon on the endoplasmic reticulum membrane (Beales et al., 2001; Suzich et al., 1993a).

The rest of the viral genome is translated as a polyprotein in the host ER membrane. The resulting polypeptide is about 3,000 amino acids long and consists of both structural and non-structural proteins. NS3, NS4A, NS4B, NS5A, and NS5B are non-structural proteins that are necessary for viral genome replication. Structural proteins include the core protein and the envelope glycoproteins E1 and E2, which are integrated into the viral particle, as well as support proteins p7 and NS2, which aid in viral particle synthesis but are not incorporated into the viral particle. After translation, the polyprotein undergoes proteolytic processing via viral proteases releasing 10 mature proteins inside the host ER membrane (Zeisel et al., 2013).

The accumulation of viral proteins on the ER membrane induces a rearrangement of the membrane, resulting in the formation of double membrane vesicles (DMV). These vesicles cluster together to form the membranous web, with each vesicle containing HCV non-structural proteins, HCV RNA, and lipid droplets. The membranous web of vesicles results in a protected environment for viral replication (Gosert et al., 2002).

The HCV genome is replicated inside DMV, initiated by RNA-dependent RNA polymerase (NS5B) as part of the replicase complex. During RNA replication a negative-strand intermediate is generated which in turn serves as a template for the synthesis of multiple progeny positive sense viral RNA. After the viral genome is replicated, it is used for translation, or be packaged into a new viral particle (Stoeck et al., 2018).

The viral particle assembly initially takes place at the host ER membrane with the help of a lipid droplet. Synthesis of a new viral particle is initiated with the aggregation of the core proteins and their interaction with the lipid droplet containing E1, E2 P7, and NS2, copies of viral RNA at the host ER membrane. After the HCV viral particle is assembled, it is transported to the Golgi apparatus to be packaged into an endosome and transported outside of the cell (Moradpour & Penin, 2013).

2.1.1. Hepatitis C Virus Genome

HCV is a member of the genus hepacivirus, which is part of the *Flaviviridae* family. As is the case with all members of the *Flaviviridae* family, HCV is an enveloped virus with surface-anchored envelope proteins. The virus's surface is composed of a lipid bilayer surrounding the core protein and viral RNA genome in a nucleocapsid. The HCV genome is 9.6 thousand nucleotides long which encodes for 3000 amino acids. The whole genome is translated from single open reading frame encoding for ten structural and non-structural proteins flanked by 5' and 3' untranslated regions (Moradpour & Penin, 2013). The untranslated region is critical for viral infection as it plays crucial role in ribosome binding, translation, and replication (Zeisel et al., 2013). The structural proteins involved in encapsulating the virus and identifying a host include the core protein and the envelope glycoproteins E1 and E2. NS3, NS4A, NS4B, NS5A, and NS5B are non-structural proteins that are essential for viral replication or play a role in the viral life cycle. The viral genome and its encoded proteins are explored in greater depth in the sections that follow, beginning with the 5' and 3' untranslated regions (Moradpour & Penin, 2013).

The 5'UTR of HCV contains four highly structured domains. Together with the core coding regions, they fold into a complex secondary structure. The complex secondary structure forms an internal ribosome entry site (IRES) that mediates the binding of ribosomes in the host cell cytosol (Honda et al., 1996). The HCV internal ribosome entry site (IRES) is able to bind directly to the 40s ribosomal subunit without needing translation initiation factors, subsequently initiating the translation of the HCV genome. The HCV 3'UTR is approximately 200 nucleotides long and varies in length. HCV 3'UTR is divided into three regions: a 30-40nt long variable region, Poly (U/UC) region and 3'X-tail. The 3' untranslated region mediates the replication of the viral genome by interacting with NS5B RNA-dependent RNA polymerase (RdRp) (Cheng et al., 1999). The 3'X tail and poly (U/UC) region of the 3' end are essential for viral replication, whereas the rest of the 3'-UTR appears to modulate the activity of viral replication (Yi & Lemon, 2003).

2.1.2. Hepatitis C Virus Structural Proteins.

The core protein is the first structural protein translated from a viral open reading frame. The mature core protein is roughly 177 amino acids in length and has a molecular mass of 21 kilodaltons. It is a membrane protein with a dimetric alpha-helical structure maintained by a disulfide bond (Boulant et al., 2005). The core protein is made of two domains: domain 1 has a high concentration of basic residues, while domain 2 is very hydrophobic. Domain 1 interacts with various cellular factors to facilitate viral particle structural remodeling and packaging (Chassey et al., 2008). The more hydrophobic domain 2 facilitates the interaction with lipid droplets at the phospholipid interface (Boulant et al., 2005).

E1 and E2 envelop glycoproteins are required for viral infection because they recognize the host and facilitate viral entry. Both E1 and E2 are transmembrane proteins containing an ectodomain at the n-terminus and a 30 amino acid-long transmembrane domain at the c-terminus. E1 measures roughly 160aa in length, while E2 measures approximately 360aa in length. On the basis of disulfide mapping, a model of the E2 ectodomain's structure was presented, implying three domains (Krey et al., 2010). Domain 1 has eight beta strands that include the bulk of known CD81 interaction determinants. Domain 2 is a fusion loop between domains 1 and 3. Domain 3 is the CD81-binding domain that is also connected to the transmembrane domain through a flexible stem loop at the C-terminal. Both E1 and E2 are required for viral infection and for binding specifically to the entry host receptors CD81 and SCARB1, as well as for interacting with occluding and claudin-1 (Miao et al., 2017).

P7 is a homo-oligomer trans membrane protein with an ion channel-like structure. According to its cation channel activity, it is believed to be a member of the viroporin family. P7 is a transmembrane protein composed of two alpha helices joined by a positively charged cytosolic loop with both N and C terminals towards the ER lumen (Carrère-Kremer et al., 2002). Although the specific function of p7 is unknown, it is likely to be an ion channel that enables proton movement inside the virion during virion assembly (Voisset & Dubuisson, 2004).

2.1.3. Hepatitis C Virus non-structural proteins.

NS2 is the first non-structural protein translated after structural proteins, and it is the only non-structural protein not necessary for viral genome replication and translation. NS2 is a cysteine protease that acts soon upon translation to cleave the NS2/NS3 junction. Cleavage at

the NS2/NS3 junction is required for fully functional NS3 protein to be produced. Additionally, in vitro studies suggests that the N-terminal of NS3 significantly enhances the catalytic activity of NS2 (Schregel et al., 2009).

NS3 and NS4A are co-translated as part of a larger polyprotein but are then cleaved into two distinct mature proteins by the NS3 serine protease. After separating NS3 and NS4A, they create the noncovalent complex NS3-NS4A, with NS4A acting as a cofactor for NS3 protease. NS3 is a 70 kDa, 630 amino acid multifunctional protein having an N-terminal serine protease and a C-terminal helicase (Raney et al., 2010). The protease found at NS3's N-terminal is 180 amino acids long, whereas the helicase located at the NS3's C-terminal is 430 amino acids long. Despite being covalently bound as part of the same nonstructural protein, both helicase and serine protease can be expressed, and isolated independently of each other. Additionally, in-vitro studies indicate that the presence of protease does affect helicase's activity (Frick et al., 2004; Kuang et al., 2004).

After viral genome translation in the ER, NS4A with NS3 protease is responsible for cleaving the remainder of the polyprotein. The NS4A cofactor for ns3 protease modulates the protease activity by directing the substrate towards its active site (Love et al., 1998). Due to the fact that protease at the n-terminus influences the activity of the C-terminal helicase. It is stated that NS4a, which binds to the protease, also stimulates the helicase activity as well. The NS3 helicase is a member of the DExH/D-box helicase superfamily (Raney et al., 2013). With the aid of ATP hydrolysis, the NS3 helicase separates double-stranded RNA (Singleton et al., 2007). Although the NS3 helicase is required and plays a critical role as a component of the replication

machinery, its specific function throughout the viral life cycle remains unknown. The helicase and its function will be explored in more detail in the next part of this chapter.

NS4B is a poorly characterized integral membrane protein with a molecular mass of 27kDa and a sequence length of 261 amino acids. Structurally, NS4B is separated into three sections: the N-terminus, which contains 70 amino acids, the middle portion, which contains 70-190 amino acids, and the C-terminus, which contains 191-261 amino acids (Lundin et al., 2003). A portion of the N-terminal is specifically associated with the ability to traverse the membrane bilayer, most likely following oligomerization, and it interacts with the replication complex, while the remainder of the protein, including the N-terminal, is associated with transmembrane domains. While NS4B is the primary inducer of membranous web or double membrane vesicles, it is not NS4b alone that participates in the formation of the membranous web it is assisted by the other viral nonstructural proteins. NS4B initiates the formation of the membranous web, whereas other nonstructural proteins contribute to its formation (Lundin et al., 2003).

NS5A is a membrane-associated multifunctional RNA-binding phosphoprotein that lacks any detectable enzymatic activity. NS5A is composed of three domains and an amphipathic alpha-helix membrane anchor at the N-terminus (Lundin et al., 2006). NS5A is a 447-amino-acid-long protein composed of a 36-amino-acid-long membrane anchor, a 36-213-amino-acid-long domain 1, a 250-342-amino-acid-long domain 2, and a 356-447-amino-acid-long domain 3. Domain 1 is important in viral replication and interacts with the lipid droplet, domain 2 is involved in viral replication, and domain 3 is required for viral assembly through interaction with the core protein (Masaki et al., 2008). The central and C-terminal regions of the HCV NS5A are

phosphorylated, indicating that the NS5A influences HCV replication, potentially via modulating the interaction with replication-specific host factors, while the C-terminal region interacts with assembly-specific host factors (Appel et al., 2005; Evans et al., 2004).

NS5B is a 591 amino acid viral RNA-dependent RNA polymerase with a molecular mass of 68 kDa. A 40-amino acid linker between the catalytic domain and the C-terminal membrane anchor, NS5B mimics the traditional finger, palm, and thumb subdomain arrangement of a right hand (Simister et al., 2009). The three-dimensional structural model indicates extensive contact between the finger and thumb subdomains, resulting in an encircling active site that is restricted by the linker through interaction with the beta-flap of the thumb domain. The replication of the viral genome utilizes positive sense RNA as a template to synthesize the complementary negative strand in a primer-dependent manner. Upon binding of positive sense RNA, initiation of viral replication directly involves the linker in the formation of first dinucleotide primer. Transition from initiation to elongation phase requires the rearrangement and removal of linker and beta flap blocking the exit. The rearrangement and RNA synthesis requires high concentration of GTP binding to the allosteric site on NS5B contributing to the switch from initiation to elongation phase. The rearrangement and significant conformational changes allow for newly formed dsRNA to exit (Harrus et al., 2010).

2.1.4. Hepatitis C Virus non-structural protein 3 helicase (NS3h).

Non-Structural Protein 3 is a critical component in HCV replication (NS3). The helicase domain of the NS3 protein is referred to as NS3h. Helicases are recognized as motor proteins that utilize the energy generated by nucleotide hydrolysis to catalyze the separation of double

stranded nucleic acids. Helicases are engaged in a wide variety of biological activities involving nucleic acids and are critical in the metabolism of DNA and RNA, including genome replication, transcription, translation, and DNA repair (Gorbalenya et al., 1989).

Helicases are divided into six superfamilies based on their sequence and structure. Helicases from superfamily 1 and 2 do not form ring type structures, whereas helicases from superfamilies 3-6 need to be assemble and oligomerize into a ring like structures (Singleton et al., 2007). Our focus is on helicase from superfamily 2 (HCV) superfamily 2 (SARS-CoV-2). Superfamily 2 is the largest and the most diverse of the helicase superfamilies and it is divided further based on sequence homology. HCV NS3h is a member of the SF2 helicase with conserved RecA-like folds. The RecA-like fold is common among proteins that have ATPase activity and utilize energy from ATP hydrolysis to translocate along the nucleic acid backbone (Suzich et al., 1993b).

Structurally, NS3h folds into three subdomains. When viewed as y-shaped molecule, subdomains 1 and 2 of NS3h are on top, and consist of RecA-like folds common among SF1 and SF2 helicases with an ATP binding site in the middle (Singleton et al., 2007). Subdomain 3 contains residues that make key contacts with the nucleic acid. Upon binding, the single stranded nucleic acid binds in the center of all three subdomains. The 5' end of the oligonucleotide sits at the interface between subdomains 2 and 3, while the 3' end sits at the interface between subdomains 1 and 3. The interaction between the subdomains and the oligonucleotide is primarily through the nucleotide backbone, with no interaction with the 2' hydroxyl group when the RNA is bound, thereby explaining the non-specific binding with both DNA and RNA (Frick,

2006). However, the mechanism by which NS3h translocates and the mechanism of strand separation is still unclear (Kim et al., 1998).

One theory for the mechanism of NS3h translocation is that the helicase harnesses the energy generated by ATP hydrolysis to exert force directly on the base pairs at the ss/ds junction, disrupting the hydrogen bond between the two strands (Amaratunga & Lohman, 1993). This model was first presented as the inchworm model, based on the conformational flexibility of subdomain 2 in comparison to subdomains 1 and 3. According to this model, NS3h subdomains 1 and 3 bind to nucleic acid with a lesser affinity than subdomain 2 (Kim et al., 1998). However, the rearrangement mediated by ATP binding modulates the binding affinity between subdomain 2 and subdomain 1-3 (Xie, 2016). When the oligonucleotide attaches in the middle of three subdomains in the absence of ATP, subdomain 2 binds more strongly to the nucleotide than subdomains 1 and 3. When ATP binds, it triggers a structural rearrangement between subdomains 1 and 3, resulting in pinching onto the nucleic acid and a shift along the nucleic acid from 3' to 5' direction. Following hydrolysis, the cleft between subdomains 1 and 3 reopens, while oligonucleotide-bound subdomain 2 turns, creating a shift in the 3' to 5' direction, culminating in unwinding or separation at the ss/ds junction (Raney et al., 2013).

2.2. Severe Acute Respiratory Syndromes Coronavirus 2 (SARS-CoV-2)

2.2.1. Introduction to Coronavirus

Coronaviruses belong to the Coronaviridea family. Coronaviridea includes viruses with a large RNA genome consisting of 25,000–32,000 nucleotides in an enveloped nucleocapsid (Artika et al., 2020a; Payne, 2017). Coronaviridea is divided into two subfamilies called coronavirinae and torovirinae, which are distinguished by their nucleocapsid shape. Coronavirinae are roughly spherical, while torovirinae are more donut-shaped (Payne, 2017). The nucleocapsid of these viruses is covered by a phospholipid bilayer with multiple glycoproteins embedded in the membrane.

Coronaviruses get their names from the spike protein found in their membrane that extends outwards and give them the appearance of a crown when seen under an electron microscope. Coronavirinae subfamily is divided into four genera: alpha, beta, gamma, and delta coronavirus (Woo et al., 2012). Bats and birds were found to be common hosts to all four genera of coronavirus, but only alpha and beta coronaviruses can infect humans (Woo et al., 2010).

For a long time, the alpha coronaviruses have been circulating in the human population, inducing mild respiratory diseases such as the common cold (Paules et al., 2020). However, outbreaks, epidemics, and pandemics of severe acute respiratory syndromes (SARS) and Middle East respiratory syndrome (MERS) have been caused by viruses belonging to the beta genus of the Coronaviridae family. These occurred in 2002, 2014, and 2019, and were caused by severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), and SARS-CoV-2, the causative agent of COVID-19.

2.2.2. Coronavirus infection and life cycle

2.2.3. Spike Protein initiating host infection

The CoV virion contains a 30 kilobase (kb) positive sense RNA genome. The RNA is coated with a nucleocapsid protein (N) and forms a helical pattern. The nucleocapsid is encased in a phospholipid bilayer membrane envelope that contains spike (S), membrane (M), and envelope (E) proteins (Siu et al., 2008). Spike is a trimeric protein (Delmas & Laude, 1990) that recognizes, binds to, and fuses with host cells. The spike protein's upper portion (furthest from the envelope) is made up of a globular receptor binding domain that recognizes and binds to the host ACE2 protein. Spike's lower section (nearest the envelope) is a fusion domain with a hydrophobic peptide in the center that assists in the entry process (Cai et al., 2020).

The spike protein of SARS-CoV and SARS-CoV-2 binds to the host cell's angiotensin converting enzyme 2 (ACE2), which acts as the viral receptor first. Following that, a cellular protein known as transmembrane serine protease 2 (TMPRSS2) is recruited. TMPRSS2 cleaves the connection between the receptor binding domain and the fusion domain. Following the exposure of the fusion domain, TMPRSS2 cleaves yet another site on the fusion domain, resulting in the expansion of the hydrophobic fusion peptide (Glowacka et al., 2011; Hoffmann et al., 2020). The hydrophobic fusion protein extends and penetrates the host cell membrane before retracting and fusing the two membranes. The viral genome is then deposited into the host cell's cytoplasm after the two membranes fuse (Lal, 2010).

2.2.4. Translation of nonstructural proteins (NSP's) and structural Proteins (SP).

After entry and uncoating, the viral genome is translated and replicated. Replication of the viral genome requires an RNA-dependent RNA Polymerase (RdRp), which the host cell lacks, so the viral genome must first be translated. The viral genome is a capped, positive-sense RNA. It is recognized by the host cell machinery and is immediately translated (Artika et al., 2020; Lal, 2010).

The SARS-CoV-2 genome possesses ~14 open reading frames encoding 27 proteins, but not all 27 proteins are translated directly from the genomic RNA. Only nonstructural proteins (nsps), which are essential for viral survival, are translated directly from genomic RNA, starting with ORF1a. Downstream of ORF1a and before its stop codon is a sequence with a complex secondary structure in which the host ribosome tends to slip 20%–25% of the time, causing one nucleotide frameshift. ORF1a encodes a single large polyprotein that is processed into 11 non-structural proteins. If the ribosome slips, a much longer RNA is translated, encoding a polyprotein processed into 16 nsps (Irigoyen et al., 2016; Plant et al., 2013). The large polyproteins are cleaved by viral proteases, releasing individual proteins (Ziebuhr et al., 2000).

Genes that encode structural proteins are present towards the 3' end of the viral genome. These genes are first transcribed into (-) RNA, which in turn is transcribed into mRNAs. One sub-genomic mRNA encodes only for each structural protein (Sola et al., 2015).

To summarize, the 5' end of the viral genome encodes for nonstructural proteins. It is translated into two large polyproteins, which are cleaved into individual nonstructural proteins by viral proteases. The 3' end of the genome is first transcribed into a smaller sub-genomic

negative sense RNA, which in turn are transcribed into sub-genomic mRNAs that are translated into various, mainly structural, proteins (Sola et al., 2015).

2.2.5. Replication and transcription complex (RTC)

A subset of proteins released from ORF1a and ORF1b form the replication and transcription complexes (RTC). The core of RTC is composed of RNA dependent RNA Polymerase (RdRp) called nsp12. Nsp12 shows minimal activity on its own, however. Its activity is greatly stimulated in the presence of nsp7 and nsp8 (Subissi et al., 2014) Nsp12 is bound to Nsp8 at the N-terminal and a second Nsp8 heterodimerized with nsp7 at the opposite end. The two nsp8 proteins bind in different structural conformations based on the location of binding sites on nsp12 (Kirchdoerfer & Ward, 2019). In addition to the RdRp core complex, replication and transcription complexes also include a helicase (nsp13), single-stranded binding protein (nsp9), and exonuclease (nsp14), as well as other cofactors and capping enzymes (Hartenian et al., 2020). The replication and transcription complexes are anchored in membranes known as convoluted membranes. These membranes are generated by reshaping the host endoplasmic reticulum in a process mediated by viral nsp3, nsp4, and nsp6 (Hagemeijer et al., 2014).

2.2.6. Formation of DMVs and assembly of virion

Nsp3, nsp4, and nsp6 remodel the host ER to create convoluted membranes and double membrane vesicles (DMVs) (Ulasli et al., 2010). These membranes and vesicles surround the viral genome, replication and transcription complex, and other proteins required for viral replication, protecting them from host pathogen sensors (Snijder et al., 2006). It was recently discovered that viral transcription occurs inside DMVs, and a new viral genome was visualized using radio labeling

and electron microscopy radiography in SARS-CoV-2 (Snijder et al., 2020). After the viral genome is fully replicated, it is then transported out of the DMVs. However, the transport mechanism is not yet identified. The nucleocapsid-coated positive sense genomic RNA buds into the endoplasmic-reticulum–Golgi intermediate compartment, an organelle in eukaryotic cells, which is rooted in structural proteins S, E, and M. The enveloped virion is then exported from the cell through an exocytic pathway (Snijder et al., 2020).

2.2.7. SARS-CoV-2 non-structural proteins:

2.2.7.1. Nsp1

SARS-CoV-2 non-structural protein 1 (nsp1) is the first protein cleaved from the Orf1a/b polyprotein that is translated upon host entry. Schubert *et al.* demonstrated that nsp1 might inhibit host gene expression by physically blocking the entry site to the host ribosome and by degrading host mRNA (Schubert et al., 2020). They used cryo-microscopy to show that nsp1 binds to the human ribosome, effectively inhibiting cellular mRNA translation by blocking the mRNA entry point (Schubert et al., 2020). The C terminal area of nsp1 appears to be the central domain that blocks and interacts with the ribosome entry point. Their findings showed that nsp1 is an important inhibitor of the host ribosome. As a result, reducing the availability of host ribosomes and increasing the concentration of viral RNA explains why infected cells are more likely to translate viral proteins (Artika et al., 2020b).

2.2.7.2. Nsp2

Graham *et al.* (2005) investigated the function of nsp2 by testing whether nsp2 is needed for viral replication. They modified the mouse hepatitis virus (MHV) and SARS-CoV nsp2 by

removing the nsp2 coding region. The mutated virus lacks not only nsp2 expression but also the nsp2-nsp3 precursor. It was hypothesized that the nsp2-3 precursor might be needed for replication, perhaps performing distinct roles, or regulating the availability of mature nsps 2 and 3. Further studies done by Graham *et al.* (2006) using peak titers showed that the nsp2 coding sequences of MHV and SARS-CoV are not required for viral replication in cell culture, but deletion of the nsp2 coding gene limits viral growth (Graham & Denison, 2006).

Next, nsp1-nsp2 and nsp2-nsp3 cleavage sites were investigated also with deletion mutants. Deleting nsp1-nsp2 cleavage site decreased viral replication by 10-fold, whereas deleting nsp2-nsp3 cleavage site delayed the initial increase in viral replication but was able to recover compared to wildtype (Graham & Denison, 2006). These findings suggest that RNA encoding the nsp2 protein and the cleavage sites nsp 1-2 and nsp 2-3 is not essential for viral replication in cell culture (Graham et al., 2008).

2.2.7.3.Nsp3

Nsp3 is the 213-kDa multifunctional polypeptide encoded by SARS-CoV genome. It is encoded as part of a large polyprotein from Orf1a or Orf1b. It is cleaved into separate protein by a papain-like protease which is also part of nsp3 (Ziebuhr et al., 2000). Nsp3 interacts with multiple host and viral proteins. It is essential in the formation of viral replication and transcriptional complexes (van Hemert et al., 2008) and helps with the formation of double membrane complex (Angelini et al., 2013) Nsp3 is a multifunctional polyprotein that consists of: Ubiquitin-like domain (UBL1), Glu-rich acidic region or (HVR), PL1Pro, Mac1 (X domain), Mac2 (SUD-N), Mac3 (SUD-M), DPUP (SUD-C), Ubl2-PL2pro (Artika et al., 2020b).

2.2.7.3.1. Ubiquitin-like domain 1 with glutamic acid rich domain.

UBL1 is the first domain located at nsp3's N-terminus. Ubl1's core is composed of residues 20–108, which form a typical ubiquitin-like structure. Following Ubl1 is a highly flexible glutamic acid-rich domain spanning residues 113–183, which is often referred to as a hypervariable area (HVR). It is well documented that the UBL1 domain is a multifunctional protein that facilitates viral genome transmission to RTC but also plays a role in suppressing the host's innate immune response.

Serrano et al. (2009) used nuclear magnetic resonance (NMR) to identify chemical shift perturbations in a sample containing both UBL1 and ssRNA. The existence of RNA caused a chemical change perturbation, which may mean an association between Ubl1 and RNA. Hurst et al. also found an association between UB1 and N-protein however deletion of UBL1 domain inhibited MHV replication. According to (Hurst et al., 2013) the association of UBL1 with N protein coated genome enables the passage of viral genome coated in N-protein to RTC.

The viral Ubl1 domain resembles the human ubiquitin domain in its usual ubiquitin-like fold. The human ubiquitin domain is essential in innate antiviral immunity (Heaton et al., 2016). It is possible for viral Ubl1 to imitate the action of the human Ubiquitin domain and modulate host immune response, causing host antiviral signaling to be disrupted (Lei et al., 2018).

2.2.7.3.2. Papain-like protease domains 1 and 2.

Papain-like-protease is an enzyme that is translated as part of nsp3. It is responsible for cleaving a long polypeptide chain and releasing nonstructural proteins such as itself. (Barretto et al., 2005). PL1pro and PL2pro are two proteases that are commonly found in coronaviruses. They

have differing substrate specificities and are used to cleave a section of nonstructural proteins. However, PL1pro is missing from coronaviruses belonging to the beta genera. In the alpha-CoV transmissible gastroenteritis virus, both PL1pro and PL2pro are responsible for cleaving and releasing different nonstructural proteins with varying substrate specificity but this is not true for SARS-CoV (Lei et al., 2018)

2.2.7.3.3. Macrodomain 1 (Mac1, X domain)

Macrodomains are domains found in the majority of living organisms, including viruses, bacteria, and humans. Due to its larger size than histone 2A, it was first recognized as a component of histone 2A and was termed macroH2A. Later, it was recognized that this was a domain present in other proteins, and the term macrodomain was established. Macrodomains are present not only as standalone proteins, but also as modules inside multidomain proteins such as PARPs (Chakravarthy et al., 2005a)

A macrodomain is a three-layered $\alpha/\beta/\alpha$ sandwich fold and it is not restricted to one function. Some macrodomains have been shown to have ADP-ribose-1"-phosphate phosphatase activity (Saikatendu et al., 2005). More recently, it was demonstrated that macrodomains from several (+) ssRNA viruses are also capable of hydrolyzing single ADPr (de-MARylation) or multiple ADPr residues from proteins (de-PARylation) (Fehr et al., 2016). MARylation and PARylation are reversible posttranslational protein modifications usually utilized by host cells as a signal for various cellular processes including initiation of antiviral immune response by PARPs (Liu & Yu, 2015).

Fehr *et al.* (2016) reported that the SARS-CoV wild-type Mac1 domain prevents the production of innate-immunity genes *in vitro*, thus inhibiting the host immune response (Fehr *et al.*, 2016). It is also possible for macrodomains to only bind to the ADPr already bound to acceptor protein blocking an essential signal. More recently, SARS-CoV-2 Mac1 domain has been shown to bind to ADPr via various binding techniques (Frick *et al.*, 2020). Based on the capacity of the viral macrodomain to simply bind to ADPr or to catalyze MARYlation, PARylation, and association with PARPS, it has been proposed that the viral macrodomain modulate the host cellular immune response (Fehr *et al.*, 2016; Fehr & Perlman, 2015).

2.2.7.3.4. Macrodomains II and III, and the DPUP (SUD-N, SUD-M, SUD-C)

Initially when the SARS CoV genome was sequenced, there were three domains suspected to be unique to the SARS virus thus they were named “SARS unique domain” (SUD-N, SUD-M, SUD-C) (Snijder *et al.*, 2003a). SUD-N, SUD-M, and SUD-C were the names given to the three subdomains, which represent the N-terminal, middle, and C-terminal regions of SUD, respectively. Later it was discovered that SUD-N and SUD-M was also present in other genera of coronavirus. Thus, Cheng *et al.* recommended to rename these three domains to macrodomain 2 (Mac2), macrodomain 3 (Mac3) and domain preceding Ubl2 and PL2pro (DPUP) (W. Cheng *et al.*, 2015).

Although Mac2 and Mac3 have the same fold as Mac1, the residues that would contact ADPr are absent, so Mac2 and Mac3 lack the ADPr-binding functionality of Mac1 (Tan *et al.*, 2009). Kusov *et al.* discovered that the Mac3 is essential for viral replication since a mutation in the lysine patch, which is involved in binding oligo(G) or G-Quadruplex RNA, prevented SARS-CoV

replication (Kusov et al., 2015). Additionally, DPUP may play a role in modulating Mac3 binding to RNA as deletion of DPUP domain greatly reduces RNA Synthesis (M. A. Johnson et al., 2010; Kusov et al., 2015).

2.2.7.3.5. Nucleic acid-binding (NAB) domain and betacoronavirus-specific marker (β SM) domain.

Nucleic acid-binding (NAB) domain and betacoronavirus- specific marker (β SM) domain only exists in genus betacoronavirus. According to (Neuman et al., 2008) the domain binds to ssRNA and unwinds dsDNA. According to (Neuman et al., 2008) the domain binds to ssRNA and unwinds dsDNA. The RNA binding behavior is like Mac3 domain of SARS CoV as it interacts with Oligo (G). No information regarding betacoronavirus- specific marker (β SM) domain is available (Tan et al., 2007).

2.2.7.4. Nsp4

Modifying host's cellular membrane is a common feature of all enveloped positive sense RNA viruses including coronavirus. Viral nsp's develop double-membrane vesicles (DMVs) and convoluted membranes (CMs), which together create a broad reticulovesicular network that is involved in viral replication and transcription (Miller & Krijnse-Locker, 2008). Coronavirus genome encodes for 16 nonstructural proteins, three of which seem to be active in cellular membrane modification (nsp3, nsp4, and nsp6) (Angelini et al., 2013). Gadlage *et al.* discovered that nsp4 plays a significant role in the organization and stability of DMVs (Gadlage et al., 2010). Gadlage *et al.* (2010) modified MHV nsp3 and discovered abnormal morphology of virus-induced double membrane vesicles as compared to wild type.

2.2.7.5. Nsp5

The coronavirus protease nsp5 is a 30 kDa, three domain cysteine protease that is structurally and functionally conserved in all identified coronaviruses (Fehr & Perlman, 2015). Nsp5 is known as Mpro, a major protease in coronavirus gene expression, or 3CLpro, a 3C-like protease based on homology to 3C proteases in picornavirus (Anand et al., 2003). While nsp5 consists of three domains, dimerization of domain 1 and 2 is essential for nsp5 function. Domain 3 has been shown to direct and stabilize the dimerization (Anand et al., 2003). The active sites of the protomers are facing outwards away from each other while domain 3 is holding and stabilizing the structure (Stobart et al., 2013). Nsp5, after it is translated as part of the long polyprotein is responsible for its own autoproteolytic cleavage. It is proposed that nsp5 forms a homodimer and anchors itself in the membrane with the assistance of nsp4 and nsp6 (Hsu et al., 2005). It is then responsible for cleaving all downstream nonstructural proteins.

2.2.7.6. Nsp6

Autophagy is a normal cellular function that aids in the maintenance of cell homeostasis. it is a biochemical mechanism that lets cells turn unnecessary or damaged components into energy which it can use to expand and divide. Autophagy needs to be initiated by initiation factors. Upon initiation a semicircular membrane known as isolation membrane is formed which is expanded by ATG family of proteins. This expanded isolation membrane then engulfs damaged components forming what is called Autophagosome (Ichimiya et al., 2020). The autophagosome then fuses with a lysosome which contains digestive enzymes that break down the autophagosome's contents the resultant molecules are the released back into the cytosol to be

recycled. Cottam *et al.* demonstrated nsp6 from MHV and SARS in infected cell generates autophagosomes (Cottam et al., 2014). They compared autophagosomes from viral nsp6 induced cells and starvation controls. They observed greater number of autophagosomes in infected cells but the with smaller diameter (Cottam et al., 2014). To summarize, nsp6 restricts the expansion of autophagosome and plays a role in DMV formation with nsp3 and nsp4.

2.2.7.7. Nsp7 and Nsp8

SARS-COV nsp7 and nsp8 were reported to form hollow ring hexadecameric structure when mixed 1:1 in solution. Velthuis et al. measured and compared the activity of nsp7-8 in probable hexadecameric form to nsp12-RdRp and their data suggests nsp12 to be about 20-fold faster (te Velthuis et al., 2012). Nsp7 and nsp8 are also part of the replication and transcription complex. Nsp7 and nsp8 form a heterodimer that binds to RNA-dependent RNAPolymerase (nsp12).

Another nsp8 molecule interacts with nsp12 but at the opposite end of nsp7-8 heterodimer. Shi et al. characterized that it is the specific interaction between nsp7 of nsp7-8 heterodimer and nsp8 monomer that undergoes conformational change, and these conformational changes allow RTC movement (Shi et al., 2020). Additionally, nsp7 and nsp8 cofactors might also play a regulatory role in polymerase activity of nsp12 (Kirchdoerfer & Ward, 2019; Q. Peng et al., 2020; te Velthuis et al., 2012).

2.2.7.8. Nsp9

Nonstructural protein 9 (nsp9) of SARS-CoV was discovered to be an essential protein with RNA/DNA binding activity. The interaction of parallel alpha-helices carrying the protein-protein

interaction motif GXXXG causes Nsp9 to form a dimer. Miknis et al., designed site-directed mutants to determine which amino acids are essential for dimerization and if the dimerization was necessary for viral replication. They found dimerization of nsp9 was not essential for RNA binding however it is essential for replication (Miknis et al., 2009).

2.2.7.9. Nsp10

SARS-CoV nsp10 is a 148-residue subunit with two zinc fingers. Nsp10 interacts with both nsp14 and nsp16 and this interaction is conserved among all four genera in *Coronaviridae* (Lugari et al., 2010). Nsp14 is an exoribonuclease and nsp16 is a 2'-O-methyltransferase and they are both part of the replication and transcription complex. Both nsp14 and nsp16 bind to the same site on nsp10 therefore nsp10 can only interact with either nsp14 or nsp16 at a time. However, it is believed that nsp14 and nsp16 do not compete for nsp10 because nsp10 is present at a much higher concentration than both nsp14 and nsp16. This is because nsp10 is encoded as part of ORF1a, while nsp14 and 16 are only translated as part of ORF1b if ribosome slips on the "frameshift site" (Sawicki et al., 2007).

Using a bioluminescence resonance energy transfer assay and pulldown assays, Bouvet *et al.* established nsp10 residues involved in the association with nsp14. They also performed an enzyme activity assay to see whether nsp10 stimulates nsp14 (Bouvet et al., 2014). They inferred that the relationship between nsp10 and nsp14 is critical because changes to the binding site on nsp10 culminated in nsp14 losing its ExoN-activity. Removing nsp14 completely cripples the replicating virus, while disrupting the association between nsp10 and nsp14 is lethal to the virus

(Donaldson et al., 2007). This suggests that there are additional unknown functions mediated by the nsp10 and nsp14 interaction that are vital for viral replication (Bouvet et al., 2014).

2.2.7.10. Nsp12

SARS-CoV nsp12 is a 932 amino acid long RNA-dependent RNA polymerase that is expressed only after ribosomal frameshift. This means that nsp12, as well as all the proteins expressed as part of ORF1b, are generated at considerably lower quantities than non-structural proteins translated as part of ORF1a. The catalytic domain of nsp12 is located at the C-terminus and contains conserved RdRp motifs creating a cupped right hand with palm-finger-thumb structure enclosing a nucleotide binding site, but the N-terminal has no known counterpart in RdRps (Xu et al., 2003). Cheng et al. demonstrated that bacterially produced nsp12 could extend an oligo primer coupled to a template (Cheng et al., 2005).

2.2.7.11. Nsp13

SARS-CoV nsp13 is a 601 amino acid long multidomain protein with a zinc binding domain at the N-terminus and conserved motifs comparable to those seen in superfamily 1 helicases at the C-terminus. This superfamily 1 helicase unwind RNA and DNA duplexes unidirectionally in the 5' to 3' direction (Seybert et al., 2000; Weber & McCullagh, 2021), and the translocation is powered by dNTP hydrolysis. Ivanov et al. (2004) demonstrated that nucleic acids promote the activity of nsp13 and that it also has RNA 5'-triphosphatase activity, which could catalyze the initial step in viral RNA 5'-capping (Ivanov et al. 2004a). The N-terminal region of nsp13 includes a zinc binding domain that is required for helicase function. This domain comprises 12 conserved

cysteine and histidine residues that are assumed to form a binuclear zinc binding cluster (Seybert et al., 2005).

2.2.7.12. Nsp14

The N-terminal of nsp14 is a 3'-to-5' exoribonuclease that is related to DEDD superfamily with zinc finger in the middle of conserved motif II and motif III (Snijder et al., 2003b). The catalytic activity of nsp14 is shown to be specific to RNA and requires two Mg^{2+} ions per molecule. It is believed that coronavirus encode for 3'-to-5' exoribonuclease to maintain their large genome and nsp14 might be involved in proofreading mechanisms during replication. Studies done by (Minskaia et al., 2006) shown modification in HCoV-229E ExoN active site failed to produce viable virus. Consistent with other studies by (Almazán et al., 2006) showing reduction in genome replication and transcription by 10-fold.

2.2.7.13. Nsp15

Nsp15 is a second conserved ribonuclease called NendoU. NendoU is a conserved domain found in the members of the order *Nidovirales* and it stands for NidoViral endoribonuclease, specific for U (Snijder et al., 2003b). It is known to be an endoribonuclease that cleaves preferentially 3' of uridylates and generates 2'-3' cyclic phosphate ends (Ivanov, Hertzog, et al., 2004). It's been proposed that nsp15 and Bovine RNase A use the same catalytic mechanism due to the homology in residues forming a catalytic triad (Ricagno et al., 2006). However, the difference between nsp15 and Bovine RNase A is that the catalytic activity of nsp15 is enhanced in the presence of Mg^{2+} and upon formation of hexameric structures (Guarino et al., 2005).

Formation of hexameric structure enhances the activity of nsp15 however it is believed that the presence of nsp15 is not essential for viral replication.

2.2.7.14. Nsp16

Nsp16 is the last non-structural protein translated as part of the ORF1ab-encoded big polypeptide. Nsp16 is believed to be ribose-2'-O-methyltransferases since it has been shown to methylate $^7\text{MeGpppAC}_n$ at the ribose-2'-O-moiety of adenosine, changing a cap-0 to a cap-1 structure (Decroly et al., 2008). Nsp16 seems to be required for viral replication, since deletion or alteration of a single residue in the catalytic region results in the inability to synthesize RNA (Almazán et al., 2006). Nsp16 is essential for viral replication because viruses that replicate in the cytoplasm must supply the enzymes necessary to produce RNA cap structures or utilize an alternate strategy.

2.3. Macrodomain from MacroH2A.

Macrodomains are found in the majority of living organisms, including viruses, bacteria, and humans. It was first discovered as part of the histone, due to its larger size than histone, it was recognized as a component of histone and was termed macroH2A. MacroH2A is one of the known variants of H2A histone. With its molecular weight of ~40kDa, it is almost three times the size of canonical H2A histone. The N-terminal of macroH2A consists of histone region with 64% homology to canonical H2A, in the middle is a linker region and at the C-terminal is the Macrodomain. In total, macroH2A consists of 371 amino acids. Amino acids 1-122 form a histone domain, amino acids 123-160 are linkers, and amino acids 161-371 are the macrodomain. This macrodomain on the histone variant of MacroH2A consists of 7 beta sheets that are sandwiched between 5 alpha helices. Most of the beta sheets are surrounded by the alpha helices. However, some of it is exposed to the surface. This exposed surface of beta sheets exhibits an extended hydrophobic region. This hydrophobic region includes two patches of three (182–185) and four residues (355–359) (Rack et al., 2016).

2.3.1. Structural differences between MacroH2A and canonical H2A.

MacroH2A's histone domain is 64% similar to nonvariant H2A from *Xenopus laevis* and *Mus musculus*. Two regions have been identified by sequence analysis as having the potential to induce structural and functional variations. One of these regions is L1 loop, which is involved in the primary interaction that forms the H2A-H2B dimer. The differences in L1 loop consist of three residues, 83, 84, and 88, which correspond to residues 80, 81, and 84 in macroH2A. In nonvariant H2A, the interaction between L1-L1 is stabilized by salt bridges, whereas in variant MacroH2A it

is stabilized by hydrophobic interaction. These changes thus influence the stability of the histone octamer. The second region of structural and functional variation is the ladle-shaped H2A docking domain that consists of residues 83–108. This region interacts with the H3-H4 heterotetrameric and guides its interaction with nucleosome DNA. The changes in docking domain does not affect the interaction between histones when forming an octamer (Gamble, 2013; Li et al., 2013; Rack et al., 2016).

2.3.2. Structural changes in MacroH2A affecting protein-protein interaction.

Chakravarthy & Luger et al. (2006) showed that the L1 loop in MacroH2A's histone domain has an important impact on the assembly and dynamics of nucleosomes. This is attributed to a four-amino-acid exchange in the L1 loop. The L1 loop stabilizes the nucleosome core particle, rendering it resistant to histone exchange facilitated by chaperones. Structural experiments confirmed that although the altered L1 loop has little effect on the overall structure of the nucleosome core, it does change the protein-protein interaction between histones in the histone domain of the MacroH2A-H2B dimer (Chakravarthy et al., 2005b; Chakravarthy & Luger, 2006).

NAP-1 is a yeast nucleosome assembly protein that is capable of exchanging H2A-H2B dimers in chaperone-mediated histone exchange. However, it was found that nucleosome core particles containing MacroH2A were resistant to chaperone mediated histone exchange by NAP-1 (Abboott et al., 2005). Additionally, it was suggested that the linker domain of MacroH2A might also interfere with chromatin remodeling machines through steric hindrance. The research found lower affinity in nucleosome core particles that contain MacroH2A for chromatin remodeling machines, reducing transcription initiation (Doyen et al., 2006).

2.3.3. MacroH2A interaction with HDAC proteins.

The nonhistone "macro" domain of macroH2A protein is composed of 130–190 amino acids and has a structure consisting of 7 beta sheets sandwiched between 3 or more alpha helices. Following genome sequencing, two of the seven beta strands are expected to be leucine zippers (Pehrson & Fuji, 1998), with one side of the beta strand pointing inwards facing alpha helices and the other side exposed to the surface. A total of 7 (182-185 [FTVL]) and 355–359 [IGIYV]) hydrophobic residues from the beta sheet are exposed to the surface, exhibiting an extended hydrophobic region. Chakravarthy et al. (2005) discovered that an expanded hydrophobic area on nonhistone domains would help in the recruitment, interaction, and stabilization of HDAC proteins in vitro. To test this hypothesis, they expressed three class I HDACs, but only HDAC1 and HDAC2 were found to associate with MacroH2A's nonhistone domain. Additionally, to confirm their results, they decided to modify the hydrophobic amino acids with hydrophilic amino acids. Mutations in the extended hydrophobic region had a major effect on HDACs and nonhistone domain interaction. They also discovered that mutations in the 160–180 area eliminated the repressive activity of this domain using transcriptional repression assays (Buschbeck & di Croce, 2010).

2.3.3.1. MacroH2A Inhibit gene expression by interacting with HDAC.

The MacroH2A variant was initially found in a high concentration on chromosome X. It was hypothesized that MacroH2A might interact with and recruit proteins that play a role in inactivating chromosome X. One of the known modifications to suppress gene expression is deacetylation, carried out by histone deacetylases (HDACs). They expressed three class I HDACs:

HDAC1, HDAC2 and HDAC3, along with histone lysine methyl transferase and Suv39h1. They found HDAC1 and HDAC2 interacted with the non-histone region of MacroH2A as these proteins were copurified together. The interaction between HDACs and MacroH2A was also confirmed with a western blot using anti-HDAC1, -2, and -3 antibodies on the supernatant of erythroleukemia cells. Following that, they mutated MacroH2A to identify regions that could be involved in association with HDACs. They found residues 182-185 and 355–359 to be involved in interaction. These are the same residues residing in the exposed beta sheet of the macrodomain, forming an extended hydrophobic region. When these residues were then substituted with polar amino acids, the association between HDACs and the nonhistone region of MacroH2A was lost. In vivo transcriptional repression experiments showed residues in 355–359 to be essential for repression activity. These and other immunoprecipitation studies revealed a connection between MacroH2A-mediated HDAC recruitment and chromatin hypoacetylation (Buschbeck & di Croce, 2010).

2.3.4. MacroH2A inhibits gene expression by masking regulatory region.

Almost every cell in our body contains the same genetic information provided by our mother and father. However, not every cell in our body looks or behaves the same. It is because of the cell-specific gene expression system that dictates what genes will be encoded. One way that the cell-specific gene expression system controls gene expression is by modifying the chromatin structure via recruiting histone variant MacroH2A (Agelopoulos & Thanos, 2006). Thanos et al. (2006) discovered a new mechanism of gene expression that uses ubiquitous transcription factors to recruit histone variants for gene regulation. Namalwa B cells and

epithelial HeLa cells were infected with the virus and the expression of INF-B and IL8 genes was observed. Activation of IL8 and INF-B genes is mediated by a simple enhancer bound by transcriptional factors such as NF-KB, C/EBP and ATF/AP1, and all these factors are present in both cells. Both genes were expressed in HeLa cells, but only INF-B was expressed in Namalwa B cells (Agelopoulos & Thanos, 2006).

The Thanos lab suspected transcription factors might be the reason the IL8 gene was not expressed in Namalwa B cells and carried out a chromatin immunoprecipitation (ChIP) experiment. A Chromatin Immunoprecipitation (ChIP) experiment was conducted to determine the interaction of these transcription factors with the promoter region of each gene. As suspected, all the basal transcription machinery was recruited to the IL8 promoter region of the HeLa cell but not the B cell. Instead, they observed an ATF-2/Jun-D heterodimer bound to the promoter region of the IL8 gene on B cells along with GCN5 histone acetyltransferase protein. GCN5 tends to acetylate K9 and K14 on H3, signaling the recruitment of the TFIID complex. The following step was to look into the nucleosomal organization in this cell type. In HeLa cells, the promoter site of IL8 was nucleosome free, whereas in B cells, the promoter site was masked by a nucleosome, making it inaccessible to transcriptional machinery (Agelopoulos & Thanos, 2006).

In other studies, it's been shown that a macro variant of H2A in a nucleosome core particle could potentially make the gene inaccessible to transcriptional factors. They suspected this might be the case with IL8 as well and decided to carry out a ChIP assay using an antibody specific to MacroH2A. Their results showed no MacroH2A on the IL8 promoter in HeLa cells, contrary to B cells. A MacroH2A-containing nucleosome was found in the coding region of the IL8 gene in B

cells. To examine the role of MacroH2A in IL-8 gene expression, siRNA was used to inhibit the expression of MacroH2A. It was confirmed beforehand that the inhibition of MacroH2A would not affect other histones. By depleting MacroH2A from B cells, the IL-8 gene was successfully expressed. This confirmed that MacroH2A does play a role in IL8 gene inhibition (Agelopoulos & Thanos, 2006).

2.3.5. MacroH2A interaction with Activating transcription factor 2 (ATF2).

The ATF/AP1 location was thought to play a role in positioning the macroH2A nucleosome to the IL-8 promoter region in B cells. Following viral infection, a change in ATF/AP1 position resulted in activation of the IL-8 promoter in Namalwa B cells but not in HeLa cells. Additional nucleosome mapping experiments identified loss of MacroH2A in the nucleosome following ATF/AP1 modification. Protein–protein interaction assays were conducted to see whether ATF-2 specifically recruits macroH2A. Histone composition was evaluated by Western blotting using antibodies specific for macroH2A and canonical histones after nucleosome reconstitution. When ATF-2 was not present, the assembled chromatin did not contain macroH2A on the IL-8 or IFN- γ promoters. On the other hand, the addition of recombinant ATF-2 during the nucleosome reconstitution reaction resulted in the induction of a macroH2A nucleosome on the IL-8 enhancer but not the IFN- γ enhancer (Agelopoulos & Thanos, 2006).

Next, ATF sites were swapped between IL8 and INF- γ as previous studies have shown proteins containing the b-zip motif acquire distinct conformations based on the sequence of the binding site. They evaluated the histone composition again via Western blotting using antibodies specific for macroH2A and canonical histones after nucleosome reconstitution. They discovered

that macroH2A-containing nucleosomes are recruited by ATF-2 on the IFN- β template with an ATF site sequence taken from IL-8 (Agelopoulos & Thanos, 2006).

These results demonstrate that the bZip DNA-binding domain of ATF-2 is responsible for the recruitment of macroH2A-containing nucleosomes and that this recruitment depends on the specific sequence of the ATF DNA site, thus implying a DNA-induced protein allosteric mechanism. These results indicate that the ATF-2 bZip DNA-binding domain is accountable for the recruitment of macroH2A-containing nucleosomes and that this recruitment is dependent on the precise sequence of the ATF DNA site (Agelopoulos & Thanos, 2006).

2.3.6. Macrodomains as standalone proteins.

Macrodomains are found in the majority of living organisms, including viruses, bacteria, and humans. Macrodomains are present not only as standalone proteins, but also as modules inside multidomain proteins such as PARPs (Chakravarthy et al., 2005a). A macrodomain is a three-layered $\alpha/\beta/\alpha$ sandwich fold and it is not restricted to one function. Some macrodomains have been shown to have ADP-ribose-1-phosphate phosphatase activity (Saikatendu et al., 2005). More recently, it was demonstrated that macrodomains from several (+) ssRNA viruses are also capable of hydrolyzing single ADPr (de-MARylation) or multiple ADPr residues from proteins (de-PARylation). (Fehr et al., 2016). MARylation and PARylation are reversible posttranslational protein modifications usually utilized by host cells as a signal for various cellular processes, including the initiation of antiviral immune responses by PARPs (Liu & Yu, 2015).

Fehr et al. reported that the SARS-CoV wild-type Mac1 domain prevents the production of innate-immunity genes in vitro, thus inhibiting the host immune response (Fehr et al., 2016).

It is also possible for macrodomains to only bind to the ADPr already bound to the acceptor protein, blocking an essential signal. The SARS-CoV-2 Mac1 domain has been shown to bind to ADPr via various binding techniques (Frick et al., 2020). Based on the capacity of the viral macrodomain to simply bind to ADPr or to catalyze MARYlation, PARYlation, and association with PARPS, it has been proposed that the viral macrodomain modulates the host cellular immune response (Kraus, 2009).

3. Literature Review: Chemistry

3.1. Background Thermodynamics

3.1.1. Energy

Energy is often defined as the capacity to do work. In chemistry, work is defined as the change in energy that occurs as a process. Potential and kinetic energy are two forms of energy, while the others are varying combinations of the two. For example, chemical energy is the potential energy stored within the structural units of chemical substances, whereas thermal energy is the kinetic energy associated with the random motion of atoms and molecules. Additionally, all forms of energy can be converted into one another. A ball rolling down a hill is an example of potential energy converted into kinetic energy. Although one form of energy can be converted into another, scientists have concluded that energy can neither be created or destroyed and this principle is called the law of conservation of energy . The first law of Thermodynamics is based on the law of conservation of energy, which states that energy can be converted from one form into another, but it cannot be created or destroyed. The First and the second law of thermodynamics were established by 1860 as formalized in the work of scientists such as Rudolf Clausius (Clausius, R., 1865) and Kevin, William T (Kevin, William T., 1849). (Chang & Overby, 2011)

3.1.2. The First Law of Thermodynamics:

The first law of thermodynamics states that the energy (U) of the universe is a constant. The universe includes the system and the surroundings. Therefore, the energy of the universe is equal to the sum of the energy of the system and the surroundings (Equation 3.1-1).

$$\Delta U_{\text{Universe}} = \Delta U_{\text{Surrounding}} + \Delta U_{\text{System}} = 0 \quad \text{Equation 3.1-1}$$

Energy change (ΔU) can occur in two different ways: via heat (q) and through work (w). When a system experiences a change in energy, it is either because the system gained or lost heat, or the work was done on or by the system (Equation 3.1-2).

$$\Delta U_{\text{System}} = q + w \quad \text{Equation 3.1-2}$$

When studying a chemical reaction, one way to demonstrate mechanical work is to examine the expansion or compression of the system. This can be illustrated in terms of volume and pressure (Equation 3.1-3).

$$w = -P\Delta V \quad \text{Equation 3.1-3}$$

Substituting equation 3.1-2 into equation 3.1-3 for work (w) yields equation 3.1-4, which can be applied to laboratory processes.

$$\Delta U_{\text{System}} = q - P\Delta V \quad \text{Equation 3.1-4}$$

The next step is to apply the first law of thermodynamics to chemical reactions. Most chemical reactions in a laboratory take place under either constant volume or constant pressure conditions. If a chemical reaction is carried out under constant volume $\Delta V = 0$, the internal energy of the system is then proportional to the heat of the reaction (Equation 3.1-5).

$$\Delta U_{\text{System}} = q_v \quad \text{Equation 3.1-5}$$

The subscript “ v ” in equation 3.1-5 indicates that it is a constant volume process. Constant volume processes are sometimes impossible to achieve, and chemical reactions must instead be performed under constant pressure. To help with this situation, a new state function, Enthalpy change (ΔH) was introduced (Equation 3.1-6).

$$H = U + PV \quad \text{Equation 3.1-6}$$

Enthalpy is defined by equation 3.1-6, where U is internal energy, P is pressure and V is the volume of the system. Enthalpy for a system is always calculated as a change in enthalpy (Equation 3.1-7) and for a system under constant pressure, enthalpy is given by Equation 3.1-8.

$$\Delta H = \Delta U + \Delta(PV) \quad \text{Equation 3.1-7}$$

$$\Delta H = \Delta U + P\Delta V \quad \text{Equation 3.1-8}$$

When Equation 3.1-4 is applied to a constant-pressure system, it is transformed and simplified to Equation 3.1-9. Equation 3.1-10 is obtained by substituting U from Equation 3.1-8 into Equation 3.1-9 and simplifying.

$$q_p = \Delta U + P\Delta V \quad \text{Equation 3.1-9}$$

$$q_p = \Delta H \quad \text{Equation 3.1-10}$$

There is now a way of applying the first law of thermodynamics to a system and quantifying energy under various conditions. If a chemical reaction is performed at a constant volume, the change in heat is equal to ΔU (Equation 3.1-5). If a reaction is performed under a constant pressure, then the change in heat is equal to ΔH (Equation 3.1-10). Calorimetry is a frequently used technique in laboratories to measure the heat changes caused by chemical and physical processes (Velázquez-Campoy et al., 2004). Constant-volume calorimetry can be used to estimate one value describing the process's thermodynamic: enthalpy change (ΔH).

3.1.3. The Second Law of Thermodynamics

The second law of thermodynamics introduces a new thermodynamic quantity called Entropy (S) in conjunction with spontaneity (Chang & Overby, 2011). According to the second

law of thermodynamics, a spontaneous reaction increases the entropy of the universe (Equation 3.1-11). Entropy is often used to indicate the degree to which a system's energy is distributed across the many potential ways that system may hold energy.

$$\text{For a spontaneous reaction: } \Delta S_{\text{Universe}} = \Delta S_{\text{System}} + \Delta S_{\text{Surroundings}} > 0$$

Equation 3.1-11

As with other thermodynamic quantities, entropy in terms of spontaneity is of primary interest. A new thermodynamic function for Gibbs free energy was introduced (Equation 3.1-12). When applied to a system, Gibbs free energy may be used to assess whether a chemical reaction will occur spontaneously.

$$\text{For a spontaneous reaction: } \Delta G = \Delta H + T\Delta S < 0 \quad \text{Equation 3.1-12}$$

As shown in equation 3.1-12, a spontaneous reaction would give a negative change in Gibbs free energy (ΔG), where ΔH is the change in enthalpy of a system, T is the temperature and ΔS is the entropy change of the system. In this case, free energy refers to the energy available to do work, and if a reaction is accompanied by the release of useable energy, i.e., ΔG is negative, this alone establishes its spontaneity (Chang & Overby, 2011). While a system's free energy change may be determined under any set of conditions, when data is gathered under standard conditions, the result is standard-state free energy of reaction (Equation 3.1-13).

$$\Delta G^{\circ} = \Delta H^{\circ} + T\Delta S^{\circ} \quad \text{Equation 3.1-13}$$

The significance of the equation describing a system's free energy lies in its ability to identify the relative relevance of the enthalpy and entropy components as driving factors in each reaction. In other words, the change in the system's free energy indicates the balance of two

driving factors that decide whether a chemical reaction is spontaneous. The following table highlights the effect of entropy and enthalpy changes on spontaneity of a chemical process (Atkins & Paula, 2006).

ΔH	ΔS	Spontaneity
Positive	Positive	At high temperatures, the reaction occurs spontaneously at low temperatures, it occurs in reverse.
Positive	Negative	The reaction occurs spontaneously in the opposite direction at all temperatures.
Negative	Positive	The reaction occurs spontaneously at all temperatures.
Negative	Negative	At low temperatures, the reaction occurs spontaneously at high temperatures, it occurs in reverse.

Table 3.1-1 Enthalpy (ΔH), Entropy (ΔS) dictates reaction spontaneity.

A summary of how Enthalpy (ΔH), Entropy (ΔS), and Temperature (T) affect a chemical reaction's spontaneity.

Using Gibbs free energy, one can also identify the direction of a chemical process that are not in standard states. In equation 3.1-14, ΔG is the Gibbs free energy change of a chemical process not in standard state, ΔG° is the Gibbs free energy of chemical process in standard state, R is the gas constant, T is the absolute temperature of the reaction and Q is the reaction quotient. At equilibrium, $\Delta G = 0$ and $Q = K$. When applied to equation 3.1-14 and simplified, this results in equation 3.1-15, where K is the equilibrium constant. Equation 3.1-15 allows us to determine the equilibrium constant of a reaction given the change in standard free energy, or vice versa. Table 3.1-2 summarizes the relationship between Gibbs free energy and equilibrium constant based on equation 3.1-15 (Chang & Overby, 2011).

$$\Delta G = \Delta G^\circ + RT \ln(Q) \quad \text{Equation 3.1-14}$$

$$\Delta G^\circ = -RT \ln(K) \quad \text{Equation 3.1-15}$$

K	ln(K)	ΔG	At Equilibrium
> 1	Positive	Negative	Products are favored
= 1	0	0	Products and reactants are equally favored
< 1	Negative	Positive	Reactants are favored

Table 3.1-2 Summary of ΔG and ln(K) based on equation 3.1-15.

3.1.4. Calorimetry

Calorimetry is a technique used for determining the quantity of heat emitted or absorbed during a chemical reaction. A calorimeter is a device that was developed especially for this purpose. Numerous calorimeters are available on the market, including differential scanning calorimeters (DSC) and isothermal titration calorimeters (ITC). In our studies, an isothermal titration calorimeter (ITC) was utilized. The ITC measures the change in heat of a solution, which can be related to a chemical reaction. For example, when one molecule is titrated with another and they react, heat is either released or absorbed, and from this, the free energy, enthalpy, and entropy of binding may be determined accurately with ITC (Velázquez-Campoy et al., 2004).

3.2.Isothermal Titration Calorimetry (ITC)

ITC is a useful tool to determine energy changes when two molecules bind in a solution. The advantages of using ITC for binding studies over other methods are that it is label-free, there is no molecular weight limitation, and it is not influenced by the color or the optical clarity of the sample (Velázquez-Campoy et al., 2004). One ITC experiment can provide whether a binding event occurs along with thermodynamic parameters of the interaction. However, accurate measurements of thermodynamic quantities associated with intermolecular interaction require careful standardization of the instrument. This can be achieved by evaluating the instrument response with the measurement of well-known physico-chemical processes such as Ca^{2+} -EDTA chelation (Ràfols et al., 2016).

ITC consists of two identical cells encased in an adiabatic jacket. These cells are constructed of a highly efficient, thermally conductive, and chemically inert material. The temperature of both cells is measured with very high accuracy and heaters are used to maintain a constant temperature.

In a typical ITC experiment, macromolecules such as proteins are kept in the sample cell and water or buffer without protein and ligand in the reference cell. Before initiating the experiment, both cells are equilibrated to ensure the same temperature. The reference cell is kept closed for the duration of the experiment, whereas the sample cell is open from the top, accessible for delivery of small molecules via a syringe. The addition of a very precisely known aliquot of small molecules into the sample cell initiates an experiment. Upon interaction between two molecules, heat is either generated or absorbed. A miniscule change in temperature is

detected and heaters are activated to equilibrate both cells (Velázquez-Campoy et al., 2004). The amount of heat required to maintain an exact temperature between the sample and reference cells is then recorded in a time-dependent manner.

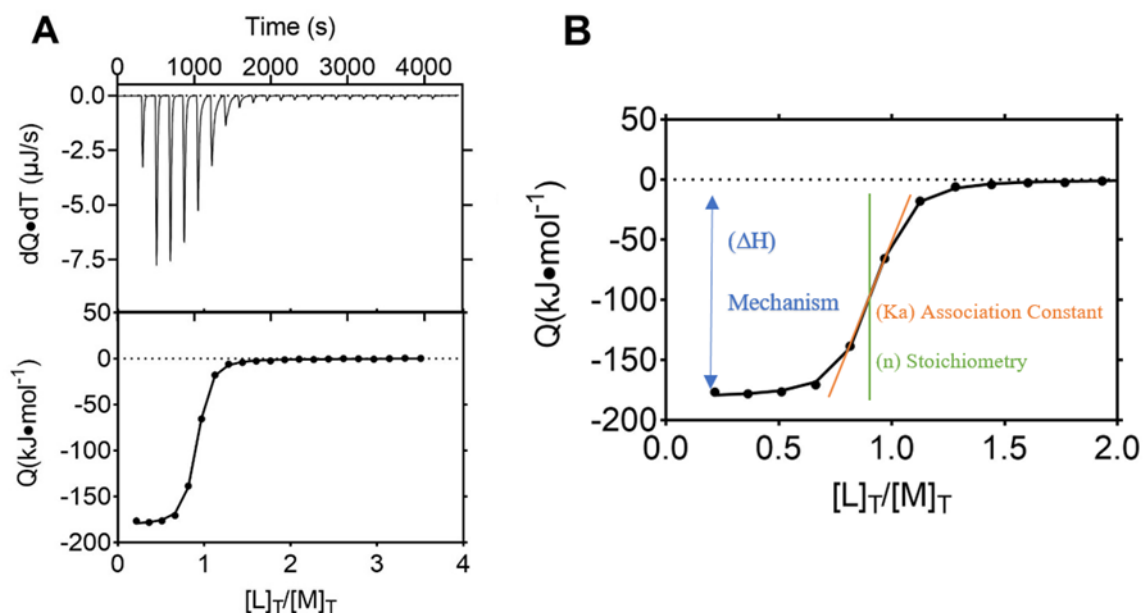


Figure 3.2-1 Example of raw data collected from ITC

(A) The top panel illustrates typical raw data from an ITC experiment, while the bottom panel illustrates the integrated heat plot produced by measuring the area under each peak. Enthalpy, binding constant, and stoichiometry may all be estimated directly from the figure as shown in (B). This data was accompanied by Ca^{2+} and EDTA titration done to measure equipment response.

An example of raw data is shown in the top section of Fig. 3.2-1a. The x-axis is time in seconds and the y-axis is the change in raw heat observed over time in micro joules per second, with each spike corresponding to an injection from a syringe into the sample cell. The size and the direction of the peak are based on the magnitude of heat that was either released or absorbed upon ligand binding in the sample cell. Each peak is then integrated and plotted against the molar ratio, which is shown as the bottom part of Fig. 3.2-1a. Initially, all the titrant added to

the sample cell binds to the macromolecule, resulting in a relatively high heat output. As the end point is reached, the quantity of ligand bound per injection diminishes beyond the end point, a continuous heat of dilution is often observed and eliminated during analysis (Velázquez-Campoy et al., 2004).

A rough estimate of molar enthalpy, stoichiometry and binding constant can be deduced from the first few peaks in the titration, the position of the equivalence point and the slope of the curve at the equivalence point, respectively (Figure 3.2-1b). Calculating parameters using ITC data requires choosing a model that is appropriate for the system under investigation and fitting a function that will be utilized in the non-linear least square regression analysis (Velázquez-Campoy et al., 2004). In our studies, a model based on the premise that a single macromolecule has just one binding site was utilized (Equation 3.2-1). Equation 3.2-2 expresses an equilibrium constant that characterizes the binding of a ligand to a macromolecule with a single binding site.



$$K_a = \frac{[ML]}{[M][L]} \quad \text{Equation 3.2-2}$$

An ITC experiment involves injecting a ligand into a sample cell containing a macromolecule, resulting in a concentration change. Equations 3.2-3 and 3.2-4 are used to determine the total concentrations of macromolecule, $[M]_t$ and ligand, $[L]_t$ in the sample cell after each injection (i).

$$[M]_{T,i} = [M]_o \left(1 - \frac{v}{V}\right)^i \quad \text{Equation 3.2-3}$$

$$[L]_{T,i} = [L]_o \left(1 - \left[1 - \frac{v}{V}\right]^i\right) \quad \text{Equation 3.2-4}$$

In the preceding equation, $\left(1 - \frac{v}{V}\right)$ is the dilution factor that considers the change in reactant concentrations caused by each sample injection, where (V) is the volume of the sample cell and (v) is the injection volume. The concentration of macromolecule and ligand in the sample cell is determined by Equations 3.2-3 and 3.2-4, and binding results in a minute change in temperature. The power required to compensate for the change in temperature is referred to as heat (q) . As stated in equation 3.2-5, the heat (q) is proportional to the difference in concentration of the complex formed upon each injection $[ML]_i$, the volume of the sample cell (V) , and the binding enthalpy (ΔH_a) .

$$q_i = V\Delta H_a \left([ML]_i - [ML]_{i-1} \left[1 - \frac{v}{V} \right] \right) \quad \text{Equation 3.2-5}$$

The binding enthalpy can therefore be determined from the heats measured upon each injection and complex concentration (Equation 3.2-5), which can be determined, based on the mass action law and conservation of mass for each species. (Velázquez-Campoy et al., 2004) (Equation 3.2-6)

$$[ML]_i = \frac{1 + n[M]_{T,i}K_a + K_a[L]_{T,i} - \sqrt{(1 + n[M]_{T,i}K_a + K_a[L]_{T,i})^2 - 4n[M]_{T,i}K_a^2[L]_{T,i}}}{2K_a}$$

Equation 3.2-6

Equation 3.2-6 describes binding events in which ligands bind to a macromolecule with n equivalent and independent binding sites and thermodynamic parameters such as K_a , ΔH_a , and n may be determined via fitting. The binding affinity is then related to the free energy ΔG of binding as shown in equation 3.2-7.

$$\Delta G = -RT\ln(k_a) \quad \text{Equation 3.2-7}$$

In Equation 3.2-7, R is the gas constant, T is the absolute temperature. Gibbs free energy is also related to other thermodynamic parameters such as Enthalpy ΔH and Entropy ΔS as shown in Equation 3.2-8.

$$\Delta G = \Delta H - T\Delta S \quad \text{Equation 3.2-8}$$

3.2.1. Example Analysis of data

In a typical ITC experiment, data is collected, analyzed, and fitted to an equation to determine Gibbs free energy, enthalpy, and entropy pertaining to a specific molecular binding. Below are three possible scenarios:

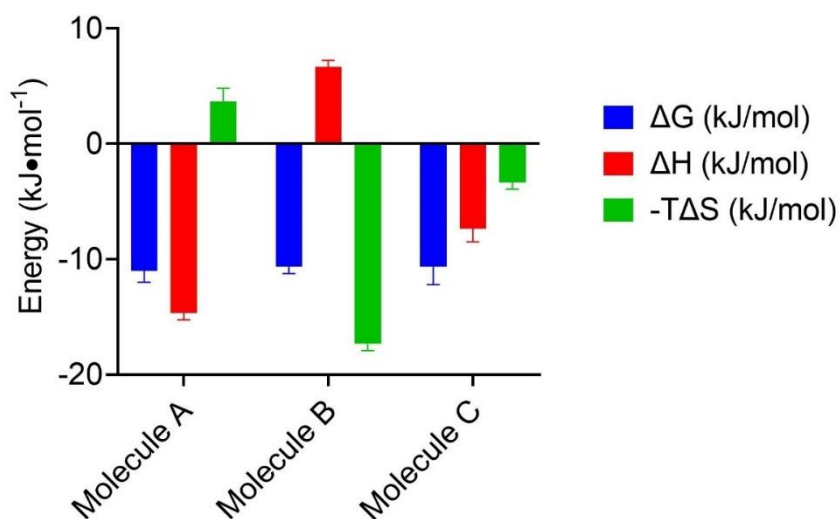


Figure 3.2-2. **Example analysis of ITC data.**

Illustration of three distinct compounds binding to the same target with the same affinity or Gibbs free energy but with varied enthalpy and entropy.

If all three molecules shown in Fig. 3.2-2 were ranked only based on their affinity, they would be identical. However, the variation lies in their enthalpy and entropy values. Entropy and enthalpy offer further information on the binding mechanism. The binding mechanism

determines how effectively the molecule adheres to the binding pocket and the interactions that are made between the ligand and the macromolecule.

Molecule A's binding is strongly enthalpically driven. Enthalpy substantially contributes to binding affinity and Gibbs free energy, suggesting that the ligand and macromolecule are closely associated through hydrogen bonds and van der Waals interactions. Additionally, there is a negative entropy contribution, which may be attributed to a reduction in the degrees of freedom of both the macromolecule and the ligand, as indicated by positive $-T\Delta S$. Compound B binding, on the other hand, is highly entropic, and therefore the binding affinity is determined by nonspecific hydrophobic contact and desolvation of water molecules upon binding. Additionally, compound B has an unfavorable enthalpy. This is due to the surface being dehydrated or loss of hydrogen bonding. Compound C binds with a balanced affinity due to the presence of both favorable hydrogen and hydrophobic contacts.

3.2.2. The significance of Entropy and Enthalpy.

ITC is often used to determine the binding energies of biological processes directly. These biological processes include, but are not limited to, protein-ligand and protein-protein binding, where the ligand is usually a small organic molecule. Binding energies of biological processes can be calculated using equation 3.2-9 in terms of binding affinity. Binding affinity is a combined function of the enthalpy (ΔH) and entropy (ΔS)

$$\Delta G = -RT \ln K_d = \Delta H - T\Delta S \quad \text{Equation 3.2-9}$$

This is an important way of evaluating a compound's binding properties, since the equilibrium dissociation constant (K_d) as determined by Gibbs free energy (ΔG), enables the

identification of the binding's dominant forces, change in Entropy (ΔS) and change in Enthalpy (ΔH). A drug molecule with extremely high binding affinity (*i.e.* a low K_d) requires favorable contribution from both enthalpy change (ΔH) and entropy change (ΔS) (Voet et al., 2018).

The total enthalpy change is calculated as the sum of the positive and negative enthalpic contributions. The positive enthalpy changes are due to the formation of hydrogen bonds and van der Waals interactions between the ligand and the protein, while the unfavorable enthalpic contribution is due to the ligand being desolvated or removed from solution in order to form a complex (Freire, 2008).

The challenge is in synthesizing ligands that contribute significantly to enthalpy changes (Ruben et al., 2006). In other words, the ligand needs to be covalently bonded to a very electronegative atom such as a N, O, or F. If these atoms can form hydrogen bonds with amino acids at the binding site, they almost certainly already do so with water or other buffer molecules. To bind to the binding site of a protein, the ligand breaks bonds with the solvent and then establishes new hydrogen bonds with amino acids in the protein's binding site. The penalty for desolvating a polar molecule is very large, therefore the beneficial contribution upon binding must be much higher to achieve a favorable enthalpic contribution overall. The hydrogen bonding is also very fastidious in terms of bond angle and distance if the bond angle and distance are not optimal, the favorable enthalpy contribution diminishes and suddenly turns unfavorable. Thus, for a molecule to have a significant favorable enthalpic contribution, it must fit precisely into the binding pocket. Loose fitting would result in a unfavorable enthalpic contribution (Freire, 2008).

A favorable enthalpy change upon complex formation is proportional to the ability of a ligand and protein to maximize van der Waals forces and hydrogen bonding while preserving an exact geometric fit inside the binding pocket. This is an indication that the molecule establishes a stable bond with the target and that the bonds are sufficiently strong to compensate for the unfavorable enthalpy associated with desolvation. Even when the binding enthalpy of a molecule is increased, this does not necessarily translate into an increase in binding affinity, since the gain in enthalpy may be countered by a loss of entropy (Freire, 2008).

On the other hand, the entropic contribution is dependent on desolvation, conformation, and the number of individual particles in solution. Desolvation of a ligand usually results in a favorable change in entropy due to the release of a water molecule from the binding site upon ligand binding. On the other hand, conformational entropy change is nearly always unfavorable to ligand binding. This is because, although the ligand has rotatable bonds and can adopt many conformations in solution, it will be limited to occupying a single conformation upon binding. This leads to a reduction in degrees of freedom and entropy. Due to the loss of conformational degrees of freedom in both the drug molecule and the protein, the conformational entropy change that occurs after ligand binding is almost always unfavorable (Freire, 2008).

To conclude, the binding affinity, enthalpy and entropy changes describing a ligand binding to a protein may be estimated by determining heat changes in a binding isotherm. Enthalpy is related to by the contribution of hydrogen bonds and van der Waals interactions, as well as the contribution of ligand desolvation from solution. As will be discussed next, it is considerably more difficult to design an enthalpic-driven ligand with a polar functional group

because if the ligand is not fitted optimally into the binding site, it will acquire a high enthalpic penalty, while the enthalpic penalty for non-polar ligands is orders of magnitude smaller. Optimizing a ligand for favorable entropic contribution is considerably simple since most negative contributions are due to conformational or degree of freedom loss in solution. A molecule may be constructed in a limited conformation that matches the binding conformation, preventing the molecule from changing shape upon binding.

3.2.3. The significance of Enthalpy in Drug Discovery

Enthalpy plays a significant role in determining the ligand attributes in terms of binding. Favorable enthalpy is contributed by having strong hydrogen bond and van der Waals interaction from a tight fit however large penalty is given for a loose fit. Favorable contribution from enthalpic energy is an important way of evaluating a compound binding property (Ohtaka & Freire, 2005). Let us consider two instances from the literature on the advancement of HIV protease inhibitors and statins, inhibitors for HMG-CoA reductase. Taken together, Fig 3.2-3 and Fig. 3.2-4 demonstrate a connection between binding affinity and thermodynamic parameters.

Another example that demonstrates the change in thermodynamics as improvements were made to the compound is β -Hydroxy β -methylglutaryl-CoA (MHG-CoA) reductase inhibitors also known as statins (Ohtaka & Freire, 2005). As shown in Fig. 3.2-5, Fluvastatin was one of the first MGH-CoA inhibitor however as improvements were made there was shift in thermodynamic signature. Fluvastatin is entropy driven whereas rosuvastatin, an improved MGH-CoA inhibitor is enthalpically driven.

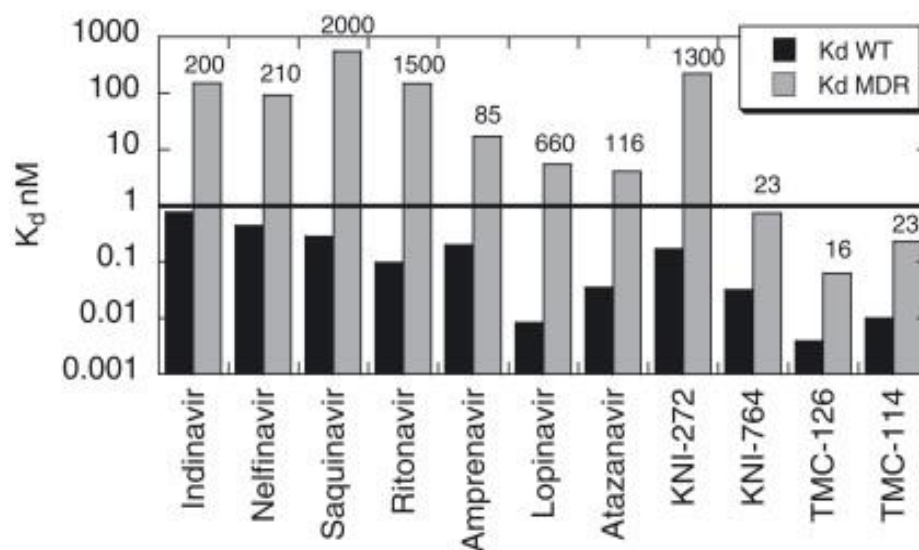


Figure 3.2-3. **The progression of HIV inhibitors' binding affinity.**

The dissociation constants of HIV inhibitors with wild type (WT) and a drug resistant mutant (MDR) HIV-Protease are shown. This figure has been adapted with permission from (Ohtaka & Freire, 2005)

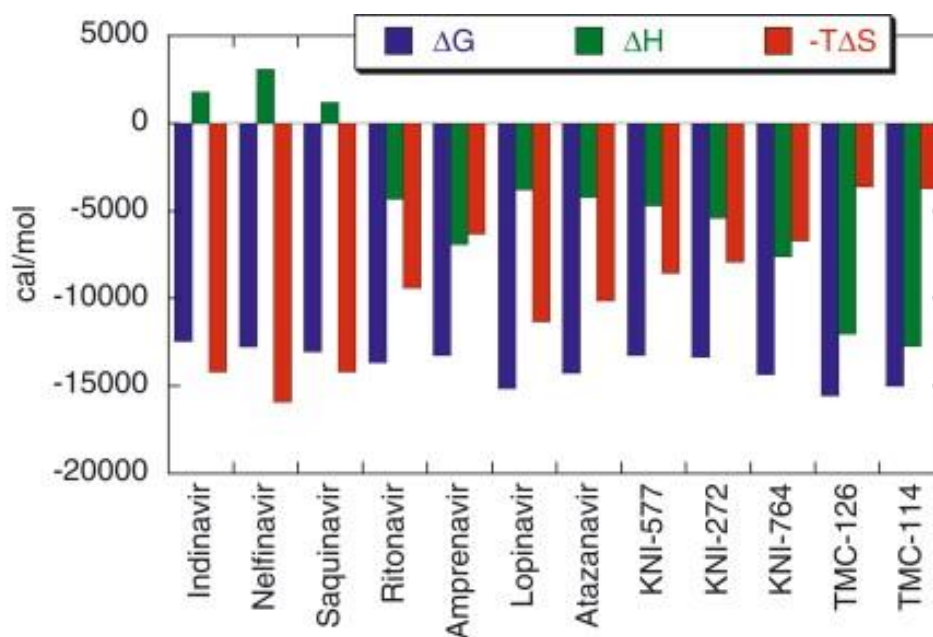


Figure 3.2-4 **The development of HIV inhibitors is measured by thermodynamic data.**

Shows thermodynamic information associated with inhibitors binding to HIV protease. This figure has been adapted with permission from (Ohtaka & Freire, 2005)

In both instances, It is seen that the modifications were made to the compounds to increase their affinity and selectivity to shift in thermodynamic signature. This shift in thermodynamic signature happens to be from first compound being entropically driven whereas the improved version being more enthalpically driven. This evidence indicates that the most effective drugs in their class are usually enthalpically driven. This does not always imply that the therapeutic molecules were developed with pure intent toward increasing enthalpy rather, advances in compounds with high selectivity and potency are generally enthalpically driven (Freire, 2008).

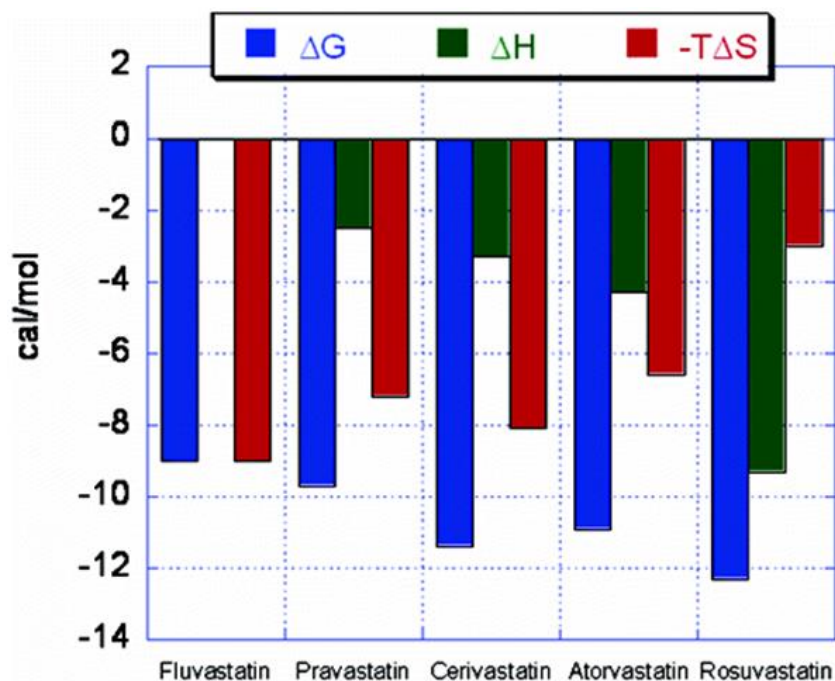


Figure 3.2-5 **The development of statins is measured by thermodynamic data.**

Carbonell & Freire et al. shows the inhibitor constant and associated thermodynamic values associates with statins, inhibitor of HMG-CoA reductase. This figure has been adapted with permission from (Carbonell & Freire, 2005)

3.3.Compound optimization in drug discovery

3.3.1.1. Thermodynamic Optimization Plot

In general, the thermodynamic optimization plot enables the correlation of structural activity with enthalpy, entropy, and affinity in a single graph (Freire, 2009). This information enables the mapping of compounds and the correlation of changes in the thermodynamic signature with changes in each molecular region. This may be visually instructive in terms of deciding which functional groupings should be implemented and which should be avoided. For instance, if a certain functional group at a particular region on the molecule produces beneficial results, the functional group should be implemented. On the other hand, if a functional group in a certain area result in a penalty, it should be avoided.

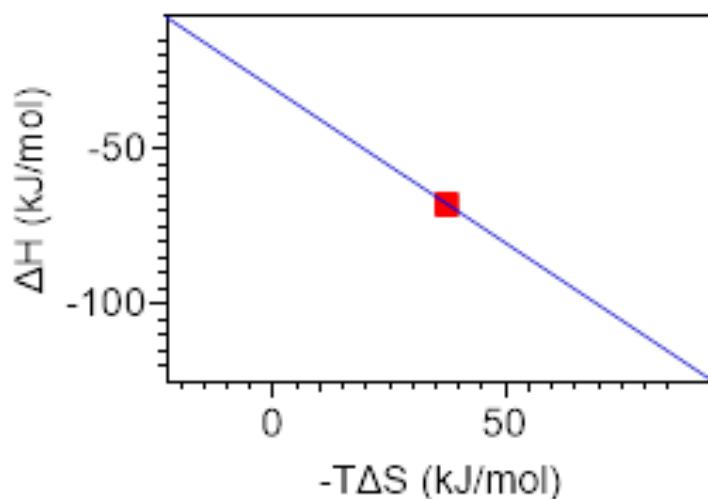


Figure 3.3-1 A **Thermodynamic Optimization Plot**.

An example of a typical Thermodynamic Optimization Plot, with the only lead chemical in the center and a diagonal optimization line.

Thermodynamic optimization plot is easy to implement, all is needed is enthalpy (ΔH) on the y-axis, and $-T\Delta S$ on the x-axis. The enthalpy and entropy data of the lead compound is placed on the graph and line is drawn through it as shown above. The optimization line intersects $(-T\Delta S, \Delta H)$ of the lead compound and a point at $(0, \Delta G)$. $(0, \Delta G)$ is obtained using $\Delta G = \Delta H - T\Delta S$ (Equation 3.3-1) and setting $-T\Delta S = 0$ (Equation 3.3-2) giving $\Delta G = \Delta H$ (Equation 3.3-3). This results in a line with a negative slope of one. In our study (Chapter 6), ADPr was used as the "lead" compound to compare alternative nucleotides with slightly different structures and different binding affinity.

$$\Delta G = \Delta H - T\Delta S \quad \text{Equation 3.3-1}$$

$$-T\Delta S = 0 \quad \text{Equation 3.3-2}$$

$$\Delta G = \Delta H \quad \text{Equation 3.3-3}$$

The optimization line is a simple line that is not fitted to any equation but only intersects $(-T\Delta S, \Delta H)$ of the lead compound and at $(0, \Delta G)$. If there are any compounds on the optimization line, that means it has a similar binding affinity same as of a lead compound but different enthalpy and entropy.

Compounds that are above the optimization line have a lower binding affinity, whereas those that are below the optimization line have a higher binding affinity. The graph is further classified by drawing horizontal and vertical lines across the lead compound. This distinguishes the contribution of enthalpy and entropy to affinity, allowing for the assignment of particular chemical modifications to a limited category. The functional group under this category may then be optimized further, implemented as-is, or ignored entirely.

The optimization plot is more helpful when there are many compounds with multiple modifications. These advantageous changes may then be pooled and evaluated to improve the design of the molecule. The thermodynamic optimization plot, as defined by (Freire, 2009), serves as a platform for identifying the region and type of compound modification that increases enthalpy and entropy contribution, thus increasing binding affinity.

Quadrant	Entropy	Enthalpy	Binding affinity
1	Better	Worse	Better
2	Better	Worse	Worse
3	Worse	Worse	Worse
4	Worse	Better	Worse
5	Worse	Better	Better
6	Better	Better	Better

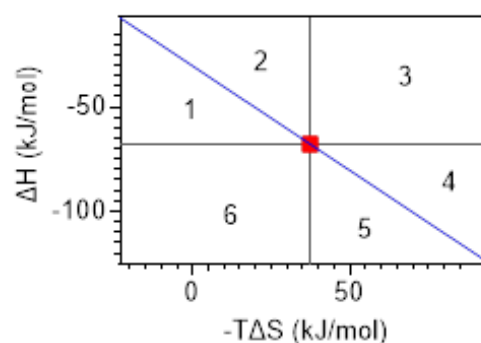


Table 3.3-1 **Summary of each quadrant of the thermodynamic optimization plot.**

3.3.1.2. LipE Plot

The logD value or lipophilicity is obtained empirically by measuring the distribution of molecules between the hydrophilic and hydrophobic phases at a physiologically appropriate pH of 7.4. Water and octanol are usually used as the hydrophilic and hydrophobic phases, respectively. Lipophilicity is a metric that quantifies the hydrophobicity, polarity, and ionic interaction of molecules. Molecular hydrophobicity is favored by dispersion or van der Waals forces, while molecular ionic interaction due to attractive/repulsive forces or polarity due to hydrogen bonding or inductive forces are favored by hydrophilicity. Typically, lipophilicity is obtained for a molecule that has not yet been synthesized. In these situations, it is possible to

predict computationally and then measure empirically after synthesizing the analogs. The calculated lipophilicity of a compound is denoted by the abbreviation cLogP (Lipinski et al., 2001).

Numerous publications discuss the optimum lipophilicity range for producing "drug-likeness," since lipophilicity influences not only ligand-protein affinity, but also absorption, distribution, metabolism, and excretion (ADME). According to Johnson et al. (2018), the optimum molecular LogD value for improving ADME features in an oral medicine is between 1 and 3. Because the goal of drug design is to optimize ADME properties while simultaneously maximizing potency, it is critical to assess potency in relation to lipophilicity.

Lipophilic efficiency (LipE) is a measure that normalizes potency according to the molecule's lipophilicity (T. W. Johnson et al., 2018). LipE can be estimated using the following equations:

$$LiPE_{(P)} = -Log(IC50) - cLogP \quad \text{Equation 3.3-4}$$

$$LiPE_{(D)} = -Log(EC50) - LogD \quad \text{Equation 3.3-5}$$

Usually, *LipE* along with *clogP* and potency are visualized together on single plot, all is needed is *clogP* on the x-axis, potency on the y-axis and LipE visualized as a diagonal line between x-y-axis with slope of 1. For example: $Y = mx + b$ is an equation for a linear line. $Y = (1)x + 1$ would identify as LipE of 1 and $Y = 1x + 2$ would identify as LipE of 2. A LipE Plot is shown below:

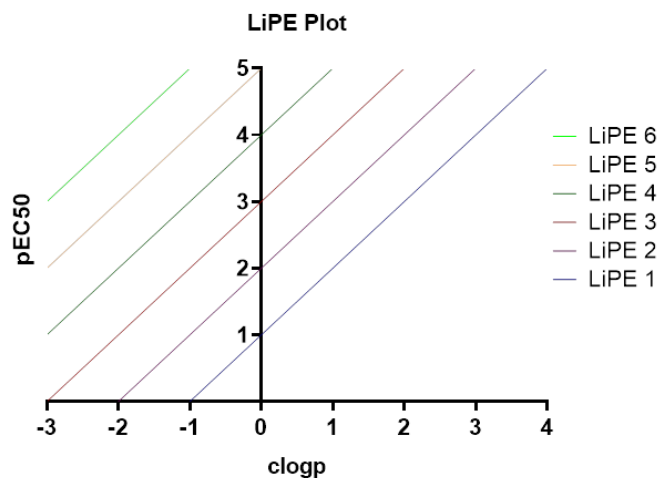


Figure 3.3-2 An example of LIPE Plot.

Illustrates an example of a typical LipE plot with calculated lipophilicity (cLogP) on the x-axis and pEC₅₀ on the Y-axis.

LipE is a measure that is used to assess the drug-likeness of a molecule by correlating its potency with lipophilicity (Lipinski et al., 2001). A therapeutic molecule's desirable characteristic is a balance of high potency and low clearance. While high potency reduces the likelihood of nonspecific binding, low clearance, in part defined by lipophilicity, has an impact on how a chemical interacts in a number of biological processes or on ADME properties (Lipinski, 2000). Additionally, Shultz (2013) published an article establishing a connection between LipE and enthalpy-driven binding (Shultz, 2013). Consequently, it is believed that compounds having a high LipE value are likely to be driven by enthalpy and have higher specificity.

3.4.Chromatography

3.4.1. Introduction

Chromatography is a popular technique that is utilized as part of a scientific method to separate individual components from a mixture (Skoog et al., 2007) Chromatography can be used alone or in tandem with other instruments. If chromatography is used to separate and purify components that could be collected for later use, it is called preparatory chromatography. An example of a preparatory chromatography in biochemistry is the use of nickel beads to purify His-tagged fused protein. If chromatography is used in tandem with other instruments, and a small volume of sample in microliters is injected. The sample is separated via chromatography and individual components are identified using other instruments, this is referred to as analytical chromatography. An example of analytical chromatography is LCMS. In LCMS, different types of columns can be used with different solvents to separate a mixture. Shortly after the mixture is separated, it is automatically injected into the mass spectrometer for identification. Chromatography is a technique that is only used to separate a mixture. However, another technique or instruments is needed for identification (Skoog et al., 2007).

3.4.2. Mobile and stationary phase.

In chromatography different techniques are utilized to separate a mixture and a mixture is separated based on its interaction with stationary phase (Skoog et al., 2007). Every chromatography includes a mobile phase and a stationary phase. A mobile phase could either be a gas or a liquid that is utilized to carry the sample through the column (among other things). A

stationary phase on the other hand could be liquid or solid and could be arranged as a packed column, adhered on the side of the column or on a flat surface.

Thin layer chromatography is an example of planar chromatography it is one of the simplest and oldest chromatography techniques used to monitor the progress of a chemical reaction. In planar chromatography, the mobile phase moves through the stationary phase by capillary action or gravity. On the other hand, if the stationary phase is packed or adhered to the side of the tube, with the mobile phase running through it under pressure or with gravity, it is called column chromatography. Column chromatography employs either a liquid or a gaseous mobile phase. Size exclusion chromatography is one type of liquid column chromatography (Skoog et al., 2007).

The type of stationary phase one can utilize depends on the purpose of the experiment and the condition requirements of the sample. For example, thin layer chromatography utilizes a flat surface that is used to monitor the progress of a chemical reaction, and it can only be used using a liquid mobile phase. The purpose of the experiment utilizing TLC was to monitor the progress of chemical reaction and the condition requirement was that the sample be stable in liquid. Furthermore, this is an analytical chromatography technique in which the sample used is no longer viable for use. On the other hand, if the purpose of the experiment is to purify a component from a mixture by size, it would be best if a packed column is used (Skoog et al., 2007).

3.4.3. Separation of components.

The separation of mixture is based on the interaction of components with stationary and mobile phase. The components with greater affinity to the stationary phase tend to elute later than the components with less affinity. After the sample is passed through the column, it is eluted and the elution passes the detector giving a chromatogram. On a typical chromatography chromatogram, the x-axis is the time and the y-axis is the signal. The eluted components in the sample are identified as peaks and the time it takes for a sample to elute with respect to when it was injected, is called retention time (t_R).

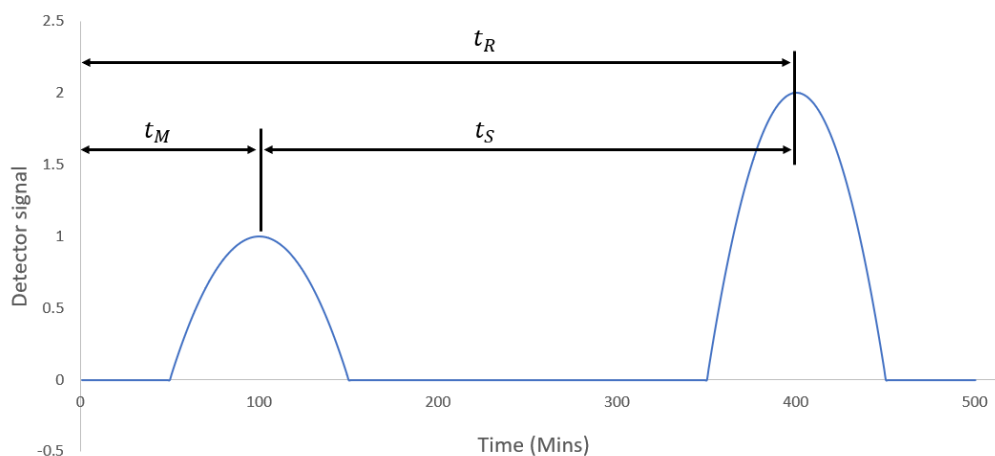


Figure 3.4-1. A typical chromatogram of a mixture containing two components.

The small peak in the chromatogram was eluted much earlier than the larger peak indicating little to no interaction with the stationary phase. Additionally, the area of each peak can be used as an indication of difference in concentration of each component eluted from column.

Figure above shows a typical chromatogram of a mixture containing two components. One of the components had little to no interaction with the stationary phase, therefore it was eluted shortly after it was injected. The elution time for a component with little to no interaction

with the stationary phase is called dead or void Time (t_M). This is an important parameter that can be used to calculate the time a component spends in the stationary phase (Skoog et al., 2007).

$$\text{Adjusted retention time: } t'_R = t_R - t_M \quad \text{Equation 3.4-1}$$

t'_R as shown in equation 3.4-1, is called adjusted retention time and it represents the actual time an analyte was retained in the column. Another experimental quantity that is widely used to compare the migration rates of solutes in columns is called retention factor (k). The retention factor is more ideal to use when comparing elution times as it doesn't depend on column geometry or volumetric flow rate. Therefore, results from different columns exhibiting different conditions should have a similar retention factor (Skoog et al., 2007).

$$\text{Retention factor: } k'_A = \frac{t_R - t_M}{t_M} \quad \text{Equation 3.4-2}$$

Retention factor (k'_A) is calculated by dividing adjusted retention time for a component by void time as shown in equation 3.4-2. Furthermore, the retention factor between two analytes from one sample can be used in comparison, providing a value called the selectivity factor (α). The selectivity factor is calculated by dividing the retention factor of the analyte that was retained in the column for a longer period of time by the retention factor of the less strongly held analyte. According to the definition, it is always the larger value of the retention factor divided by the smaller value, therefore the selectivity factor is always greater than unity. The equation used to calculate the selectivity factor is shown below.

$$\text{Selectivity factor: } \alpha = \frac{k'_B}{k'_A} \quad \text{Equation 3.4-3}$$

The selectivity factor α and the Retention factor k are typically used to calculate and to compare the resolving power or resolution of a column. The resolving power or the resolution of a column

is based on the spacing and the width between two peaks. This brings in the idea of column efficiency (Skoog et al., 2007)

3.4.4. Calculations for Column efficiency.

Column efficiency is the measure of the dispersion of a peak. It is usually explained using the concept of theoretical plates. This model assumes that a column is split into a large number of plates, with the sample being equilibrated between the stationary and mobile phases within each plate. The analyte descends the column as the equilibrated mobile phase is transferred from one plate to the next. A column with a high efficiency would have a high number of theoretical plates, so more peaks can be separated. The number of theoretical plates is related to the length of the column and the height of the individual plates, as shown in equation 3.4-4.

$$\text{Numbers of Plates: } N = \frac{L}{H} \quad \text{Equation 3.4-4}$$

The number of plates in a column is directly proportional to the length of the column (L) and inversely proportional to the height of each plate (H). Since the plate is theoretical and not visible to the eye, it is suggested that the height of the theoretical plate can be visualized and measured using a peak shaped like a gaussian distribution. The suggested measurement of plate height is proportional to the variance squared divided by the length of a column.

$$H = \frac{\sigma^2}{L} \quad \text{Equation 3.4-5}$$

Unfortunately, Equation 3.4-5 cannot be applied directly to an experimental chromatogram since the length of a column is not one of the variables. Therefore, it needs to be converted. One can obtain the variance of a solute peak in terms of time (τ) and relate it to the variance in terms of

length (σ). The proposed correlation between the variance in terms of time (τ) and the variance in terms of length (σ) is as follows:

$$\tau = \frac{\sigma}{L/t_R} \quad \text{Equation 3.4-6}$$

L/t_R is the average linear velocity of the solute. Using figure 3.4-3 (τ) can be approximated as:

$$W = 4\tau \quad \text{Equation 3.4-7}$$

Solving equation 3.4-7 for (τ) and replacing (τ) in equation 3.4-6 and solving for (σ) gives:

$$\sigma = \frac{LW}{4t_R} \quad \text{Equation 3.4-8}$$

Replacing variance (σ) in equation 3.4-5 with Equation 3.4-8 gives:

$$H = \frac{LW^2}{16t_R^2} \quad \text{Equation 3.4-9}$$

Lastly plate height in equation 3.4-4 is replaced with equation 3.4-9 giving us:

$$N = 16 \left(\frac{t_R}{W} \right)^2 \quad \text{Equation 3.4-10}$$

Equation 3.4-10 provides a way of calculating the number of theoretical plates (N) from an experimental chromatogram using the width (W) and retention time (t_R) of an analyte peak. (Skoog et al., 2007)

3.4.5. Factors affecting column efficiency

So far, it has been discussed how the analyte peak's shape may be utilized to determine the number of theoretical plates. However, the factors that can influence the peak's shape have yet to be discussed. The peak of an analyte is generally shaped like a gaussian distribution, in which the band width is attributed to random fluctuations in the duration that the analyte spends in the stationary or mobile phase. The Van Deemter equation is focused on identifying the causes

of random fluctuation, as well as how these factors vary depending on flow rate and their influence on theoretical plate height. The Van Deemter equation 3.4-11 that relates the flow rate to theoretical plate height.

$$H \cong A + \frac{B}{u_x} + Cu_x \quad \text{Equation 3.4-11}$$

According to Van Deemter Equation, the height of the theoretical plates is affected by the flow rate u_x due to: Multiple path term (A), longitudinal Diffusion (B), and Resistance to mass transfer (C) (Skoog et al., 2007).

The multiple path term in the Van Deemter equation assumes that a sample while passing through the column takes multiple paths. If there are irregularities in the path, the path length will vary and the analyte will spread, increasing the width of the band and elution time, and therefore raising the value of the multiple path term. This could be minimized if the column is packed properly without any irregularities in the bed and if the smaller bead size is used. Multiple path term only applies to packed columns, typically in liquid chromatography. Additionally, multiple path term is independent of the flow rate, as shown in Figure 3.4-4.

The term longitudinal diffusion (B) defines how a sample or analyte diffuses as it travels through a column. Diffusion rate is strongly reliant on mobile phase speed. The higher the flow rate, the less time a sample must diffuse. Longitudinal diffusion can be reduced by raising the flow rate so that components have less time to diffuse across the column. Figure 3.4-4 depicts the longitudinal diffusion term and its impact on the height of the theoretical plate as flow rate increases. This term is almost insignificant in liquid chromatography it is primarily used in gas chromatography.

Lastly, resistance to mass transfer consists of two terms: stationary phase and mobile phase mass transfer. When applied to a column with a solid stationary phase and a liquid mobile phase, the stationary phase mass transfer term is directly proportional to the time required for a species to be adsorbed and desorbed, whereas the mobile phase mass transfer is proportional to the square of the particle diameter of the packing material. To simplify, resistance to mass transfer is based on the equilibration time of an analyte and it is highly dependent on the flow rate. If the flowrate is very high, not all particles would have enough time to interact with the stationary phase and some would even leave behind causing an increase in band width. The adsorption and desorption of the sample can be controlled with temperature, decreasing the equilibration time by increasing the temperature would lead to smaller resistance to mass transfer term.

To summarize, the Van Deemter equation suggests that to reduce plate height and increase the number of plates, one can use packing material with smaller diameters and choose an optimum flow rate that is slow enough for the entire sample to interact with the stationary phase but fast enough to avoid unnecessary diffusion.

3.4.6. Calculations for resolution from a chromatogram.

Resolution is how well our peaks are separated. The basis of calculating resolution is the estimate of standard deviation of the peak with respect to time. Resolution or resolving power for a column can be calculated using the equations below:

$$R_S = \frac{2[(t_R)_B - (t_R)_A]}{W_B + W_A} = \frac{\Delta t_R}{W_{av}} \quad \text{Equation 3.4-12}$$

The retention time between two peaks is compared using Equation 3.4-12, which divides the retention time by the sum of the widths of each peak. Equation 3.4-12 may be simplified further by dividing both the top and bottom by 2, with the denominator becoming the average width of the two peaks and the numerator change in retention time.

3.4.7. Ion Exchange Chromatography

Ion exchange chromatography is one of the ways of separating a mixture into its components. It is a form of liquid column chromatography in which the mobile phase is liquid, and the stationary phase is a solid packed into a column. Ion exchange chromatography in our studies is used to extract proteins of interest from a mixture containing contaminants such as other proteins or nucleotides.

Ion exchange chromatography separates mixtures based on charge. Therefore, the number of options that are available in ion exchange chromatography is governed by the extent to which the charge can be manipulated on the protein. Proteins are delicate molecules that require specific buffer conditions to remain active and avoid denaturation. As a result, it is crucial to establish these conditions in order to avoid protein damage (Marshak, 1996)

The very first factor that needs to be considered before utilizing protein purification via ion exchange is the pH range for the protein of interest. An initial rough estimate can be made based on the nature of the natural environment of the protein. For a protein from a cell at physiological pH, a pH of 7.4 is a good guess and the choices of buffers can then be determined. Empirical activity experiments at pH of 7.4 must be performed using different choices of buffers to determine the best buffer with the most activity. Furthermore, more experiments need to be

done to test the stability of proteins under different pHs to determine the pH range (Marshak, 1996).

The charge on the protein is the next parameter that must be obtained. This is influenced by the pH of the buffer and the protein's isoelectric point (pI). The isoelectric point of a protein is ruled by the amino acid content. This may be calculated by averaging the pK_a values of each amino acid within the protein, or sometimes it can be obtained empirically. The protein is positively charged if the pH of the buffer is lower than the pI. The protein is negatively charged if the pH of the buffer exceeds the protein's pI.

The following step is to identify which type of stationary phase is appropriate for our protein of interest. The stationary phase for ion exchange chromatography is composed of a polymer material having functional groups attached to the surface. For cation exchange, the functional groups are acidic, while for anionic exchange, they are basic. If the protein's surface charge is positive, a cation exchange stationary phase is employed, and if it is negative, an anion exchange column is used. It is also advised to employ a buffer with a buffering ion with the same charge as the charge on the column matrix. The TRIS buffering ion, for example, is positively charged and is a suitable candidate for anion exchange chromatography because it does not compete with negatively charged proteins for positively charged stationary phase.

So far, we've classified the optimum pH range and buffer compatibility, which would allow proper interaction between the matrix and the protein while also maintaining protein stability. The next step is to identify the ideal conditions that can be used to remove the protein from the matrix. As previously described, in ion exchange chromatography, the protein is separated based

on the surface charge. The greater the charge difference, the tighter the binding. The pH of the buffer regulates (to a certain extent) the strength of the interaction and aids in protein binding to the matrix. However, once the protein is bound to the matrix, salt is used to overcome the matrix-protein interaction and to replace the protein with its ion. Mini purification procedures in an Eppendorf tube can be used to experimentally test the range of salt concentrations needed to elute a protein of interest.

Typically, a 1.5ml Eppendorf tube with 1 ml of matrix and a total 1.5 ml sample size can be used with a variation in salt concentration. After spinning the Eppendorf tubes, the supernatant from the top can be collected and tested for protein activity. Typically, the protein should be absent at a low salt concentration and present at a high salt concentration. If the results do not give good yield or separation, they can be further modified by using different types of salt, changing the buffer pH or by using a stronger or weaker ionic matrix. After the necessary optimizations are made, the protein of interest can then be separated from a mixture and elution with the protein of interest can then be detected using a UV detector for preparatory ion exchange chromatography (Marshak, 1996)

Ion exchange chromatography separates a mixture based on charge. The charge of a species depends on pH, and the pH determines the choices of available buffers that can be used. From the multiple choice of buffers, buffers with the same charge on buffering ion as the matrix are recommended. Lastly, a salt gradient or isocratic conditions can be used to purify and extract proteins of interest.

3.4.8. Gel Filtration Chromatography

Gel filtration chromatography, or size exclusion chromatography, is another term for this technique. As the name implies, the mixture is separated according to the size of the components. The larger molecules are eluted faster than the smaller molecules. This is due to the longer path taken by the smaller molecule through the stationary phase. After the sample mixture is loaded, the sample is pushed by the mobile phase where smaller molecules are pushed into the pores of the porous beads and larger molecules are pushed into the space between the beads. The space between the beads is referred to as the void volumes. Molecules travelling in void volume tend to travel at the same speed as the mobile phase. The large molecule's elution time may be used to determine the spacing between the beads, whilst the smaller molecule's elution can be used to calculate the column's total volume. Typically, the void volume of the majority of gel filtration resins is $1/3^{\text{rd}}$ of the resin's total packed volume (Marshak, 1996)

3.4.9. Gel Filtration Selection of stationary phase.

Gel filtration chromatography separates mixtures by utilizing the pores in the matrix hence, the resin is the most critical component of gel filtration. The type of resin that may be employed is dictated by the equipment, the size of the target protein, and the desired resolution. Before utilizing the protein purification via size exclusion, it is important to determine the optimum resin. There are different resins available with variable range of particulates that can be separated. The first step is to ensure that the target protein is within the resin range of purification. It is recommended that the size of the target protein is somewhere in the middle of the recommended resin range of purification. Next it is to determine the size of the beads.

Usually, smaller beads tend to give better separation however they also have higher back pressure with slow flow rate. therefore, it depends on the equipment and the backpressure a column can handle. Additionally, all beads have a recommended range of pressure, exceeding the range of pressure may destroy the beads, the packing of the column or both. Lastly, there are different types of beads available with composition implemented hydrophobic and hydrophilic properties.

3.4.10. Gel Filtration Tips for better separation.

Lastly, when utilizing gel chromatography for protein purification, certain things can be done to improve the resolution or separation. Among them is column packing. A gel filtration column must be evenly packed without bubbles. Having bubbles in the column would increase the width of the elution peak. As mentioned before, the choice of beads should be based on the size of the target protein and the size of the target protein should be in the middle of the recommended range of purification. However, this is not always true. It is possible for a molecule smaller in size to elute faster than a larger molecule. This is possibly due to the shape of the molecule. The better way of determining the optimum beads for purification is to run the sample and the target protein should elute somewhere in the middle of the void and total volume. Another thing that can impact the resolution is the dimensions of the column. To improve the resolution, it is usually recommended to use a longer and taller column than a wider column (Marshak, 1996)

3.5.Förster resonance energy transfer (FRET).

Fluorescence resonance energy transfer or Förster resonance energy transfer (FRET) is a method that makes use of fluorescence to identify the interaction of two molecules (Lakowicz, 2006a). The level of interaction is identified based on the amount of energy transferred. The energy is transferred without the use of photons and is the consequence of a long-range dipole-dipole interaction between two molecules. These two molecules are identified as donor and acceptor. The donors and acceptors are chosen depending on their excitation and emission characteristics (Piston & Kremers, 2007). The donor's emission should be at the same wavelength as the acceptor's absorption. As a result, as the donor is about to emit, the energy is transferred to the acceptor, and the acceptor, rather than the donor, emits light (Lakowicz, 2006b). The quantity of energy transferred between two molecules is inversely proportional to their distance. As a result, if two molecules contact, the interaction may be detected using FRET. FRET is characterized by a decline in donor and a gain in acceptor emission fluorescence.

3.5.1. FRET as a molecular ruler

As mentioned above, FRET is utilized to identify interactions between two fluorescently labeled molecules, although it cannot be directly quantified. FRET or FRET efficiency (E_{app}) is determined by monitoring the magnitude of the acceptor's and donor's intensity changes (Lakowicz, 2006b). FRET efficiency (E_{app}) is defined in general as the proportion of excited donors that transfer energy to the acceptor. More specifically, FRET efficiency is calculated using the equation below.

$$E_{app} = \frac{R_0^6}{R_0^6 + r^6} \quad \text{Equation 3.5-1}$$

FRET efficiency is shown to be equal to the ratio of Forster radius (R_0) divided by the sum of Forster radius (R_0) and the distance between acceptor and donor (r). This equation is significant because it establishes a relationship between changes in the magnitude of fluorescence and the distance between acceptor and donor, and therefore between two molecules.

Forster Radius (R_0) used to calculate FRET Efficiency represents the distance at which half the excited donor molecule transfers energy to the acceptor. It is calculated specifically using the equation below:

$$R_0 = 9.7 \cdot 10^2 (\kappa^2 \cdot n^{-4} \cdot j \cdot q_d)^{1/6} \quad \text{Equation 3.5-2}$$

Forster Radius is affected by the donor's orientation factor (κ^2), index of reflection (n^{-4}), overlap integral (j), and quantum yield (q_d) (Lakowicz, 2006b; RM, 1995). The orientation factor refers to the angle of the dipole moment between acceptor and donor. The orientation factor ranges between 0 to 4. If the dipole moments of the acceptor and donor are perpendicular, the value is zero if they are parallel, the value is four. In other words, if the dipole moments are aligned, the transfer of energy becomes very efficient. Typically, a constant value of 2/3 is used for the orientation factor. Overlap integral (j) is the calculated overlap of the area between donor fluorescence emission and acceptor excitation. The greater the overlap, the more efficient the energy transfer is (Lakowicz, 2006a).

To recap, FRET is a technique used to determine the interaction of two molecules. This interaction is quantified in terms of energy transfer between acceptor and donor. The acceptor and donor are chosen based on the maximum overlap of donor emission and acceptor excitation, as well as geometry, a high quantum yield for the donor and a high extinction coefficient for the acceptor. FRET can only detect an interaction between an acceptor and a donor if they are within 10 nm of one another (Lakowicz, 2006b). This means that a solution containing both acceptor and donor at concentrations of up to 1 mM can be used before both donor and acceptor are less than 10 nm apart introducing significant background, making this technique useful in studying molecular interaction in biological studies.

3.5.2. FRET Efficiency.

As previously stated, FRET is a process in which energy is transferred from donor to acceptor. FRET is determined indirectly by comparing the magnitude of the donor and acceptor's fluorescence. If there is an interaction, the donor's fluorescent emission will decrease and the acceptor's fluorescent emission will increase. As a result, FRET efficiency may be determined solely by the change in the magnitude of the donor or acceptor fluorescence intensity (Lakowicz, 2006a).

If the FRET efficiency is measured solely based on the increase in the magnitude of the acceptor fluorescence, it is called sensitized emission. In a perfect world, one would excite the donor and the increase in acceptor emission would be a pure measurement of the transfer of energy. However, this is not true. The observed change in the magnitude of acceptor emission is contaminated by the direct excitation of the acceptor and bleed through of the donor

emission into the acceptor detection channel. In the studies presented here, FRET efficiency was calculated based only on donor quenching with negligible bleed through of the acceptor at this observed wavelength therefore corrections that are typically needed for sensitized emission are no longer necessary. FRET efficiency based on donor quenching is calculated using the following equation:

$$E_{app} = \left(1 - \frac{F_{DA}(\lambda_D^{ex} \lambda_D^{em})}{F_D(\lambda_D^{ex} \lambda_D^{em})} \right), \quad \text{Equation 3.5-3,}$$

where $F_{DA}(\lambda_D^{ex} \lambda_D^{em})$ is the emission of the donor in the presence of acceptor integrated over the entire emission wavelength range; similarly, $F_D(\lambda_D^{ex} \lambda_D^{em})$ is the integrated emission of the donor in the absence of the acceptor. Using the expressions $F_{DA}(\lambda_D^{ex} \lambda_D^{em}) = k_{DA} w^D$ and $F_D(\lambda_D^{ex} \lambda_D^{em}) = k_D w^D$ with k_{DA} and k_D being adjustable parameters and w^D is the area under the normalized donor emission we obtain (Raicu et al., 2009):

$$FRET_{app} = \left(1 - \frac{k_{DA}(\lambda_D^{ex} \lambda_D^{em})}{k_D(\lambda_D^{ex} \lambda_D^{em})} \right) \quad \text{Equation 3.5-4}$$

In our analysis hereafter, in the absence of spectral unmixing of the composite donor plus acceptor emission spectra, we make the approximation that k_{DA} is equal to the fluorescence intensity detected at maximum of the donor emission. After calculating $FRET_{app}$, the data is fitted to equation 3.5-5 defined in Raicu et al. (Raicu & Singh, 2013) using non-linear regression (GraphPad Prism). Fits were used to estimate pairwise FRET efficiency (E_{AD}), based on total concentration of the donor (D_T), and the concentration of the complex ($[AD]$). The concentration of the complex is calculated using equation 3.5-11.

$$FRET_{app} = \frac{1}{D_T} \cdot ([AD] \cdot E_{DA}) \quad \text{Equation 3.5-5}$$

The equation above is used to calculate FRET between donor acceptor under the assumption that the interaction is between one donor and one acceptor and the donor always binds next to the acceptor within in the allowable space (Raicu et al., 2007).

3.5.3. Estimating complex concentration [AD].

The interaction between helicase and DNA was monitored using FRET. The helicase is part of HCV NS3 fused with yellow fluorescent protein (YFP-NS3h) titrated with fluorescently labeled single stranded or partially double stranded oligonucleotide. YFP-NS3h used in this study is identified as a donor while fluorescent molecule on the oligonucleotide is an acceptor.

Titration were performed by adding 1-2 μ l of fluorescently labeled oligonucleotide to YFP-NS3h in the cuvette. The sample is excited at 450 nm and an initial emission spectrum is collected, with subsequent spectra taken after adding oligonucleotide. Typically, if there is an interaction between helicase and oligonucleotide the emission of the donor YFP-NS3h would decrease with an increase in acceptor emission. The data is then collected as a batch of spectra with a decrease in donor emission and an increase in acceptor emission. We assume that the observed decrease in donor emission is due to the helicase binding with the oligonucleotide. The correlation of acceptor and donor binding and forming a complex can be interpreted as:



Equation above, does not account for all different type of interaction possible between acceptor and donor but assumes that only one form of donor acceptor complex is formed when n donors bind to one acceptor. The “n” in equation 3.5-6 is the stoichiometry coefficient.

Unfortunately, equations above cannot be applied as-is since it is impossible to measure the concentration of each free species in the cuvette. However, while performing the experiment, the total concentration of donor and acceptor is known. Using the known total concentration of donor and acceptor added to the cuvette, the concentration of free donor and acceptor can be calculated using the relationship below.

$$[A] = [A_T] - [AD_n] \quad \text{Equation 3.5-7}$$

$$[D] = [D_T] - n[AD_n] \quad \text{Equation 3.5-8}$$

Furthermore, the interaction between acceptor and donor is identified by calculating the binding affinity or dissociation constant as shown in Equation 3.5-9 and Equation 3.7-10 respectively.

$$K_a = \frac{[AD_n]}{[A][D]^n} \quad \text{Equation 3.5-9}$$

$$K_d = \frac{[A][D]^n}{[AD_n]} \quad \text{Equation 3.5-10}$$

In our studies, dissociation constant K_d is typically used which is the reciprocal of binding affinity K_a . Equation 3.5-8 through 3.5-10 were then used to derive the Equation 3.5-11 to calculate complex concentration in the cuvette from the known total acceptor and donor concentration added to the cuvette.

$$[AD_n] = \frac{(K_d + [A_T] + n[D_T]) \pm \sqrt{(K_d + [A_T] + n[D_T])^2 - 4n[D_T][A_T]}}{2} \quad \text{Equation 3.5-11}$$

Lastly, complex concentration $[AD]$ in Eq. 3.5-5 is replaced with Eq. 3.5-11, where FRET efficiency data from Eq. 3.5-4 is fitted in Eq. 3.5-5 and pairwise FRET efficiency is then estimated.

3.5.4. FRET chapter summary.

All FRET experiments are done in a cuvette containing 500 µl of sample with donor (YFP) covalently bonded to the helicase from HCV NS3. The sample is excited at 450 nm and the emission spectrum is collected with maximum emission at 520 nm. In the same cuvette, YFP-NS3h is titrated with 1-2 µl of acceptor (TAMRA, MAX, Hexachlorofluorescein or TYE) labeled oligonucleotide and subsequent spectrum is collected. An interaction between helicase and oligonucleotide is assumed if decrease in donor emission at 520nm is observed. Using Eq. 3.5-4, the $FRET_{app}$ is calculated. Where $k_D(\lambda_D^{ex} \lambda_D^{em})$ is the maximum fluorescence emission at 520 nm from the sample containing only the donor and $k_{DA}(\lambda_D^{ex} \lambda_D^{em})$ is the maximum fluorescence emission at 520 nm from the sample containing both donor and acceptor. The calculated $FRET_{app}$ using Eq. 3.5-4 along with the acceptor concentration is then plotted and the data is fitted to Eq. 3.5-12 or 3.5-5 to estimate pairwise FRET efficiency (E_{DA}).

$$FRET_{app} = \frac{1}{D_T} \cdot \left[\left(\frac{(K_d + [A_T] + n[D_T]) \pm \sqrt{(K_d + [A_T] + n[D_T])^2 - 4n[D_T][A_T]}}{2} \right) \cdot E_{DA} \right]$$

Equation 3.5-12

FRET is assumed based on the estimated value of E_{DA} . An experiment with lower value for E_{DA} is assumed to conduct less FRET than an experiment with higher E_{DA} and since FRET is used to estimate interaction between helicase and oligonucleotide, with more FRET we assume that the helicase is binding closer or near the acceptor whereas with less FRET, helicase is binding farther away from the acceptor.

Additionally, it's important to emphasize that the FRET theory used to fit the data using Eq 3.5-12 or Eq. 3.5-5 assumes that only one helicase molecule binds to the oligonucleotide at a time, i.e. the helicase and oligonucleotide forms heterodimers. At the same time, in relating AD_n to the dissociation constant, we allowed the heterooligomer size to take an arbitrary value n . To reconcile this two apparently contradicting assumptions with one another, we would add the third assumption, which is that the helicase that binds closest to the acceptor is the only one involved in FRET; any additional helicase up to order n is assumed to be so far away that it does not transfer energy to the donor. (Please refer to reference (Raicu et al., 2007) for additional details regarding the FRET theory.) As such, this analysis is to be regarded as preliminary, and future work should attempt to refine the FRET theory.

3.6. Differential scanning fluorimetry (DSF)

Typically, the DSF or thermal shift assay is used to detect the interaction of a protein and a ligand. The interaction between protein and ligand is quantified on the assumption that the protein is stabilized upon binding. The stability of the protein is then determined by measuring the temperature at which it unfolds, and the unfolding of the protein is determined by the degree of the fluorescence change. The observed fluorescence is due to a linear correlation between unfolded protein and the SYPRO orange dye's measured fluorescence. The change in fluorescence of the SYPRO orange dye employed in the test is proportional to the concentration of hydrophobic groups exposed during protein unfolding (Chaires, 2008; Ericsson et al., 2006).

The raw data is simply the change in fluorescence as a function of temperature, with fluorescence plotted on the Y-axis and temperature recorded on the X-axis. The data acquired initially exhibits a low fluorescence, but as the temperature rises, the protein unfolds, increasing the fluorescence. Numerous samples with varying ligand concentrations give data relating the unfolding temperature to the ligand concentration. Multiple samples' data are then displayed on a single graph, clearly illustrating the relationship between melting temperature and ligand concentration. This is useful in determination of EC_{50} or compound needed to cause a change in the melting temperature by 50% (EC_{50}). The determination of EC_{50} is discussed in detail later in Chapter 3.8.2.

3.6.1. Estimating melting temperature (T_m)

The raw data collected from the instrument with fluorescence on the y-axis and temperature on the X-axis typically presents a sigmoidal curve. The inflection point in the middle

of the sigmoidal curve corresponds to the halfway of the protein unfolding process, commonly referred to as the melting temperature (T_m). T_m is at the inflection point with the steepest slope, which could be calculated by taking the first derivative of the curve. Unfortunately, this method doesn't differentiate between noise, data, or an artifact with tendencies of providing false T_m values. Therefore, curve fitting method proposed by Ng *et al.*(2019) was used via a software called TSA CRAFT (<https://sourceforge.net/projects/tsa-craft/>). The curve fitting model based on the Boltzmann equation is a sigmoidal function that is fitted to the sigmoidal curve from the raw data providing T_m values. TSA-CRAFT, is open source and takes raw data from the instrument modified into a CSV file as an input. The raw data as CSV file is then fitted to the Equation 3.6-1 and if the data fits the overall shape of a typical sigmoidal curve, a T_m value is then estimated.

$$F_{obs}(T) = F_{min} \frac{F_{max}-F_{min}}{1+e^{\left(\frac{T_m-T}{a}\right)}} \quad \text{Equation 3.6-1}$$

In equation above, F_{obs} is the observed fluorescence at each temperature (T), F_{min} is the minimum observed fluorescence, F_{max} is the maximum observed fluorescence, and a is the hill slope.

3.6.2. Estimation of EC_{50}

The T_m values collected from fitting Equation 3.6-1 were plotted against the ligand concentration and fitted to equation below via prism:

$$T_{m\ obs} = \frac{\Delta T_{m\ max} [L]}{EC_{50}+[L]} + T_{m\ 0} \quad \text{Equation 3.6-2}$$

Equation 3.6.2 relates $T_{m\ obs}$ to [L], EC_{50} , and $\Delta T_{m\ max}$. Data fit to this equation can be used to estimate the amount of compound needed to cause a change in the melting temperature by 50% (EC_{50}). Again, $T_{m\ obs}$ is the T_m value calculated from Equation 3.6-1 at a specific ligand

concentration $[L]$, $T_{m-\text{Max}}$ is the maximum change in T_m , and T_{m-0} is the melting temperature of Mac1 in the absence of ligand in equation 3.6-2.

3.6.3. Estimating binding affinity (K_d)

While it is straightforward to explore the interaction of a protein and a ligand using DSF, obtaining the binding constant is not as straightforward. As a result, a novel approach suggested by Bai et al. (2019) was employed to extract and calculate the binding constant from DSF data. From unfolding curves, a single temperature is chosen near the unfolding transition and the fraction of protein that is still folded. This results in constant temperature system based on protein folding/unfolding and ligand binding/unbinding coupled equilibria as shown below:



The above equation describes the relationship between a protein in its unfolded $[U]$, folded $[F]$, or complexed $[FL]$ state with a ligand $[L]$. The unfolded state of the protein can be related to the folded state using an equilibrium constant for unfolding proteins, denoted as K_U (Bai et al., 2019; Lee et al., 2019). The folded state of the protein and ligand is related to the formation of the protein-ligand complex via its equilibrium constant, denoted as K_D , or dissociation constant. Furthermore, the concentration of protein and ligand is related to the total concentration as shown below.

$$[P]_T = [U] + [F] + [FL] \quad \text{Equation 3.6-4}$$

$$[L]_T = [L] + [FL] \quad \text{Equation 3.6-5}$$

$$K_U = \frac{[U]}{[F]} \quad \text{Equation 3.6-6}$$

$$K_D = \frac{[F][L]}{[FL]} \quad \text{Equation 3.6-7}$$

Once the data is collected, it is first normalized then it is fitted to following equation:

$$f_u = \frac{[U]}{[P]_T} = \frac{[U]}{[U]+[F]+[FL]} \quad \text{Equation 3.6-8}$$

Equation 3.6-8 cannot be applied since the concentration of free ligand [F], free unfolded protein [U] or bound complex [FL] is not known. Therefore equations 3.6-4, to 3.6-7 are applied to equation 3.6-8 giving:

$$f_u = \frac{[U]}{[U]+[F]+\left(\frac{[F][L]}{K_d}\right)} = \frac{[U]}{[U]+[F]\left(1+\left(\frac{[L]}{K_d}\right)\right)} = \frac{[U]}{[U]+\frac{[U]}{K_u}\left(1+\left(\frac{[L]}{K_d}\right)\right)} = \frac{1}{1+\frac{1}{K_u}\left(1+\frac{[L]}{K_d}\right)}$$

$$\text{Equation 3.6-9}$$

Unfortunately, Equation 3.6-9 still needs to be further simplified because concentration of Free ligand [L] is not known. The concentration of Free ligand [L] is calculated using quadratic equation:

$$[L] = \frac{([L]_T - [P]_T - K_d(1+K_u)) + \sqrt{([L]_T - [P]_T - K_d(1+K_u))^2 - 4([L]_T K_d(1+K_u))}}{2}$$

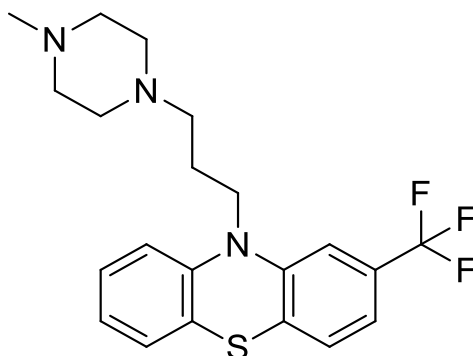
$$\text{Equation 3.6-10}$$

The proportion of unfolded protein (f_u) at various temperatures in the presence of varied ligand concentrations may then be used to determine the dissociation constant using equation 3.6-9. This equation is then plotted, with the fraction of unfolded protein on the Y-axis and ligand concentration on the X-axis, and each point on the spectra represents fraction of unfolded protein (f_u) with ligand concentration [L].

3.7. Selected Drugs from DSF Hits

Due to the recent pandemic, several labs around the world are trying to find drugs that are already approved by the FDA to expediate the clinical trials. Below selected compounds from the hits are described: Trifluoperazine (3.7.1), irinotecan (3.7.2), nifedipine (3.7.3), cefaclor and cefatrizine (3.7.4), omeprazole and rabeprazole (3.7.5):

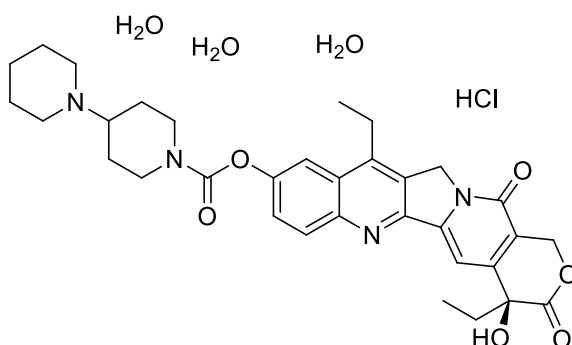
3.7.1. Trifluoperazine (antipsychotic drug)



Trifluoperazine is a widely used antipsychotic drug that is used to treat people with schizophrenia. It is effective in small doses and considered safe since 1960s. It is believed to work by blocking D1 and D2 receptors minimizing symptoms of schizophrenia (Marques et al., 2004). Xiao *at al.* performed HT screening of approximately 1700 US FDA approved compounds against human coronavirus strain OC43 and SARS-CoV-2. LLC-MK2 cells (Rhesus monkey kidney cells), was transfected with human coronavirus (HCoV) strain OC43. SARS-CoV-2 virus was propagated in Vero cells (African green monkey kidney cell). They measured IC₅₀ (the Half-Maximal Inhibitory Concentration), CC₅₀ (the Half-Maximal Cytotoxic Concentration), and SI (Selectivity Index) via immunofluorescence, quantitative RT-PCR and cell-cell fusion assay. They reported

trifluoperazine 2HCl as one of compounds effective against OC43 and SARS-COV-2. They reported CC_{50} , IC_{50} and SI of >10, 3.36 and 3.08 μM respectively for OC43 and 29, 12 and 2.5 μM respectively for SARS-CoV-2.

3.7.2. Irinotecan hydrochloride trihydrate (topoisomerase 1 inhibitor)



The goal of cancer chemotherapy is to eradicate cancer cells. [note chemotherapy actually means “the treatment of disease by the use of chemical substances”]. Most cancer chemotherapy, however, uses cytotoxic agents meaning they are toxic to all living cells and their mechanism of actions target essential part of cell cycle. During the cell cycle, a cell prepares itself for cell division by replicating its genome. It is a continues process with checkpoints during which the cell ensures there are no abnormalities before it progresses to next phase. During replication, the DNA double helix supercoils when it is separated into two strands. Supercoiled DNA is overwound, and if left unchanged, this torsion would eventually stop DNA replication. This torsion in supercoiled DNA is relieved by enzymes called topoisomerases. There are two types of topoisomerases called topoisomerases I and topoisomerases II. Topoisomerase II enzymes cleave only one strand of the DNA, unwinds it and attaches it closing with one less coil. Topoisomerase 2 enzymes cleave both strands of DNA relaxing supercoil during DNA replication (Champoux, 2001).

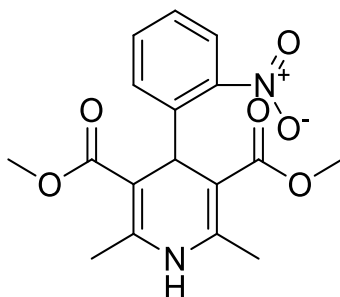
Irinotecan has been extensively studied as a topoisomerase I inhibitor since the late 1960s. It was approved for the treatment of cervical, lung, and ovarian cancer. Irinotecan inhibits topoisomerase I by binding and forming ternary irinotecan–topoisomerase I–nicked DNA complex and preventing topoisomerase release from DNA. Topoisomerases are usually active during replication and inhibition of its release cause a collision with replication forks which results in lethal double strand break. This DNA lesion acts as an initiator or apoptotic signal (Kciuk et al., 2020).

The first protein recruited by this apoptotic signal is the ATM serine/threonine kinase. The kinase then activated p53 tumor suppressor protein. P53 then activates the p53 upregulated modulator of apoptosis (PUMA) protein, which is a pro-apoptotic protein. PUMA then activates the BAX protein in cytosol and activated BAX protein binds to mitochondrial membrane with Bcl-2 homologous antagonist BAK and opens VDAC voltage dependent anorectic channels. This allows the release of Cytochrome C from mitochondria into cytosol (Kciuk et al., 2020).

After the release of Cytochrome C from mitochondria, it binds to apoptotic protease activating factor 1 APAF-1 in presence of dATP and this forms a complex called apoptosome. this apoptosome then interacts with Pro-Caspase-9 and activates it converting it into Caspase 9. Caspase 9 then activates caspase 3 which acts as an endonuclease, cleaves cytosolic and nuclear proteins and activates caspase-dependent DNase (Kciuk et al., 2020). This then leads to cleaved DNA and cell death. Since cancer cells divide more frequently, they go through cell cycle more frequently therefore they are affected more by this chemotherapy. The side effect of

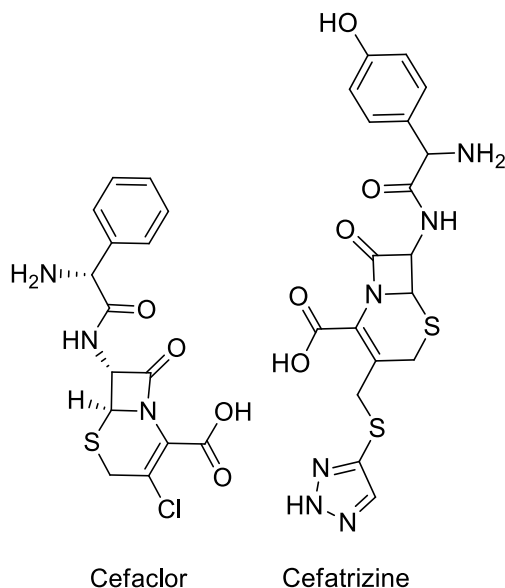
chemotherapy is caused by cells with similar growth rate as cancer cell however these side effects are usually not permanent.

3.7.3. Nifedipine (calcium channel blocker)



Nifedipine is a calcium channel blocker in the dihydropyridine subclass that is prescribed to patients as an antihypertensive medication (Pontremoli et al., 2005). During the depolarization phase of smooth muscle cells, calcium ions enter via voltage-gated channels. Calcium ions are taken up by cells, resulting in an increase in Ca^{2+} ion (Curtis & Scholfield, 2001). Intracellular calcium then initiates a cascade that causes vessel contraction. When the concentration of calcium ion (Ca^{2+}) increases, it binds to calmodulins, forming a calcium-calmodulin complex (Ca^{2+} calmodulin). After that, the calcium-calmodulin complex interacts with myosin light chain kinases (MLCK). When MLCK interacts with Ca^{2+} calmodulin, Ca^{2+} ions are transferred to MLCK, forming Ca^{2+} MLCK. Ca^{2+} MLCK is a kinase that is activated and catalyzes the transfer of a phosphate group from ATP to myosin, resulting in phosphorylated myosin. Phosphorylated myosin then interacts with actin, causing smooth muscle cells to contract (Andersson & Högestätt, 2009). Nifedipine is a calcium channel blocker that binds to voltage-gated channels, inhibiting cell Ca^{2+} uptake. A decrease in intracellular Ca^{2+} ion effectively blocks the activation of the contraction cascade, resulting in smooth muscle cell relaxation, dialysis, and decrease in blood pressure.

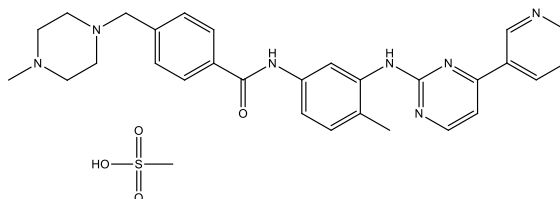
3.7.4. Lactams: cefaclor and cefatrizine (antibiotics)



Cefaclor and cefatrizine belong to broad-spectrum cephalosporin class of β -lactam antibiotics. β -lactam antibiotics contain β -lactam ring in their molecular structure. β -lactam antibiotics inhibiting the biosynthesis of cell walls which are composed of Peptidoglycan. Peptidoglycan is a stacked layer of polysaccharides consist of alternating β -(1,4) linked N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM) (Sauvage & Terrak, 2016). Stacked peptidoglycans are also cross-linked between the last glycine residue of one peptidoglycan and the terminal alanine on the other strand. Formation of peptidoglycan is catalyzed by bifunctional penicillin binding protein (PBP), with transglycosylase domain producing glycan strands and the transpeptidase domain performing the crosslinking reaction. The mechanism for crosslinking reaction includes serine at the active site forming an acyl intermediate with alanine. This acyl intermediate is then attacked by glycine forming a cross-linked peptidoglycan (Shalaby et al., 2020).

The mechanism by which the PBP transpeptidase domain conducts the crosslinking reaction is the same mechanism employed by the β -lactam family of antibiotics to inhibit this enzyme. The catalytic serine attacks the carbonyl group of the β -lactam ring, causing ring opening and formation of a stable covalent acyl-enzyme complex. The presence of an acyl-enzyme complex blocks the catalytic site, causing peptidoglycan cross-linking to be inhibited (Sauvage & Terrak, 2016). Bacterial cell walls become thinner due to a lack of cross linking between peptidoglycan. Bacteria are thus unable to withstand osmotic impact, resulting in rupture and death.

3.7.5. Imatinib methane sulfonate (tyrosine kinase inhibitor)



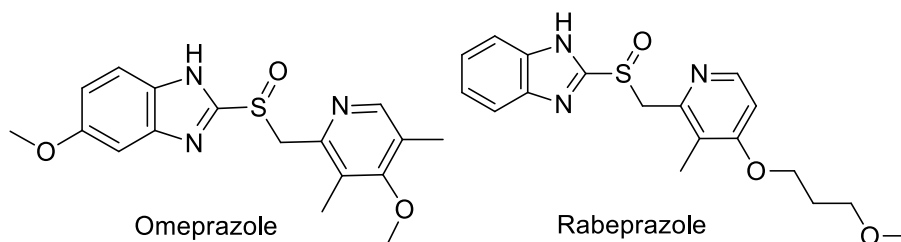
Imatinib Methane Sulfonate

Imatinib is an inhibitor specific to BCR-ABL tyrosine kinase. It is effective against chronic myelogenous leukemia (CML) and gastrointestinal stromal (GIST) (Willig et al., 2020). Both CML and GIST are caused by a genetic abnormality known as the Philadelphia Chromosome, which is a cause of an abnormally expressed tyrosine kinase, resulting in uncontrolled cell proliferation (Epstein et al., 2010).

The Philadelphia Chromosome is an abnormal chromosome 22 formed when a portion of chromosome 9 is translocated to chromosome 22. The portion of chromosome 9 that is translocated is an ABL protooncogene and it is translocated to chromosome 22 at the breakpoint

cluster region (BCR), forming the BCR-ABL oncogene (Collins & Groudine, 1983). The chromosome 22 that consists of the BCR-ABL fusion oncogene is called the Philadelphia chromosome. The BCR-ABL oncogene encodes for a BCR-ABL tyrosine kinase. Tyrosine Kinases are a family of enzymes that catalyze the phosphorylation of tyrosine residues or the transfer of phosphate group from ATP to a tyrosine residue on a target protein. This target protein may play an important role in cell proliferation, differentiation, metabolism, or apoptosis. Therefore, tyrosine kinases are mediators of the signal transduction process that leads to the activation or inactivation of important cell functions. In this case, BCR-ABL encodes for a BCR-ABL Tyrosine kinase that allows the signal transduction process in infected cells to bypass the regulatory signals controlling cell growth and differentiation (Epstein et al., 2010). This then results in uncontrolled and unregulated cell proliferation and malignant transformation to become leukemic cells, which causes the pathogenesis of CML. Imatinib is an inhibitor that specifically binds to the of BCR-ABL gene encoded tyrosine kinase. It inhibits BCR-ABL activity by competitively interacting with the ATP binding site on the tyrosine kinase blocking ATP from binding (Willig et al., 2020).

3.7.6. Benzimidazoles: omeprazole and rabeprazole (proton pump inhibitors)



Omeprazole was one of the first proton pump inhibitor approved for medical use in 1988, and it was intended to address acid-related disorders. Subsequently several other proton pump inhibitors (PPIs) were introduced including rabeprazole. Most PPIs are structurally and

pharmacologically similar. They are benzimidazole derivatives in which a benzimidazole moiety is linked to pyridine by a methyl sulfinyl group, with various substitutions on the pyridine or benzimidazole rings. All PPIs are given in an inactive form which are modified and activated in strong acidic conditions found in canaliculate of parietal cells. Parietal cells are epithelial cells in the stomach that secrete hydrochloric acid (HCl) and intrinsic factor.

The mechanism of PPI activation is as follows: Due to acidic environment, second nitrogen on benzimidazole gets protonated, forming a positive charge. Next, the lone pair on nitrogen of pyridine transfers electron and forms a bond with 2nd carbon of imidazole on benzimidazole. This leads to formation of cyclic intermediate between benzimidazole and pyridine and loss of aromatic characteristic of imidazole. Next, the lone electron on nitrogen of imidazole transfer its electrons reforming aromatic characteristic and breaking the sulfoxide ring leading to formation of sulfonic acid intermediate (Figure 3.7-1). This sulfonic acid intermediate undergoes intramolecular rearrangement to form pyridinium sulfonamide structure. This structure then forms disulfide bond with proton pump leading to permanent inactivation and blocking the release of stomach acid (Ward & Kearns, 2013).

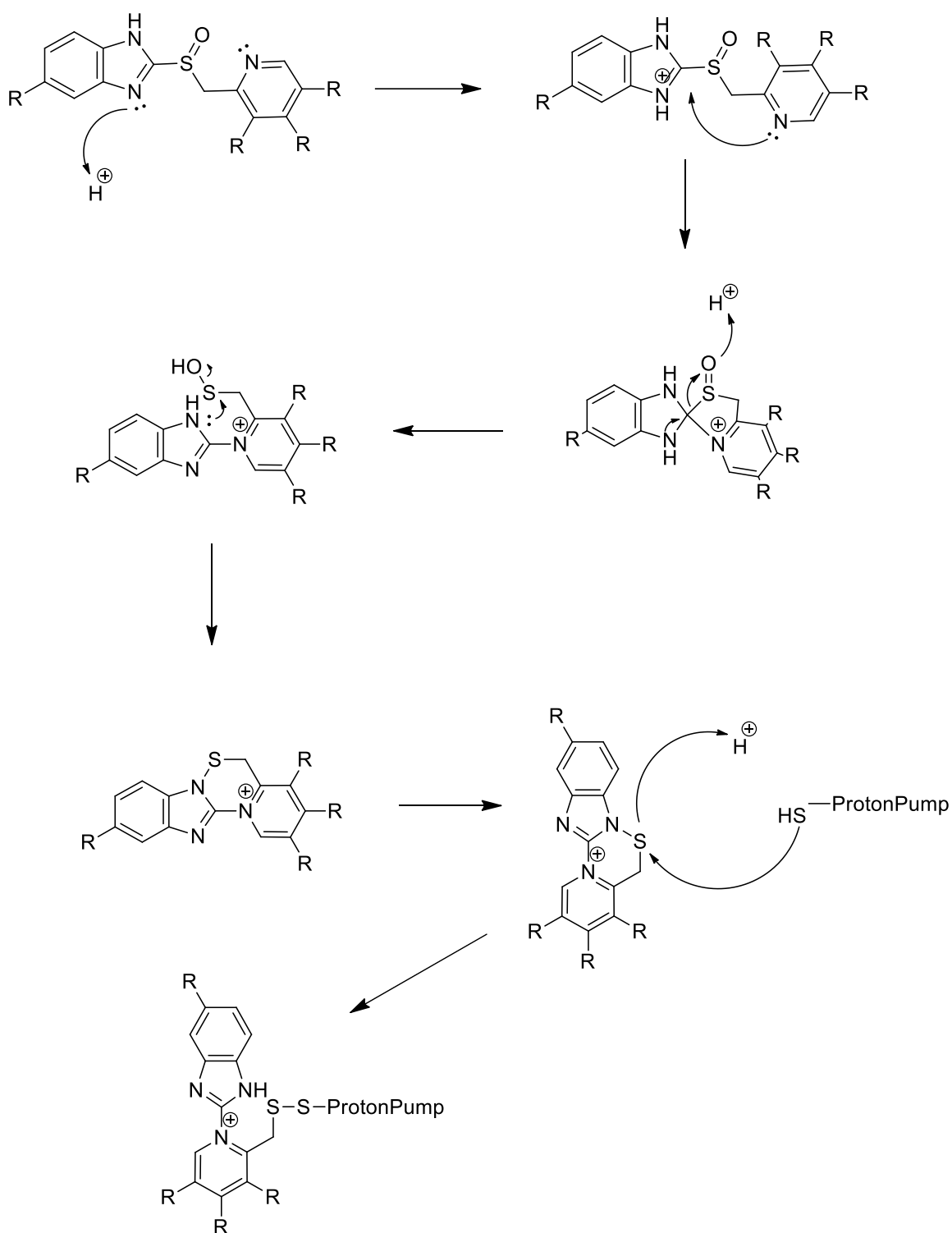


Figure 3.7-1: **PPI Activation Mechanism: Omeprazole and Rabeprazole.**

The mechanism of action of proton pump inhibitor was presented in (Ward & Kearns, 2013)

4. Materials & Methods

4.1. Gene synthesis

To facilitate comparison between SARS-CoV-1 and SARS-CoV-2, a protein expression vector was generated similar to the one used by Eglott *et al.* (2006). To this end, a codon optimized open reading frame was synthesized by GenScript (Piscataway, NJ) that encodes the Mac1 flanked by *NheI* and *BamHI* restriction sites. This open reading frame was cloned into pET21b to give plasmid pET21-COVID-Mac1. The pET11-COVID-Mac1 plasmid was used to transform BL21(DE3) cells.

4.2. Protein purification (Mac1)

Colonies of BL21(DE3) cells harboring the pET21-COVID-Mac1 plasmid were used to inoculate 3 ml of LB medium containing 100 mg/ml ampicillin. The starter culture was incubated at 37 °C with shaking at 225 rpm. After the cells grew to an OD₆₀₀ of 1.0, they were transferred to 1 liter of fresh medium containing ampicillin. After the cells reached an OD₆₀₀ of 1.0 again, protein production was induced with 1 mM isopropyl- β -D-thiogalactoside. After growing 16 h at 23 °C, the cells were harvested by centrifugation at 4,000 rpm, 4°C. The resulting cell pellet was suspended in 25 mL of IMAC buffer (20 mM Tris pH 8, 0.5 M NaCl), sonicated on ice for five 1 min bursts, with 2 min rests between, and clarified by centrifugation at 10,000g for 30 min. The supernatant was loaded onto a 5 ml Ni-NTA column and the fractions were eluted with a step gradient from 5 to 500 mM imidazole. Fractions containing the Mac1 domain protein (5 ml total) were loaded on a 250 ml Sephacryl S300 gel filtration column and eluted with 10 mM MOPS, 150 mM NaCl. Concentration of the purified protein was determined by measuring absorbance at 260

nm using a molar extinction coefficient calculated with the ProtParam tool (<https://web.expasy.org/protparam/>).

4.3. Protein purification (Nsp13-His-tag)

To produce Nsp13 with a His-tag on the N-terminal, a Nsp13 vector identical to that used by Mickolajczyk et al. was employed (Mickolajczyk et al., 2020). This pET21b-COVID-N-His-nsp13 plasmid was used to transform BL21(DE3) cells. Transformed BL21(DE3) were used to start a 3 ml starter culture that was grown overnight at 37 °C. Next day the starter culture was transferred to a 1-liter fresh medium containing 1 ml of 50 mg/ml kanamycin. After the OD600 reached 0.7, the temperature was then lowered to 23 °C (room temperature) and the culture was induced with 1 mL of 200 mM IPTG per liter of growth media. The cultures were left overnight (~16 hrs) at 23 °C and harvested the next day by spinning down at ~5000 g for 15 minutes and the pellets was resuspended in IMAC buffer (20 mM Tris pH 7, 0.5 M NaCl, 5mM imidazole). The resuspended cells were sonicated on ice for five 1 min bursts, with 2 min rests between, and clarified by centrifugation at 10,000g for 30 min. The supernatant was loaded onto a 5 ml Ni-NTA column and the fractions were eluted with a step gradient from 5 to 300 mM imidazole. The fraction collected at 300 mM imidazole was then dialyzed in GF buffer (20 mM Tris, pH 7, 50 mM NaCl, 1 mM EDTA, 0.1 mM DTT) + 20% glycerol and concentrated. The nsp13 protein (2 ml total) were loaded on a 250 ml Sephacryl S300 gel filtration column and eluted with (20 mM Tris, pH 7, 50 mM NaCl, 1 mM EDTA, 0.1 mM DTT) + 20% glycerol. The protein was concentrated again and the concentration of pure protein was estimated using an absorbance measurement at 260 nm and

a molar extinction coefficient of $68785 \text{ M}^{-1} \text{ cm}^{-1}$ calculated using the ProtParam tool (<https://web.expasy.org/protparam/>).

4.4. Isothermal Titration Calorimetry (ITC)

Binding of ADP-ribose to the SARS-CoV-2 Mac1 domain was measured using a Nano ITC (TA Instruments). Before starting the measurement, samples of both ligand and protein were diluted in 10 mM MOPS, 150 mM NaCl (pH 7) and were degassed at 400 mmHg for 30 minutes. Measurements were taken at 20 °C by injecting 2.0 μl aliquots of 500 μM ADP-Ribose (Sigma) to 50 μM protein (175 μl initial volume) with 250 rpm stirring rate. Using Nano Analyze Software (v. 3.11.0), data were fitted by non-linear regression to an independent binding model. Briefly, after baseline correction, background heats from ligand-to-buffer titrations were subtracted, and the corrected heats from the binding reaction were used to find best fit parameters for the stoichiometry of the binding (n), free energy of binding (ΔG), apparent enthalpy of binding (ΔH), and entropy change (ΔS). Dissociation constants (K_d) were calculated from the ΔG .

4.5. Differential Scanning Fluorimetry (DSF)

Assays were performed in 96-well PCR plates using an Eppendorf Mastercycler ep Realplex Quantitative Realtime PCR thermocycler with each well containing 19 μl of Master-mix and 1 μl of compound stock at 10 mM. The master-mix was prepared by adding 20 μl of 500 μM Mac1, 2.5 μl 5000x SPYRO Orange protein gel stain (Sigma-Aldrich catalog #S5692) and 1977.5 μl of buffer (20 mM MOPS, 25 mM NaCl, pH 7) with a final volume of 2 ml. The 96-well PCR plate was then sealed with clear adhesive film and centrifuged at 1,100 RPM for 5 mins. In the thermocycler, the temperature was raised from 20 °C to 95 °C at a rate of 2 °C per min while

measuring fluorescence in the “TAMRA” channel. Each plate included both negative (DMSO) and positive (ADP-ribose) controls. T_m 's were calculated either by finding the peaks in plots of dF/dT vs T , or by fitting the data to equation 1 using either GraphPad Prism or TSA-craft (<https://sourceforge.net/projects/tsa-craft/>). (Lee et al., 2019)

$$F_{obs} = F_{min} \frac{F_{max} - F_{min}}{1 + e^{\left(\frac{T_m - T}{a}\right)}} \quad (\text{Equation 1})$$

In Eq. 1, F_{obs} is the observed fluorescence at each temperature (T), F_{min} is the minimum observed fluorescence, F_{max} is maximum observed fluorescence, and a is a *Hill* slope.

Compound EC_{50} value were determined in a similar way as for the initial screening but by varying the concentration of each compound. Concentration-response assays were also performed in 96-well PCR plates with 19 μ l of Mastermix with 1 μ l of the highest concentration of the compound in the first row and lowest concentration in last row. Observed T_m values ($T_{m obs}$) at each compound concentration (L) were fit to equation 2.

$$T_{m obs} = \frac{\Delta T_{m max} * [L]}{EC_{50} + [L]} + T_m \quad (\text{Equation 2})$$

In equation, $\Delta T_{m max}$ is the maximum change in T_m , where T_m is the melting temperature of the protein in the absence of compounds, and EC_{50} is the ligand concentration causing a change in T_m that is half of $\Delta T_{m max}$.

4.6. Molecular modeling & virtual screening

Computational ligand screening was performed on various forms of the Mac1 protein, including both the unliganded (6WEY and 6VXS) and ADP-ribose-bound forms (6W02), using the program AutoDock VINA. (Trott and Olson., et al. 2010) These protein structures were computationally screened with NCI mechanistic set (library 1) and the NIH clinical collection NCI

(library 2). The protein files were downloaded directly from the protein data bank and processed as follows before submitting for screening. All solvent molecules (HETATM) were removed from the files. Polar hydrogen atoms were added and Kollman charges were included in the protein files. The converted protein and ligand file libraries (both pdbqt files) were uploaded to a parallel computing cluster and run with the following parameters: Energy difference = 4 Number of recorded modes = 20 Exhaustiveness was set to 12. Docking box location was configured prior to using AutoDock tools. After the docking calculation was complete, the locations, orientations, and binding affinities of the top candidates were examined by the aid of UCSF Chimera and tabulated for comparison.

4.7.Nucleotide sequence of pET28-COVID-Nsp13

ATGGGCGCGGTGGGTGCGTGCGTTCTGTGCAATAGCCAAACCAGCCTGCGCTGCGGTGCGTGCG
 ATTCGTCGTCCGTTCTGTGCTGCAAGTGCTGCTACGATCACGTTATCAGCACCAGCCACAAGCTGGTGCG
 TGAGCGTTAACCCGTATGTGTGCAACGCGCCGGGTGCGACGTGACCGATGTTACCCAGCTGTACCTGG
 GTGGCATGAGCTACTATTGCAAGAGCCACAAACCGCCGATTAGCTTCCCGCTGTGCGCGAACGGTCAAG
 TTTTGGCCTGTATAAGAACACCTGCGTGCGGTAGCGACAACGTTACCGATTCAACGCGATTGCGACCTG
 CGACTGGACCAACGCGGGTGATTACATTCTGGCGAACACCTGCACCGAACGTCTGAAACTGTTGCGGGC
 GGAAACCCTGAAGGCGACCGAGGAAACCTTTAAACTGAGCTACGGTATCGCGACCGTGCGTGAGGTTTC
 TGAGCGACCGTGAACCTGCACCTGAGCTGGGAAGTGGGTAAACCGCGTCCGCCGCTGAACCGTAACTAC
 GTGTTACACGGTTATCGTGTTACCAAGAACAGCAAAGTGCGAGTATACCTTTGAAAAGGGT
 GACTACGGCGATGCGGTGGTTTATCGTGTTACCACCACCTACAACTGAACGTGGGCGATTACTTCGTT
 CTGACCAGCCACACCGTGATGCCGCTGAGCGCGCCGACCCTGGTTCCGCAGGAACACTATGTGCGTATC
 ACCGGTCTGTACCCGACCCTGAACATTAGCGACGAGTTCAGCAGCAACGTTGCGAACTATCAGAAAGTG
 GGTATGCAAAAATATAGCACCTGCAAGGTCCGCCGGGTACCGGCAAGAGCCACTTTGCGATTGGCCTG
 GCGCTGTACTATCCGAGCGCGCGTATTGTTTATACCGCGTGCGAGCCATGCGGCGGTGGATGCGCTGTGC
 GAAAAGGCGCTGAAATACCTGCCGATCGACAAATGCAGCCGTATCATTCCGGCGCGTGCGCGTGTTGAA
 TGCTTCGACAAGTTTAAAGTGAACAGCACCTGGAGCAGTATGTGTTCTGCACCGTTAACGCGCTGCCG
 GAAACCACCGCGGACATCGTGTTTTGATGAGATTAGCATGGCGACCAACTACGATCTGAGCGTGTT
 AACGCGCGTCTGCGTGCGAAGCACTACGTTTATATTGGTGACCCGGCGCAACTGCCGGCGCCGCGTACC
 CTGCTGACCAAGGGTACCCTGGAGCCGGAATATTTCAACAGCGTGTGCCGTCTGATGAAAACCATCGGT
 CCGGATATGTTTCTGGGTACCTGCCGTCGTTGCCCGCGGAAATTGTGGACACCGTTAGCGCGCTGGTT
 TATGATAACAAGCTGAAAGCGCACAAGGACAAAAGCGCGCAGTGCTTCAAGATGTTTTACAAAGGTGT
 GATCACCCACGACGTTAGCAGCGCGATCAACCGTCCGCAAATTGGCGTGGTTCGTGAGTTCTGACCCG

TAACCCGGCGTGCGCTAAGGCGGTTTTTATCAGCCCGTACAACAGCCAGAACGCGGTGGCGAGCAAAA
TTCTGGGTCTGCCGACCCAGACCGTTGATAGCAGCCAAGGCAGCGAATACGACTATGTGATCTTCACCC
AAACCACCGAAACCGCGCACAGCTGCAACGTGAACCGTTTTAACGTTGCGATTACCCGTGCGAAAGTGG
GTATCCTGTGCATTATGAGCGACCGTGATCTGTATGATAAACTGCAATTCACCAGCCTGGAAATCCCGCG
TCGTAATGTGGCGACCCTGCAACTCGAGCACCACCACCACCACCTGA

4.8.Amino acid sequence of Nsp13-Histag encoded by pET28-COVID-Nsp13

Translation 611 a.a. MW=68110.32 g/mol

MGAVGACVLCNSQTSRLRCGACIRRPFLCCKCCYDHVISTSHKLVLVSNPYVCNAPGCDVTDVTQLYL
GGMSYYCKSHKPPISFPLCANGQVFGLYKNTCVGSDNVTDFNAIATCDWTNAGDYILANTCTERLKLFAET
LKATEETFKLSYGIATVREVLSDRELHLSWEVGKPRPPLNRNYVFTGYRVTKNSKVQIGEYTFEKGDYGDVAVV
RGTTTTYKLVNGDYFVLTSHTVMPLSAPTLVPQEHYVRITGLYPTLNISDEFSSNVANYQKVGMQKYSTLQGPP
GTGKSHFAIGLALYYPSARIVYTACSHAAMDALCEKALKYLPIDKCSRIIPARARVECFDKFVNSTLEQYVFCTV
NALPETTADIVVFDEISMATNYDLSVNNARLRKHVYVIGDPAQLPAPRTLTKGTLEPEYFNSVCRLMKTIGP
DMFLGTCRRCPAEIVDTVSALVYDNKLKAHKDKSAQCFKMFYKGVITHDVSSAINRPQIGVVREFLTRNPAW
RKAVFISPYNSQNAVASKILGLPTQTVDSQSEYDYVIFTQTETAHSCNVNRFNVAITRAKVGILCIMSDRD
LYDKLQFTSLEIPRRNVATLQLEHHHHHH*

4.9.Nucleotide sequence of pET21-COVID-Nsp13

The Nsp13 vector was generated similar to the one used by Mickolajczyk et al.
(Mickolajczyk et al., 2020).

ATGGGCAGCAGCCATCATCATCATCACAGCAGCGGCCTGGAGGTGCTGTTCCAGGGACCTC
ATATGGCTGTGGGTGCCTGCGTCCTGTGCAACTCTCAGACTTCTCTGAGGTGCGGTGCTTGCATCAGGC
GTCCCTTCCTGTGCTGCAAGTGCTGCTACGACCACGTCATCTCTACTTCACACAAGCTGGTGCTGTCCGTC
AACCCATACGTGTGCAACGCTCCTGTTGCGACGTGACTGACGTACCCAGCTGTACCTGGGTGGCATG
TCCTACTACTGCAAGAGCCACAAGCCTCCCATCAGCTTCCCACTGTGCGCCAACGGTCAGGTGTTCCGGCC
TGTAACAAGAACACCTGCGTGGGTTCTGACAACGTCACTGACTTCAACGCTATCGCCACCTGCGACTGGAC
TAACGCTGGCGACTACATCCTGGCCAACACTTGCACCGAACGTCTGAAGCTGTTGCTGCCGAGACCCT
GAAGGCTACTGAGGAAACCTTCAAGCTGTCATACGGAATCGCCACTGTGAGAGAGGTCTGTCTGACCG
CGAACTGCACCTGTCATGGGAGGTGCGCAAGCCTAGGCCACCTCTGAACAGAACTACGTGTTACCCGG
CTACCGTGTCATAAGAACTCCAAGGTGCAGATCGGAGAGTACACCTTCGAAAAGGGCGACTACGGAG
ACGCTGTGGTCTACAGGGGCACCACTACCTACAAGCTGAACGTGGGAGACTACTTCGTCCTGACTTCCC
ACACCGTGATGCCACTGAGCGCCCCTACTCTGGTCCCCCAGGAACACTACGTGCGCATCACCGGCCTGTA
CCCAACTCTGAACATCAGCGACGAGTTCTCCAGCAACGTGCTAACTACCAGAAAGTGGGTATGCAGAA
GTACTCTACCCTGCAGGGTCTCTGGAAGTGGCAAGTCACACTTCGCTATCGGACTGGCCCTGTACTAC
CCTTCGCTAGAAATCGTCTACACTGCTTGCTCTCACGCTGCTGTGGACGCTCTGTGCGAAAAGGCCCTGA

AGTACCTGCCCATCGACAAGTGCTCACGCATCATCCCAGCTAGGGCCAGAGTCGAATGCTTCGACAAGT
TCAAGGTGAACTCCACCCTGGAGCAGTACGTGTTCTGCACTGTCAACGCTCTGCCCCGAACTACCGCCGA
CATCGTGGTCTTCGACGAGATCTCTATGGCTACCAACTACGACCTGTCAGTGGTCAACGCTCGCCTGCGT
GCCAAGCACTACGTCTACATCGGAGACCCTGCTCAGCTGCCTGCCCCAGAACTCTGCTGACCAAGGGT
ACTCTGGAGCCAGAATACTTCAACTCTGTGTGCCGCCTGATGAAGACCATCGGACCAGACATGTTCTG
GGTACTTGACAGGAGATGCCCTGCTGAAATCGTGGACACCGTCTCTGCCCTGGTGTACGACAACAAGCTG
AAGGCTCACAAGGACAAGTCAGCCCAGTGCTTCAAGATGTTCTACAAGGGCGTGATCACCCACGACGTC
TCTTCAGCCATCAACCGTCCTCAGATCGGAGTGGTCCGCGAGTTCCTGACTCGTAACCTGCTTGGAGGA
AGGCCGTCTTCATCTCTCCCTACAACCTCACAGAACGCTGTGGCCTCCAAGATCCTGGGACTGCCACTCA
GACCGTCGACTCCAGCCAGGGTAGCGAATACGACTACGTGATCTTCACCCAGACTACCGAGACTGCTCA
CTCTTGCAACGTGAACCGTTTCAACGTCGCTATCACCAGGGCCAAGGTCGGCATCCTGTGCATCATGTCC
GACAGGGACCTGTACGACAAGCTGCAGTTCACCAGCCTGGAGATCCCTCGCCGTAACGTGGCCACTCTG
CAGTAA

4.10. Amino acid sequence of His-Nsp13 encoded by pET21-COVID-Nsp13

Translation 624 a.a. MW=69425.81 g/mol

MGSSHHHHHHSSGLEVLFGPHMAVGACVLCNSQTSRLCGACIRRPFLCCKCCYDHVISTHKLVL
VNPYVCNAPGCDVTDVTQLYLGGMSYYCKSHKPPISFPLCANGQVFGLYKNTCVGSDNVTDFNAIATCDWT
NAGDYILANTCTERLKLFAAETLKATEETFKLSYGIATVREVLSDRELHLSWEVGKPRPPLNRNYVFTGYRVT
NSKVQIGEYTFEKGDYGDVAVYRGTTTYKLVNGDYFVLTSHTVMPLSAPTLVPQEHYVRITGLYPTLNISDEF
SNVANYQKVGMQKYSTLQGPPGTGKSHFAIGLALYPSARIVYTACSHAAMDALCEKALKYLPIDKCSRIIPAR
ARVECFDKFKVNSTLEQYVFCTVNALPETTADIVVFDEISMATNYDLSVVNARLRKHYVYIGDPAQLPAPRT
LLTKGTLEPEYFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSAVYDNKLKAHKDKSAQCFKMFYKGVITHD
VSSAINRPQIGVVREFLTRNPAWRKAVFISPYNSQNAVASKILGLPTQTVDSQSEYDYVIFTQTTETAHSCN
VNRFNVAITRAKVGILCIMSDDRDLYDKLQFTSLEIPRRNVATLQ*

4.11. Malachite green assay

4.11.1. Theory

Malachite green assay can only be used to quantify a reaction in which inorganic phosphate is released or used as part of the reaction. In this assay a complex is formed between malachite green, ammonium molybdate and free inorganic phosphate under acidic conditions. Free inorganic phosphate binds to ammonium molybdate forming a phosphomolybdate complex which then interacts with three molecules of malachite green dye causing a change in color.

Under acidic conditions ATP can be acid catalyzed, to prevent this from happening Sodium citrate is added to neutralize the sample. The sample is then left to develop color for 10 minutes and the absorbance at 630nm is collected. The resulting change in absorbance at 630nm is directly correlated to the concentration of inorganic phosphate in the solution.

4.11.2. Protocol

Materials:

1. 3x ml 4.2% Ammonium Molybdate in 4.0N HCl.
2. 1x ml 0.045% Malachite Green
3. 0.1x 35% Sodium Citrate.

Mix 3x ammonium Molybdate with 1x Malachite Green making a phosphate reagent with 0.002% Tween 20. To test a 100 μ l sample for free phosphate, add 800ul of phosphate reagent and immediate add 0.1x sodium citrate. Let the reaction sit for 10 minutes until the color develops and take the absorbance reading at 630nm.

4.12. FRET oligonucleotide sequence

Substrate	Name
5'-TGGCGACGGCAGCGAT[TAM]	3TAM
5'-[TAM]TGGCGACGGCAGCGA	5TAM
5'-[HEX]TGGCGACGGCAGCGAGGC-3'	HEX18
T[TAM] 5'-TGGCGACGGCAGCGA 3'-ACCGCTGCCGTCGCTTTTTTTTTTTTTTT-5'	3TAM:5'15
5'-[TAM]TGGCGACGGCAGCGA 3'-ACCGCTGCCGTCGCTTTTTTTTTTTTTTT-5'	5TAM:5'15
5'-TGGCGACGGCAGCGAT[TAM] 3'-TTTTTTTTTTTTTTTACCGCTGCCGTCGCT-5'	3TAM:3'15
5'-[TAM]TGGCGACGGCAGCGA 3'-TTTTTTTTTTTTTTTACCGCTGCCGTCGCT-5'	5TAM:3'15
5'-TGGCGACGGCAGCGAT[TAM] 3'-TTTTTTTTTTTTTTTACCGCTGCCGTCGCT-5'	3TAM:3'10
5'-[TAM]TGGCGACGGCAGCGA 3'-TTTTTTTTTTTTTACCGCTGCCGTCGCT-5'	5TAM:3'10
5'-TGGCGACGGCAGCGAT[TAM] 3'-TTTTTACCGCTGCCGTCGCT-5'	3TAM:3'5
5'-[TAM]TGGCGACGGCAGCGA 3'-TTTTTACCGCTGCCGTCGCT-5'	5TAM:3'5
5'-[HEX]TGGCGACGGCAGCGAGGC-3' 3'-TTTTTTTTTTTTTTTACCGCTGCCGTCGCT-5'	HF18:3'15
5'-[HEX]TGGCGACGGCAGCGAGGC-3' 3'-TTTTTTTTTTTTTACCGCTGCCGTCGCT-5'	HF18:3'10
[TYE]TGGCGACGGCAGCGAGGCACCGATCACATGTTTTTTTTTTTTTT-3'	TYE45
3'-ACCGCTGCCGTCGCT [TYE]TGGCGACGGCAGCGAGGCACCGATCACATGTTTTTTTTTTTTTT-3'	TYE45:comp15
3'-ACCGCTGCCGTCGCTCCGTG [TYE]TGGCGACGGCAGCGAGGCACCGATCACATGTTTTTTTTTTTTTT-3'	TYE45:comp20
3'-ACCGCTGCCGTCGCTCCGTGGCTAG [TYE]TGGCGACGGCAGCGAGGCACCGATCACATGTTTTTTTTTTTTTT-3'	TYE45:comp25
3'-ACCGCTGCCGTCGCTCCGTGGCTAGTGTAC [TYE]TGGCGACGGCAGCGAGGCACCGATCACATGTTTTTTTTTTTTTT-3'	TYE45:comp30
3'-ACCGCTGCCGTCGCTCCGTGGCTAGTGTACAAAAA [TYE]TGGCGACGGCAGCGAGGCACCGATCACATGTTTTTTTTTTTTTT-3'	TYE45:comp35

5. Using FRET to determine the binding of NS3h to ssDNA and dsDNA

5.1.Introduction

The Hepatitis C virus is a blood-borne pathogen that infects the liver. It is transmitted via blood and it is transported to the liver via the bloodstream. The focus of this study is the hepatitis C virus nonstructural protein 3 (NS3). NS3 and NS4A are co-translated as part of a larger polyprotein but are then cleaved into two distinct mature proteins. NS3 is a 70 kDa, 630aa multifunctional protein having an N-terminal serine protease and a C-terminal helicase (Raney et al., 2010). The protease found at NS3's N-terminal is 180 amino acids long, whereas the helicase located at the NS3's C-terminal is 430 amino acids long. Despite being covalently bound as part of the same nonstructural protein, both helicase and serine protease can be expressed, and isolated independently of each other.

Non-Structural Protein 3 is a critical component in HCV replication. The helicase domain of the NS3 protein is referred to as NS3h. Helicases are recognized as motor proteins that utilize the energy generated by nucleotide hydrolysis to catalyze the separation of double-stranded nucleic acids. In the work presented here, the binding of NS3h with DNA was monitored using FRET. FRET is a method that makes use of fluorescence to identify the interaction between two molecules. In all the FRET assays studied here, the donor is Yellow Fluorescent Protein (YFP), covalently attached to the helicase domain of nonstructural protein 3 from the Hepatitis C Virus. FRET in our studies is characterized as pairwise FRET efficiency and is estimated based on donor quenching. Estimated pairwise FRET efficiency was used to examine factors effecting binding,

such as the length of the duplex, the length of a single-stranded overhang, and whether the overhang had a 5' or 3' end.

5.2.Examining factors that could indirectly affect the donor emission.

Typically, the magnitude of a chemical reaction or an experiment is determined by the measurable change occurring due to the chemical process. The measurable change is then observed and collected as data, and the data is analyzed, adjusted, and fitted to an equation to give parameters that provide information pertaining to the chemical reaction. One of the serious mistakes one can make during this process is collecting data that is contaminated by the factors effecting the measurable change other than the actual chemical reaction. In a typical FRET experiment, binding between two molecules is estimated based on the observed fluorescence emission of both the acceptor and donor. However, it is possible that fluorescence emission might be contaminated by factors other than the transfer of energy between acceptor and donor. Of course, it is already known that some of the factors that contaminate the results of FRET assays are direct excitation and bleed through, but the correction for these contaminations is not needed as FRET in this study is measured based on donor quenching. In this section of this chapter, factors that could indirectly affect the fluorescence emission of the donor are examined. Such factors include: effect on donor fluorescence emission due to helicase binding onto the unlabeled oligonucleotide possible direct interaction of the oligonucleotide with yellow fluorescent protein and effect on donor fluorescence in the presence of an acceptor. Additionally, for comparison, a positive control titration of YFP-NS3h with a TAMRA labeled oligonucleotide was also included.

The effect on donor fluorescence emission due to helicase binding onto the oligonucleotide or possible interaction of the oligonucleotide with yellow fluorescent protein was examined by adding an unlabeled oligonucleotide into a cuvette with a sample containing YFP-NS3h. In brief, concentrated and purified YFP-NS3h was dissolved in helicase dilution buffer (25 mM MOPS, 5 mM MgCl_2 , 1 mM DTT, 0.01 mg/ml BSA, and 0.01% Triton X 100 at pH 7) to make a 500 μl sample with a YFP-NS3h concentration of 100 nM. This sample was then transferred into the cuvette. In a separate Eppendorf tube, an 18 nucleotide long unlabeled oligonucleotide was also dissolved in the helicase dilution buffer, making a 2 μM unlabeled oligonucleotide solution. The cuvette was placed in the Varian Cary Eclipse Fluorescence Spectrophotometer and the sample was excited at 450 nm. Emission was collected in the range of 460-650 nm. At 450 nm, the yellow fluorescent protein is excited and it emits between 520-530 nm. After the initial scan, for each subsequent scan, a 1 μl aliquot of sample containing unlabeled oligonucleotide was added to the cuvette, the sample was excited, and emission between 460-650 nm was collected. After taking 17 subsequent spectrums, no change in YFP-NS3h was observed (fig. 5.2-1A).

Next, the effect on donor fluorescence in the presence of an acceptor was examined by adding an 18 nucleotides long TAMRA labeled oligonucleotide into a cuvette with a sample containing free yellow fluorescent protein. In brief, concentrated and purified yellow fluorescent protein was dissolved in helicase dilution buffer, making a 500 μl sample with a YFP concentration of 100 nM. This sample was then transferred into the cuvette. In a separate Eppendorf tube, an 18 nucleotides long TAMRA labeled oligonucleotide was also dissolved in the helicase dilution buffer, making a 2 μM labeled oligonucleotide solution. The cuvette with the YFP sample was

excited at 450 nm and emission was collected in the range of 460-650 nm. After the initial scan, for each subsequent scan, a 1 μ l aliquot of sample containing labeled oligonucleotide was added to the cuvette, the sample was excited, and emission between 460-650 nm was collected. After taking 10 subsequent spectrums, no change in YFP-NS3h was observed (data not shown).

Lastly, the donor fluorescence was observed by adding an 18 nucleotides long TAMRA labeled oligonucleotide to the cuvette with a sample containing 100 nM YFP-NS3h. In brief, concentrated and purified YFP-NS3h was dissolved in helicase dilution buffer, making a 500 μ l sample with a YFP-NS3h concentration of 100 nM. This sample was then transferred into the cuvette. In a separate Eppendorf tube, an 18 nucleotides long TAMRA labeled oligonucleotide was also dissolved in the helicase dilution buffer, making a 2 μ M labeled oligonucleotide solution. The cuvette with the YFP sample was excited at 450 nm and emission was collected in the range of 460-650 nm. After the initial scan, for each subsequent scan, a 1 μ l aliquot of sample containing TAMRA labeled oligonucleotide was added and emission was collected. After taking 16 subsequent spectrums, a decrease in YFP-NS3h emission was observed (fig. 5.2-1B).

The FRET_{app} for each experiment was then calculated by using equation 3.5-4 and the data is plotted in fig.5.2-1C. The data that was calculated using equation 3.5-4 was then fitted to equation 3.5-5 to estimate pairwise FRET efficiency (E_{DA}), which is shown in the form of a histogram in fig. 5.2-1D.

The purpose of this section was to identify factors that could indirectly affect the fluorescence emission of the donor. The first experiment was done to examine the effect on donor fluorescence emission due to helicase binding onto the oligonucleotide or possible

interaction of the oligonucleotide with yellow fluorescent protein. Upon adding 17 1 μ l aliquots of unlabeled oligonucleotide, no change in YFP-NS3h was observed. This data suggests that when the YFP-helicase binds onto the oligonucleotide, the binding of the oligonucleotide doesn't affect YFP emission. This data could also be interpreted as either no interaction between YFP and unlabeled oligonucleotide or no effect on donor emission due to the presence of unlabeled oligonucleotide.

Next, the effect on donor fluorescence in the presence of an acceptor was examined by adding an 18 nucleotides long TAMRA labeled oligonucleotide into a cuvette with a sample containing free yellow fluorescent protein. Upon adding 17 1 μ l aliquots of TAMRA labeled oligonucleotide, no change in YFP-NS3h was observed.

This data suggests that the presence of an acceptor molecule in the same sample as the donor doesn't affect the fluorescence emission. This data could also be interpreted to mean that there is no interaction between donor and acceptor, or at least not enough to actually affect donor emission.

Lastly, the donor fluorescence was observed by adding an 18 nucleotides long TAMRA labeled oligonucleotide to the cuvette with a sample containing 100 nM YFP-NS3h. A decrease in YFP-NS3h emission was observed after adding 17 1 μ l aliquots of TAMRA labeled oligonucleotide. This data suggests that both YFP-NS3h (donor) and TAMRA labeled oligonucleotide (acceptor) are needed to see a change in donor fluorescence emission. The change in donor emission is due to helicase binding onto the oligonucleotide, bringing both acceptor and donor close enough so that the energy is transferred between donor and acceptor.

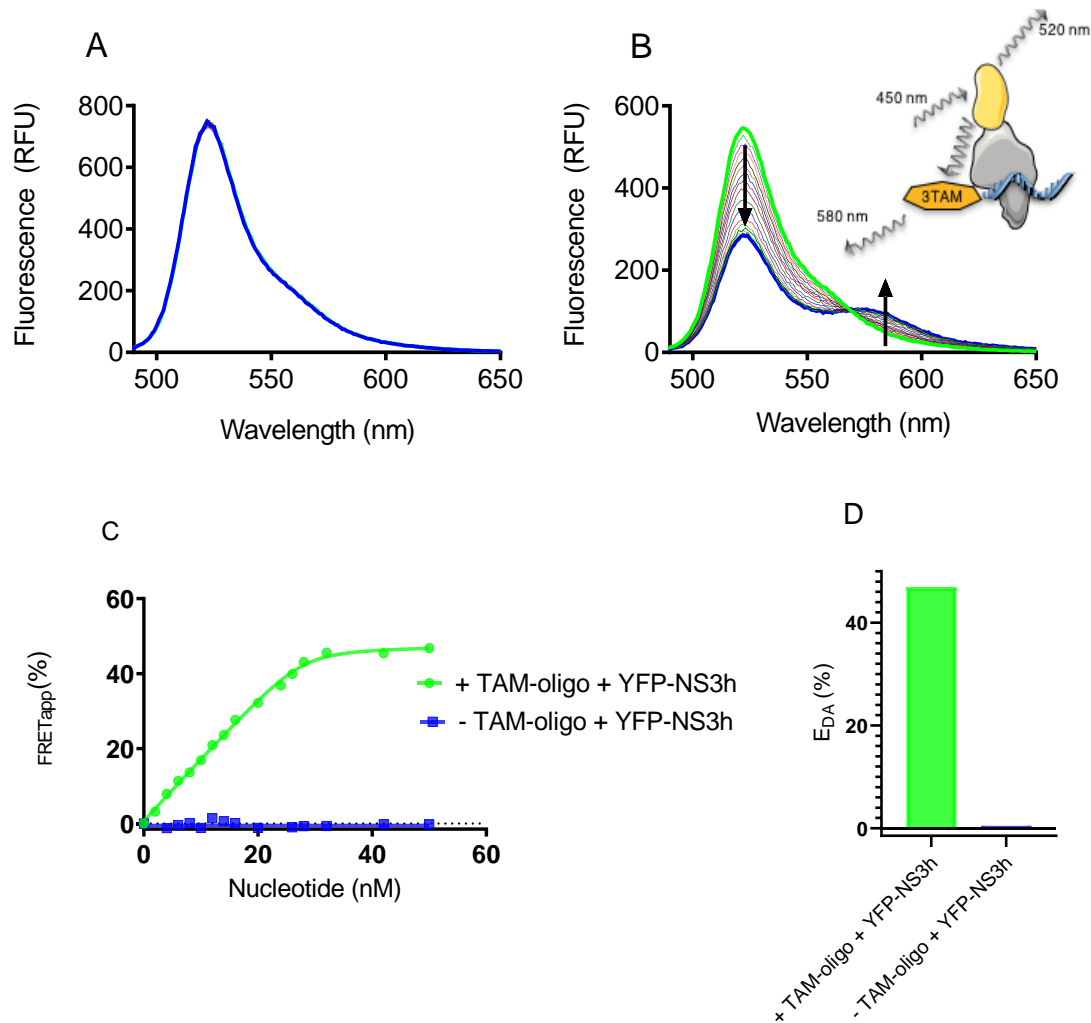


Figure 5.2-1 NS3h Titration with unlabeled DNA, and a TAMRA labeled oligonucleotide.

Panel A depicts a series of emission spectra taken with each addition of an unlabeled oligonucleotide into 100 nM of YFP-NS3h. The experiment in panel B was carried out in the same way as in panel A, except that the unlabeled oligonucleotide was replaced with a TAMRA-labeled 18 nucleotide long oligonucleotide. These experiments were conducted in a cuvette containing a 500 μ l sample of YFP-NS3h dissolved in helicase buffer at a concentration of 100 nM. After adding 1 μ l of 2 μ M unlabeled or labeled oligonucleotide, the sample was excited at 450 nm and the emission spectrum was collected. In total, 17 spectra were collected after each addition of oligonucleotide. FRET_{app} for both experiments were calculated using equation 3.5-4 and the plot is presented in panel C. The data in Panel C was then fitted to Equation 3.5-5 to give the Pairwise FRET efficiency E_{DA} shown in Panel D as a histogram representing energy transfer between donor and acceptor, and next to it is a cartoon representation of how binding of YFP-NS3h to 3'TAMRA labeled oligonucleotide is imagined.

5.3. Estimating dissociation constant and stoichiometry constant.

The objective of this experiment is to estimate dissociation constant and the stoichiometry constant from FRET_{app} using Eq. 3.5-5 or 3.5-12. The experiment was done in a similar manner as other FRET titration assays. The donor fluorescence was observed by adding an 18 nucleotides long HEX labeled oligonucleotide to the cuvette with a sample containing YFP-NS3h. In brief, concentrated and purified YFP-NS3h was dissolved in helicase dilution buffer, yielding a 500 μl sample with a YFP-NS3h concentration of 50 nM. This sample was then transferred into the cuvette. In a separate Eppendorf tube, an 18 nucleotides long HEX labeled oligonucleotide was also dissolved in the helicase dilution buffer, making a 2 μM labeled oligonucleotide solution. The cuvette with the YFP sample was excited at 450 nm and emission was collected in the range of 460-600 nm. After the initial scan, for each subsequent scan, a 1 μl aliquot of sample containing HEX labeled oligonucleotide was added and emission was collected. In total, 17 subsequent spectrums were collected and a decrease in YFP-NS3h emission was observed with each addition of an oligonucleotide. The assay was then repeated with 75 and 100 nM concentrations of YFP-NS3h.

In total, three samples containing 50 nM, 75 nM, and 100 nM YFP-NS3h were titrated and their emission spectra were recorded as shown in Fig. 5.3-1 A, B, and C, respectively. A decrease in donor emission at 520nm was observed with each titration, and the maximums at 520 nm were used to calculate the FRET_{app} using Eq. 3.5-4. In Eq. 3.5-4, $k_D(\lambda_D^{ex} \lambda_D^{em})$ is the maximum fluorescence emission at 520 nm from the first sample containing only the donor and $k_{DA}(\lambda_D^{ex} \lambda_D^{em})$ is the maximum fluorescence emission at 520 nm from the subsequent samples

containing both donor and acceptor. The calculated FRET_{app} using Eq. 3.5-4 along with the acceptor concentration from each titration is then plotted (Fig. 5.3-1 D) and the data is fitted to Eq. 3.5-5 or 3.5-12 to estimate pairwise FRET efficiency (E_{DA}) (Fig. 5.3-1 E) dissociation constant and stoichiometry constant. The resultant dissociation constants of 18.3 ± 5.9 nM and stoichiometry coefficients of 3.8 ± 0.3 were estimated using equations 3.5-5 or 3.5-12.

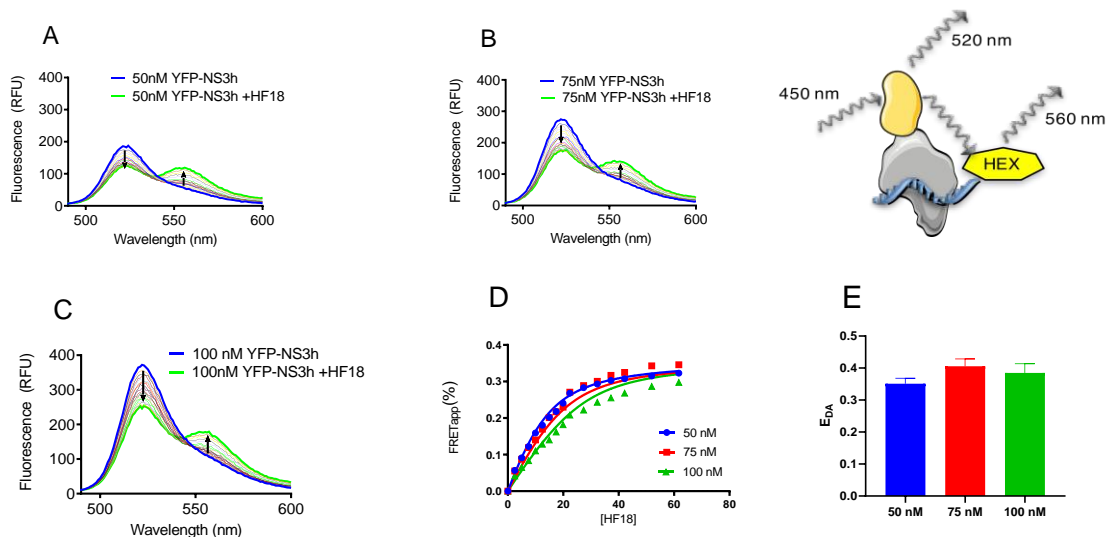


Figure 5.3-1. Estimated dissociation constant and stoichiometry coefficients.

Three different concentrations of YFP-NS3h (donor) were titrated with HF18 (acceptor) and emission spectra were recorded as shown in Panels A, B, and C. The maximums at 520 nm in Panels A, B, and C were used to calculate FRET_{app} with equation 3.5-4. FRET_{app} and oligonucleotide concentration were plotted (Panel D), and the data was fitted to Equation 3.5-5 to estimate FRET efficiency (E_{DA}), dissociation constant, and stoichiometry constant. The resultant dissociation constants of 18.3 ± 5.9 nM and stoichiometry coefficients of 3.8 ± 0.3 were estimated using equations 3.5-5 or 3.5-12. Panel E is the estimated E_{DA} for each experiment shown in Panel D.

Its important to emphasize that the FRET theory used to fit the data in panel E of fig. 5.3-1 assumes that only one helicase molecule binds to the oligonucleotide at a time, i.e. the helicase and oligonucleotide forms heterodimers. At the same time, in relating ADn to the dissociation

constant, we allowed the heterooligomer size to take an arbitrary value, n . To reconcile these two apparently contradicting assumptions with one another, we would add the third assumption, which is that the helicase that binds closest to the acceptor is the only one involved in FRET; any additional helicase up to order n is assumed to be so far away that it does not transfer energy to the donor. This allows us, as the first approximation, to extract approximating information on n from the law of mass action without unnecessarily complicating the FRET theory represented by equations 3.5-5. (Please refer to the reference Raicu et al., (2007) for additional details regarding the FRET theory.) As such, this analysis is to be regarded as preliminary, and future work should attempt to refine the FRET theory so as to include linear oligomers of size n , thereby improving the fit of the experimental data shown in 5.3-1 Panel D.

5.4.Examine NS3h binding with partially dsDNA.

The purpose of this assay was to examine if different FRET_{app} are observed with TAMRA labeled on either end of the junction of partial dsDNA. The experiment was done using two different partial double-stranded oligonucleotides labeled with TAMRA either next to or farther from the junction. Typically, more FRET should be observed with an oligonucleotide that has an acceptor molecule located right next to the helicase binding site than farther away. If this turns out to be false, one explanation for the deviation could be that the oligonucleotide is wrapping around the helicase. This appears unlikely because the helicase is a motor protein that moves the DNA. Wrapping the DNA around would hinder its movement. Another explanation for the deviation could be helicase binding to the double-stranded DNA. However, this is also improbable because NS3h is an SF2 helicase and they are known to bind to single-stranded DNA or RNA.

The effect of titrating YFP-NS3h with a partial dsDNA with an acceptor located on either end of the junction was examined by adding 5'TAM-3'tail-10 or 3'TAM-3'tail-10 oligonucleotide into a cuvette with a sample containing 125 nM YFP-NS3h. The experiment was conducted by dissolving concentrated and purified yellow fluorescent protein in helicase dilution buffer, making a 500 μ l sample with a YFP concentration of 125 nM. This sample was then transferred into the cuvette. In a separate Eppendorf tube, 5'TAM-3'tail-10 TAMRA labeled oligonucleotide was also dissolved in the helicase dilution buffer, making a 5 μ M TAMRA labeled oligonucleotide solution. The cuvette with the YFP sample was excited at 450 nm and emission was collected in the range of 460–650 nm. After the initial scan, for each subsequent scan, a 1 μ l aliquot of sample containing labeled oligonucleotide was added to the cuvette, the sample was excited, and emission between 460-650 nm was collected. In total, 6 subsequent spectrums were collected and a decrease in YFP-NS3h emission was observed with each addition of an oligonucleotide (fig. 5.4-1 A and B). The assay was then repeated with 3'TAM-3'tail-10.

The maximums at 520 nm from each titration were extracted and FRET_{app} was calculated using equation 3.5-4. Calculated FRET_{app} was then plotted against the oligonucleotide concentration in fig. 5.4-1C. This data was then fitted to equation 3.5-5 in order to obtain pairwise FRET efficiency (E_{DA}). As expected, more FRET_{app} was estimated from the oligonucleotide with TAMRA located next to the junction so that the helicase binds onto the ssDNA next to the acceptor, causing more donor fluorescence quenching. Lower FRET_{app} was estimated with an acceptor located at the far end of the junction or further away from the ssDNA helicase binding site.

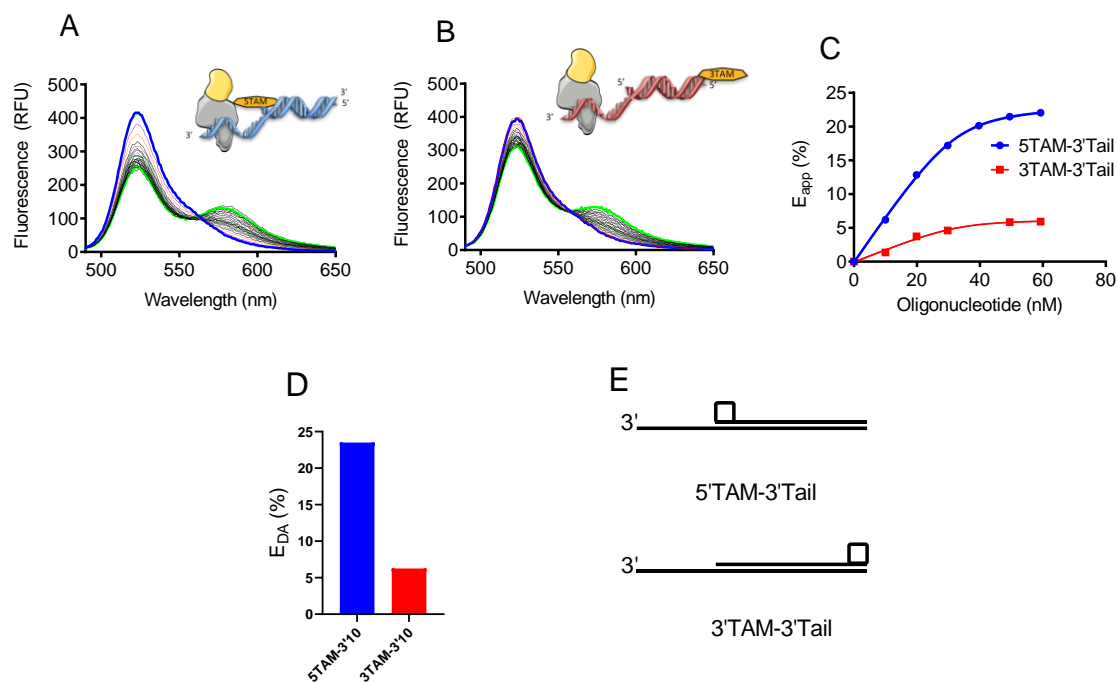


Figure 5.4-1 Titration of YFP labeled NS3h with TAMRA-labeled on 3' or 5' end of partial dsDNA. The emission spectrums of YFP-NS3h titration with 5'TAM-3'tail-10 or 3'TAM-3'tail-10 oligonucleotide are shown in panels A and B, respectively. A decrease in fluorescence at 520 nm and an increase at 580 nm were observed. The maximums at 520 nm from each titration were extracted and $FRET_{app}$ was calculated using equation 3.5-4. Calculated $FRET_{app}$ was then plotted against the oligonucleotide concentration in panel C. This data was then fitted to equation 3.5-5 in order to obtain pairwise FRET efficiency (E_{DA}). The E_{DA} value for each oligonucleotide is shown as a histogram in panel D. Panel E depicts the line structure of two different oligonucleotides used in this experiment.

5.5. NS3h seems to bind to ssDNA in a specific orientation.

NS3h is a SF2 helicase that translocates unidirectionally in a 3' to 5' direction. In order to translocate on DNA, the helicase needs to make specific conformational changes upon ATP hydrolysis in order to move forward. Since the SF2 helicase family moves unidirectionally in the 3' to 5' direction, it must also bind to the ssDNA with a specific orientation. To examine if helicase

binds to ssDNA with a specific orientation, YFP-NS3h was titrated with a TAMRA-labeled oligonucleotide at either 3' or 5' end. If no difference in FRET_{app} is observed, this would suggest that YFP-NS3h doesn't differentiate among these nucleotides upon binding. However, if a different FRET_{app} is observed, this would suggest the helicase binds to the DNA with a specific orientation.

To examine if YFP-NS3h binds to the DNA in a specific orientation, the helicase was titrated with TAMRA labeled on either the 5' end or 3' end of an 18 nucleotides long oligonucleotide. The experiment was done in a similar manner as others. The experiment was conducted by dissolving concentrated and purified yellow fluorescent protein in helicase dilution buffer, making a 500 μl sample with a YFP concentration of 100 nM. This sample was then transferred into the cuvette. In a separate Eppendorf tube, 5'TAM oligonucleotide was also dissolved in the helicase dilution buffer, making a 2 μM TAMRA labeled oligonucleotide solution. The cuvette with the YFP sample was excited at 450 nm and emission was collected in the range of 460–650 nm. After the initial scan, for each subsequent scan, a 1 μl aliquot of sample containing labeled oligonucleotide was added to the cuvette, the sample was excited, and emission between 460-650 nm was collected. The assay was then repeated with 5'TAM. In total, 17 subsequent spectrums were collected with each titration, and a decrease in YFP-NS3h emission was observed with each addition of an oligonucleotide (fig. 5.5-1 A and B).

The maximums at 520 nm from each titration were extracted and FRET_{app} was calculated using equation 3.5-4. Calculated FRET_{app} was then plotted against the oligonucleotide concentration in fig. 5.5-1C. This data was then fitted to equation 3.5-5 in order to obtain E_{DA} fig. 5.5-1D.

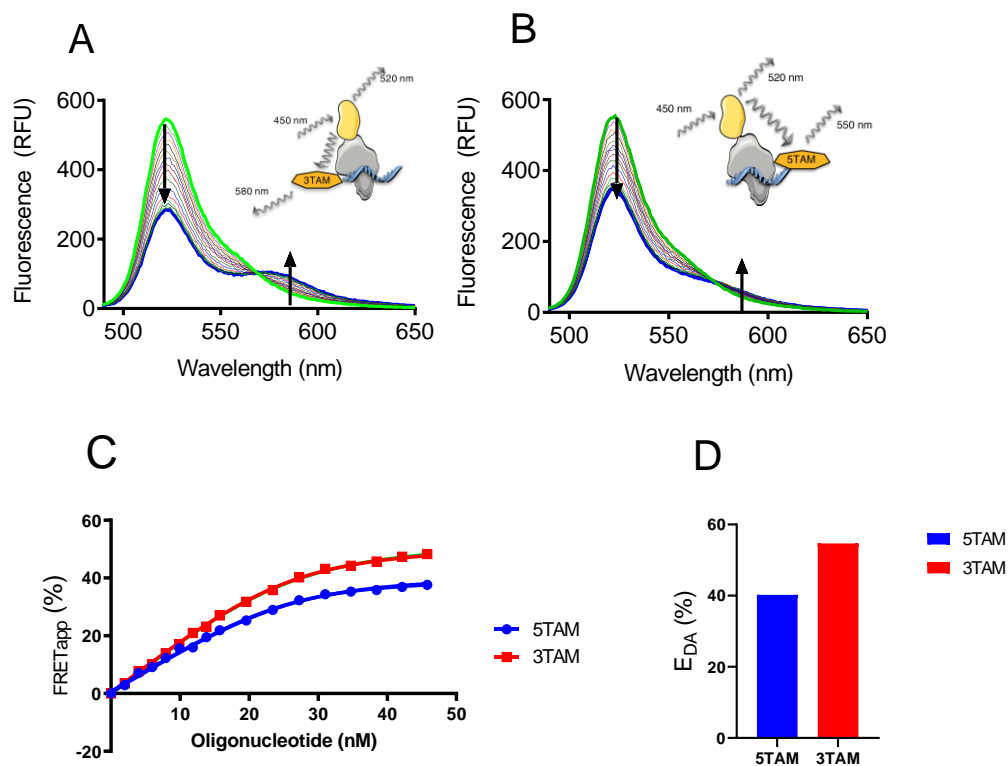


Figure 5.5-1 YFP-NS3h Titration with 3' and 5' TAM

The emission spectrums of YFP-NS3h titration with 3'TAM and 5'TAM oligonucleotides are shown in panels A and B, respectively. The maximums at 520 nm from each titration were extracted and $FRET_{app}$ was calculated using equation 3.5-4. Calculated $FRET_{app}$ was then plotted against the oligonucleotide concentration in panel C. This data was then fitted to equation 3.5-5 in order to obtain pairwise FRET efficiency, E_{DA} . The E_{DA} value for each oligonucleotide is shown as a histogram in panel D.

The experiment was conducted to examine if the helicase binds to ssDNA with a specific orientation. For this experiment, YFP-NS3h was titrated with a TAMRA-labeled oligonucleotide covalently bound to the ssDNA at either 3' or 5' end. Based on different estimates of pairwise FRET efficiency (E_{DA}) from each nucleotide, it seems that helicase was not bound randomly but with a specific orientation.

5.6. Using FRET to examine the effect of increasing the overhang on partial dsDNA.

The experiment was designed to determine if the size of the overhang on a partial dsDNA affects the pairwise FRET efficiency. In other words, the location of the junction at which the ssDNA meets with dsDNA remains the same while the length of the ssDNA changes. It was expected that by increasing the available binding space for the helicase, more helicase would bind and hence more FRET would be detected.

To investigate how the available size of the overhang on a partial dsDNA oligonucleotide affects the helicase binding, YFP-NS3h was titrated with several partial dsDNA oligonucleotides with varying overhang sizes while maintaining the oligonucleotide's double strand size constant. The experiment was conducted in a similar fashion to previous ones. In brief, the experiment was conducted by dissolving concentrated and purified yellow fluorescent protein in helicase dilution buffer, making a 500 μ l sample with a YFP concentration of 100 nM. This sample was then transferred into the cuvette. In a separate Eppendorf tube, 5'TAM-3'tail-5 oligonucleotide was also dissolved in the helicase dilution buffer, making a 2 μ M TAMRA labeled oligonucleotide solution. The cuvette with the YFP sample was excited at 450 nm and emission was collected in the range of 460–650 nm. After the initial scan, for each subsequent scan, a 1 μ l aliquot of sample containing labeled oligonucleotide was added to the cuvette, the sample was excited, and emission between 460-650 nm was collected. The assay was then repeated with 5 more samples: 5'TAM-3'tail-10, 5'TAM-3'tail-15, 3'TAM-3'tail-5, 3'TAM-3'tail-10, and 3'TAM-3'tail-5. In total, 10-20 subsequent spectrums were collected with each titration, and a decrease in YFP-NS3h emission was observed with each addition of an oligonucleotide.

The maximums at 520 nm from each titration were extracted and FRET_{app} was calculated using equation 3.5-4. Calculated FRET_{app} was then plotted against the oligonucleotide concentration in fig. 5.6-1A and C. This data was then fitted to equation 3.5-5 in order to obtain E_{DA} fig. 5.6-1B and D.

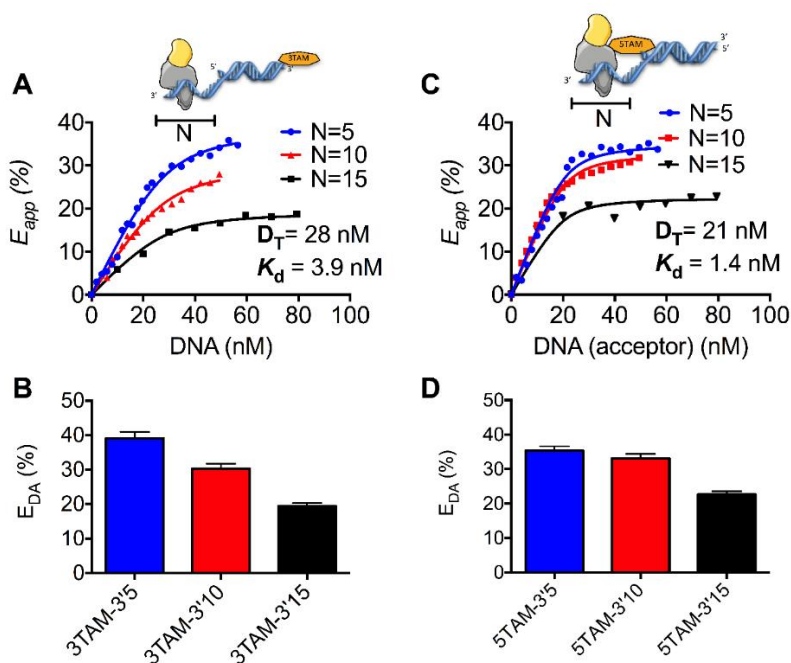


Figure 5.6-1 Titration of YFP-NS3h with various length ssDNA overhang on partial dsDNA. YFP-NS3h was titrated with various length overhangs at the 3' end of partial dsDNA with TAMRA labeled on either end of the duplex. Emission spectra was collected and FRET_{app} was calculated from donor quenching. (A, C) FRET_{app} was then fitted to binding isotherms to determine E_{DA} as shown in panel B and D.

The experiment was conducted to evaluate the effect of increasing the overhang size on a partial dsDNA without increasing the dsDNA size. In other words, the junction where ssDNA and dsDNA meet doesn't change. It was expected that by increasing the available binding space for the helicase, more helicase would bind and hence more FRET would be detected. Contrary to our expectations, longer overhangs while keeping the junction at the same length resulted in

reduced FRET. This might be because extra helicase proteins may be binding but they are orientated in such a manner that they cannot participate in FRET.

5.7. Using FRET to identify the effect of increasing the duplex size

The experiment was conducted to determine the impact of moving the junction on a partial dsDNA without increasing the total DNA size. It was assumed that more FRET would be observed with smaller duplex. Chapter 4.12 contains the sequences of all the oligonucleotides utilized in this study.

To evaluate the impact of increasing the duplex size on a partial dsDNA without increasing the total DNA size, helicase was titrated using 45 nucleotides long TYE labeled oligonucleotide duplexed with 15, 20 and 25 bases long complementary strands. The experiment was conducted in a similar fashion to previous ones. The experiment was conducted by dissolving concentrated and purified yellow fluorescent protein in helicase dilution buffer, making a 500 μ l sample with a YFP concentration of 100 nM. This sample was then transferred into the cuvette. In a separate Eppendorf tube, 5'TYE oligonucleotide was also dissolved in the helicase dilution buffer, making a 5 μ M TYE labeled oligonucleotide solution. The cuvette with the YFP sample was excited at 450 nm and emission was collected in the range of 460–600 nm. After the initial scan, for each subsequent scan, a 1 μ l aliquot of sample containing labeled oligonucleotide was added to the cuvette, the sample was excited, and emission between 460-600 nm was collected. The assay was then repeated using three more sample containing: 5'TYE-comp15, 5'TYE-comp20, and 5'TYE-comp25. In total, 7 subsequent spectrums were collected with each titration and a

decrease in YFP-NS3h emission was observed with each addition of an oligonucleotide (fig. 5.7-1 A, B, C, and D).

The maximums at 520 nm from each titration were extracted and FRET_{app} was calculated using equation 3.5-4. Calculated FRET_{app} was then plotted against the oligonucleotide concentration in fig. 5.7-1E. This data was then fitted to equation 3.5-5 in order to obtain E_{DA} fig. 5.7-1F.

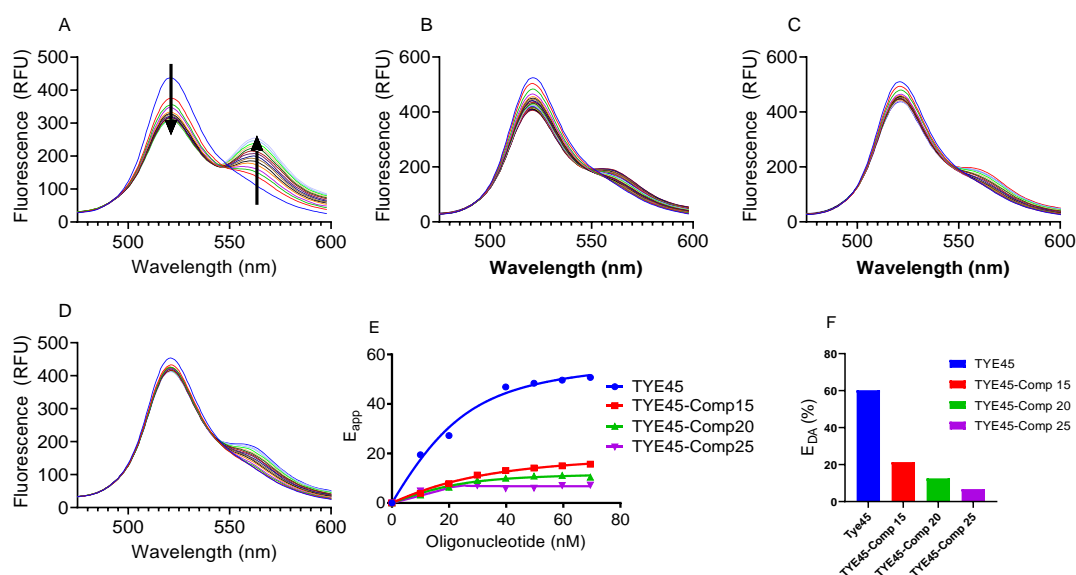


Figure 5.7-1 Titration of YFP-NS3h with varied duplex size on partial dsDNA.

Titration of YFP-NS3h were performed using 1 μl aliquots of 5 μM dsDNA. TYE45, TYE45-Comp15, TYE45-Comp20, and TYE45-Comp25 dsDNA were used for titration, and samples were excited at 450 nm and emission spectra were obtained as shown in panels A, B, C, and D, respectively. Panel E shows the calculate FRET_{app} from donor quenching. Panel F is calculated pairwise FRET efficiency.

The purpose of this experiment was to assess the effect of increasing the duplex size on a partial dsDNA without increasing the overall DNA size. With smaller duplexes, more FRET was

observed suggesting helicase is binding towards the junction or multiple helicases are binding on the ssDNA occupying the free space leading to helicase binding closer to the acceptor.

5.8.Summary

In this chapter, FRET was utilized to examine the interaction between helicase and oligonucleotide, and it was calculated as pairwise FRET efficiency based on the decrease in donor intensity due to fluorescence quenching. But before any FRET measurements were taken, factors that could indirectly affect the fluorescence emission of the donor were examined. Those factors include: change in donor fluorescence emission due to helicase binding onto the oligonucleotide, possible interaction of the oligonucleotide with yellow fluorescent protein, and effect on donor fluorescence in the presence of an acceptor alone.

Based on our data, no donor quenching was observed when YFP-NS3h was titrated with an unlabeled oligonucleotide or when free YFP was titrated with a TAMRA-labeled oligonucleotide, suggesting the donor quenching is not affected by oligo binding into the binding site of the helicase and there is no interaction between donor and acceptor (YFP and TAMRA) by themselves. Donor quenching is observed only when both donor and acceptor are bound to their respective molecules (helicase and DNA) and donor quenching is the result of helicase binding to the oligonucleotide.

The next assay was to examine whether less FRET is observed from an acceptor and donor located farther apart than those located closer to each other. This experiment was done using two different partially dsDNA with TAMRA labeled on either end of the junction. This assisted us in determining that the oligonucleotide behaves as expected upon binding. The results show that

an acceptor located further away from a partial dsDNA junction exhibits less FRET than an acceptor located directly adjacent to the junction or closer to helicase binding site. Afterwards, the binding orientation of the helicase was examined. For this experiment, YFP-NS3h was titrated with a TAMRA-labeled oligonucleotide covalently bound to the ssDNA at either 3' or 5' end. Based on different estimates of pairwise FRET efficiency (EDA) from each nucleotide, it seems that helicase must be bound with an orientation facing one way on the ssDNA.

Lastly, two experiments were done to examine how the size of the overhang versus the duplex affects pairwise FRET efficiency. It was expected that increasing the size of the overhang would allow more helicase to bind and hence more FRET would be detected. Contrary to our expectations, longer overhangs resulted in reduced FRET. This might be because multiple helicase proteins are bound but they are not orientated in such a manner that they can participate in FRET. Whereas as expected, the size of the duplex correlated with FRET efficiency, greater pairwise FRET efficiency was estimated with a smaller duplex and lower pairwise FRET efficiency with a longer duplex.

6. Variable nsp3 Mac1 Domain of SARS-CoV-2 retains the ability to bind ADP-ribose

6.1.Introduction

Antivirals designed to combat the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) will most likely target viral non-structural proteins required for viral replication. The positive sense RNA genome of SARS-CoV-2 is 30,000 nucleotides long and contains numerous open reading frames. The majority of the proteins that comprise the viral replicase are encoded by the "rep 1ab" reading frame. Rep 1ab encodes for a 7,096 amino acid-long polyprotein that is eventually cleaved by the viral proteases into 15 mature proteins. Parts of the SARS-CoV-2 rep 1ab polyprotein are remarkably similar to the rep 1ab protein of other coronaviruses (Figure 6.1-1), indicating that drugs targeting the SARS-CoV nonstructural proteins may be effective against SARS-CoV-2. The largest variance occurs in a Nsp3 protein domain suspected of binding ADP-ribose, referred here as the "mac1" domain to distinguish it from the two downstream "SARS unique macrodomains (SUDs)," which do not bind ADP-ribose.

Nsp3 is a large membrane-bound multidomain protein known to be essential for viral replication and transcriptional complex formation (van Hemert et al., 2008), as well as for the development of double membrane complexes (Angelini et al., 2013) It is encoded by Orf1a or Orf1b as part of a large polyprotein and is cleaved into mature non-structural proteins by a papain-like protease (Ziebuhr et al., 2000). The majority of variation in nsp3 between Coronavirus 1 and Coronavirus 2 occurs in the macrodomains near the protein's N-terminus. Macrodomains in general are three-layered $\alpha/\beta/\alpha$ / sandwich folds that are not limited to one function. While

SARS-CoV-1 has three macrodomains in tandem, only the MAC1 macrodomain binds ADP ribose.

The Mac1 domain's amino acid sequences differ by 26% between SARS-CoV-1 and SARS-CoV-2.

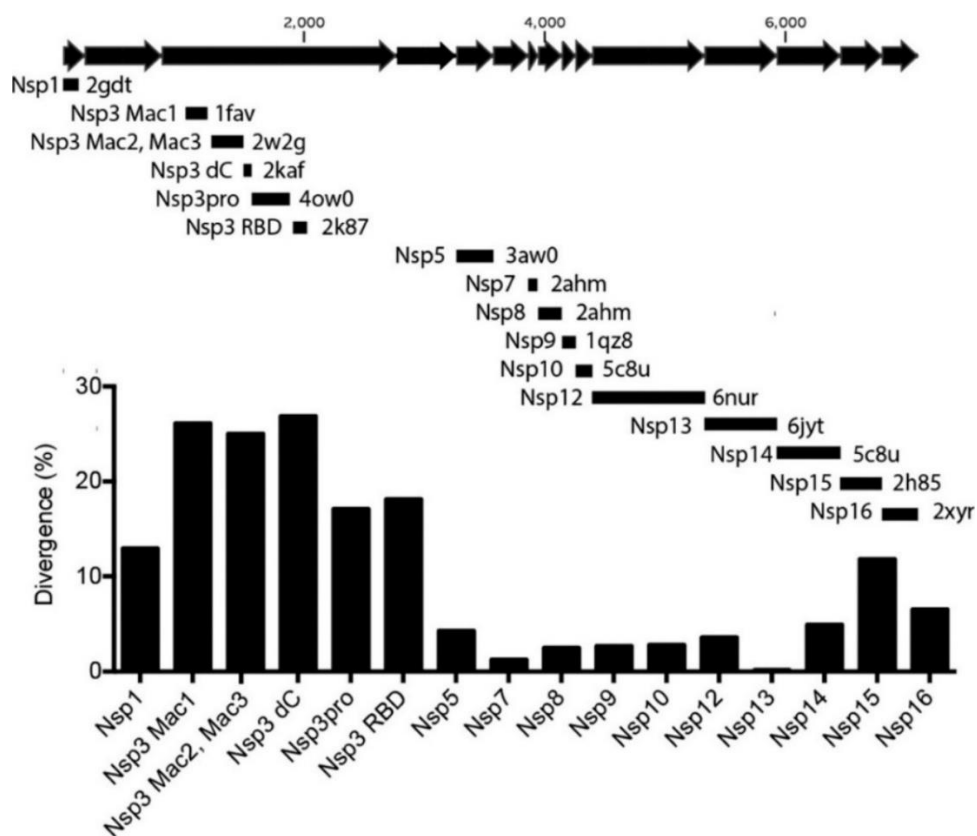


Figure 6.1-1 **Sequence divergence between potential drug targets in SARS-CoV and SARS-CoV-2.**

The SARS-CoV-2 rep 1ab peptide sequence was aligned with each of the PDB files listed, which describe an atomic structure of a homologous region of the SARS-CoV-1 rep 1ab polyprotein. Nsp's are shown in sequence as black arrows (note: there is no "nsp11", and the translational frameshift occurs after nsp10). The percent of amino that differ in each protein is plotted. Nsp1 is and interferon antagonist. (Almeida et al., 2007) The Nsp3 Mac1 domain is studied here, the Nsp3 SUD consists of tandem macrodomains that bind G-quadruplex structures, (Kusov et al., 2015) Nsp dC is the C-terminus of the SUD, (M. A. Johnson et al., 2010) Nsp3pro is a papain-like protease, (Báez-Santos et al., 2014) and Nsp3 RBD is another possible RNA binding domain. (Serrano et al., 2009) Nsp5 is the main viral protease. (Akaji et al., 2011) Nsp7 and Nsp8 are polymerase cofactors (Kirchdoerfer & Ward, 2019) Nsp9 is an RNA binding protein (M. P. Egloff et al., 2004) Nsp10 is a zinc-binding cofactor for Nsp14 and Nsp16 (Ma et al., 2015) Nsp12 is the RNA polymerase (Kirchdoerfer & Ward, 2019). Nsp13 is a helicase (Jia et al., 2019). Nsp14 is a 3'-5' exonuclease and a 7-methyltransferase (Ma et al., 2015). Nsp15 is an RNA endonuclease (Ricagno et al., 2006). Nsp16 is an RNA cap 2'-O-Methyltransferase (Decroly et al., 2011).

Many of the 47 variant residues in the 180 amino acid long SARS Mac1 domain are clustered near its N-terminus in a region that is particularly variable in the three Coronaviridae genera (Figure 6.1-2). Mac1 domains from coronaviruses that cause the common cold (alpha coronaviruses) and the beta coronaviruses like SARS-CoV and Middle East respiratory syndrome corona-virus (MERS-CoV) all bind ADP ribose. However, the same domain in gamma coronaviruses does not bind ADP-ribose.

Although the Mac1 domain of SARS-CoV-1 is known to bind and catalyze the hydrolysis of ADP-ribose 1" phosphate, the significant variation prevents the use of the SARS-CoV-1 Mac1 domain structures as scaffolds to design compounds that might target this nsp3 region in SARS CoV-2. This is especially important given that the same nsp3 domain from gamma coronaviruses or even the Mac2 and Mac3 domains of the same coronaviruses do not bind or hydrolyze ADP-ribose. This chapter examines the binding of ADP-ribose and similar nucleotides to a nonstructural protein 3 (nsp3) macrodomain which might be utilized to find antiviral drugs. The binding between SARS-CoV-2 Mac1 domain and ADP-ribose was investigated utilizing recombinant pure protein and isothermal titration calorimetry (ITC).

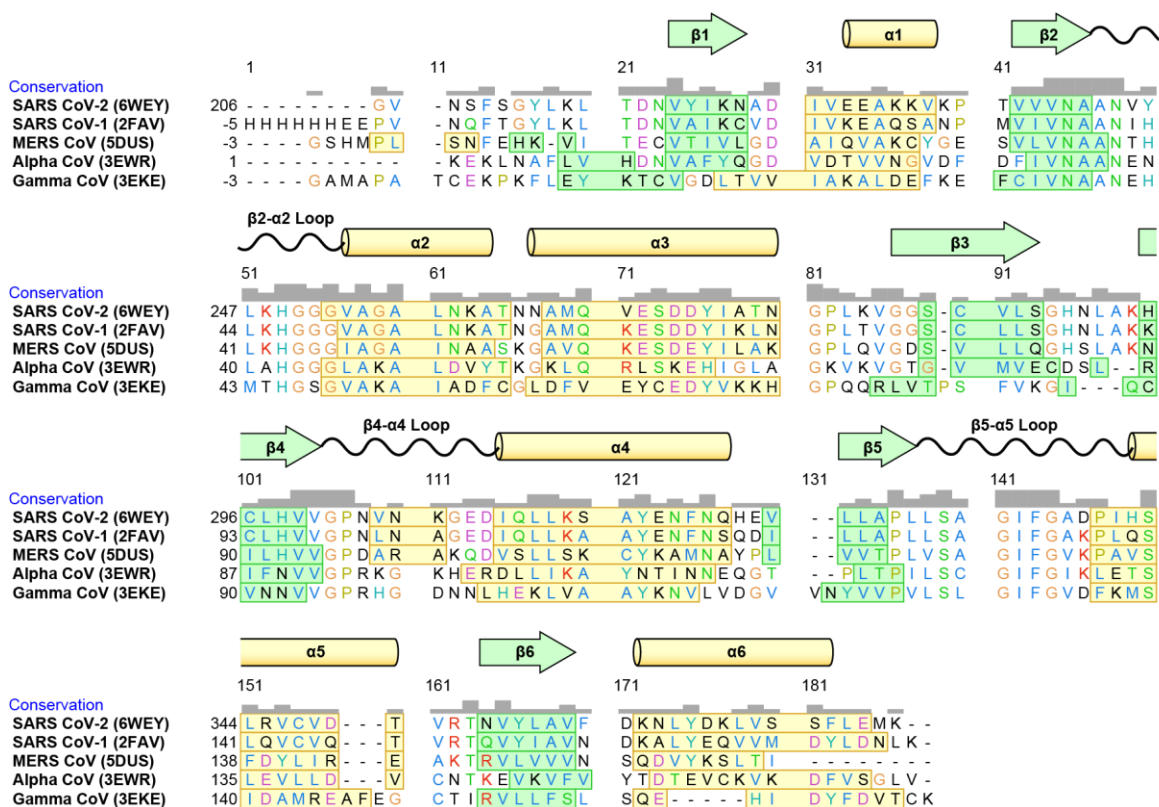


Figure 6.1-2 **Variation in the Mac1 domains of coronaviruses.**

Variation in the Mac1 domains of coronaviruses. Mac1 structures were aligned using the “MatchMaker” function of UCSF Chimera (version 1.14) (Pettersen et al., 2004a). Amino acids are colored by class. β sheets are denoted with green boxes, and α helices are denoted with yellow boxes.

6.2. Purification and Isolation of Mac1 domain

In order to determine the binding of ADP-ribose to the Mac1 domain, the Mac1 domain was isolated and purified using a NiNTA and a gel filtration column. The detailed explanation of gene synthesis and the protocol for protein purification are discussed in the materials and methods section (Chapter 4.2). Briefly, the pET11-COVID-MacroX plasmid was used to transform BL21 (DE3) cells. Transformed BL21(DE3) were used to start a 3 ml starter culture until the OD₆₀₀ reached 1 and then transferred to a 1L fresh medium containing 100 mg/ml ampicillin. After the

OD₆₀₀ reached 1, the cells were induced using 1 mM IPTG and harvested the next morning. The protein was purified using a step gradient on a NiNTA column with a final concentration of 500 mM Imidazole. To verify the expression of Mac1 domain, 10 µl sample from each IMAC fraction was loaded on to 15% SDS gel (Figure 6.1-1).

After the expression of mac1 domain was verified using SDS gel, The eluted fractions at 500 mM imidazole were then concentrated and loaded onto a 250 ml Sephacryl S300 gel filtration column for further purification and eluted with 10 mM MOPS and 150 mM NaCl. Concentration of the purified protein was determined by measuring absorbance at 260 nm using a molar extinction coefficient calculated with the ProtParam tool (<https://web.expasy.org/protparam/>).

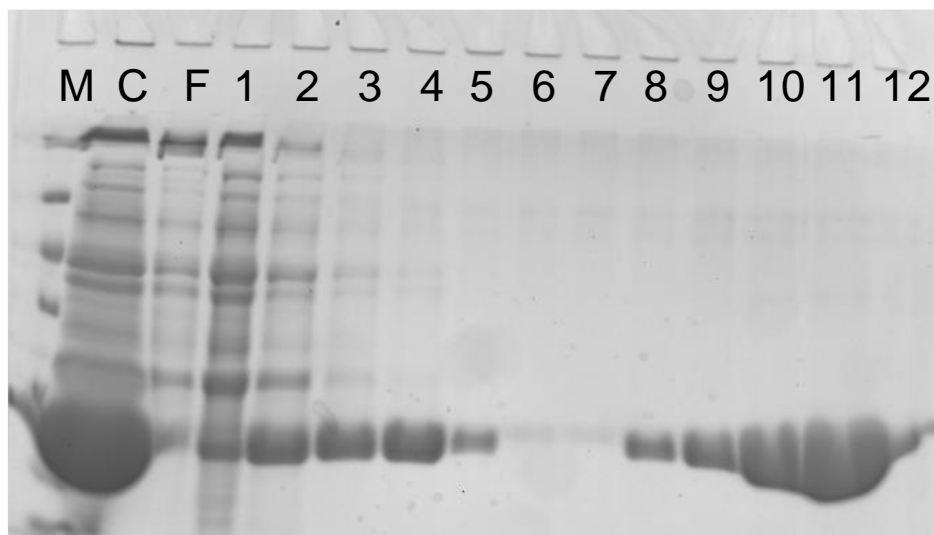


Figure 6.2-1 15% SDS PAGE of SARS-CoV-2 Mac1 domain.

Panel A: 15% SDS PAGE showing 10 µL samples of: a soluble crude lysate of induced BL21(DE3) cells harboring the plasmid p21-COVID-Mac1 (lane C), proteins the do not bind a Ni-NTA column (F), and fractions eluted from a NiNTA column during an imidazole step gradient from 0 mM (lanes 1-3), 5 mM (lanes 4-6), 40 mM (lanes 7-9), and 500 mM (lanes 10-12). Protein markers (lane M) are 116, 66.2, 45, 35, and 25 kDa.

6.3.Determine optimal conditions for ITC experiments.

Purified Mac1 was used in all ITC experiments to determine the binding of the Mac1 to ADPr. The starting ligand and proteins concentration were based on the instructions provided in (Velázquez-Campoy et al., 2004) However multiple conditions were employed to determine optimum ratio between ligand and protein concentrations. Prior to the start of the experiment, samples of ligand and protein were diluted in 10 mM MOPS and 150 mM NaCl at pH 7 and degassed for 30 minutes at 400 mm Hg. At a temperature of 20 °C, measurements were conducted by injecting 2.0 μ L aliquots of 500 μ M ADP-ribose into various concentrations of protein in a 175 μ L starting volume stirred at 250 rpm. The data gathered is depicted below.

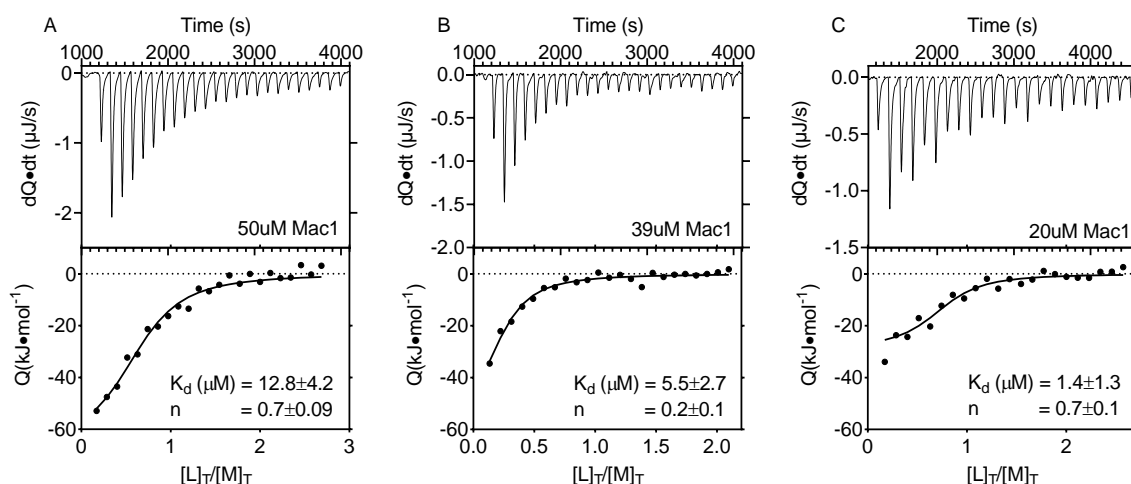


Figure 6.3-1 **Mac1 titration with ADP-ribose.**

Panel A, B and C top: Titration thermograph of 500 μ M ADP added to 50 μ M Mac1 dissolved in 10 mM MOPS 150 mM NaCl pH7. The titration thermogram represented as the amount of heat released per unit of time after each injection. The relationship between the quantity of heat generated by each injection and the ratio of total ligand to total protein concentration is shown underneath it. Panel A and B bottom: The circles plot experimental data, while the line represents model's best fit to n identical and independent binding sites. Data fit best to the independent model with dissociation constant of 6.6 ± 3.3 μ M, 5.5 ± 2.7 μ M and 1.4 ± 1.3 μ M, stoichiometric constant of 0.7 ± 0.1 , 0.2 ± 0.1 , and 0.7 ± 0.1 enthalpy of -66.8 ± 12.9 (kJ/mol), -63.8 ± 34.7

(kJ/mol) and -72.7 ± 20.1 (kJ/mol), $-T\Delta S$ of 37.7 (kJ/mol), 34.3 (kJ/mol) and 39.9 (kJ/mol) Gibbs free energy of -29.1 (kJ/mol), -29.5 (kJ/mol) and -32.8 (kJ/mol) for panel A and B respectively.

According to the experiment conducted with various concentrations of the Mac1, 50 μM for the Mac1 and 500 μM for ADP-ribose appears to be the optimal conditions that were used for the remainder of the ITC binding studies

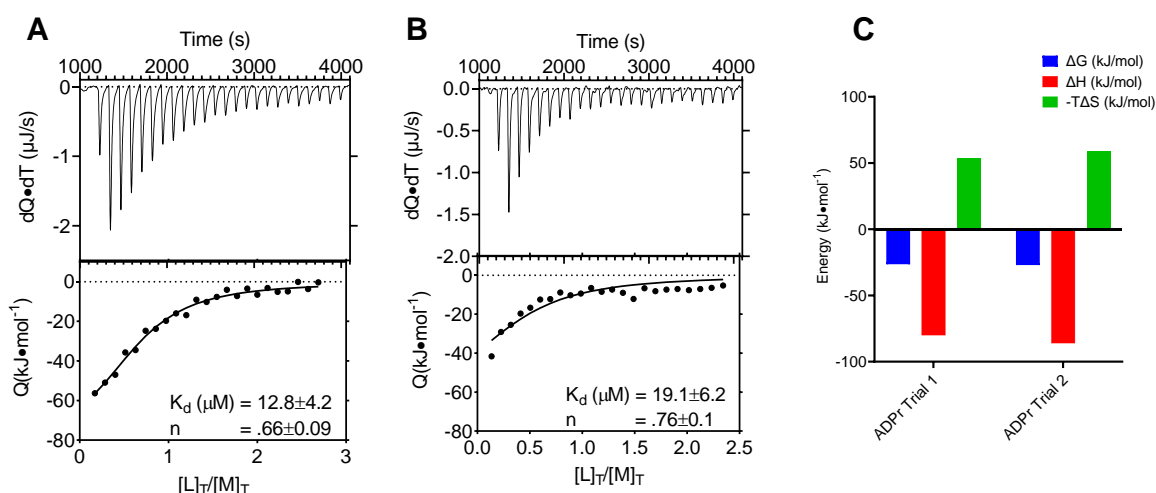


Figure 6.3-2 **Mac1 titration with ADP-ribose.**

Panel A and B top: 50 μM Mac1 titrated with 500 μM ADP-ribose dissolved in 10mM MOPS 150mM NaCl pH7. The titration thermogram is represented as the amount of heat released per unit of time after each injection. The relationship between the quantity of heat generated by each injection and the ratio of total ligand to total protein concentration is shown underneath it. In Panel A and B, bottom, the circles plot experimental data, while the line represents the model's best fit to n identical and independent binding sites. Fitting the data to an independent model gives us the dissociation constant of 12.8 ± 4.2 μM and 19.1 ± 6.2 μM , stoichiometric constant of 0.66 ± 0.1 and 0.76 ± 0.1 , enthalpy of -86.0 ± 5.2 (kJ/mol) and -70.0 ± 22.6 (kJ/mol), $-T\Delta S$ of 59.0 (kJ/mol) and 43.6 (kJ/mol) Gibbs free energy of -27.0 (kJ/mol) and -26.4 (kJ/mol) for panel A and B respectively.

Repeated ITC experiments revealed that the purified recombinant protein bound ADP-ribose Figure 6.3-3 with a dissociation constant of 16 ± 4 μM (uncertainty is the standard

deviation of K_d 's from two independent titrations). To examine binding specificity, similar titrations were repeated with related nucleotides such as: ATP (Fig. 6.3-3), ATP-glucose (Fig. 6.3-4), AMP (Fig. 6.3-5), Adenosine (Fig. 6.3-6), cAMP (Fig. 6.3-7) and ADP (Fig. 6.3-8).

6.3.1. ATP binding to Mac1

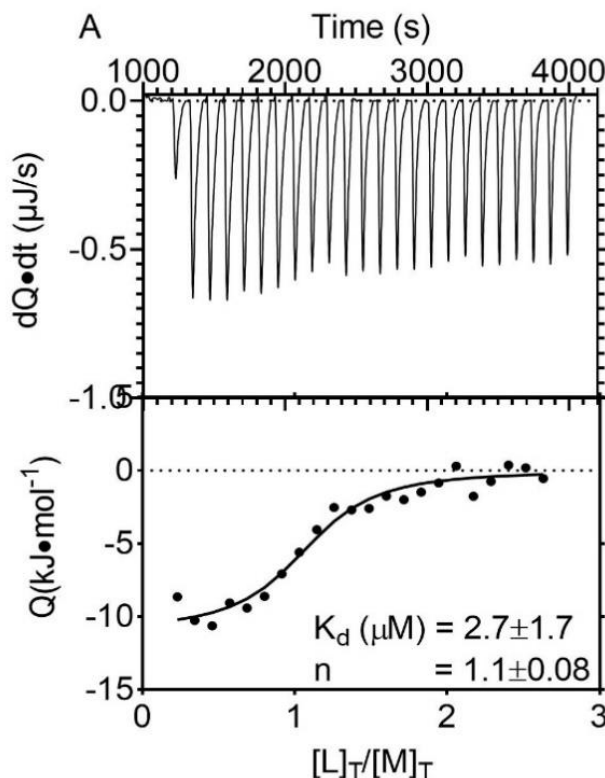


Figure 6.3-3 ATP titration with Mac1 domain.

Top: 50 μM Mac1 titrated with 500 μM ATP dissolved in 10mM MOPS 150mM NaCl pH7. The titration thermogram is represented as the amount of heat released per unit of time after each injection. The relationship between the quantity of heat generated by each injection and the ratio of total ligand to total protein concentration is shown underneath it. In bottom panel, the circles plot experimental data, while the line represents the model's best fit to n identical and independent binding sites. Fitting the data to an independent model gives us the dissociation constant of $2.7 \pm 1.7 \mu\text{M}$, stoichiometric constant of 1.1 ± 0.08 , enthalpy of $-11.0 \pm 1.3 \text{ (kJ/mol)}$, $-\Delta S$ of -20.23 (kJ/mol) and Gibbs free energy of -31.24 (kJ/mol) .

6.3.2. ATP-Glucose binding to Mac1

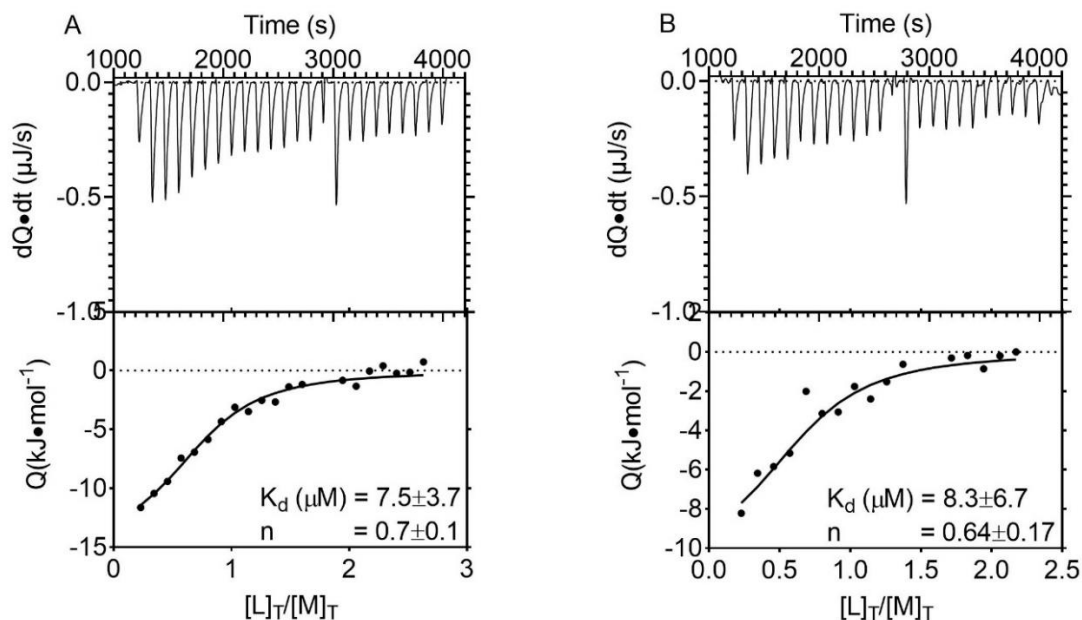


Figure 6.3-4 Mac1 titration with ATP-Glucose.

Panel A and B top: 50 μM Mac1 titrated with 500 μM ATP-glucose dissolved in 10mM MOPS 150mM NaCl pH7. The titration thermogram is represented as the amount of heat released per unit of time after each injection. The relationship between the quantity of heat generated by each injection and the ratio of total ligand to total protein concentration is shown underneath it. In Panel A and B, bottom, the circles plot experimental data, while the line represents the model's best fit to n identical and independent binding sites. Fitting the data to an independent model gives us the dissociation constant of 7.5 ± 3.7 μM and 8.3 ± 6.7 μM , Stoichiometric constant of 0.7 ± 0.1 and 0.64 ± 0.17 , Enthalpy of -15.2 ± 3.1 (kJ/mol) and -11.0 ± 5.6 (kJ/mol), $-\Delta S$ of -13.6 (kJ/mol) and -17.5 (kJ/mol) Gibbs free energy of -28.8 (kJ/mol) and -28.8 (kJ/mol) for panel A and B respectively.

6.3.3. AMP binding to Mac1

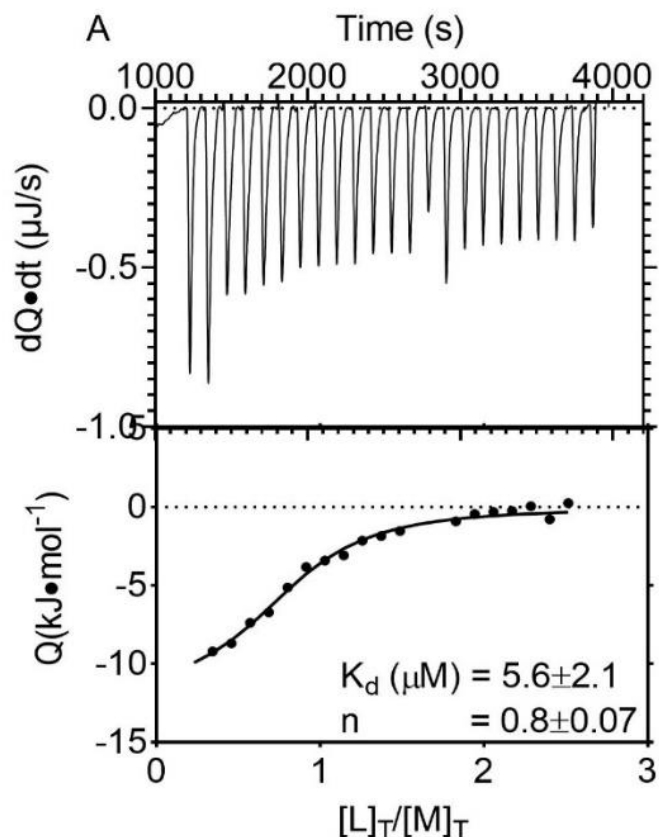


Figure 6.3-5 **Mac1 titration with AMP.**

Top: 50 μM Mac1 titrated with 500 μM AMP dissolved in 10mM MOPS 150mM NaCl pH7. The titration thermogram is represented as the amount of heat released per unit of time after each injection. The relationship between the quantity of heat generated by each injection and the ratio of total ligand to total protein concentration is shown underneath it. In bottom panel, the circles plot experimental data, while the line represents the model's best fit to n identical and independent binding sites. Fitting the data to an independent model gives us the dissociation constant of 5.6 ± 2.1 μM , stoichiometric constant of 0.8 ± 0.07 , enthalpy of -12.2 ± 1.7 (kJ/mol), $-\Delta S$ of -17.3 (kJ/mol) and Gibbs free energy of -29.4 (kJ/mol).

6.3.4. Adenosine binding to Mac1

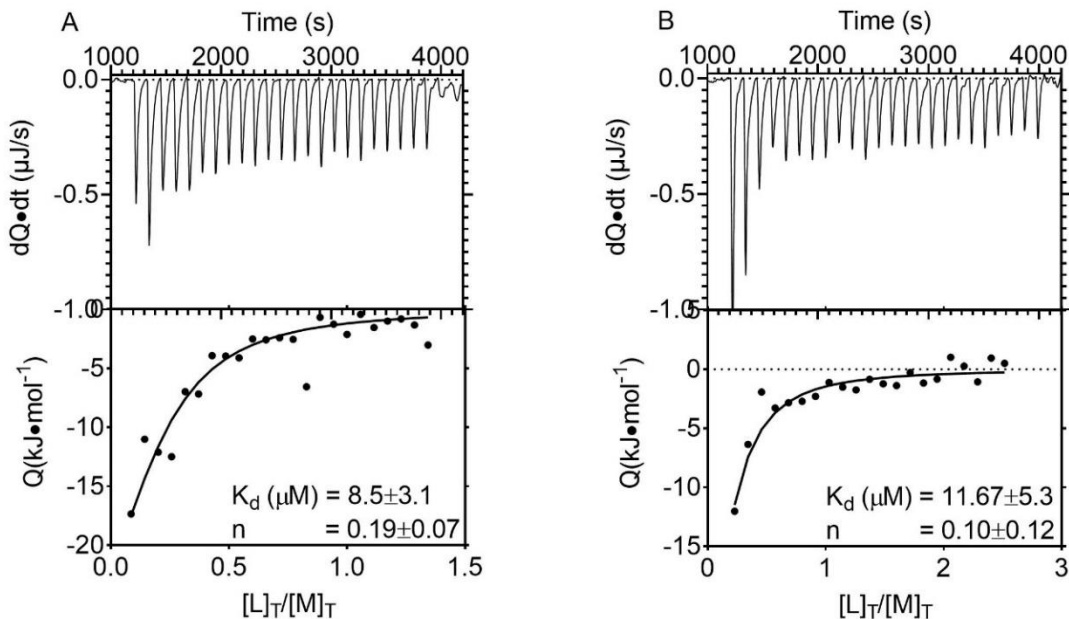


Figure 6.3-6 **Mac1 titration with Adenosine.**

Panel A and B top: 50 μM Mac1 titrated with 500 μM adenosine dissolved in 10mM MOPS 150mM NaCl pH7. The titration thermogram is represented as the amount of heat released per unit of time after each injection. The relationship between the quantity of heat generated by each injection and the ratio of total ligand to total protein concentration is shown underneath it. In Panel A and B, bottom, the circles plot experimental data, while the line represents the model's best fit to n identical and independent binding sites. Fitting the data to an independent model gives us the dissociation constant of 8.5 ± 3.1 μM and 11.7 ± 5.4 μM, stoichiometric constants of 0.2 ± 0.1 and 0.1 ± 0.1 , enthalpy of -34.5 ± 15.9 (kJ/mol) and -33.99 ± 20.6 (kJ/mol), $-T\Delta S$ of 2.9 (kJ/mol) and 0.8 (kJ/mol) Gibbs free energies of -31.5 (kJ/mol) and -34 (kJ/mol) for panel A and B respectively.

6.3.5. cAMP binding to Mac1

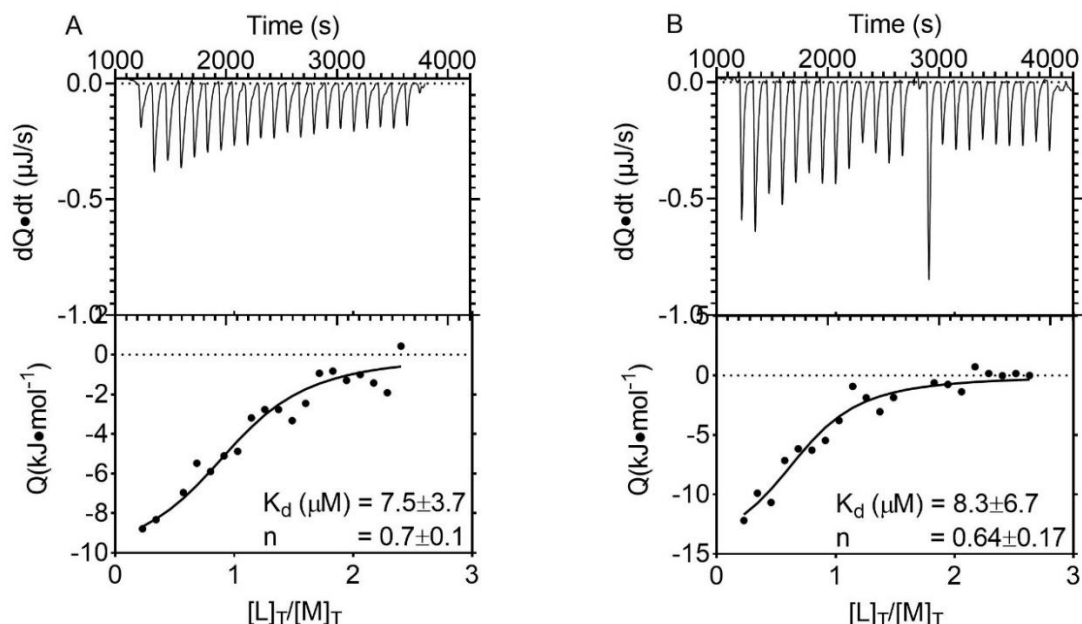


Figure 6.3-7 Mac1 titration with cAMP.

Panel A and B top: 50 μM Mac1 titrated with 500 μM cAMP dissolved in 10mM MOPS 150mM NaCl pH7. The titration thermogram is represented as the amount of heat released per unit of time after each injection. The relationship between the quantity of heat generated by each injection and the ratio of total ligand to total protein concentration is shown underneath it. In Panel A and B, bottom, the circles plot experimental data, while the line represents the model's best fit to n identical and independent binding sites. Fitting the data to an independent model gives us the dissociation constant of 0.7 ± 0.1 and 0.7 ± 0.1 , enthalpy of -10.5 ± 1.9 (kJ/mol) and -15.1 ± 5.9 (kJ/mol), $-T\Delta S$ of -18.4 (kJ/mol) and -14.0 (kJ/mol) Gibbs free energy of -28.8 (kJ/mol) and -29.1 (kJ/mol) for panel A and B respectively.

6.3.6. ADP binding to Mac1

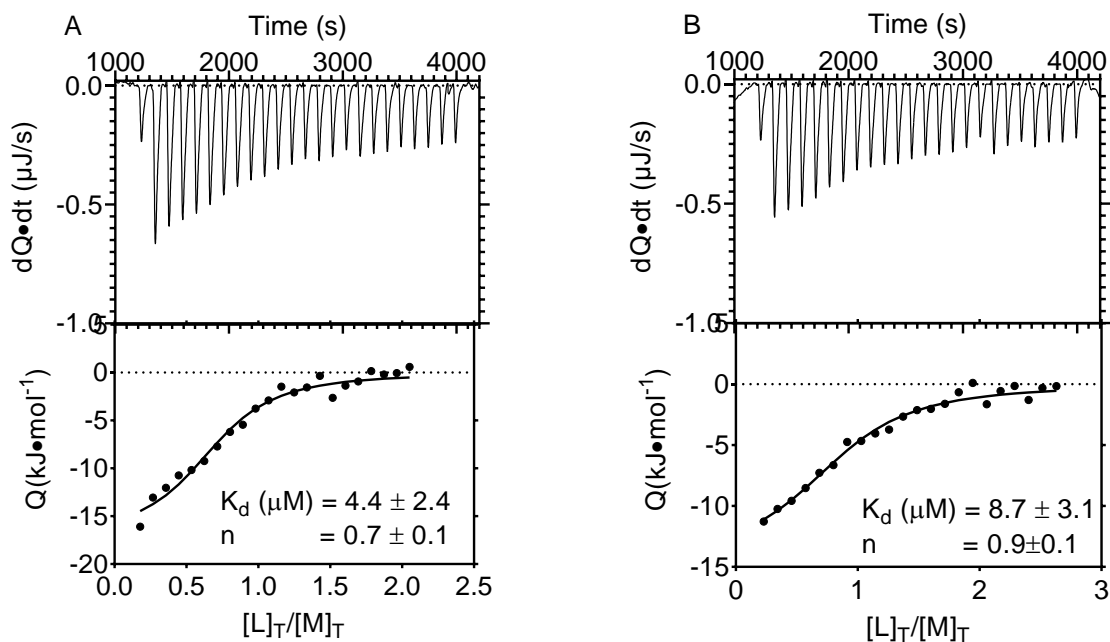


Figure 6.3-8 Mac1 titration with ADP.

Panel A and B top: 50 μM Mac1 titrated with 500 μM ADP dissolved in 10mM MOPS 150mM NaCl pH7. The titration thermogram is represented as the amount of heat released per unit of time after each injection. The relationship between the quantity of heat generated by each injection and the ratio of total ligand to total protein concentration is shown underneath it. In Panel A and B, bottom, the circles plot experimental data, while the line represents the model's best fit to n identical and independent binding sites. Fitting the data to an independent model gives us the dissociation constant of $4.4 \pm 2.4 \mu M$ and $8.7 \pm 3.1 \mu M$, stoichiometric constant of 0.7 ± 0.1 and 0.9 ± 0.1 , enthalpy of -17.3 ± 2.7 (kJ/mol) and -14.5 ± 1.9 (kJ/mol), $-T\Delta S$ of -12.8 (kJ/mol) and -13.9 (kJ/mol) Gibbs free energy of -30.05 (kJ/mol) and -28.4 (kJ/mol) for panel A and B respectively.

6.4. Summary of ITC Data with nucleotides.

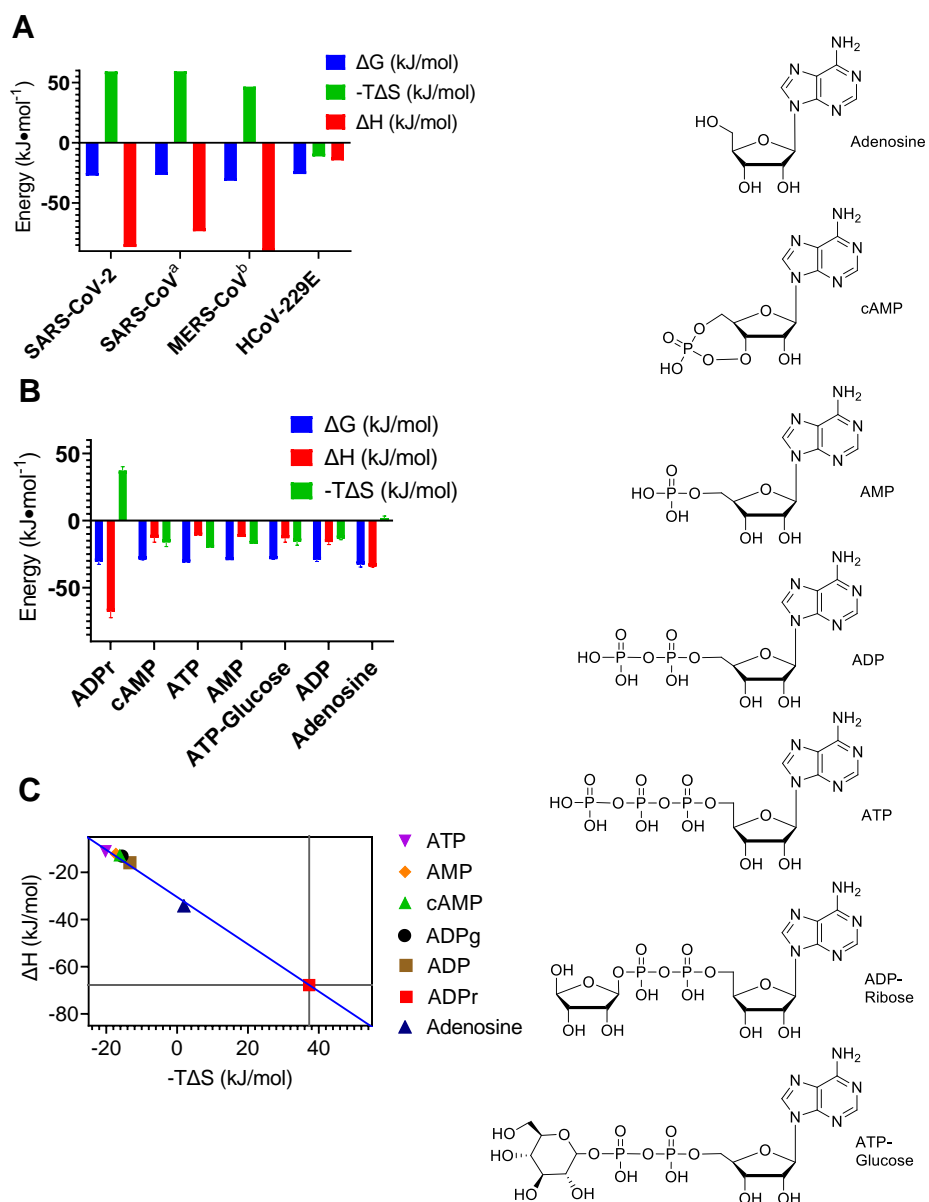


Figure 6.4-1 **Thermodynamic data for all nucleotides were obtained using ITC.**

Panel A. Comparison of the thermodynamics of ADP-ribose binding by Mac1 domains from SARS-CoV-2, SARS-CoV-1 (M.-P. Egloff et al., 2006), MERS-CoV (Cho et al., 2016a), and an alpha coronavirus (Piotrowski et al., 2009). ITC experiments like those shown in panel A were repeated three times with each of the nucleotide listed. Means are plotted and error bars are standard deviations. Average (\pm SD) dissociation constants were $10 \pm 4 \mu\text{M}$ for ADP-ribose, $7 \pm 3 \mu\text{M}$ for ADP, $3 \pm 3 \mu\text{M}$ for ATP, $6 \pm 2 \mu\text{M}$ for AMP, $8 \pm 5 \mu\text{M}$ for ADP-glucose, $8 \pm 5 \mu\text{M}$ for cAMP, and $10 \pm 4 \mu\text{M}$ for adenosine. The same data from panel B is shown in panel C but as part of a

thermodynamic optimization plot. On the right, structures of all the nucleotides used specifically in these ITC studies.

All nucleotides lacking the ribose moiety bound with similar affinities, but none bound with an enthalpy change like that seen with ADP-ribose, suggesting specific interaction between ADP ribose and Mac1 domain.

To recap, the total enthalpy change is calculated as the sum of the positive and negative enthalpic contributions. The positive enthalpy changes are due to the formation of hydrogen bonds and van der Waals interactions between the ligand and the protein, while the unfavorable enthalpic contribution is due to the ligand being desolvated or removed from solution in order to form a complex (FREIRE, 2008). Based on the structure (PDB ID 6W02) of the SARS-CoV-2 Mac1 domain bound to ADP-ribose determined by Michalska. et al. (2020) and in comparison, to unliganded Mac1, notable conformational differences were observed. The notable conformational changes occur in three surface-exposed loops in or near the ligand-binding pocket. These loops connect strand $\beta 2$ with helix $\alpha 2$ (the $\beta 2$ - $\alpha 2$ loop), strand $\beta 4$ with helix $\alpha 4$ ($\beta 4$ - $\alpha 4$ loop), and strand $\beta 5$ with helix $\alpha 5$ ($\beta 5$ - $\alpha 5$ loop). Not all changes in conformation are due to the direct interaction of protein with the ADP ribose however these conformational changes lead to an opening of binding pocket allowing the amide N atom of G252 to participate in a hydrogen bonding interaction with the 1'-hydroxyl of the ribose moiety of ADP-ribose. G252 on SARS-CoV2 Mac1 Domain corresponds to V33 residue in Chikungunya macrodomain, which Eckeï et al. (2017) proposed is needed for de-ADP-ribosylation activity based on results for site directed mutagenesis. N244 on the loop that connects strand $\beta 2$ with helix $\alpha 2$ (the $\beta 2$ - $\alpha 2$ loop) also directly interacts with the distal ribose moiety.

The direct interaction between both ribose moieties of ADP ribose and the pocket of Mac1 domain leads to very favorable enthalpic interaction seen in ITC data. On the opposite end, the binding of ADP ribose also leads to very unfavorable entropic energy due to restriction in degrees of freedom of the ligand when bound to the macrodomain.

The energetics of ADPr binding to the Mac1 domain are similar among beta coronavirus but significantly different from alpha coronavirus. There is some favorable Enthalpic driven binding of ADP-ribose with macrodomain of alpha coronavirus however it is significantly less than beta coronavirus suggesting that the binding of ADPr in beta coronavirus is enthalpically driven. Additionally, the entropic penalty of binding for beta coronavirus is significantly more whereas it is favorable for alpha coronavirus.

6.5. Optimization of DSF Assay conditions.

DSF has been used previously to study ligand binding to viral macrodomains (Cho et al., 2016b; Malet et al., 2009). In DSF experiments using SARS-CoV-2 Mac1, the presence of ADPr raised the Mac1 melting temperature in a concentration-dependent manner (Fig. 6.6-1a). To estimate the ligand concentrations needed to alter melting temperatures by 50% (EC₅₀), melting temperatures were fit to Eq. 3.6-2 (Fig. 6.6-1b). Such EC₅₀ values do not, however, describe protein-ligand affinity because DSF assays do not directly measure binding. The isothermal analysis recently described by Bai et al. was therefore used to estimate binding affinities. Fits of the fraction of protein unfolded at various temperatures in the presence of various ligand concentrations Eq. 3.6-9 (Fig. 6.6-1C) were used to estimate a dissociation constant (K_d) and equilibrium unfolding constant (K_u). K_d estimated using Eq 3.6-9 from DSF data at high

temperature when fitted to the Van't Hoff relationship (Fig. 6.6-1D), were in good agreement with the dissociation constant describing the interaction of ADPr and Mac1 (10 μ M) determined at 23 °C (Frick et al., 2020).

Dimethyl sulfoxide (DMSO) did not change the melting curve even at concentrations as high as 10% (v/v) and similar results were also obtained when titrations with ADPr were repeated in various buffers, with the pH ranging from 6.5 to 8.0, or in the presence of various concentrations of divalent metal cations (Mg^{2+} or Mn^{2+}). In contrast, the ionic strengths of the assay buffers influenced the results, with the largest ΔT_m values and lowest EC_{50} values being obtained at the lowest ionic strengths (Fig. 6.5-1C). Based on these results, DSF assays were subsequently performed in 20 mM MOPS pH 7 buffer containing 25 mM NaCl to reduce possible non-specific interactions with ligands. Z' factors were always above 0.5 for each plate and typically above 0.7. Plate-to-plate variability was negligible (Fig. 6.5-1D).

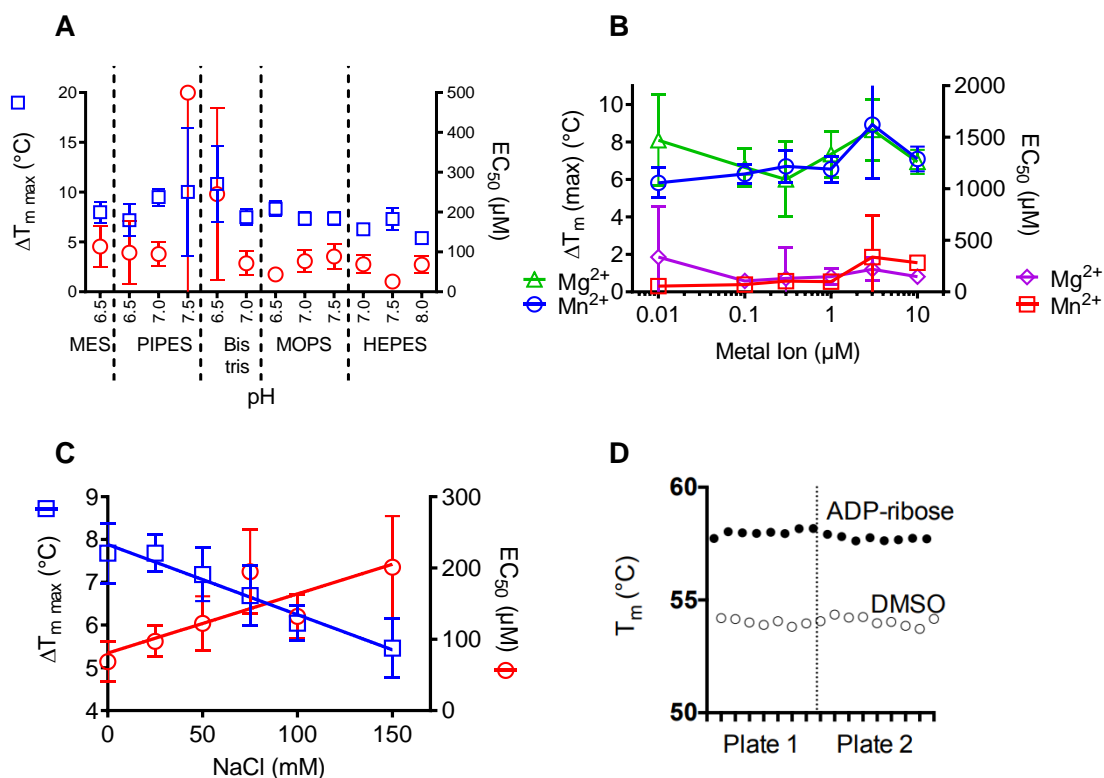


Figure 6.5-1 Optimization of DSF assay conditions.

The DSF assay was optimized by measuring ADP-ribose binding to the MAC1 domain under a variety of conditions, beginning with the use of various buffers ranging in pH from 6.5 to 8.0. (Panel A). Similarly, the impact of two divalent cations, manganese and manganese, on the EC₅₀ or T_m values for ADP-ribose was determined with no obvious effect (Panel B). Finally, the binding was influenced by the concentration of salt, with the lowest ionic strengths producing the greatest T_m values and the lowest EC₅₀ values (Panel C). Z' factors were always above 0.5 for each plate and typically above 0.7. Plate-to-plate variability was negligible (Panel D).

6.6.DSF to monitor the binding of ADP-Ribose with Mac1.

Purified SARS-CoV-2 Mac1 protein was prepared as described previously (Section 6.2). The conditions of DSF assay and the protocol is discussed in Section 6.5 and 4.1.4. Briefly, All DSF studies were carried out in a 96-well PCR plate using Eppendorf thermofluor in a similar manner. Mac1 was mixed in buffer containing SYPRO Orange dye, and an equal amount of protein-SYPRO

mixture was then added to each well on a plate containing a different concentration of compound. The temperature was raised from 20 °C to 95 °C at a rate of 2 °C per min while measuring the fluorescence in the “TAMRA” channel.

Each assay was repeated multiple times and data was normalized and plotted with fluorescence on y-axis and temperature on x-axis as shown on Figure 6-6-1 Panel A. T_m 's were calculated by fitting the data to equation 3.6.1 using either GraphPad Prism or TSA-CRAFT

Two methods were used to estimate the binding affinity of Mac1 domain. First the T_m values that were obtained using equation 3.6-1 were plotted against ADP-ribose concentration. This data was then fitted to Equation 3.6-2 to obtain EC_{50} and non-linear regression to estimate $\Delta T_{m \max}$ from the melting temperature of Mac1 in the absence of ligand (T_{m0}). $\Delta T_{m \max}$ is the maximum change in the melting temperature, and EC_{50} is the concentration of ligand needed to cause a change in the melting temperature by 50%. The second method used to estimate binding affinity is based on calculating the dissociation constant (K_d) of Mac1 and ADPr. The dissociation constant (K_d) of Mac1 and ADPr and the equilibrium constant describing protein unfolding (K_u) were estimated using isothermal analysis as described by Bai et al. (2019). Briefly, normalized melting curves were used to calculate the fraction of protein unfolded at a particular temperature (f_u), and those values fitted to the total ligand (L_t) and protein (P_t) concentrations using non-linear regression and Equation 3.6-9.

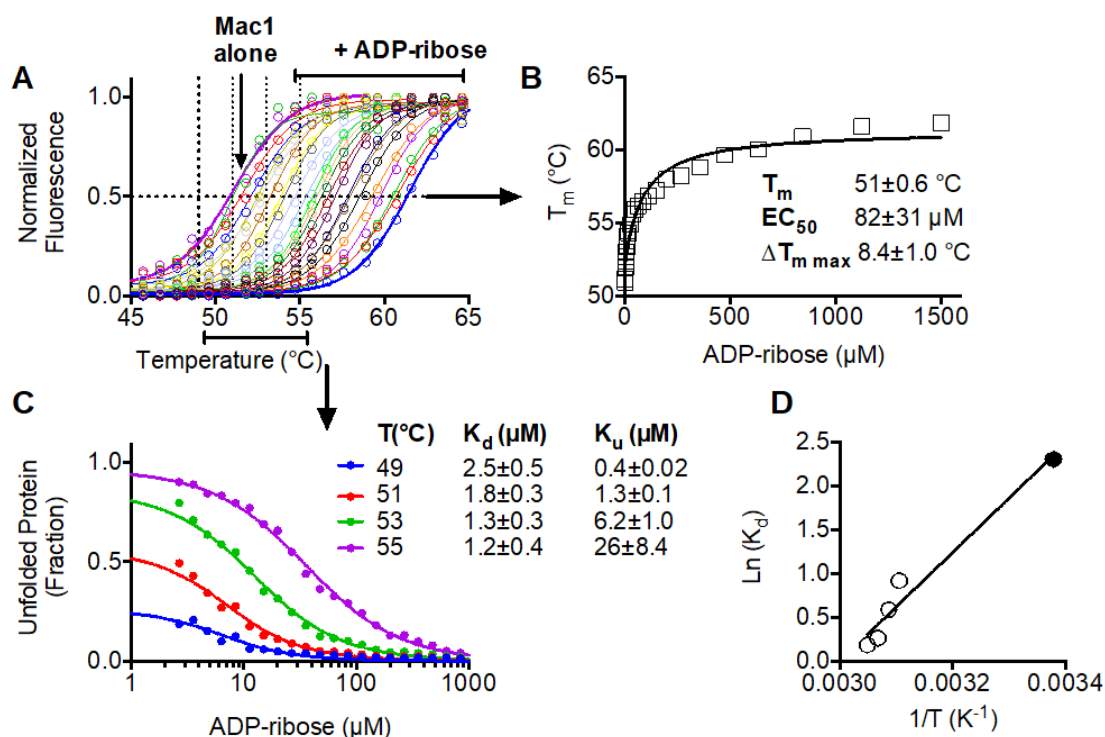


Figure 6.6-1 Methods used to estimate the binding affinity of Mac1 domain.

(A) Normalized SPYRO Orange fluorescence in the presence of 5 μM Mac1 protein at various temperatures in the presence of indicated concentrations of ADPr. Data are fit to equation 3.6-1 using non-linear regression with GraphPad Prism. (B) T_m values were obtained from direct fitting to Eq. 3.6-1 and this data is then plotted and fitted to Eq. 3.6-2. (B) Isothermal analysis of percent unfolded protein at each indicated temperature was collected and data was fit to Eq. 3.6-9 to estimate K_u and K_d . Uncertainties are standard errors of the curve fits. (C) Van't Hoff plot of estimated K_d values from panel C (open circles) and the K_d for ADPr binding to Mac1 that was previously determined at 23°C using isothermal titration calorimetry (filled circle) (Frick et al., 2020).

6.7.DSF to estimate the binding of various nucleotides with Mac1.

Optimized DSF conditions of 20 mM MOPS at pH 7 buffer containing 25 mM NaCl were used in the assay to compare the binding of various nucleotides using ADP-Ribose as a positive and DMSO as a negative control. The purified SARS-CoV-2 Mac1 protein was then titrated with a variety of nucleotides comparable to ADP-ribose to assess if the DSF assay accurately reproduces

the structure-activity relationships established in previous ITC investigations (Section 6.3 and Table 6.7-1) (Frick et al., 2020). All nucleotides tested were previously shown to bind Mac1. The main differences observed with ITC reflected the fact that binding of only ADP-ribose was significantly enthalpy driven, while the other nucleotides were bound by Mac1 with smaller favorable enthalpy, but with counterbalancing decreases in binding entropy so that the free energy changes (and K_d) were similar to ADP ribose (Frick et al., 2020). The majority of nucleotides altered T_m , except for ADP-glucose and AMP, which did not until the sample was saturated with ligand. In contrast to the K_d values determined in ITC tests, the EC_{50} values for each nucleotide varied significantly, with ADP-ribose having a much lower EC_{50} value than the other nucleotides (Table I).

<i>Compounds</i>	$\Delta T_{m\ max} (^{\circ}C)$	$EC_{50} (\mu M)$	$K_d (ITC)$
<i>ADP-ribose</i>	7±0.9	63 ± 21	10±6
<i>ADP-glucose</i>	~2 @ 500 μM	> 500	9.5±5
<i>ADP</i>	6.2±1	314 ± 43	3.4±0.6
<i>ATP</i>	5.6 ± 0.7	483 ± 129	2.4±1
<i>AMP</i>	~2 @ 500 μM	> 500	5.6±2
<i>cAMP</i>	1.9 ± 0.8	746 ± 340	6.1±3
<i>NADH</i>	5.8 ± 0.4	410 ± 76	44±20

Table 6.7-1 Effects of various nucleotides on SARS-CoV-2 Mac1.

$T_{m\ max}$ and EC_{50} values obtained in concentration response assays using the DSF assay and dissociation constant obtained via ITC. Most nucleotides caused T_m to change, with the notable exception of ADP-glucose and AMP, which did not unless the sample was saturated with 500 μM ligand concentration.

6.8. Summary

The study's relevance arises mostly from the discovery that the SARS-CoV-2 Mac1 domain binds ADP-ribose. While SARS-CoV and SARS-CoV-2 have a 26 percent amino acid sequence

difference, their structures are quite similar, which may account for their ability to bind ADP-ribose. This is the first step in justifying antiviral search for compounds that bind in lieu of ADP-ribose. However, further research is needed to fully grasp the antiviral potential of such drugs, since the biological function of ADP-ribose binding remains unknown. Some work with alpha coronaviruses suggests that ADP-ribose binding by the Mac1 domain is not required for viral replication (Keep et al., 2018). However, studies with other (+) RNA viruses suggest that macrodomains are essential for virulence (Abraham et al., 2018). Additionally, this study is significant because the synthetic codon-optimized plasmid yields up to 100 mg of soluble Mac1 domain protein per liter of *E. coli* culture, and this protein maintains a strong affinity for ADP-ribose. The protein might be employed in structural and screening research. Screening experiments using the SARS-CoV-2 protein may be more efficient due to the SARS-CoV-2 protein's somewhat stronger affinity for ADP-ribose ($K_d = 10 \text{ M}$) than the SARS-CoV protein's affinity for ADP-ribose ($K_d = 24 \text{ M}$). The recombinant protein reported here together with detailed structural information might also be useful to others developing SARS-CoV-2 diagnostics and/or therapeutics.

7. Discovery of Drug-like Ligands for the Mac1 Domain of SARS-CoV-2

7.1. Introduction

Direct acting antivirals (DAAs) are desperately needed to treat COVID-19 patients and stem the devastation caused by the current Severe Acute Respiratory Sndrome Coronavirus 2 (SARS-CoV-2) pandemic. DAAs are typically developed from potent inhibitors of viral enzymes or high-affinity ligands of viral proteins. For example, the first FDA-approved DAA against SARS-CoV-2, remdesivir (Agostini et al., 2018) is a phosphonamidite prodrug that is converted to a nucleoside triphosphate non-obligate chain terminator that inhibits the SARS-CoV-2 RNA dependent RNA polymerase, halting SARS-CoV-2 replication (Shannon et al., 2020). Based on past experiences with other viruses, any effective DAA therapy will likely involve a cocktail of more than one antiviral agent because drug resistance evolves rapidly (de Clercq, 2004, 2013b, 2013a). It is therefore essential that methods are developed to rapidly identify small-molecule drug-like ligands for as many SARS-CoV-2 proteins as possible. Even if such ligands are not developed into drugs, they will be useful as molecular probes to study the biological role of SARS-CoV-2 proteins.

In the previous chapter, the Mac1 of SARS-CoV-2 was characterized, primarily because it is the most different nsp3 domain from its nearest relative, SARS-CoV (Egloff et al., 2004). This domain binds ADP-ribose and has a weak hydrolytic activity (Saikatendu et al., 2005) that removes ADP-ribose that is covalently attached to proteins (Alhammad & Fehr, 2020). Claverie recently proposed that this activity might be linked to the cytokine storm syndrome commonly found in COVID-19 patients (Claverie, 2020). The results of a screening campaign designed to find

drug-like Mac1 ligands are reported here, which might facilitate DDA design that could be useful molecular probes.

7.2. Screening of drugs and drugs-like compounds via DSF

Two high-throughput screens were designed to identify such ligands. Both were used to screen libraries of drugs and drug-like compounds, including the National Institutes of Health (NIH) clinical collection (726 compounds), the National Cancer Institute (NCI) mechanistic set (540 compounds) and Sigma-Aldrich's 1280 compound Library of Pharmacologically Active Compounds (LOPAC).

The first assay used differential scanning fluorimetry (DSF, aka the thermal shift, or ThermoFluor assay) (Huynh & Partch, 2015) methods to examine the melting temperature (T_m) of SPYRO Orange-saturated SARS-CoV-2 Mac1 in the presence of each compound in an assay that can be performed in standard quantitative PCR thermocyclers. In an orthogonal screen, various high-resolution Mac1 domain structures were used with AutoDock VINA19 to identify compounds bound with the highest affinity to all structures, and lastly, ITC was used as a second orthogonal study to examine protein-ligand binding.

The conditions of the DSF assay and the protocol used to purify Mac1 is discussed in detail in Chapter 3. Briefly, DSF assays were performed in 96-well PCR plates using a Quantitative Realtime PCR System with each well containing 19 μ l of Master-mix (5 μ M Mac1) and 1 μ l of either a compound stock (10 mM for screening) or DMSO. The 96-well PCR plate was then sealed, centrifuged, and placed in QR-PCR. The temperature was raised from 20°C to 95°C while measuring the fluorescence in the "TAMRA" channel. Each plate included both negative (DMSO)

and positive (ADPr) controls. The DSF assay was optimized for HTS to yield Z'-Factors between 0.5 and 0.75. In the screening of the NCI, NIH and LOPAC collections, about 1%, 5%, and 1% of compounds tested influenced apparent melting temperature, respectively (Fig. 7.2-1).

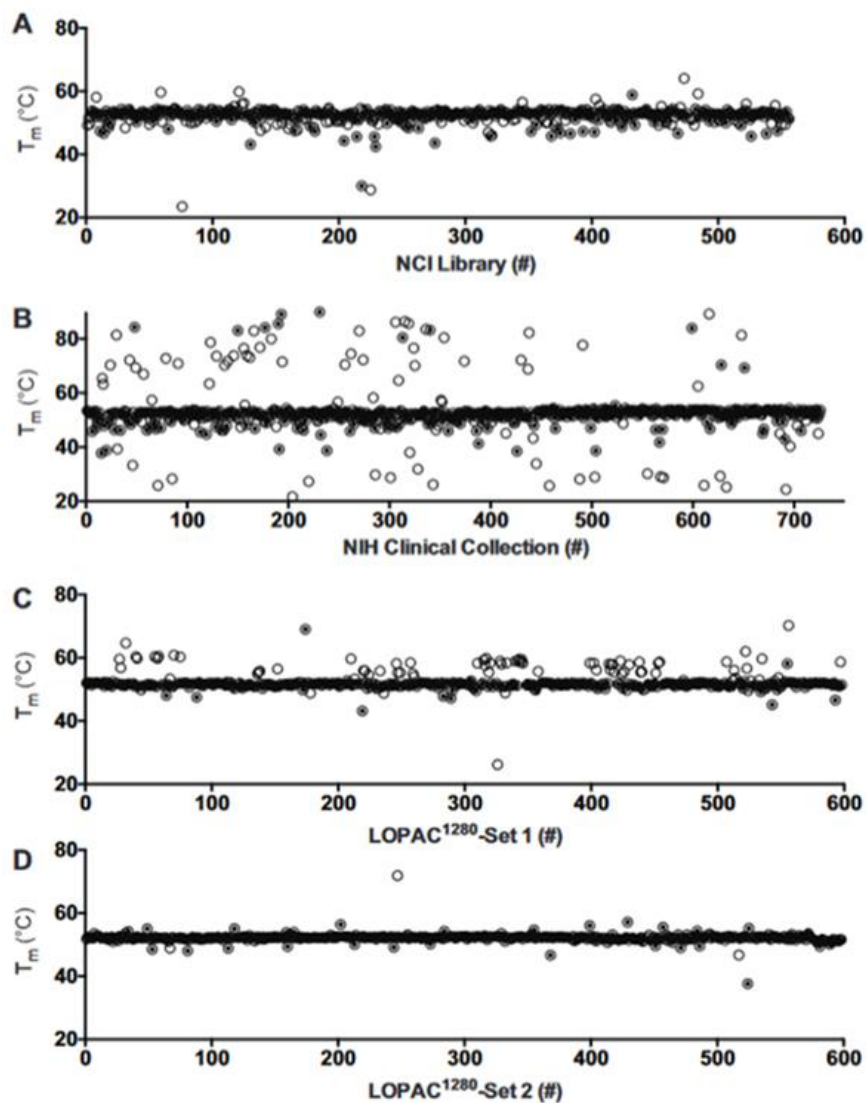


Figure 7.2-1 DSF screens of FDA-approved drugs and drug-like compounds for SARS-CoV-2 Mac1 ligands.

T_m values calculated by fitting melting curves to Eq.3.6-1 (open circles) obtained for Mac1 in the presence of each compound in (A) the NCI library, (B) the NIHcc, and (C, D) the LOPAC1280. Assays yielding a “typical” melting curve, as defined by the TSA-CRAFT algorithm, are noted (filled circles).

The results of the screening show that a substantial proportion of compounds in various libraries impact the measured protein melting temperature, which was especially noticeable with the NIHcc. These abnormal melting curves are associated with substances that interfere with the assay's fluorescence, either directly or indirectly by their enhancement of the SPYRO Orange reporter dye's fluorescence. To exclude the possibility of such compounds in further trials, the TSA-CRAFT software application was used to determine what it refers to as "typical" curves (filled-circles, Figs. 7.2-2). Later, it was discovered that such interfering substances may also be detected by plotting T_m values against the initial fluorescence seen in the melting curve (Fig. 7.2-2).

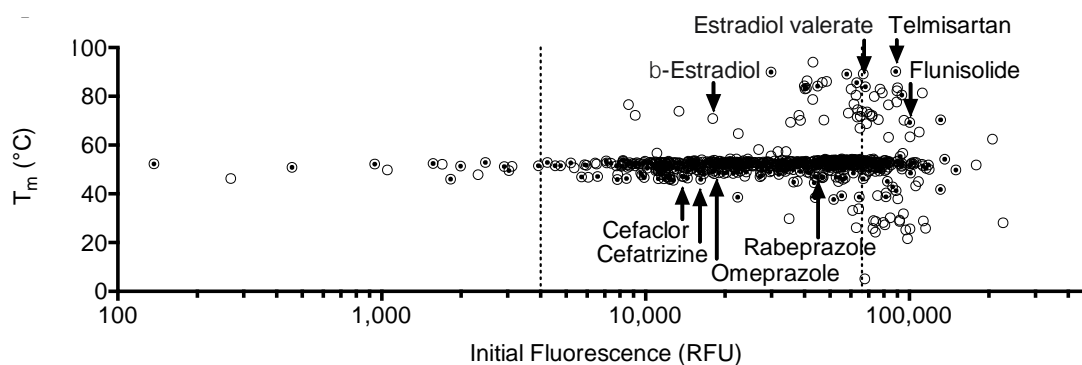


Figure 7.2-2 Identifying interfering compounds.

Plot of T_m values for samples in the NIHcc plotted versus fluorescence observed at the beginning of each melt (*i.e.* @ 20 °C). "Typical" melting curves are highlighted (filled circles). The dotted lines are arbitrary cutoffs drawn at three times more and less than the average fluorescence intensity recorded in all assays.

After interfering compounds were eliminated, only one compound in the NCI set raised the T_m value of SARS-CoV-2 Mac1 protein NSC 635448, which is a copper bromide complex (Copper,bromo[2-[1-(2-pyridinyl)ethylidene][N,N-dimethyl-hydrazinecarbothioamidato-N, N,S]-

Hydrazine carbothioamide). In contrast, a few compounds in the NIH collection significantly increased the melting temperature of Mac1. Two noteworthy hits relevant to current COVID-19 research were the ACE inhibitor telmisartan (Rothlin et al., 2020) and two steroids, estradiol valerate and the anti-inflammatory corticosteroid flunisolide (Fig. 7.2-3) (Breithaupt-Faloppa et al., 2020b; Grandi et al., 2020).

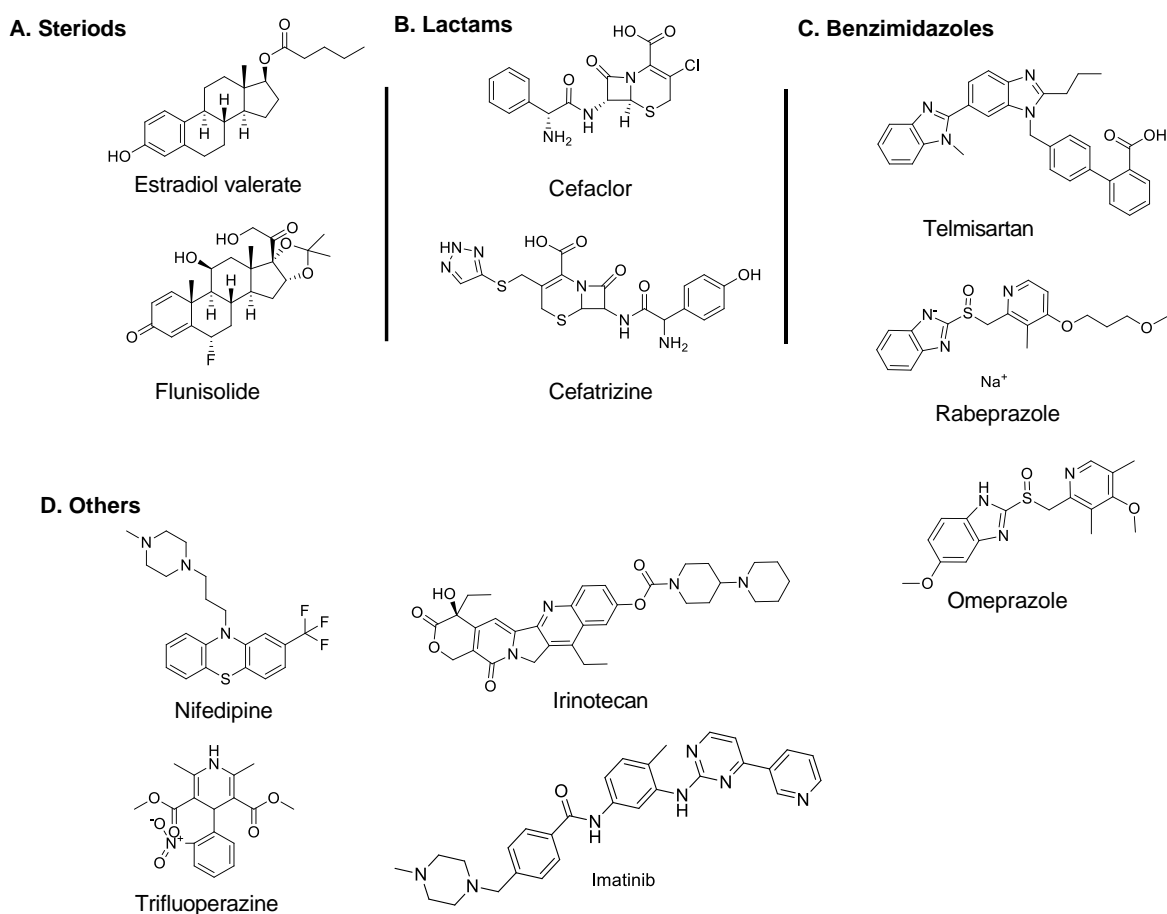


Figure 7.2-3 Selected hit compounds from the NIH clinical collection.

(A) Steroids raise the apparent T_m of SARS-CoV-2 Mac1. (B) Two lactam antibiotics that lowered the apparent T_m of SARS-CoV-2 Mac1. (C) Benzimidazoles altering the SARS-CoV-2 Mac1 T_m (D) four additional hits for which binding could be confirmed using ITC.

Selected compounds identified to induce a change in melting temperature were followed by concentration response assays (red and green circles, Fig. 7.3-1). Although many of the selected compounds that yielded high T_m values were reproduced (green circles, Fig. 7.3-1) upon cherry-picking, their concentration-response curves often looked unusual. For example, the addition of estradiol valerate led to flattened regions with sharp peaks (Fig. 7.2-4A).

One possibility for erratic results is that compounds may be unstable when stored as DMSO solutions in microplates. However, when new aliquots of selected compounds were purchased, abnormal melting curves were again observed with estradiol valerate, and no effects were observed with flunisolide (Fig. 7.2-4A and B, respectively). In contrast, fresh batches of both lactams (Cefaclor and Cefatrizine) and two benzimidazoles (rabeprazole and omeprazole) yielded the same effects seen with the screening library (Fig. 7.2-4C, D, F, G, respectively). However, fresh telmisartan yielded a different effect, lowering the apparent T_m as was seen with related compounds (Fig. 7.2-4E), suggesting that a possible degradation product led to the T_m increase observed using the library sample.

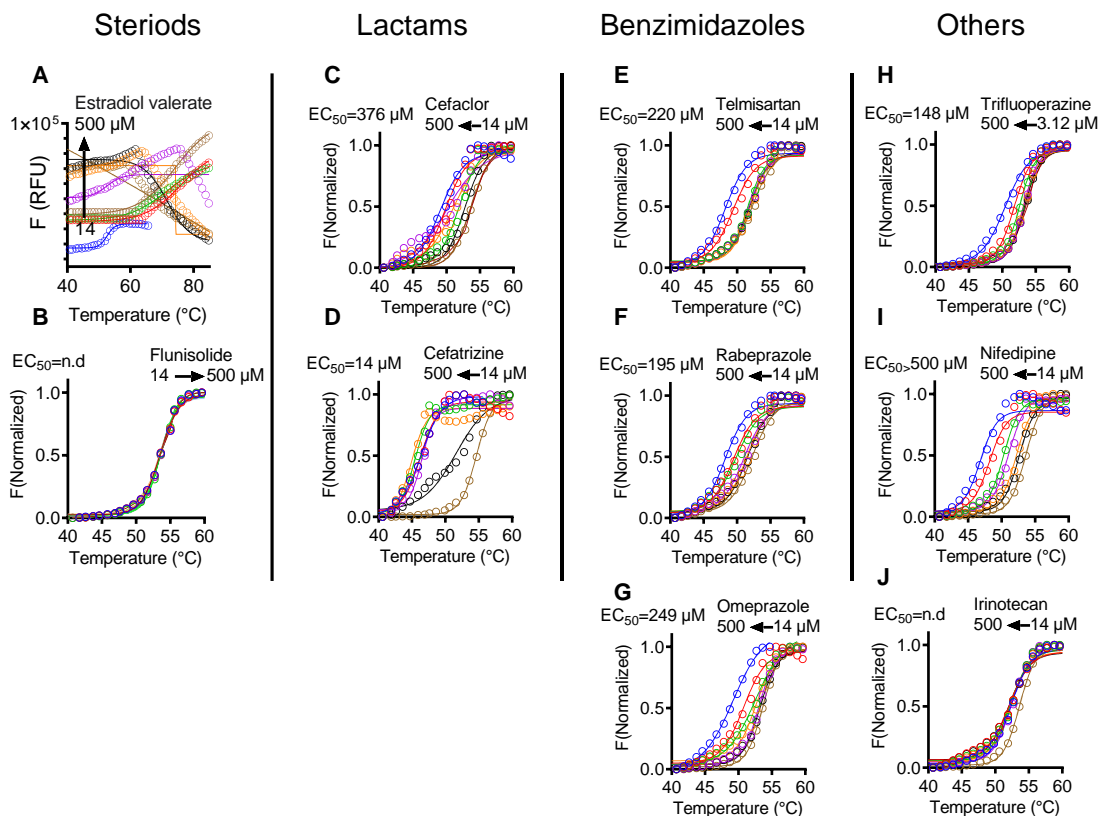


Figure 7.2-4 Concentration response assays.

Concentration response analysis of each compound in DSF assays are shown, along with EC₅₀ values. The direction of the arrow represents increase (to the right) or decrease (to the left) in T_m upon increase in ligand concentration.

Although ligands that reduce a protein's T_m are often assumed to bind and stabilize unfolded structures, (Layton & Hellenga, 2010), nevertheless some of these hits were examined in more detail. Interestingly, many of the hits that lowered T_m values were members of similar chemical classes. For example, two were lactam antibiotics (Fig. 7.2-3B), and two others were benzimidazoles chemically resembling telmisartan (Fig. 7.2-3C). Many reproduced well upon cherry-picking and depressed the T_m values when interacting with SARS-CoV2 Mac1 in a concentration-dependent manner (Fig. 7.2-4).

7.3.Orthogonal “Virtual” Screen.

The orthogonal "virtual" screen used the SARS-CoV-2 Mac1 crystal structures (PDB files 6WEY, 6W02, and 6VXS) as targets in the AutoDock Vina program. PDB file 6W02 is a co-structure with ADP-ribose, but the other two were done in the absence of ADP-ribose. Each was searched free of ligands. Binding sites were not restricted but, for most compounds, minimum binding energy (best fit) values were obtained for structures in which the compound docked near the ADPr binding site. Plots of AutoDock Vina scores vs. T_m could be used to identify compounds for follow-up assays (Fig. 7.3-1). First, there was a clear correlation between the number of hits and AutoDock Vina scores, with more hits clustering at lower energies. Second, when cherry-pick assays were performed on hits, those with lower energy scores (12/15) were more likely to be reproducible than compounds with higher-energy scores (3/9) (Fig. 7.3-1).

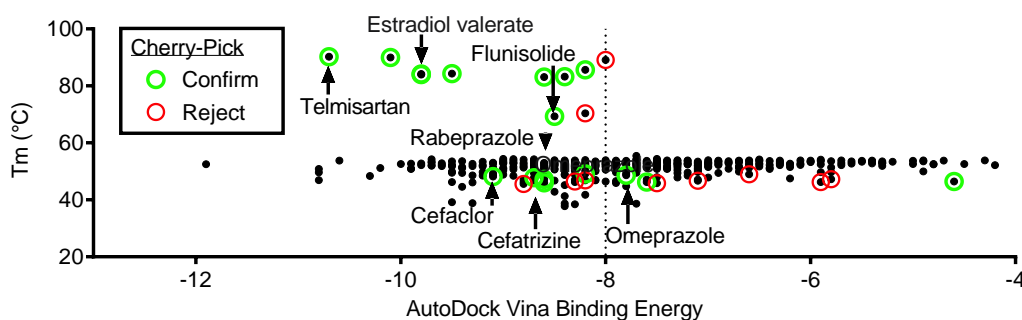


Figure 7.3-1 **Comparison of DSF and AutoDock screening results.**

Plot of T_m values derived using TSA-CRAFT for samples in the NIHcc plotted versus AutoDock Vina binding energies obtained for each compound after docking with PDB file 6WEY. “Typical” melting curves are filled circles. Compounds selected for follow-up (cherry pick) analysis are highlighted. The dotted line is arbitrary cutoffs differentiating high or low energy scores.

The two methods outlined above could therefore be used in future screening campaigns to identify compounds that more likely bind Mac1 and could be differentiated from those that likely do not. As evidence, all nucleotides that were hit yielded results similar to those seen with ADPr.

7.4. ITC with selected compounds

After confirming the results with fresh compounds, ITC was used as a second orthogonal assay to examine protein-ligand binding. Unfortunately, most of these compounds had little to no solubility in the buffer conditions used for ITC, therefore results labeled with *asterisk are compounds with low solubility and compounds with no solubility were not included. The conditions and protocol followed to do the experiment are covered in-depth in Chapter 3. Briefly, samples of ligand and protein were diluted in 10 mM MOPS and 150 mM NaCl at pH 7 and degassed for 30 minutes at 400 mmHg. At a temperature of 20 °C, 2.0 μ L aliquots of 500 μ M compounds were injected into 50 μ M of protein in a 175 μ L starting volume stirred at 250 rpm. The ITC was performed and data was collected on select compounds, including *nifedipine (Fig. 7.4.1), trifluoperazine (Fig. 7.4.2), irinotecan (Fig. 7.4.3), *imatinib methane sulfonate (Fig. 7.4.4), two steroids: *estradiol (Fig. 7.4.5) and *flunisolide (Fig. 7.4.6), and three benzimidazoles: *telmisartan (Fig. 7.4.7), rabeprazole (Fig. 7.4.8) and omeprazole (Fig. 7.4.9).

7.4.1. *Nifedipine binding to Mac1

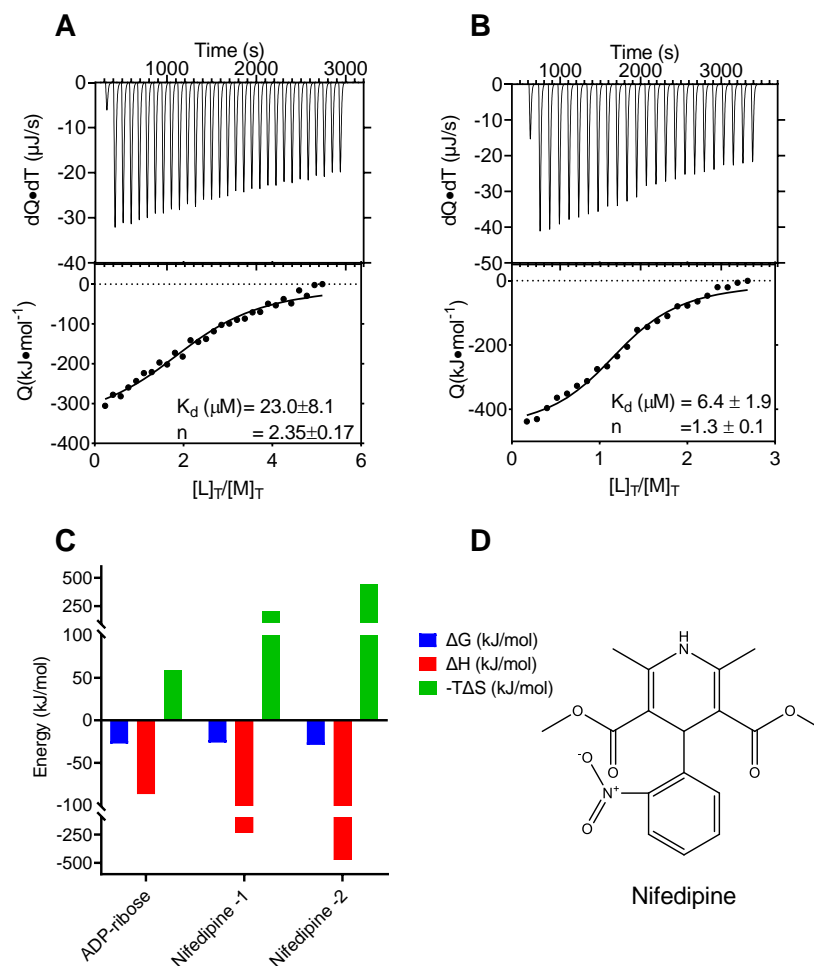


Figure 7.4-1 **Mac1 titration with nifedipine.**

Panel A and B top: 50 μM Mac1 titrated with 500 μM nifedipine dissolved in 10mM MOPS 150mM NaCl pH7. The titration thermogram is represented as the amount of heat released per unit of time after each injection. The relationship between the quantity of heat generated by each injection and the ratio of total ligand to total protein concentration is shown underneath it. In Panel A and B, bottom, the circles plot experimental data, while the line represents the model's best fit to n identical and independent binding sites. Fitting the data to an independent model gives us the dissociation constant of 23 ± 8.1 μM and 6.4 ± 1.9 μM , the stoichiometric constant of 2.4 ± 0.2 and 1.3 ± 0.1 , the enthalpy of -232 ± 25 (kJ/mol) and -471 ± 32 (kJ/mol), $-\Delta S$ of 205 (kJ/mol) and 443 (kJ/mol) Gibbs free energy of -26.2 (kJ/mol) and -28.9 (kJ/mol) for panel A and B respectively. Panel C is the histogram of energetics from Panel A and B in comparison to ADP-ribose. Panel D is the chemical structure of nifedipine.

7.4.2. Trifluoperazine binding to Mac1

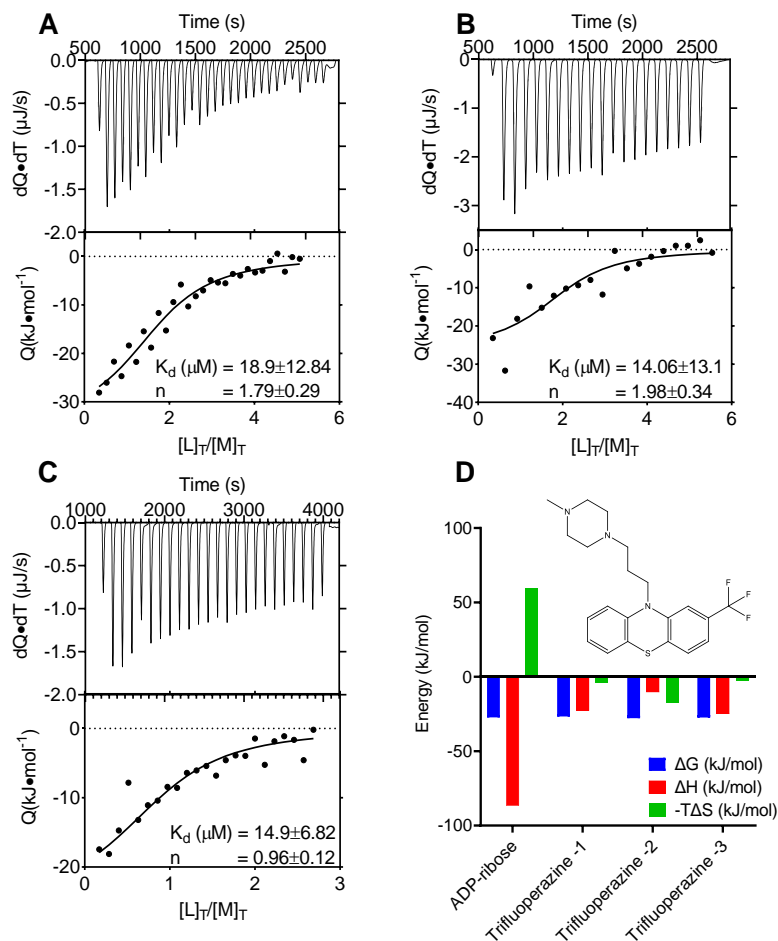


Figure 7.4-2 Mac1 titration with trifluoperazine.

Panel A and B top: 50 μM Mac1 titrated with 500 μM trifluoperazine dissolved in 10mM MOPS 150mM NaCl pH7. The thermogram represented as the amount of heat released per unit of time after each injection. The relationship between the quantity of heat generated by each injection and the ratio of total ligand to total protein concentration is shown underneath it. Panel A and B bottom: The circles plot experimental data, while the line represents model's best fit to n identical and independent binding sites. Fitting the data to independent model gives us the dissociation constant of $18.9 \pm 12.8 \mu M$, $14.1 \pm 13.1 \mu M$ and $14.9 \pm 6.8 \mu M$, stoichiometric constant of 1.8 ± 0.3 , 1.9 ± 0.4 and 1.0 ± 0.1 , enthalpy of -22.6 ± 6.3 (kJ/mol), -10.3 ± 2.7 (kJ/mol) and -24.9 ± 5.2 (kJ/mol), $-T\Delta S$ of -4.1 (kJ/mol), -17.5 (kJ/mol) and -2.5 (kJ/mol) Gibbs free energy of -26.7 (kJ/mol), -27.8 (kJ/mol) and -27.4 (kJ/mol) for panel A, B and C respectively. Panel D is the histogram of energetics from Panel A, B and C in comparison to ADP-ribose with the chemical structure of trifluoperazine.

7.4.3. Irinotecan binding to Mac1

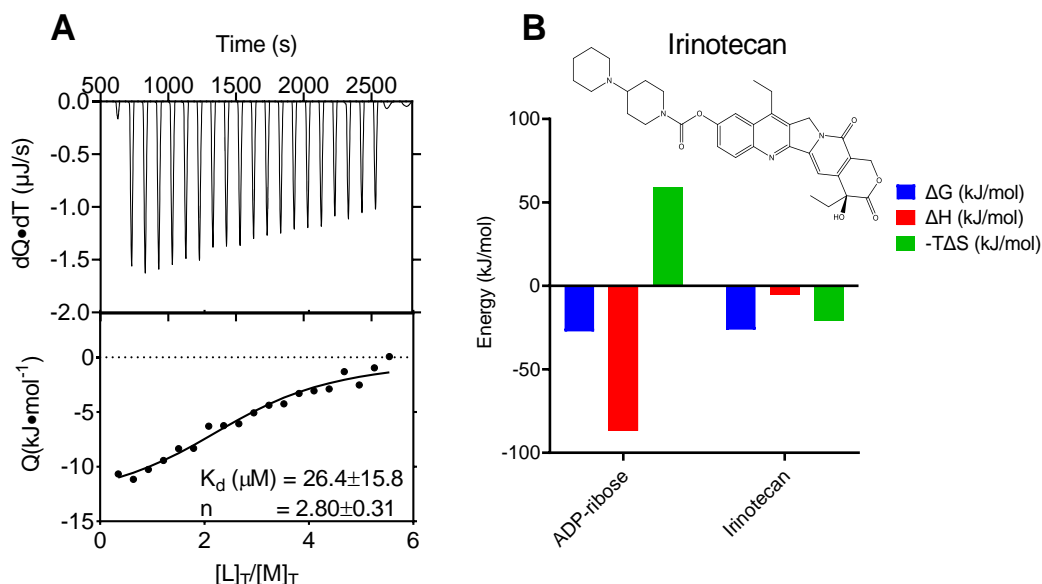


Figure 7.4-3 Mac1 titration with irinotecan.

Panel A: 50 μM Mac1 titrated with 500 μM irinotecan dissolved in 10mM MOPS 150mM NaCl pH7. The titration thermogram represented as the amount of heat released per unit of time after each injection. The relationship between the quantity of heat generated by each injection and the ratio of total ligand to total protein concentration is shown underneath it. Panel A bottom: The circles plot experimental data, while the line represents model's best fit to n identical and independent binding sites. Fitting the data to independent model gives us the dissociation constant of $26 \pm 16 \mu M$, Stoichiometric constant of 2.8 ± 0.3 , Enthalpy of -5.3 ± 0.9 (kJ/mol), $-T\Delta S$ of -20.9 (kJ/mol) Gibbs free energy of -26.2 (kJ/mol). Panel B is the histogram of energetics from Panel A in comparison to ADP-ribose with the chemical structure of irinotecan.

7.4.4. Imatinib methane sulfonate binding to Mac1

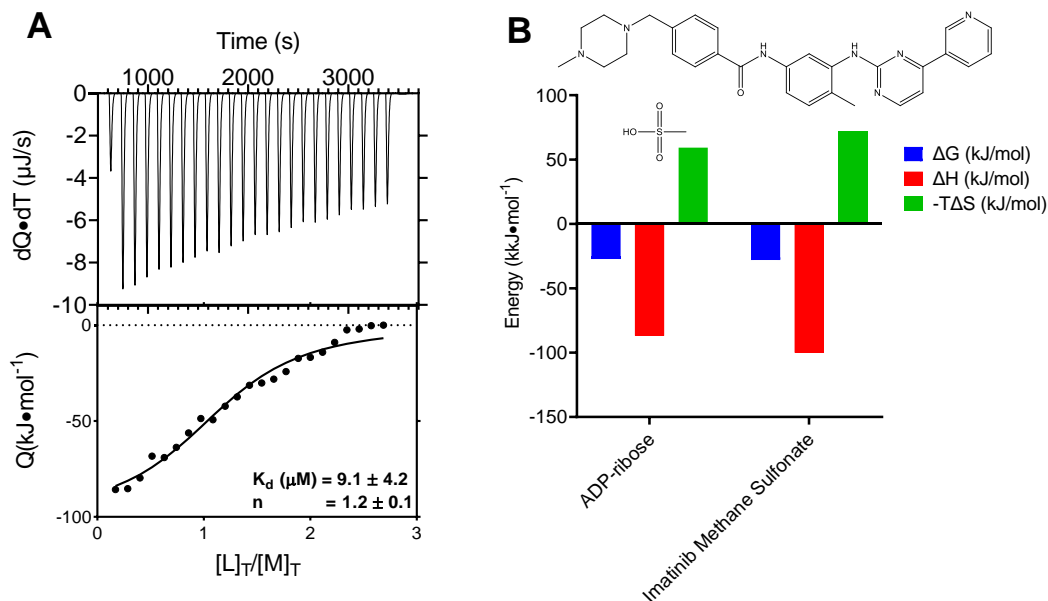


Figure 7.4-4 **Mac1 titration with imatinib methane sulfonate.**

Panel A top: 50 μM Mac1 titrated with 500 μM imatinib methane sulfonate dissolved in 10mM MOPS 150mM NaCl pH7. The titration thermogram represented as the amount of heat released per unit of time after each injection. The relationship between the quantity of heat generated by each injection and the ratio of total ligand to total protein concentration is shown underneath it. Panel A bottom: The circles plot experimental data, while the line represents model's best fit to n identical and independent binding sites. Fitting the data to independent model gives us the dissociation constant of 9.1 ± 4.2 μM , Stoichiometric constant of 1.2 ± 0.1 , Enthalpy of -99.8 ± 12.47 (kJ/mol), $-T\Delta S$ of 71.81 (kJ/mol) and Gibbs free energy of -28.0 (kJ/mol). Panel B is the histogram of energetics from Panel A in comparison to ADP-ribose with the chemical structure of imatinib methane sulfonate.

7.4.5. *Estradiol binding to Mac1

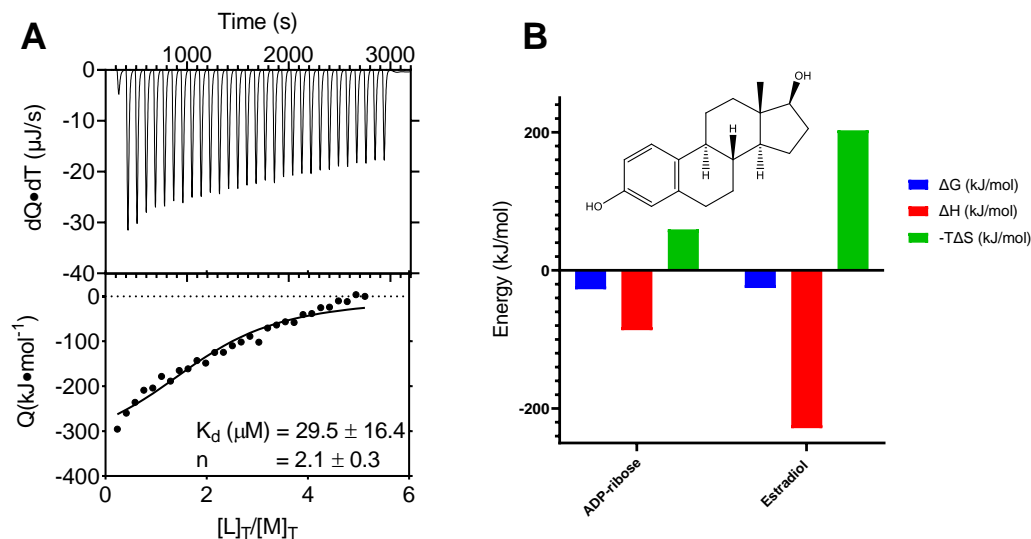


Figure 7.4-5 **Mac1 titration with estradiol.**

Panel A top: 50 μM Mac1 titrated with 500 μM estradiol dissolved in 10mM MOPS 150mM NaCl pH7. The titration thermogram represented as the amount of heat released per unit of time after each injection. The relationship between the quantity of heat generated by each injection and the ratio of total ligand to total protein concentration is shown underneath it. Panel B bottom: The circles plot experimental data, while the line represents model's best fit to n identical and independent binding sites. Fitting the data to independent model gives us the dissociation constant of $29.5 \pm 16.4 \mu M$, Stoichiometric constant of 2.1 ± 0.3 , Enthalpy of -228.4 ± 55.5 (kJ/mol), $-T\Delta S$ of 203 (kJ/mol) and Gibbs free energy of -25.6 (kJ/mol). Panel B is the histogram of energetics from Panel A in comparison to ADP-ribose with the chemical structure of estradiol.

7.4.6. *Flunisolide binding to Mac1

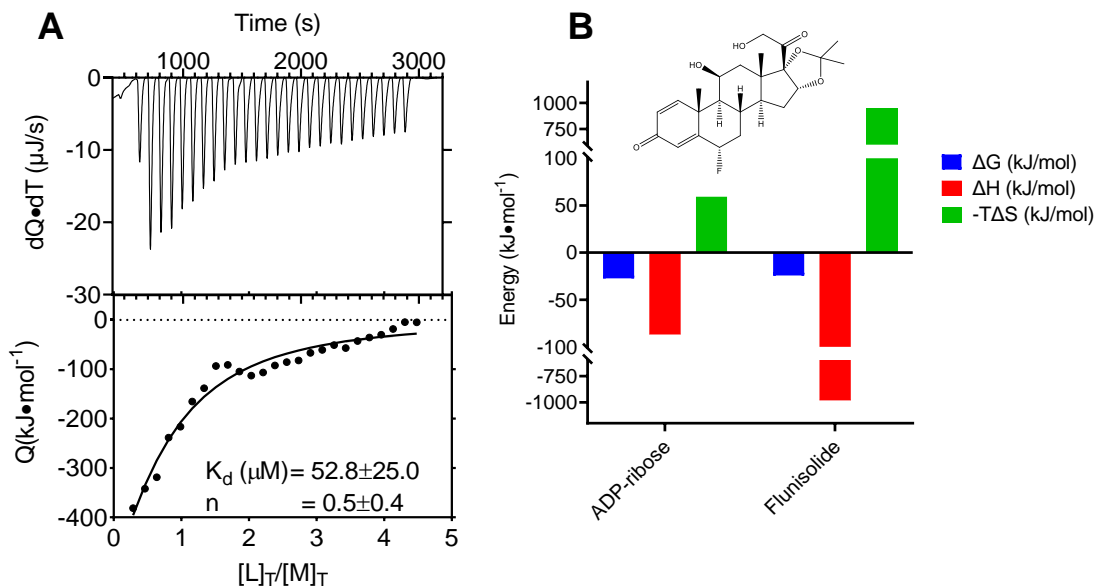


Figure 7.4-6 Mac1 titration with flunisolide.

Panel A top: 50 μM Mac1 titrated with 500 μM flunisolide dissolved in 10mM MOPS 150mM NaCl pH7. The titration thermogram represented as the amount of heat released per unit of time after each injection. The relationship between the quantity of heat generated by each injection and the ratio of total ligand to total protein concentration is shown underneath it. Panel B bottom: The circles plot experimental data, while the line represents model's best fit to n identical and independent binding sites. Fitting the data to independent model gives us the dissociation constant of 53 ± 25 μM , Stoichiometric constant of 0.5 ± 0.4 , Enthalpy of -980 ± 1972 (kJ/mol), $-T\Delta S$ of 950 (kJ/mol) and Gibbs free energy of -24.2 (kJ/mol). Panel B is the histogram of energetics from Panel A in comparison to ADP-ribose with the chemical structure of flunisolide.

7.4.7. *Telmisartan binding to Mac1

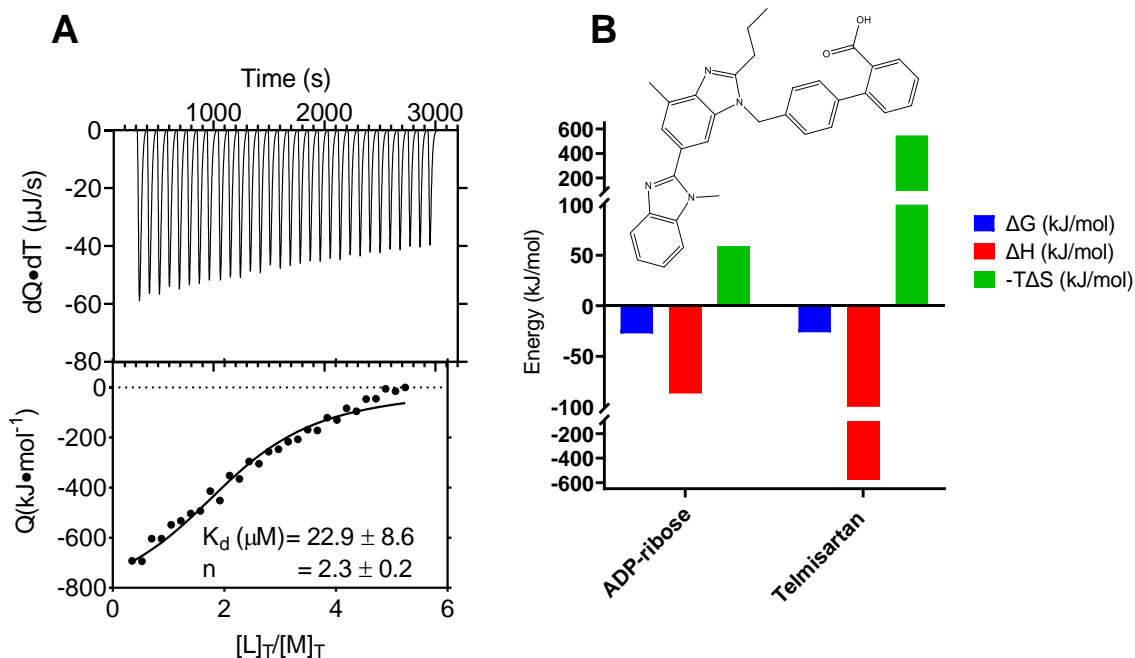


Figure 7.4-7 Mac1 titration with telmisartan.

Panel A top: 50 μM Mac1 titrated with 500 μM telmisartan dissolved in 10mM MOPS 150mM NaCl pH7. The titration thermogram represented as the amount of heat released per unit of time after each injection. The relationship between the quantity of heat generated by each injection and the ratio of total ligand to total protein concentration is shown underneath it. Panel B bottom: The circles plot experimental data, while the line represents model's best fit to n identical and independent binding sites. Fitting the data to independent model gives us the dissociation constant of $22.9 \pm 8.6 \mu M$, Stoichiometric constant of 2.3 ± 0.2 , enthalpy of -574.0 ± 69.3 (kJ/mol), $-T\Delta S$ of 547.8 (kJ/mol) and Gibbs free energy of -26.2 (kJ/mol). Panel B is the histogram of energetics from Panel A in comparison to ADP-ribose with the chemical structure of telmisartan.

7.4.8. Rabeprazole binding to Mac1

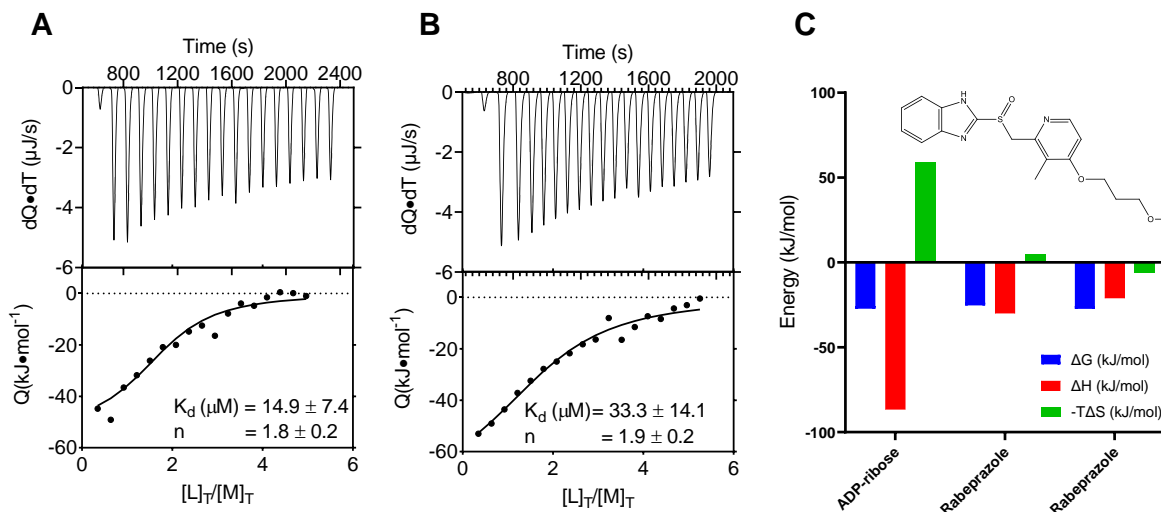


Figure 7.4-8 **Mac1 titration with rabeprazole.**

Panel A, B and C top: 50 μM Mac1 titrated with 500 μM rabeprazole dissolved in 10mM MOPS 150mM NaCl pH7. The titration thermogram represented as the amount of heat released per unit of time after each injection. The relationship between the quantity of heat generated by each injection and the ratio of total ligand to total protein concentration is shown underneath it. Panel A, B and C bottom: The circles plot experimental data, while the line represents model's best fit to n identical and independent binding sites. Fitting the data to independent model gives us the dissociation constant of 16.2 ± 6.8 μM, 33.3 ± 14.1 μM and 14.9 ± 7.4 μM, Stoichiometric constant of 0.9 ± 0.1 , 1.8 ± 0.2 and 1.8 ± 0.2 , Enthalpy of -48.9 ± 9.5 (kJ/mol), -30.05 ± 6.2 (kJ/mol) and -21.1 ± 3.4 (kJ/mol), $-T\Delta S$ of 22.33 (kJ/mol), 4.7 (kJ/mol) and -6.3 (kJ/mol) Gibbs free energy of -26.6 (kJ/mol), -25.3 (kJ/mol) and -27.4 (kJ/mol) for panel A, B and C respectively. Panel D is the histogram of energetics from Panel A, B and C in comparison to ADP-ribose with the chemical structure of rabeprazole.

7.4.9. Omeprazole binding to Mac1

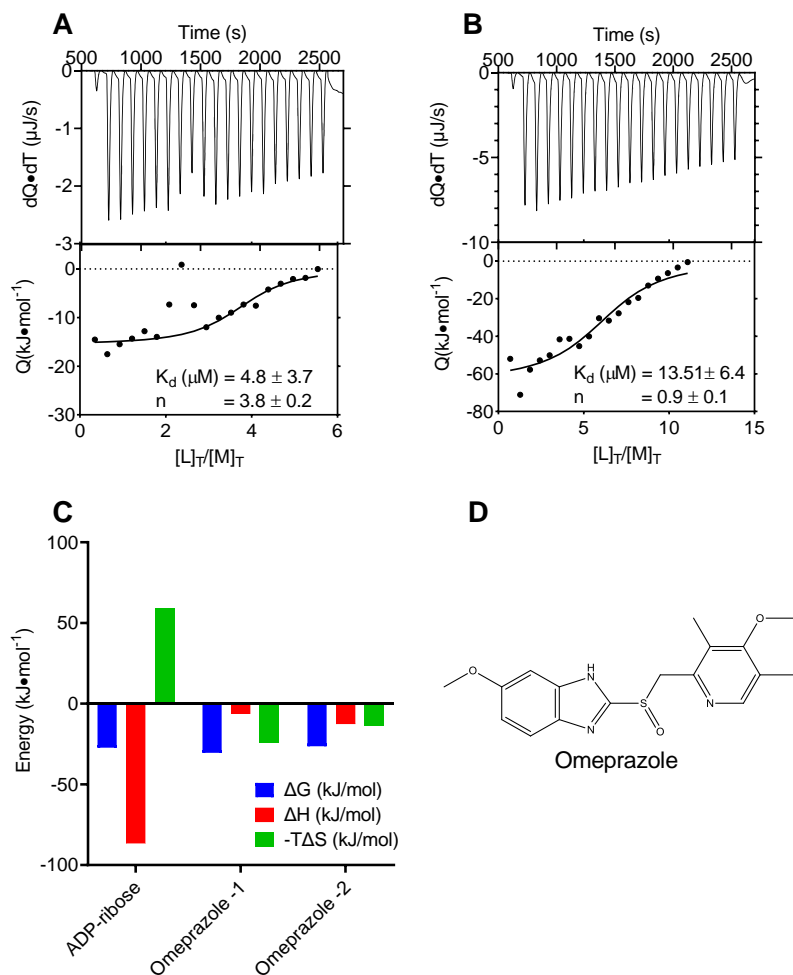


Figure 7.4-9 Mac1 titration with omeprazole.

Panel A and B top: 50 μM Mac1 titrated with 500 μM omeprazole dissolved in 10mM MOPS 150mM NaCl pH7. The titration thermogram represented as the amount of heat released per unit of time after each injection. The relationship between the quantity of heat generated by each injection and the ratio of total ligand to total protein concentration is shown underneath it. Panel A and B bottom: The circles plot experimental data, while the line represents model's best fit to n identical and independent binding sites. Fitting the data to independent model gives us the dissociation constant of $4.8 \pm 3.7 \mu\text{M}$ and $14 \pm 6.4 \mu\text{M}$, Stoichiometric constant of 3.8 ± 0.2 and 0.9 ± 0.1 , Enthalpy of -6.2 ± 0.5 (kJ/mol) and -24 ± 5.0 (kJ/mol), $-T\Delta S$ of -24.2 (kJ/mol) and 3.7 (kJ/mol) Gibbs free energy of -30.4 (kJ/mol) and -27.6 (kJ/mol) for panel A and B respectively. Panel C is the histogram of energetics from Panel A and B in comparison to ADP-ribose. Panel D is the chemical structure of omeprazole.

7.5. Discussion

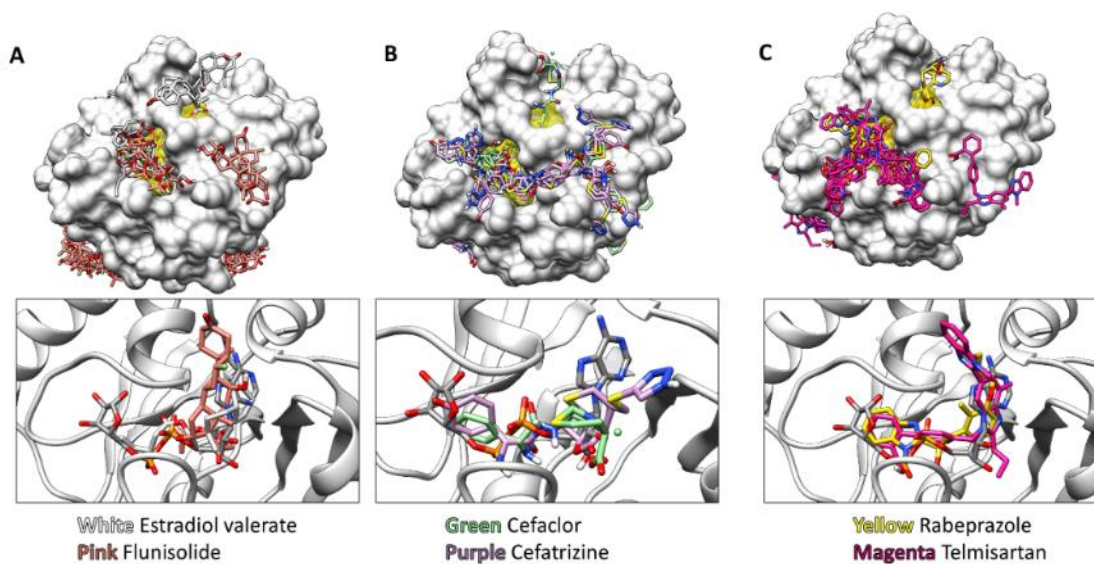
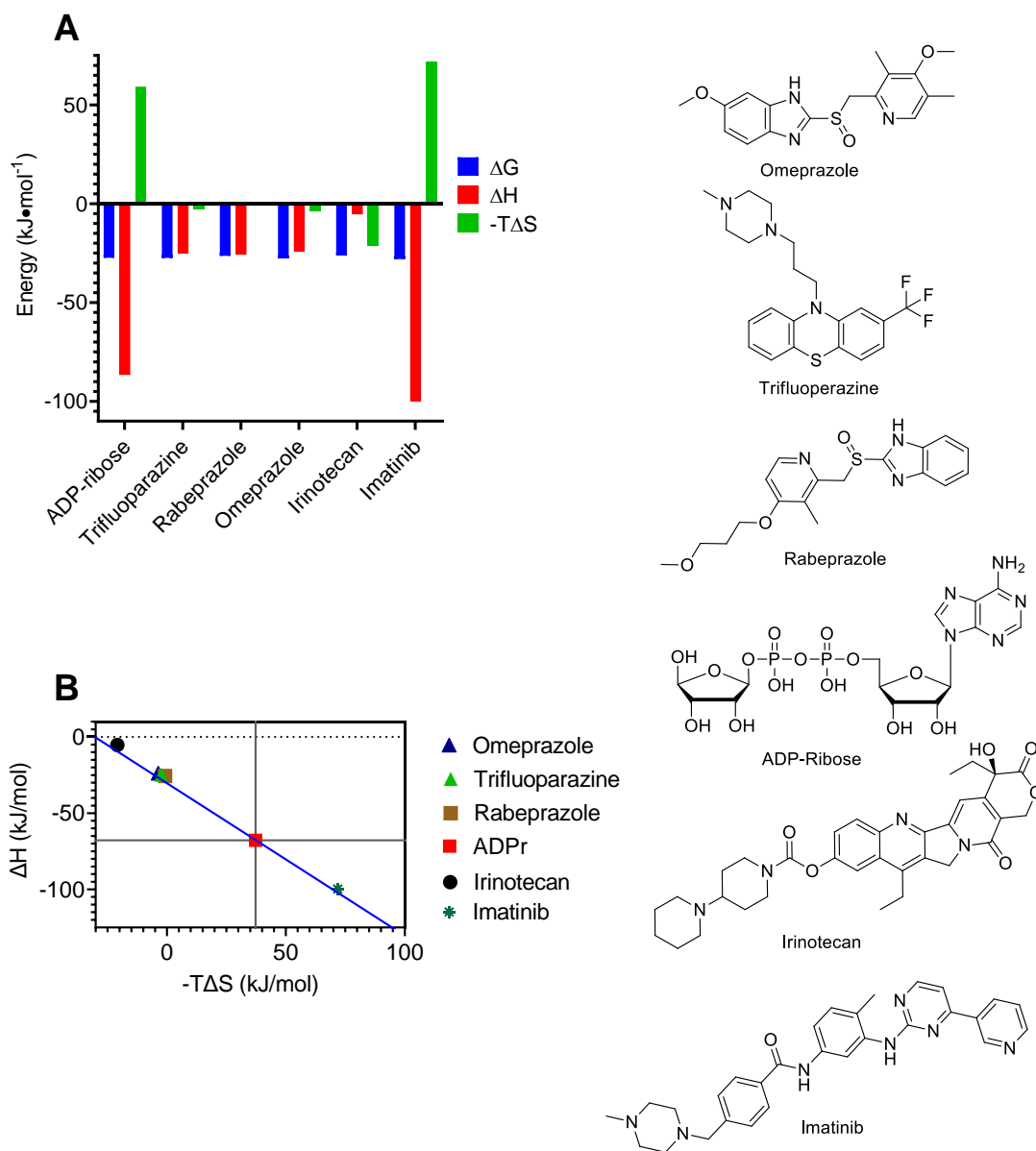


Figure 7.5-1 **Representative structures obtained using AutoDock Vina Virtual Screening.**

Top panels show the top 20 binding modes for selected compounds, and the bottom panels show the top binding mode for each compound compared to the ADP-ribose bound in PDB file 6W02. The ADP-ribose binding cleft is highlighted in yellow on the protein surfaces shown in the top panels. Results are shown for (A) two steroids, (B) two lactam antibiotics, and (C) two benzimidazoles in the NIH clinical collection. Structures are displayed using *UCSF-Chimera* (Pettersen et al., 2004b).

Nifedipine, estradiol, flunisolide, and telmisartan were among the compounds with a poor solubility. Regardless, the binding of compounds with poor solubility was estimated, and the resulting enthalpy and entropy energies were quite high. The elevated enthalpy values are assumed to be due to non-specific hydrophobic interaction between the protein and ligand. It is possible that these hydrophobic compounds are binding in the ADP-ribose binding site however it is more likely they are binding into hydrophobic area on the protein surface. The binding of hydrophobic molecule to a hydrophobic area on the protein results in hydrophobic interaction or Van der Waals

forces (van Oss et al., 1980) resulting in favorable enthalpy. Additionally, binding of hydrophobic compounds such as lactams and steroids was also virtually obtained showing different binding modes (Fig. 7.5-2).



Trifluoperazine, rabeprazole, omeprazole, irinotecan, and imatinib were among the more soluble drugs studied in ITC. These molecules, with the exception of imatinib, were soluble and bound with comparable affinities to the other nucleotides examined in Chapter 6. Imatinib is a chemotherapeutic drug that is used to treat cancer. It is a tyrosine kinase inhibitor that inhibits the enzyme by occupying a position near to the ATP binding site, resulting in a closed conformation and inhibition of ATP binding to the active site, hence lowering the enzyme's activity and preventing the cell from proliferating (B. Peng et al., 2005).

According to imatinib's thermodynamic data, it binds to the Mac1 domain with a greater enthalpy and entropy than ADP-ribose (Fig. 7.5-1). The favorable enthalpy and unfavorable entropy values do reflect ligand-protein binding through favorable hydrogen bonding or van der Waals forces and an unfavorable entropic interaction owing to degree of freedom constraint caused by ligand binding. However, activity tests will be required to determine if Imatinib affects Mac1 binding to ADP-ribose. Additionally, imatinib is an intriguing hit since it has been indicated as a potential therapeutic agent in COVID studies (Emadi et al., 2020.; Gasmi et al., 2020; Morales-Ortega et al., 2020).

7.6.Summary

In previous chapter, binding between ADP-ribose and Mac1 was identified using ITC and DSF along with binding of other nucleotides. This chapter focused on the results of a screening campaign designed to find drug-like Mac1 ligands, which might facilitate DDA design that could be useful molecular probes. Two high-throughput screens were designed to identify such ligands. Both were used to screen libraries of drugs and drug-like compounds, including the National

Institutes of Health (NIH) clinical collection (726 compounds), the National Cancer Institute (NCI) mechanistic set (540 compounds) and Sigma-Aldrich's 1280 compound Library of Pharmacologically Active Compounds (LOPAC).

All the compound libraries were screened using DSF, which was optimized previously, yet further optimization was needed to exclude compounds that interfere with the SYPRO orange dye. Two methods were used to identify interfering compounds: TSA-Craft and a plot of T_m values against the initial fluorescence seen in the melting curve. After excluding interfering compounds, a dose response assay was done on the selected few remaining compounds. After confirming that the compound induces a change in the melting temperature of Mac1, new compounds were bought and dose response was repeated to confirm the results. Unfortunately, not all compounds were able to reproduce the results, probably due to degradation in DMSO solution.

The most interesting hits in these screens were the steroids and telmisartan (Fig. 7.2-3). There seems to be a clear sex difference in COVID-19 disease severity, and an interaction between a sex hormone and a SARS-CoV-2 protein might help explain this difference (Grandi et al., 2020), (Breithaupt-Faloppa et al., 2020a). Similarly intriguing is the possibility that the ACE inhibitor telmisartan, which shares the receptor needed for SARS-CoV-2 to infect cells, might be an antiviral (Rothlin et al., 2020). Unfortunately, the poor aqueous solubility of the steroids and telmisartan confounded the ability to examine their interaction with SARS-CoV-2 Mac1. However, other benzimidazoles were also hit (Fig. 7.2-3C), and they could be verified as ligands using ITC. Rabeprazole and omeprazole elicit their antacid effects by targeting proton pumps, but there is

no evidence that proton pump inhibitors affect the severity of COVID-19 (Patel & Sengupta, 2020). Similarly, no peer-review studies have been published reporting antiviral effects for the other three hits confirmed using ITC: nifedipine, trifluoperazine, and irinotecan. Caution should also be exercised because each verified compound lowered the apparent T_m of Mac1, indicating that they might bind the protein's unfolded state. It's worth noting, however, that similar destabilizing compounds were found to inhibit macrodomains in other assays not based on thermal shifts, (Wazir et al., 2021) suggesting that the destabilizing compounds might bind a folded protein that assumes a different conformation. X-ray crystallography is presently being used to examine that possibility.

In an orthogonal virtual study, AutoDock virtual binding energy was estimated for all the compounds in the library. Interestingly, most of the cherry-picked compounds that induce change in the melting temperature had lower AutoDock energy. Close examination of molecular models generated by AutoDock Vina reveals that the steroids (Fig. 7.5-2A), β -lactams (Fig. 7.5-2B) and benzimidazoles (Fig. 7.5-2C) can each occupy the ADP-ribose-binding cleft on SARS-CoV-2 Mac1. The larger compounds in each class make more contacts with amino acids in the cleft, explaining their higher binding energies.

After confirming the results with fresh compounds, ITC was used as a second orthogonal assay to examine protein-ligand binding. Unfortunately, most of the compounds that were purchased had little to no solubility in the buffer conditions used for ITC. Regardless, methods such as dissolving compounds in DMSO first and sonication was used to force solvation of the compounds along with running ITC immediately after solvation. Therefore, the energetics of

these compounds should be interpreted with caution. However, there were other compounds that were able to dissolve readily. Imatinib is one of the compounds that is soluble in conditions used with ITC and has favorable enthalpy and unfavorable entropy values comparable to ADP-Ribose.

The next step in this project will be to examine the effects of hit compounds on cells harboring SAR-CoV-2 or surrogate reporter viruses. Some of the compounds above might inhibit SARS-CoV-2 replication because Shimizu *et al.* showed that small molecules, found in virtual screens targeting the homologous nsp3 macrodomain from Chikungunya virus, inhibit replication of Chikungunya replicons (Shimizu et al., 2020). However, further optimization would most likely be necessary for these probes to be useful in cellular studies. Fortunately, all hits reported here are FDA approved drugs with hundreds of analogs available to facilitate such work. Alternatively, this optimized DSF assay could be used to screen larger, more diverse libraries for more attractive probe candidates.

8. SARS-CoV-2 Helicase

8.1.Introduction

The raging COVID-19 pandemic caused by SARS-CoV-2 has killed millions of patients worldwide. Currently, there are no effective drugs available for treating coronavirus infections. Antivirals designed to combat the SARS-CoV-2 will most likely target viral non-structural proteins required for viral replication. SARS-CoV-2 viral genome is a capped, 30,000 nucleotides long positive sense RNA with numerous open reading frames encoding total of 27 proteins. Upon infection viral genome is directly translated encoding for nonstructural proteins (nsps), which are essential for viral survival, starting with ORF1a. ORF1a encodes a single large polyprotein that is processed into 11 nonstructural proteins. If the ribosome slips, 5 more non-structural proteins are translated by ORF1ab (Irigoyen et al., 2016 Plant et al., 2013). In total, ORF1ab encodes for 16 non-structural proteins that are involved in forming a replication and transcription complex essential for viral replication.

The SARS-CoV-2 replication and transcription complex is essential for viral replication and consists of several non-structural proteins, including RNA-dependent RNA polymerase (Nsp12), a helicase (nsp13), and two auxiliary factors: nsp7 and nsp8. Among the 16 nonstructural proteins, Nsp13 is one of the proteins with the greatest degree of sequence conservation. The high degree of sequence conservation for nsp13 within the Coronavirus family suggests its critical role in the viral lifecycle. As a result, nsp13 is a prospective target for the development of antiviral drugs (Lehmann et al., 2015).

The Nsp13 helicase is a member of the SF1 helicase superfamily that unwinds RNA and DNA duplexes unidirectionally in the 5' to 3' direction (Seybert et al., 2000; Weber & McCullagh, 2021), with dNTP hydrolysis facilitating the translocation. Nsp13 is composed of a zinc-binding domain (ZBD) at the N-terminus, two RecA1 (1A) and RecA2 (2A) helicase core domains, and an inserted domain 1B connected to the ZBD by a "stalk" region (Hao et al., 2017; Jia et al., 2019). Additionally, nsp13 is thought to be involved in mRNA capping and to play several functions throughout the coronavirus life cycle (Ivanov, Thiel, et al., 2004; Ivanov & Ziebuhr, 2004). This study examines the activity of the SARS-CoV-2 helicase (nsp13) and if helicase is stimulated in the presence of oligonucleotides.

8.2.Purification of His-Nsp13 and Nsp13-His.

Initially, a codon optimized open reading frame was synthesized by GenScript (Piscataway, NJ) that encodes for nsp13 with a His-tag on the C-terminal. This open reading frame was cloned into pET28 to give the plasmid pET28-COVID-Nsp13. The nucleotide sequence of the nsp13, the protein sequence of the nsp13, and the more detailed protocol for protein purification are discussed in the materials and methods section (Chapter 4). Briefly, the pET28-COVID-Nsp13 plasmid was used to transform BL21 (DE3) cells. Transformed BL21(DE3) bacteria were used to start a 3 mL starter culture, which was then transferred to a 1-liter fresh medium containing 1 mL of 50 mg/mL kanamycin. After the OD₆₀₀ reached 0.7, the cells were induced using 1 mM IPTG and harvested the next morning.

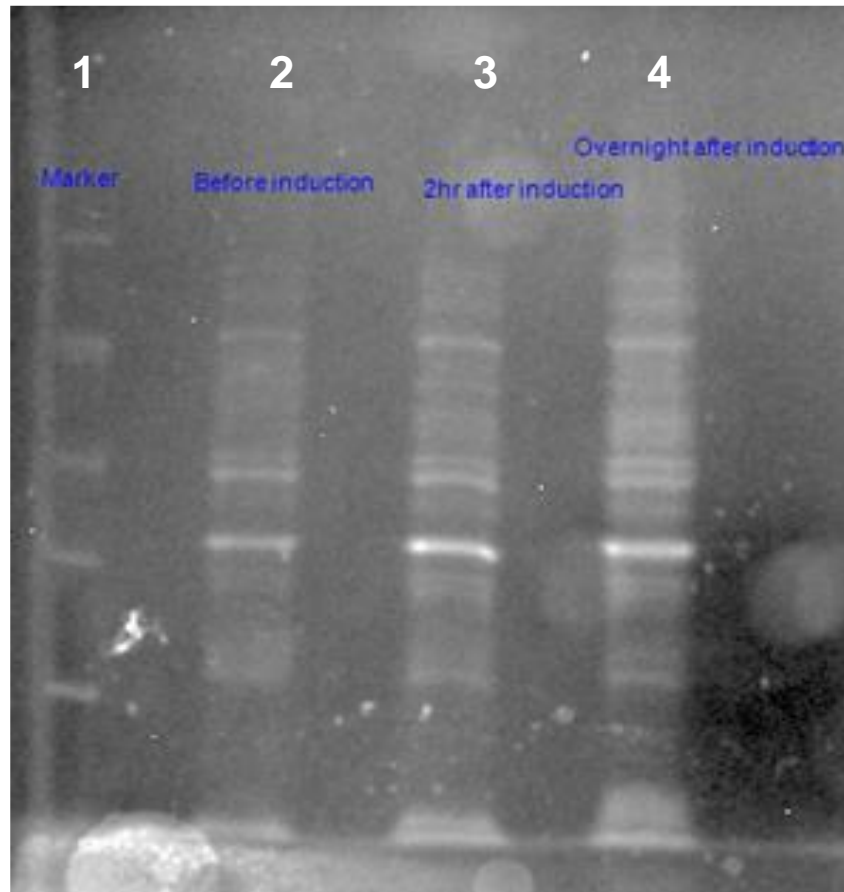


Figure 8.2-1 15% SDS gel on BL21(DE3) whole cell before and after induction.

The plasmid pET28-COVID-Nsp13 carrying BL21(DE3) whole cell was loaded onto a 15% SDS gel. The samples were taken before and after induction with IPTG to examine the protein expression. Lane 1 is the protein marker with a molecular weight of 116 kDa, 66.2 kDa, 45.0 kDa, 35.0 kDa, 25.0 kDa, 18.4 kDa, and 13.4 kDa. Lane 2 is the containing cells before induction. Lane 3 is the sample containing cells taken 2 hours after induction and Lane 3 is the sample containing cells taken 16 hours after induction.

To verify the expression of the nsp13, 100 μ l samples of media containing uninduced cells at an O.D. of 0.6 were reserved along with other samples taken 2 hours and 16 hours after induction. A reserved sample of the whole cell was then centrifuged, and the pellet was resuspended with a loading buffer containing 1x SDS. The mixture was then heated at 95 $^{\circ}$ C for

10 minutes. 10 μ l of the mixture was aliquoted out and mixed with 2 μ l of 6x SDS loading dye. 10 μ l sample was then loaded on to a 15% SDS gel. The molecular mass of the protein of interest is 70 kDa and the ladder used for the gel was Thermo Scientific Unstained Protein Molecular Weight Marker (Part number 26610). The Thermo Scientific Unstained Protein Molecular Weight Markers are a mixture of seven purified proteins ranging from 14.4 kDa to 116 kDa. Unfortunately, it seemed that there was no difference in the samples taken before or after induction and the protein band at 68 kDa was missing suggesting no protein expression. Other cell types such as Origami (DE3) and Rosetta (DE3) were also used in an attempt to express pET28-COVID-Nsp13, but no protein expression was detected.

Throughout the protein prep, 500–100 μ l samples of cells, culture medium (LB), pellet, and crude extract were collected. 10 μ l of each sample with 2 μ l of 6xSDS dye was heated and then loaded onto the 15% SDS gel. The gel ran at 50 volts for 2.5 hours, or until the blue dye reached 15% of the way from the bottom. The gel was then stained using Coomassie Blue Staining solution. Additionally, the samples that were loaded onto the gel were either lysed using a bug-buster or sonicated to examine the difference. The sample lysed using bug-buster seemed to have more protein than the sample lysed using sonication. But the difference was not significant, so sonication was used to lyse the cells.

SDS gel analysis revealed the presence of a big band with a comparable size to our protein of interest, in addition to some leaky expression. The fraction collected at 300 mM imidazole (Sample 11 in Fig. 8.2-2) was then dialyzed in GF buffer (20 mM Tris, pH 7, 50 mM NaCl, 1 mM EDTA, 0.1 mM DTT) + 20% glycerol and concentrated. The concentration of pure protein was

estimated using an absorbance measurement at 260 nm and a molar extinction coefficient calculated using the ProtParam tool (<https://web.expasy.org/protparam/>). The next step was to verify if the protein was active via an ATPase assay.

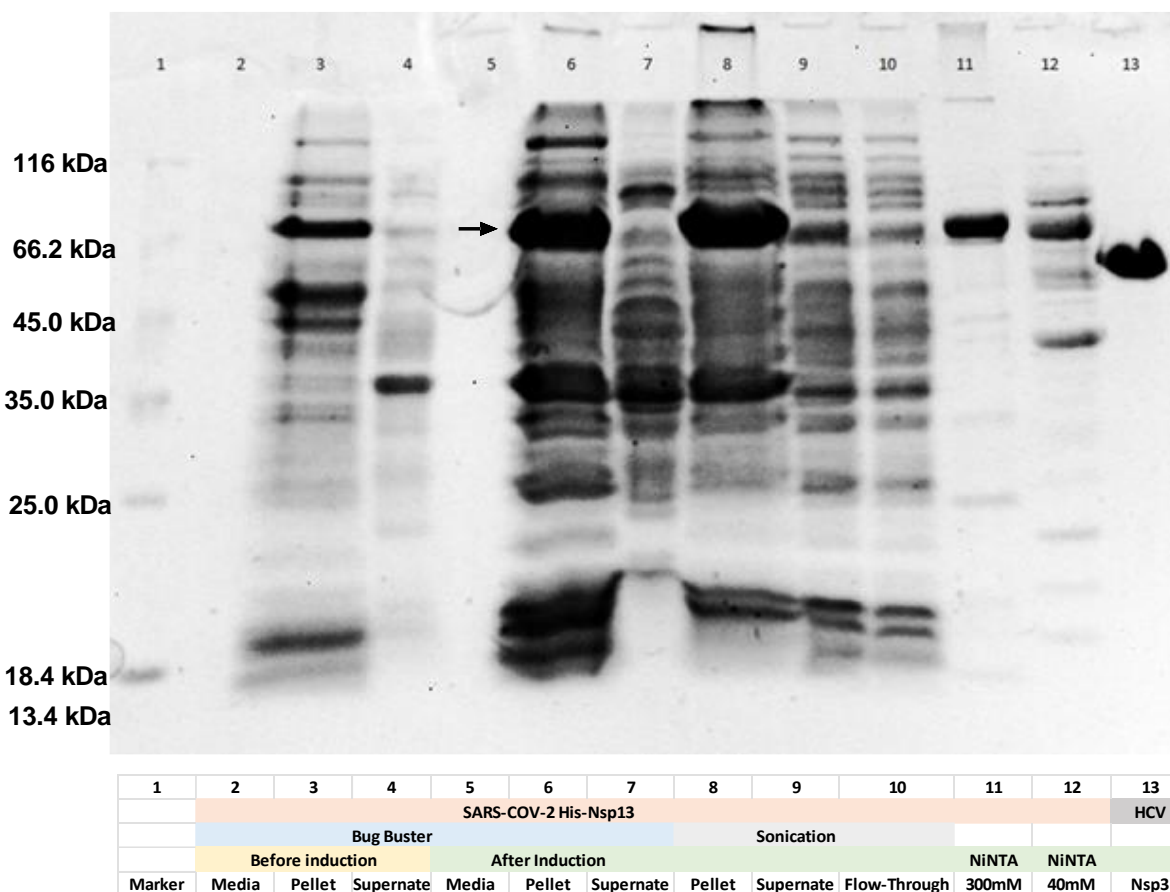


Figure 8.2-2. 15% SDS gel SARS-CoV-2 Nsp13.

15% SDS gels were loaded with 13 samples, each containing 10 μ l of sample with 2 μ l of 6x SDS dye. The gel ran at 50 volts for 2.5 hours, or until the blue dye reached 15% of the way from the bottom. Afterwards, the gel was stained with Coomassie Blue Staining solution thereby identifying our protein in each sample. The contents of the sample are represented in the table below the gel. Sample 1 is the protein marker with a molecular weight of 116 kDa, 66 kDa, 45 kDa, 35 kDa, 25 kDa, 18 kDa, and 13 kDa. Samples 3-7 were lysed using a bug-buster, whereas samples 8 and beyond were lysed using a sonicator. Sample 10 was flowthrough from NiNTA, and sample 11 contained purified protein eluted at 300 mM Imidazole, Sample 12 was protein eluted at 40 mM imidazole and sample 13 was purified HCV helicase (NS3h). The arrow in the gel represents the band with 69kDa protein, suggesting a protein of interest.

8.3.ATPase assay to evaluate His-nsp13 activity.

Purified protein was then utilized to evaluate the activity of nsp13. The activity of nsp13 was assessed using the malachite green assay. A detailed protocol for malachite green assay is discussed in detail in Chapter 4. Briefly, 3 ml of master mix was made containing 2 μ M nsp13 dissolved in 25mM of MOPS, 5mM $MgCl_2$, 1mM DTT, 0.01 mg/ml of BSA, and 0.01% Triton X 100 at pH 7. Separately, in 8 200 μ l PCR tubes, ATP was serial diluted and dissolved in the same buffer mentioned above. 50 μ l of master mix was then transferred into the first three columns of a 96-well plate. The next step was to add 50 μ l of ATP solution as the timer was started. The reaction was stopped after 5, 10 and 20 minutes upon adding phosphate reagent and immediately adding sodium citrate to each sample using a multi-channel pipette. After malachite green reagents were added, the plate was put on a shaker for 5–10 min as the color developed. After the color was developed, absorbance was taken at 630 nm. The initial reaction rate was approximated by taking the slope from 0 to 5 minutes. The slope calculated from 0 to 5 mins correlated with substrate concentration was then plotted and the data points were fitted to the Michaelis-Menten equation, giving V_{max} of 0.09018 mM/min, K_m of 0.94 mM, and k_{cat} of 90 min^{-1} .

K_m is the Michaelis-Menten constant with units of mM it is defined as the substrate concentration at half of the V_{max} . The K_m in this experiment could be considered as an estimate of dissociation constant assuming the dissociation rate from enzyme substrate complex to enzyme product is very small, or in other words, if there is a rapid equilibrium between enzyme, substrate, and enzyme substrate complex.

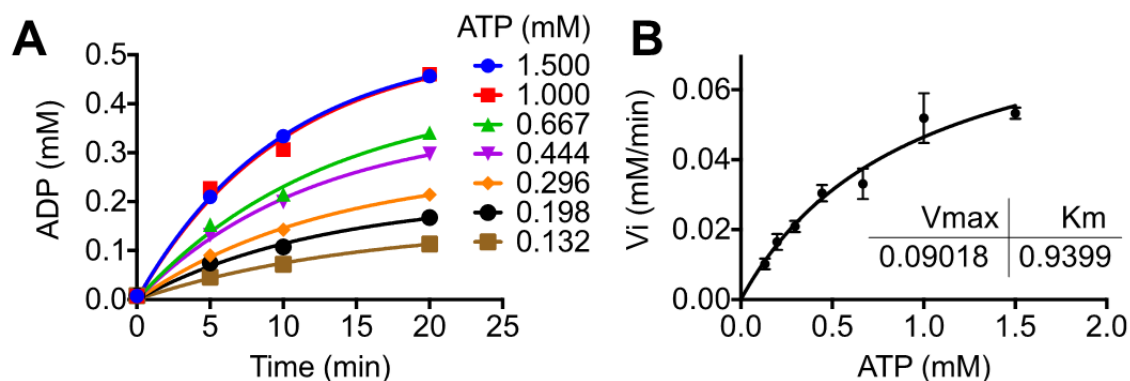


Figure 8.3-1 **Kinetic assays with nsp13.**

(A) The activity of nsp13 was measured by adding different concentrations of ATP and stopping the reaction upon adding malachite green reagent. The reaction containing different concentrations of ATP was stopped at 5, 10, and 15 min. The initial velocity was calculated by taking the slope between zero and five minutes. (B) The velocity along with the substrate concentration was plotted and the data was fitted to the Michaelis-Menten equation to give V_{max} and K_m .

The V_{max} for this reaction is calculated in terms of mM/min. V_{max} is the rate of the reaction at which the enzyme is fully saturated with substrate and increasing the substrate concentration further would not impact the rate at which the product is formed. The shape of the curve is hyperbolic, typical of enzyme kinetic studies, suggesting two different reaction orders. Unfortunately, it is not very apparent in this specific case, even at the 1.5 mM ATP concentration. A further increase in concentration of up to 5 or 10 mM of ATP may be required to fully see the hyperbolic shape of the initial first order reaction transitioning into zero order reaction as substrate concentration increases.

K_{cat} was calculated by dividing the V_{max} with the total enzyme concentration in the reaction, giving 90 min^{-1} . K_{cat} is the catalytic rate constant, or turnover number, that tells us how many substrate molecules are converted into products by a single enzyme per unit of time. In

this case, the K_{cat} of 90 min^{-1} tells us that 90 molecules of ATP are converted to ADP in a minute. Lastly, the enzyme specificity constant (K_{cat}/K_m) was measured to be $96 \text{ mM}^{-1} \text{ min}^{-1}$. In this case, the specificity constant is not very useful as it is used to compare relative rates of an enzyme acting on an alternating, competing substrate, and in this assay, no alternative substrate to ATP was used or compared.

8.4. Nsp13 is stimulated in the presence of 18-20 bases long oligonucleotides.

In the last section, the activity of purified nsp13 was measured, in this section the influence of oligonucleotide on helicase activity was investigated. The assay was performed in a similar manner as any ATPase activity assay but with an addition of oligonucleotide before adding ATP. Briefly, master mix 1 was made containing $2 \text{ }\mu\text{M}$ nsp13 dissolved in 25 mM MOPS, 5 mM MgCl_2 , 1 mM DTT, 0.01 mg/ml BSA and 0.01% Tritron X 100 at pH 7. Master mix 2 was made containing 5 mM ATP dissolved in 25 mM MOPS, 5 mM MgCl_2 , 1 mM DTT, 0.01 mg/ml BSA and 0.01% Tritron X at pH 7. Lastly, in separate tubes, $100 \text{ }\mu\text{l}$ of sample containing $15 \text{ }\mu\text{M}$ of each nucleotide dissolved in the same buffer mentioned above. The reaction was assembled by first transferring $50 \text{ }\mu\text{l}$ of master mix into first 3 columns of 96 well plate, Next, $25 \text{ }\mu\text{l}$ of individual oligonucleotide was added to each row and lastly $25 \text{ }\mu\text{l}$ of master mix 2 containing ATP was added. The reaction was stopped after 5-10 minutes upon adding malachite green reagents to each sample using multi-channel pipette. After malachite green reagents were added, the plate. was put on a shaker for 5-10 mins as the color developed. After the color was developed, absorbance was taken at 630 nm .

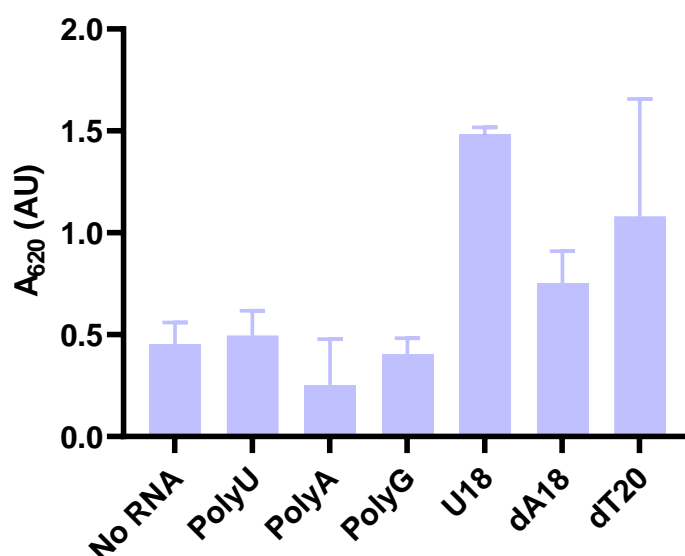


Figure 8.4-1 **Nsp13 is stimulated in the presence of 18-20 bases long oligonucleotide.** Different oligonucleotides were added to the sample containing nsp13 to examine its effect on the enzyme activity. Each assay was done in triplicate and measurements were taken 5-10 minutes after the reaction was initiated after adding ADP to the reaction mixture containing nsp13 with one oligonucleotide.

From the data, it seems that oligonucleotides between 18 and 20 base pairs tend to stimulate the enzyme, whereas longer chains of the same bases have no effect. It is possible that the oligonucleotides forms a convoluted structure restricting the interaction between helicase and bases, leading to no change in enzyme activity.

8.5. Nsp13 Chapter Summary

Antivirals designed to combat SARS-CoV-2 will most likely target viral non-structural proteins required for viral replication. Nsp13 is part of the replication and transcription complex with high sequence conservation among coronaviruses. As a result, nsp13 is a prospective target for the development of antiviral drugs. In this chapter, nsp13 was expressed in BL21(DE3) and

purified using a NiNTA column. The variant that was successfully expressed was similar to the one used by Mickolajczyk et al. (Mickolajczyk et al., 2020). This variant had his-tag added towards the n-terminal, whereas the other variant His-tag on the c-terminus. The concentration of the purified protein was then measured by taking the absorbance at 260 nm using a molar extinction coefficient calculated with the ProtParam tool (<https://web.expasy.org/protparam/>).

After concentrating, the purified protein was then utilized to evaluate the activity of nsp13. The activity of nsp13 was assessed using the malachite green assay. The enzyme activity was measured based on ATP concentration, and the reaction of ATP hydrolysis was measured based on the linear correlation of inorganic phosphate concentration in the sample and the color change of malachite green dye used in the malachite green assay. Different concentrations of ATP affect the rate at which the helicase hydrolyzes ATP. Measurements of inorganic phosphate were taken at 5-, 10-, and 20-minute intervals. The time interval and concentration of liberated phosphate were then plotted, and the rate of ATP hydrolysis was calculated based on the first-time interval between 0 and 5 minutes. The initial rates of hydrolysis with various concentration of ATP were then plotted and fitted to the Michalis-Menten equation to estimate V_{\max} of 0.09 mM/min and K_m of 0.94 mM values. K_m and helicase concentration were used to calculate k_{cat} of 90 min⁻¹. Lastly, the effect of oligonucleotides was measured on helicase activity, and based on the data, it seems that oligonucleotides between 18 and 20 base pairs tend to stimulate the enzyme, whereas longer chains of the same bases have no effect.

These ATPase assays were just the first step taken to make sure that the helicase activity can be measured in a quantitative manner and, using oligonucleotides, based on the data

oligonucleotides do in fact stimulate helicase activity. The next step in this study is to screen compounds to see if the activity of the helicase can be influenced. If the compound does seem to inhibit helicase activity, that compound might not only affect the SARS-CoV-2 helicase but all coronavirus family helicases due to high sequence conservation.

9. References

- Abraham, R., Hauer, D., McPherson, R. L., Utt, A., Kirby, I. T., Cohen, M. S., Merits, A., Leung, A. K. L., & Griffin, D. E. (2018). ADP-ribosyl-binding and hydrolase activities of the alphavirus nsP3 macrodomain are critical for initiation of virus replication. *Proceedings of the National Academy of Sciences of the United States of America*, 115(44), E10457–E10466. <https://doi.org/10.1073/PNAS.1812130115>
- Agelopoulos, M., & Thanos, D. (2006). Epigenetic determination of a cell-specific gene expression program by ATF-2 and the histone variant macroH2A. *The EMBO Journal*, 25, 4843–4853. <https://doi.org/10.1038/sj.emboj.7601364>
- Akaji, K., Konno, H., Mitsui, H., Teruya, K., Shimamoto, Y., Hattori, Y., Ozaki, T., Kusunoki, M., & Sanjoh, A. (2011). Structure-based design, synthesis, and evaluation of peptide-mimetic SARS 3CL protease inhibitors. *Journal of Medicinal Chemistry*, 54(23), 7962–7973. <https://doi.org/10.1021/jm200870n>
- Alhammad, Y., & Fehr, A. R. (2020). The Viral Macrodomain Counters Host Antiviral ADP-Ribosylation. *Viruses*, 12(4), 384. <https://doi.org/10.3390/v12040384>
- Almazán, F., DeDiego, M. L., Galán, C., Escors, D., Álvarez, E., Ortego, J., Sola, I., Zuñiga, S., Alonso, S., Moreno, J. L., Nogales, A., Capiscol, C., & Enjuanes, L. (2006). Construction of a Severe Acute Respiratory Syndrome Coronavirus Infectious cDNA Clone and a Replicon To Study Coronavirus RNA Synthesis. *Journal of Virology*, 80(21), 10900–10906. <https://doi.org/10.1128/JVI.00385-06/>
- Almeida, M. S., Johnson, M. A., Herrmann, T., Geralt, M., & Wüthrich, K. (2007). Novel β -Barrel Fold in the Nuclear Magnetic Resonance Structure of the Replicase Nonstructural Protein 1 from the Severe Acute Respiratory Syndrome Coronavirus. *Journal of Virology*, 81(7), 3151–3161. <https://doi.org/10.1128/jvi.01939-06>
- Amaratunga, M., & Lohman, T. M. (1993). Escherichia coli rep helicase unwinds DNA by an active mechanism. *Biochemistry*, 32(27), 6815–6820. <https://doi.org/10.1021/BI00078A003>
- Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., & Hilgenfeld, R. (2003). Coronavirus main proteinase (3CLpro) Structure: Basis for design of anti-SARS drugs. *Science*, 300(5626), 1763–1767. <https://doi.org/10.1126/science.1085658>
- Andersson, K.-E., & Högestätt, E. D. (2009). On the Mechanism of Action of Calcium Antagonists. *Acta Medica Scandinavica*, 215(S681). <https://doi.org/10.1111/j.0954-6820.1984.tb08672.x>
- Angelini, M. M., Akhlaghpour, M., Neuman, B. W., & Buchmeier, M. J. (2013). Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. *mBio*, 4(4), e00524-13. <https://doi.org/10.1128/mBio.00524-13>
- Appel, N., Pietschmann, T., & Bartenschlager, R. (2005). Mutational Analysis of Hepatitis C Virus Nonstructural Protein 5A: Potential Role of Differential Phosphorylation in RNA Replication and

- Identification of a Genetically Flexible Domain. *Journal of Virology*, 79(5), 3187.
<https://doi.org/10.1128/JVI.79.5.3187-3194.2005>
- Artika, I. M., Dewantari, A. K., & Wiyatno, A. (2020). Molecular biology of coronaviruses: current knowledge. *Heliyon*, 6(8), e04743. <https://doi.org/10.1016/J.HELIYON.2020.E04743>
- Atkins, P., & De Paula, J. (2006). The First Law. In *Atkins' Physical chemistry* (11th ed., pp. 28 to 47). W.H. Freeman.
- Bai, N., Roder, H., Dickson, A., & Karanicolas, J. (2019). Isothermal Analysis of ThermoFluor Data can readily provide Quantitative Binding Affinities. *Scientific reports*, 9(1), 2650.
<https://doi.org/10.1038/s41598-018-37072-x>
- Barretto, N., Jukneliene, D., Ratia, K., Chen, Z., Mesecar, A. D., & Baker, S. C. (2005). The papain-like protease of severe acute respiratory syndrome coronavirus has deubiquitinating activity. *Journal of virology*, 79(24), 15189–15198. <https://doi.org/10.1128/JVI.79.24.15189-15198.2005>
- Beales, L. P., Rowlands, D. J., & Holzenburg, A. (2001). The internal ribosome entry site (IRES) of hepatitis C virus visualized by electron microscopy. *RNA*, 7(5), 661.
<https://doi.org/10.1017/S1355838201001406>
- Boivin, S., Kozak, S., & Meijers, R. (2013). Optimization of protein purification and characterization using Thermofluor screens. *Protein Expression and Purification*, 91(2), 192–206.
<https://doi.org/10.1016/j.pep.2013.08.002>
- Boulant, S., Vanbelle, C., Ebel, C., Penin, F., & Lavergne, J.-P. (2005). Hepatitis C Virus Core Protein Is a Dimeric Alpha-Helical Protein Exhibiting Membrane Protein Features. *Journal of Virology*, 79(17), 11353. <https://doi.org/10.1128/JVI.79.17.11353-11365.2005>
- Bouvet, M., Lugari, A., Posthuma, C. C., Zevenhoven, J. C., Bernard, S., Betzi, S., Imbert, I., Canard, B., Guillemot, J. C., Lécine, P., Pfefferle, S., Drosten, C., Snijder, E. J., Decroly, E., & Morelli, X. (2014). Coronavirus Nsp10, a critical co-factor for activation of multiple replicative enzymes. *Journal of Biological Chemistry*, 289(37), 25783–25796.
<https://doi.org/10.1074/jbc.M114.577353>
- Breithaupt-Faloppa, A. C., Correia, C. J., Prado, C. M., Stilhano, R. S., Ureshino, R. P., & Moreira, L. (2020). 17β-Estradiol, a potential ally to alleviate SARS-CoV-2 infection. *Clinics (Sao Paulo, Brazil)*, 75, e1980. <https://doi.org/10.6061/clinics/2020/e1980>
- Breithaupt-Faloppa, A. C., Correia, C. de J., Prado, C. M., Stilhano, R. S., Ureshino, R. P., & Moreira, L. F. P. (2020b). 17β-Estradiol, a potential ally to alleviate SARS-CoV-2 infection. *Clinics*, 75, 1–8.
<https://doi.org/10.6061/CLINICS/2020/E1980>
- Buschbeck, M., & di Croce, L. (2010). Approaching the molecular and physiological function of macroH2A variants. *Epigenetics*, 5(2), 118–123. <https://doi.org/10.4161/EPI.5.2.11076>

- Cai, Y., Zhang, J., Xiao, T., Peng, H., Sterling, S. M., Walsh, R. M., Rawson, S., Rits-Volloch, S., & Chen, B. (2020). Distinct conformational states of SARS-CoV-2 spike protein. *Science*, 369(6511), 1586+. <https://doi.org/10.1126/science.abd4251>
- Carbonell, T., & Freire, E. (2005). Binding thermodynamics of statins to HMG-CoA reductase. *Biochemistry*, 44(35), 11741–11748. <https://doi.org/10.1021/bi050905v>
- Carrère-Kremer, S., Montpellier-Pala, C., Cocquerel, L., Wychowski, C., Penin, F., & Dubuisson, J. (2002). Subcellular Localization and Topology of the p7 Polypeptide of Hepatitis C Virus. *Journal of Virology*, 76(8), 3720. <https://doi.org/10.1128/JVI.76.8.3720-3730.2002>
- Chaires J. B. (2008). Calorimetry and thermodynamics in drug design. *Annual review of biophysics*, 37, 135–151. <https://doi.org/10.1146/annurev.biophys.36.040306.132812>
- Chakravarthy, S., Gundimella, S. K. Y., Caron, C., Perche, P.-Y., Pehrson, J. R., Khochbin, S., & Luger, K. (2005a). Structural Characterization of the Histone Variant macroH2A. *Molecular and Cellular Biology*, 25(17), 7616–7624. <https://doi.org/10.1128/mcb.25.17.7616-7624.2005>
- Chakravarthy, S., Gundimella, S. K. Y., Caron, C., Perche, P.-Y., Pehrson, J. R., Khochbin, S., & Luger, K. (2005b). Structural characterization of the histone variant macroH2A. *Molecular and Cellular Biology*, 25(17), 7616–7624. <https://doi.org/10.1128/MCB.25.17.7616-7624.2005>
- Chakravarthy, S., & Luger, K. (2006). The histone variant macro-H2A preferentially forms “hybrid nucleosomes.” *Journal of Biological Chemistry*, 281(35), 25522–25531. <https://doi.org/10.1074/jbc.M602258200>
- Champoux J. J. (2001). DNA topoisomerases: structure, function, and mechanism. *Annual review of biochemistry*, 70, 369–413. <https://doi.org/10.1146/annurev.biochem.70.1.369>
- Chang, R., & Overby, J. (2011). Thermodynamics. In *General Chemistry: The Essential Concepts* (6th ed., pp. 628 to 650). McGraw-Hill Education.
- de Chasse, B., Navratil, V., Tafforeau, L., Hiet, M. S., Aublin-Gex, A., Agaugué, S., Meiffren, G., Pradezynski, F., Faria, B. F., Chantier, T., Le Breton, M., Pellet, J., Davoust, N., Mangeot, P. E., Chaboud, A., Penin, F., Jacob, Y., Vidalain, P. O., Vidal, M., André, P., ... Lotteau, V. (2008). Hepatitis C virus infection protein network. *Molecular systems biology*, 4, 230. <https://doi.org/10.1038/msb.2008.66>
- Cheng, A., Zhang, W., Xie, Y., Jiang, W., Arnold, E., Sarafianos, S. G., & Ding, J. (2005). Expression, purification, and characterization of SARS coronavirus RNA polymerase. *Virology*, 335(2), 165–176. <https://doi.org/10.1016/J.VIROL.2005.02.017>
- Cheng, J. C., Chang, M. F., & Chang, S. C. (1999). Specific interaction between the hepatitis C virus NS5B RNA polymerase and the 3' end of the viral RNA. *Journal of virology*, 73(8), 7044–7049. <https://doi.org/10.1128/JVI.73.8.7044-7049.1999>

- Cheng, W., Chen, S., Li, R., Chen, Y., Wang, M., & Guo, D. (2015). Severe acute respiratory syndrome coronavirus protein 6 mediates ubiquitin-dependent proteosomal degradation of N-Myc (and STAT) interactor. *Virologica Sinica*, 30(2), 153–161. <https://doi.org/10.1007/s12250-015-3581-8>
- Cho, C. C., Lin, M. H., Chuang, C. Y., & Hsu, C. H. (2016a). Macro Domain from Middle East Respiratory Syndrome Coronavirus (MERS-CoV) Is an Efficient ADP-ribose Binding Module: CRYSTAL STRUCTURE AND BIOCHEMICAL STUDIES. *The Journal of Biological Chemistry*, 291(10), 4894–4902. <https://doi.org/10.1074/JBC.M115.700542>
- Cho, C. C., Lin, M. H., Chuang, C. Y., & Hsu, C. H. (2016b). Macro Domain from Middle East Respiratory Syndrome Coronavirus (MERS-CoV) Is an Efficient ADP-ribose Binding Module: CRYSTAL STRUCTURE AND BIOCHEMICAL STUDIES. *The Journal of Biological Chemistry*, 291(10), 4894–4902. <https://doi.org/10.1074/JBC.M115.700542>
- Clausius, R. (1867). *The Mechanical Theory of Heat – with its Applications to the Steam Engine and to Physical Properties of Bodies*. London: John van Voorst. Retrieved 13 December 2021.
- Claverie J. M. (2020). A Putative Role of de-Mono-ADP-Ribosylation of STAT1 by the SARS-CoV-2 Nsp3 Protein in the Cytokine Storm Syndrome of COVID-19. *Viruses*, 12(6), 646. <https://doi.org/10.3390/v12060646>
- Collins, S. J., & Groudine, M. T. (1983). Rearrangement and amplification of c-abl sequences in the human chronic myelogenous leukemia cell line K-562. *Proceedings of the National Academy of Sciences of the United States of America*, 80(15), 4813. <https://doi.org/10.1073/PNAS.80.15.4813>
- Cottam, E. M., Whelband, M. C., & Wileman, T. (2014). Coronavirus NSP6 restricts autophagosome expansion. *Autophagy*, 10(8), 1426–1441. <https://doi.org/10.4161/auto.29309>
- Curtis, T. M., & Scholfield, C. N. (2001). Nifedipine blocks Ca²⁺ store refilling through a pathway not involving L-type Ca²⁺ channels in rabbit arteriolar smooth muscle. *The Journal of Physiology*, 532(3). <https://doi.org/10.1111/j.1469-7793.2001.0609e.x>
- Decroly, E., Imbert, I., Coutard, B., Bouvet, M., Selisko, B., Alvarez, K., Gorbalenya, A. E., Snijder, E. J., & Canard, B. (2008). Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2′O)-methyltransferase activity. *Journal of Virology*, 82(16), 8071–8084. <https://doi.org/10.1128/JVI.00407-08>
- Delmas, B., & Laude, H. (1990). Assembly of coronavirus spike protein into trimers and its role in epitope expression. *Journal of Virology*, 64(11), 5367–5375. <https://doi.org/10.1128/jvi.64.11.5367-5375.1990>
- Donaldson, E. F., Sims, A. C., Graham, R. L., Denison, M. R., & Baric, R. S. (2007). Murine hepatitis virus replicase protein nsp10 is a critical regulator of viral RNA synthesis. *Journal of virology*, 81(12), 6356–6368. <https://doi.org/10.1128/JVI.02805-06>

- Doyen, C.-M., An, W., Angelov, D., Bondarenko, V., Miettinen, F., Studitsky, V. M., Hamiche, A., Roeder, R. G., Bouvet, P., & Dimitrov, S. (2006). Mechanism of Polymerase II Transcription Repression by the Histone Variant macroH2A. *Molecular and Cellular Biology*, 26(3), 1156–1164. <https://doi.org/10.1128/mcb.26.3.1156-1164.2006>
- Egloff, M. P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H., Snijder, E. J., Gorbalenya, A. E., Cambillau, C., & Canard, B. (2004). The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. *Proceedings of the National Academy of Sciences of the United States of America*, 101(11), 3792–3796. <https://doi.org/10.1073/pnas.0307877101>
- Egloff, M.-P., Malet, H., Putics, Á., Heinonen, M., Dutartre, H., Frangeul, A., Gruez, A., Campanacci, V., Cambillau, C., Ziebuhr, J., Ahola, T., & Canard, B. (2006). Structural and Functional Basis for ADP-Ribose and Poly(ADP-Ribose) Binding by Viral Macro Domains. *Journal of Virology*, 80(17), 8493. <https://doi.org/10.1128/JVI.00713-06>
- Emadi, A., Chua, J. V., Talwani, R., Bentzen, S. M., & Baddley, J. (2020). Safety and Efficacy of Imatinib for Hospitalized Adults with COVID-19: A structured summary of a study protocol for a randomised controlled trial. *Trials*, 21(1), 897. <https://doi.org/10.1186/s13063-020-04819-9>
- Epstein, F. H., Kurzrock, R., Gutterman, J. U., & Talpaz, M. (2010). The Molecular Genetics of Philadelphia Chromosome–Positive Leukemias. <http://dx.doi.org/10.1056/NEJM198810133191506>, 319(15), 990–998. <https://doi.org/10.1056/NEJM198810133191506>
- Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N., & Nordlund, P. (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry*, 357(2), 289–298. <https://doi.org/10.1016/j.ab.2006.07.027>
- Evans, M. J., Rice, C. M., & Goff, S. P. (2004). Phosphorylation of hepatitis C virus nonstructural protein 5A modulates its protein interactions and viral RNA replication. *Proceedings of the National Academy of Sciences of the United States of America*, 101(35), 13038. <https://doi.org/10.1073/PNAS.0405152101>
- Fehr, A. R., Channappanavar, R., Jankevicius, G., Fett, C., Zhao, J., Athmer, J., Meyerholz, D. K., Ahel, I., & Perlman, S. (2016). The Conserved Coronavirus Macrodome Promotes Virulence and Suppresses the Innate Immune Response during Severe Acute Respiratory Syndrome Coronavirus Infection. *mBio*, 7(6), e01721-16. <https://doi.org/10.1128/mBio.01721-16>
- Fehr, A. R., & Perlman, S. (2015). Coronaviruses: an overview of their replication and pathogenesis. *Methods in molecular biology* (Clifton, N.J.), 1282, 1–23. https://doi.org/10.1007/978-1-4939-2438-7_1
- Freire E. (2008). Do enthalpy and entropy distinguish first in class from best in class?. *Drug discovery today*, 13(19-20), 869–874. <https://doi.org/10.1016/j.drudis.2008.07.005>

- Freire, E. (2009). A Thermodynamic Approach to the Affinity Optimization of Drug Candidates. *Chemical Biology & Drug Design*, 74(5). <https://doi.org/10.1111/j.1747-0285.2009.00880.x>
- Frick, D. N. (2006). HCV Helicase: Structure, Function, and Inhibition. In S. L. Tan (Ed.), *Hepatitis C Viruses: Genomes and Molecular Biology*. Horizon Bioscience. <http://www.ncbi.nlm.nih.gov/pubmed/21250378>
- Frick, D. N., Virdi, R. S., Vuksanovic, N., Dahal, N., & Silvaggi, N. R. (2020). Molecular Basis for ADP-Ribose Binding to the Mac1 Domain of SARS-CoV-2 nsp3. *Biochemistry*, 59(28), 2608–2615. <https://doi.org/10.1021/acs.biochem.0c00309>
- Gadlage, M. J., Sparks, J. S., Beachboard, D. C., Cox, R. G., Doyle, J. D., Stobart, C. C., & Denison, M. R. (2010). Murine Hepatitis Virus Nonstructural Protein 4 Regulates Virus-Induced Membrane Modifications and Replication Complex Function. *Journal of Virology*, 84(1), 280–290. <https://doi.org/10.1128/jvi.01772-09>
- Gamble, M. J. (2013). Expanding the functional repertoire of macrodomains. *Nature Structural & Molecular Biology* 20:4, 20(4), 407–408. <https://doi.org/10.1038/nsmb.2552>
- Gasmi, A., Noor, S., Tippairote, T., Dadar, M., Menzel, A., & Bjørklund, G. (2020). Individual risk management strategy and potential therapeutic options for the COVID-19 pandemic. *Clinical Immunology (Orlando, Fla.)*, 215, 108409. <https://doi.org/10.1016/J.CLIM.2020.108409>
- Glowacka, I., Bertram, S., Muller, M. A., Allen, P., Soilleux, E., Pfefferle, S., Steffen, I., Tsegaye, T. S., He, Y., Gnirss, K., Niemeyer, D., Schneider, H., Drosten, C., & Pohlmann, S. (2011). Evidence that TMPRSS2 Activates the Severe Acute Respiratory Syndrome Coronavirus Spike Protein for Membrane Fusion and Reduces Viral Control by the Humoral Immune Response. *Journal of Virology*, 85(9), 4122–4134. <https://doi.org/10.1128/jvi.02232-10>
- Gorbalenya, A. E., Koonin, E. v, Donchenko, A. P., & Blinov, V. M. (1989). Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Research*, 17(12), 4713. <https://doi.org/10.1093/NAR/17.12.4713>
- Gosert, R., Kanjanahaluethai, A., Egger, D., Bienz, K., & Baker, S. C. (2002). RNA Replication of Mouse Hepatitis Virus Takes Place at Double-Membrane Vesicles. *Journal of Virology*, 76(8), 3697–3708. <https://doi.org/10.1128/jvi.76.8.3697-3708.2002>
- Graham, R. L., & Denison, M. R. (2006). Replication of murine hepatitis virus is regulated by papain-like proteinase 1 processing of nonstructural proteins 1, 2, and 3. *Journal of virology*, 80(23), 11610–11620. <https://doi.org/10.1128/JVI.01428-06>
- Graham, R. L., Sparks, J. S., Eckerle, L. D., Sims, A. C., & Denison, M. R. (2008). SARS coronavirus replicase proteins in pathogenesis. *Virus research*, 133(1), 88–100. <https://doi.org/10.1016/j.virusres.2007.02.017>

- Grandi, G., Facchinetti, F., & Bitzer, J. (2020). The gendered impact of coronavirus disease (COVID-19): do estrogens play a role? *The European Journal of Contraception & Reproductive Health Care : The Official Journal of the European Society of Contraception*, 25(3), 233–234. <https://doi.org/10.1080/13625187.2020.1766017>
- Guarino, L. A., Bhardwaj, K., Dong, W., Sun, J., Holzenburg, A., & Kao, C. (2005). Mutational analysis of the SARS virus Nsp15 endoribonuclease: identification of residues affecting hexamer formation. *Journal of Molecular Biology*, 353(5), 1106–1117. <https://doi.org/10.1016/J.JMB.2005.09.007>
- Hagemeijer, M. C., Monastyrska, I., Griffith, J., van der Sluijs, P., Voortman, J., van Bergen en Henegouwen, P. M., Vonk, A. M., Rottier, P. J. M., Reggiori, F., & de Haan, C. A. M. (2014). Membrane rearrangements mediated by coronavirus nonstructural proteins 3 and 4. *Virology*, 458–459(1), 125–135. <https://doi.org/10.1016/j.virol.2014.04.027>
- Hao, W., Wojdyla, J. A., Zhao, R., Han, R., Das, R., Zlatev, I., Manoharan, M., Wang, M., & Cui, S. (2017). Crystal structure of Middle East respiratory syndrome coronavirus helicase. *PLoS pathogens*, 13(6), e1006474. <https://doi.org/10.1371/journal.ppat.1006474>
- Harrus, D., Ahmed-El-Sayed, N., Simister, P. C., Miller, S., Triconnet, M., Hagedorn, C. H., Mahias, K., Rey, F. A., Astier-Gin, T., & Bressanelli, S. (2010). Further Insights into the Roles of GTP and the C Terminus of the Hepatitis C Virus Polymerase in the Initiation of RNA Synthesis. *The Journal of Biological Chemistry*, 285(43), 32906. <https://doi.org/10.1074/JBC.M110.151316>
- Hartenian, E., Nandakumar, D., Lari, A., Ly, M., Tucker, J. M., & Glaunsinger, B. A. (2020). The molecular virology of coronaviruses. *Journal of Biological Chemistry*, 295(37), 12910–12934. <https://doi.org/10.1074/jbc.REV120.013930>
- Heaton, S. M., Borg, N. A., & Dixit, V. M. (2016). Ubiquitin in the activation and attenuation of innate antiviral immunity. *The Journal of experimental medicine*, 213(1), 1–13. <https://doi.org/10.1084/jem.20151531>
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N. H., Nitsche, A., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, 181(2), 271-280.e8. <https://doi.org/10.1016/j.cell.2020.02.052>
- Honda, M., Ping, L. H., Rijnbrand, R. C. A., Amphlett, E., Clarke, B., Rowlands, D., & Lemon, S. M. (1996). Structural requirements for initiation of translation by internal ribosome entry within genome-length hepatitis C virus RNA. *Virology*, 222(1), 31–42. <https://doi.org/10.1006/VIRO.1996.0395>
- Hsu, M. F., Kuo, C. J., Chang, K. T., Chang, H. C., Chou, C. C., Ko, T. P., Shr, H. L., Chang, G. G., Wang, A. H., & Liang, P. H. (2005). Mechanism of the maturation process of SARS-CoV 3CL protease. *The Journal of biological chemistry*, 280(35), 31257–31266. <https://doi.org/10.1074/jbc.M502577200>

- Hurst, K. R., Koetzner, C. A., & Masters, P. S. (2013). Characterization of a Critical Interaction between the Coronavirus Nucleocapsid Protein and Nonstructural Protein 3 of the Viral Replicase-Transcriptase Complex. *Journal of Virology*, 87(16), 9159–9172. <https://doi.org/10.1128/jvi.01275-13>
- Huynh, K., & Partch, C. L. (2015). Analysis of protein stability and ligand interactions by thermal shift assay. *Current Protocols in Protein Science*, 79, 28.9.1-28.9.14. <https://doi.org/10.1002/0471140864.PS2809S79>
- Ichimiya, T., Yamakawa, T., Hirano, T., Yokoyama, Y., Hayashi, Y., Hirayama, D., Wagatsuma, K., Itoi, T., & Nakase, H. (2020). Autophagy and Autophagy-Related Diseases: A Review. *International journal of molecular sciences*, 21(23), 8974. <https://doi.org/10.3390/ijms21238974>
- Irigoyen, N., Firth, A. E., Jones, J. D., Chung, B. Y. W., Siddell, S. G., & Brierley, I. (2016). High-Resolution Analysis of Coronavirus Gene Expression by RNA Sequencing and Ribosome Profiling. *PLoS Pathogens*, 12(2), e1005473. <https://doi.org/10.1371/journal.ppat.1005473>
- Ivanov, K. A., Hertzog, T., Rozanov, M., Bayer, S., Thiel, V., Gorbalenya, A. E., & Ziebuhr, J. (2004). Major genetic marker of nidoviruses encodes a replicative endoribonuclease. *Proceedings of the National Academy of Sciences of the United States of America*, 101(34), 12694–12699. <https://doi.org/10.1073/PNAS.0403127101>
- Ivanov, K. A., Thiel, V., Dobbe, J. C., van der Meer, Y., Snijder, E. J., & Ziebuhr, J. (2004). Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *Journal of Virology*, 78(11), 5619–5632. <https://doi.org/10.1128/JVI.78.11.5619-5632.2004>
- Ivanov, K. A., & Ziebuhr, J. (2004). Human Coronavirus 229E Nonstructural Protein 13: Characterization of Duplex-Unwinding, Nucleoside Triphosphatase, and RNA 5'-Triphosphatase Activities. *Journal of Virology*, 78(14), 7833–7838. <https://doi.org/10.1128/JVI.78.14.7833-7838.2004>
- Jia, Z., Yan, L., Ren, Z., Wu, L., Wang, J., Guo, J., Zheng, L., Ming, Z., Zhang, L., Lou, Z., & Rao, Z. (2019). Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis. *Nucleic Acids Research*, 47(12), 6538–6550. <https://doi.org/10.1093/NAR/GKZ409>
- Johnson, M. A., Chatterjee, A., Neuman, B. W., & Wüthrich, K. (2010). SARS coronavirus unique domain: Three-domain molecular architecture in solution and RNA binding. *Journal of Molecular Biology*, 400(4), 724–742. <https://doi.org/10.1016/j.jmb.2010.05.027>
- Johnson, T. W., Gallego, R. A., & Edwards, M. P. (2018). Lipophilic Efficiency as an Important Metric in Drug Design. *Journal of medicinal chemistry*, 61(15), 6401–6420. <https://doi.org/10.1021/acs.jmedchem.8b00077>
- Kciuk, M., Marciniak, B., & Kontek, R. (2020). Irinotecan—still an important player in cancer chemotherapy: A comprehensive overview. In *International Journal of Molecular Sciences* (Vol. 21, Issue 14, pp. 1–21). MDPI AG. <https://doi.org/10.3390/ijms21144919>

- Keep, S., Bickerton, E., Armesto, M., & Britton, P. (2018). The ADRP domain from a virulent strain of infectious bronchitis virus is not sufficient to confer a pathogenic phenotype to the attenuated Beaudette strain. *The Journal of General Virology*, 99(8), 1097–1102. <https://doi.org/10.1099/JGV.0.001098>
- Kelvin, William T. (1849) "An Account of Carnot's Theory of the Motive Power of Heat – with Numerical Results Deduced from Regnault's Experiments on Steam." *Transactions of the Edinburg Royal Society*, XVI.
- Kim, J. L., Morgenstern, K. A., Griffith, J. P., Dwyer, M. D., Thomson, J. A., Murcko, M. A., Lin, C., & Caron, P. R. (1998). Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure*, 6(1), 89–100. [https://doi.org/10.1016/S0969-2126\(98\)00010-0](https://doi.org/10.1016/S0969-2126(98)00010-0)
- Kirchdoerfer, R. N., & Ward, A. B. (2019). Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nature Communications*, 10(1), 2342. <https://doi.org/10.1038/s41467-019-10280-3>
- Kraus, W. L. (2009). New functions for an ancient domain. *Nature Structural & Molecular Biology* 2009 16:9, 16(9), 904–907. <https://doi.org/10.1038/nsmb0909-904>
- Krey, T., d'Alayer, J., Kikuti, C. M., Saulnier, A., Damier-Piolle, L., Petitpas, I., Johansson, D. X., Tawar, R. G., Baron, B., Robert, B., England, P., Persson, M. A. A., Martin, A., & Rey, F. A. (2010). The Disulfide Bonds in Glycoprotein E2 of Hepatitis C Virus Reveal the Tertiary Organization of the Molecule. *PLOS Pathogens*, 6(2), e1000762. <https://doi.org/10.1371/JOURNAL.PPAT.1000762>
- Kusov, Y., Tan, J., Alvarez, E., Enjuanes, L., & Hilgenfeld, R. (2015). A G-quadruplex-binding macrodomain within the “SARS-unique domain” is essential for the activity of the SARS-coronavirus replication-transcription complex. *Virology*, 484, 313–322. <https://doi.org/10.1016/j.virol.2015.06.016>
- Lakowicz, J. R. (2006a). Principles of fluorescence spectroscopy. *Principles of Fluorescence Spectroscopy*, 1–954. <https://doi.org/10.1007/978-0-387-46312-4>
- Lal, S. K. (2010). Molecular biology of the SARS-coronavirus. *Molecular Biology of the SARS-Coronavirus*, 1–328. <https://doi.org/10.1007/978-3-642-03683-5>
- Layton, C. J., & Hellinga, H. W. (2010). Thermodynamic analysis of ligand-induced changes in protein thermal unfolding applied to high-throughput determination of ligand affinities with extrinsic fluorescent dyes. *Biochemistry*, 49(51), 10831–10841. <https://doi.org/10.1021/B1101414Z>
- Lee, P. H., Huang, X. X., Teh, B. T., & Ng, L. M. (2019). TSA-CRAFT: A Free Software for Automatic and Robust Thermal Shift Assay Data Analysis. *SLAS Discovery*, 24(5), 606–612. <https://doi.org/10.1177/2472555218823547>

- Lehmann, K. C., Snijder, E. J., Posthuma, C. C., & Gorbalenya, A. E. (2015). What we know but do not understand about nidovirus helicases. *Virus Research*, 202, 12–32. <https://doi.org/10.1016/J.VIRUSRES.2014.12.001>
- Lei, J., Kusov, Y., & Hilgenfeld, R. (2018). Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Research*, 149, 58–74. <https://doi.org/10.1016/j.antiviral.2017.11.001>
- Li, X. L., Wu, Z. Q., & Han, W. D. (2013). The macrodomain family: Rethinking an ancient domain from evolutionary perspectives. *Chinese Science Bulletin = Kexue Tongbao*, 58(9), 953–960. <https://doi.org/10.1007/S11434-013-5674-9>
- Li, X., & Zhang, C. (2021). Using Differential Scanning Fluorimetry (DSF) to Detect Ligand Binding with Purified Protein. *Methods in Molecular Biology*, 2213, 183–186. https://doi.org/10.1007/978-1-0716-0954-5_16
- Lipinski, C. A. (2000). Drug-like properties and the causes of poor solubility and poor permeability. *Journal of Pharmacological and Toxicological Methods*, 44(1), 235–249. [https://doi.org/10.1016/S1056-8719\(00\)00107-6](https://doi.org/10.1016/S1056-8719(00)00107-6)
- Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, 46(1-3), 3–26. [https://doi.org/10.1016/s0169-409x\(00\)00129-0](https://doi.org/10.1016/s0169-409x(00)00129-0)
- Liu, C., & Yu, X. (2015). ADP-ribosyltransferases and poly ADP-ribosylation. *Current protein & peptide science*, 16(6), 491–501. <https://doi.org/10.2174/1389203716666150504122435>
- Love, R. A., Parge, H. E., Wickersham, J. A., Hostomsky, Z., Habuka, N., Moomaw, E. W., Adachi, T., Margosiak, S., Dagostino, E., & Hostomska, Z. (1998). The conformation of hepatitis C virus NS3 proteinase with and without NS4A: a structural basis for the activation of the enzyme by its cofactor. *Clinical and Diagnostic Virology*, 10(2–3), 151–156. [https://doi.org/10.1016/S0928-0197\(98\)00036-1](https://doi.org/10.1016/S0928-0197(98)00036-1)
- Lugari, A., Betzi, S., Decroly, E., Bonnaud, E., Hermant, A., Guillemot, J. C., Debarnot, C., Borg, J. P., Bouvet, M., Canard, B., Morelli, X., & Lécine, P. (2010). Molecular mapping of the RNA Cap 2'-O-methyltransferase activation interface between severe acute respiratory syndrome coronavirus nsp10 and nsp16. *The Journal of biological chemistry*, 285(43), 33230–33241. <https://doi.org/10.1074/jbc.M110.120014>
- Lundin, M., Lindström, H., Grönwall, C., & Persson, M. A. A. (2006). Dual topology of the processed hepatitis C virus protein NS4B is influenced by the NS5A protein. *The Journal of General Virology*, 87(Pt 11), 3263–3272. <https://doi.org/10.1099/VIR.0.82211-0>
- Lundin, M., Monné, M., Widell, A., Heijne, G. von, & Persson, M. A. A. (2003). Topology of the Membrane-Associated Hepatitis C Virus Protein NS4B. *Journal of Virology*, 77(9), 5428–5438. <https://doi.org/10.1128/JVI.77.9.5428-5438.2003>

- Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., Lou, Z., Yan, L., Zhang, R., & Rao, Z. (2015). Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. *Proceedings of the National Academy of Sciences of the United States of America*, 112(30), 9436–9441. <https://doi.org/10.1073/pnas.1508686112>
- Malet, H., Coutard, B., Jamal, S., Dutartre, H., Papageorgiou, N., Neuvonen, M., Ahola, T., Forrester, N., Gould, E. A., Lafitte, D., Ferron, F., Lescar, J., Gorbalenya, A. E., de Lamballerie, X., & Canard, B. (2009). The crystal structures of Chikungunya and Venezuelan equine encephalitis virus nsP3 macro domains define a conserved adenosine binding pocket. *Journal of Virology*, 83(13), 6534–6545. <https://doi.org/10.1128/JVI.00189-09>
- Marques, L. O., Lima, M. S., & Soares, B. G. (2004). Trifluoperazine for schizophrenia. *The Cochrane database of systematic reviews*, 2004(1), CD003545. <https://doi.org/10.1002/14651858.CD003545.pub2>
- Marshak, D. R. (1996). *Strategies for protein purification and characterization : a laboratory course manual*. 396.
- Masaki, T., Suzuki, R., Murakami, K., Aizaki, H., Ishii, K., Murayama, A., Date, T., Matsuura, Y., Miyamura, T., Wakita, T., & Suzuki, T. (2008). Interaction of Hepatitis C Virus Nonstructural Protein 5A with Core Protein Is Critical for the Production of Infectious Virus Particles. *Journal of Virology*, 82(16), 7964. <https://doi.org/10.1128/JVI.00826-08>
- Miao, Z., Xie, Z., Miao, J., Ran, J., Feng, Y., & Xia, X. (2017). Regulated Entry of Hepatitis C Virus into Hepatocytes. *Viruses*, 9(5), 100. <https://doi.org/10.3390/v9050100>
- Mickolajczyk, K. J., Shelton, P., Grasso, M., Cao, X., Warrington, S. E., Aher, A., Liu, S., & Kapoor, T. M. (2021). Force-dependent stimulation of RNA unwinding by SARS-CoV-2 nsp13 helicase. *Biophysical journal*, 120(6), 1020–1030. <https://doi.org/10.1016/j.bpj.2020.11.2276>
- Miknis, Z. J., Donaldson, E. F., Umland, T. C., Rimmer, R. A., Baric, R. S., & Schultz, L. W. (2009). Severe Acute Respiratory Syndrome Coronavirus nsp9 Dimerization Is Essential for Efficient Viral Growth. *Journal of Virology*, 83(7), 3007–3018. <https://doi.org/10.1128/jvi.01505-08>
- Miller, S., & Krijnse-Locker, J. (2008). Modification of intracellular membrane structures for virus replication. *Nature reviews. Microbiology*, 6(5), 363–374. <https://doi.org/10.1038/nrmicro1890>
- Minskaia, E., Hertzog, T., Gorbalenya, A. E., Campanacci, V., Cambillau, C., Canard, B., & Ziebuhr, J. (2006). Discovery of an RNA virus 3'→5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proceedings of the National Academy of Sciences*, 103(13), 5108–5113. <https://doi.org/10.1073/PNAS.0508200103>
- Moradpour, D., & Penin, F. (2013). Hepatitis C virus proteins: from structure to function. *Current Topics in Microbiology and Immunology*, 369, 113–142. https://doi.org/10.1007/978-3-642-27340-7_5

- Morales-Ortega, A., Bernal-Bello, D., Llarena-Barroso, C., Frutos-Pérez, B., Duarte-Millán, M. Á., García de Viedma-García, V., Farfán-Sedano, A. I., Canalejo-Castrillero, E., Ruiz-Giardin, J. M., Ruiz-Ruiz, J., & San Martín-López, J. V. (2020). Imatinib for COVID-19: A case report. *Clinical Immunology* (Orlando, Fla.), 218, 108518. <https://doi.org/10.1016/J.CLIM.2020.108518>
- Nelson, D. L. (2012). Bioenergetics and Biochemical Reaction types. In *Lehninger Principles of Biochemistry*. (6th ed., pp. 505–510). W.H. Freeman.
- Neuman, B. W., Joseph, J. S., Saikatendu, K. S., Serrano, P., Chatterjee, A., Johnson, M. A., Liao, L., Klaus, J. P., Yates, J. R., Wüthrich, K., Stevens, R. C., Buchmeier, M. J., & Kuhn, P. (2008). Proteomics Analysis Unravels the Functional Repertoire of Coronavirus Nonstructural Protein 3. *Journal of Virology*, 82(11), 5279–5294. <https://doi.org/10.1128/jvi.02631-07>
- Ohtaka, H., & Freire, E. (2005). Adaptive inhibitors of the HIV-1 protease. *Progress in biophysics and molecular biology*, 88(2), 193–208. <https://doi.org/10.1016/j.pbiomolbio.2004.07.005>
- Patel, P., & Sengupta, N. (2020). PPIs and Beyond: A Framework for Managing Anticoagulation-Related Gastrointestinal Bleeding in the Era of COVID-19. *Digestive diseases and sciences*, 65(8), 2181–2186. <https://doi.org/10.1007/s10620-020-06408-x>
- Paules, C. I., Marston, H. D., & Fauci, A. S. (2020). Coronavirus Infections-More Than Just the Common Cold. *JAMA - Journal of the American Medical Association*, 323(8), 707–708. <https://doi.org/10.1001/jama.2020.0757>
- Payne, S. (2017). Family Coronaviridae. In S. Payne (Ed.), *Viruses* (pp. 149–158). Academic Press. <https://doi.org/10.1016/b978-0-12-803109-4.00017-9>
- Pehrson, J. R., & Fuji, R. N. (1998). Evolutionary conservation of histone macroH2A subtypes and domains. *Nucleic Acids Research*, 26(12), 2837–2842. <https://doi.org/10.1093/nar/26.12.2837>
- Peng, B., Lloyd, P., & Schran, H. (2005). Clinical pharmacokinetics of imatinib. *Clinical Pharmacokinetics*, 44(9), 879–894. <https://doi.org/10.2165/00003088-200544090-00001>
- Peng, Q., Peng, R., Yuan, B., Zhao, J., Wang, M., Wang, X., Wang, Q., Sun, Y., Fan, Z., Qi, J., Gao, G. F., & Shi, Y. (2020). Structural and Biochemical Characterization of the nsp12-nsp7-nsp8 Core Polymerase Complex from SARS-CoV-2. *Cell reports*, 31(11), 107774. <https://doi.org/10.1016/j.celrep.2020.107774>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004a). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. <https://doi.org/10.1002/JCC.20084>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004b). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. <https://doi.org/10.1002/JCC.20084>

- Piotrowski, Y., Hansen, G., Boomaars-van Der Zanden, A. L., Snijder, E. J., Gorbalenya, A. E., & Hilgenfeld, R. (2009). Crystal structures of the X-domains of a Group-1 and a Group-3 coronavirus reveal that ADP-ribose-binding may not be a conserved property. *Protein Science : A Publication of the Protein Society*, 18(1), 6–16. <https://doi.org/10.1002/PRO.15>
- Piston, D. W., & Kremers, G. J. (2007). Fluorescent protein FRET: the good, the bad and the ugly. *Trends in Biochemical Sciences*, 32(9), 407–414. <https://doi.org/10.1016/J.TIBS.2007.08.003>
- Plant, E. P., Sims, A. C., Baric, R. S., Dinman, J. D., & Taylor, D. R. (2013). Altering SARS coronavirus frameshift efficiency affects genomic and subgenomic RNA production. *Viruses*, 5(1), 279–294. <https://doi.org/10.3390/v5010279>
- Pontremoli, R., Leoncini, G., & Parodi, A. (2005). Use of nifedipine in the treatment of hypertension. Expert review of cardiovascular therapy, 3(1), 43–50. <https://doi.org/10.1586/14779072.3.1.43>
- Rack, J. G. M., Perina, D., & Ahel, I. (2016). Macrodomains: Structure, Function, Evolution, and Catalytic Activities. *Annual Review of Biochemistry*, 85, 431–454. <https://doi.org/10.1146/ANNUREV-BIOCHEM-060815-014935>
- Ràfols, C., Bosch, E., Barbas, R., & Prohens, R. (2016). The Ca²⁺–EDTA chelation as standard reaction to validate Isothermal Titration Calorimeter measurements (ITC). *Talanta*, 154. <https://doi.org/10.1016/j.talanta.2016.03.075>
- Raicu, V. (2007). Efficiency of resonance energy transfer in homo-oligomeric complexes of proteins. *Journal of Biological Physics*, 33(2), 109–127. <https://doi.org/10.1007/S10867-007-9046-Z>
- Raicu, V., & Singh, D. R. (2013). FRET Spectrometry: A New Tool for the Determination of Protein Quaternary Structure in Living Cells. *Biophysical Journal*, 105(9), 1937–1945. <https://doi.org/10.1016/J.BPJ.2013.09.015>
- Raicu, V., Stoneman, M. R., Fung, R., Melnichuk, M., Jansma, D. B., Pisterzi, L. F., Rath, S., Fox, M., Wells, J. W., & Saldin, D. K. (2009). Determination of supramolecular structure and spatial distribution of protein complexes in living cells. *Nature Photonics* 2008 3:2, 3(2), 107–113. <https://doi.org/10.1038/nphoton.2008.291>
- Raney, K. D., Byrd, A. K., & Aarattuthodiyil, S. (2013). Structure and Mechanisms of SF1 DNA Helicases. In Spies, M (Ed.), *Advances in experimental medicine and biology* (Vol. 973, p. E1). https://doi.org/10.1007/978-1-4614-5037-5_14
- Raney, K. D., Sharma, S. D., Moustafa, I. M., & Cameron, C. E. (2010). Hepatitis C Virus Non-structural Protein 3 (HCV NS3): A Multifunctional Antiviral Target. *Journal of Biological Chemistry*, 285(30), 22725–22731. <https://doi.org/10.1074/JBC.R110.125294>
- Ricagno, S., Egloff, M. P., Ulferts, R., Coutard, B., Nurizzo, D., Campanacci, V., Cambillau, C., Ziebuhr, J., & Canard, B. (2006). Crystal structure and mechanistic determinants of SARS coronavirus

- nonstructural protein 15 define an endoribonuclease family. *Proceedings of the National Academy of Sciences*, 103(32), 11892–11897. <https://doi.org/10.1073/PNAS.0601708103>
- RM, C. (1995). Fluorescence resonance energy transfer. *Current Opinion in Biotechnology*, 6(1), 103–110. [https://doi.org/10.1016/0958-1669\(95\)80016-6](https://doi.org/10.1016/0958-1669(95)80016-6)
- Rothlin, R. P., Vetulli, H. M., Duarte, M., & Pelorosso, F. G. (2020). Telmisartan as tentative angiotensin receptor blocker therapeutic for COVID-19. *Drug Development Research*, 81(7), 768–770. <https://doi.org/10.1002/DDR.21679>
- Ruben, A. J., Kiso, Y., & Freire, E. (2006). Overcoming roadblocks in lead optimization: a thermodynamic perspective. *Chemical biology & drug design*, 67(1), 2–4. <https://doi.org/10.1111/j.1747-0285.2005.00314.x>
- Saikatendu, K. S., Joseph, J. S., Subramanian, V., Clayton, T., Griffith, M., Moy, K., Velasquez, J., Neuman, B. W., Buchmeier, M. J., Stevens, R. C., & Kuhn, P. (2005). Structural basis of severe acute respiratory syndrome coronavirus ADP-ribose-1"-phosphate dephosphorylation by a conserved domain of nsP3. *Structure*, 13(11), 1665–1675. <https://doi.org/10.1016/j.str.2005.07.022>
- Sauvage, E., & Terrak, M. (2016). Glycosyltransferases and Transpeptidases/Penicillin-Binding Proteins: Valuable Targets for New Antibacterials. *Antibiotics (Basel, Switzerland)*, 5(1), 12. <https://doi.org/10.3390/antibiotics5010012>
- Sawicki, S. G., Sawicki, D. L., & Siddell, S. G. (2007). A Contemporary View of Coronavirus Transcription. *Journal of Virology*, 81(1), 20–29. <https://doi.org/10.1128/jvi.01358-06>
- Schregel, V., Jacobi, S., Penin, F., & Tautz, N. (2009). Hepatitis C virus NS2 is a protease stimulated by cofactor domains in NS3. *Proceedings of the National Academy of Sciences*, 106(13), 5342–5347. <https://doi.org/10.1073/PNAS.0810950106>
- Schubert, K., Karousis, E. D., Jomaa, A., Scaiola, A., Echeverria, B., Gurzeler, L. A., Leibundgut, M., Thiel, V., Mühlemann, O., & Ban, N. (2020). SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nature Structural and Molecular Biology*, 27(10), 959–966. <https://doi.org/10.1038/s41594-020-0511-8>
- Serrano, P., Johnson, M. A., Chatterjee, A., Neuman, B. W., Joseph, J. S., Buchmeier, M. J., Kuhn, P., & Wüthrich, K. (2009). Nuclear Magnetic Resonance Structure of the Nucleic Acid-Binding Domain of Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein 3. *Journal of Virology*, 83(24), 12998–13008. <https://doi.org/10.1128/jvi.01253-09>
- Seybert, A., Hegyi, A., Siddell, S. G., & Ziebuhr, J. (2000). The human coronavirus 229E superfamily 1 helicase has RNA and DNA duplex-unwinding activities with 5'-to-3' polarity. *RNA (New York, N.Y.)*, 6(7), 1056–1068. <https://doi.org/10.1017/s1355838200000728>

- Seybert, A., Posthuma, C. C., van Dinten, L. C., Snijder, E. J., Gorbalenya, A. E., & Ziebuhr, J. (2005). A Complex Zinc Finger Controls the Enzymatic Activities of Nidovirus Helicases. *Journal of Virology*, 79(2), 696–704. <https://doi.org/10.1128/JVI.79.2.696-704.2005>
- Shalaby, M.-A. W., Dokla, E. M. E., Serya, Rabah. A. T., & Abouzid, K. A. M. (2020). Penicillin binding protein 2a: An overview and a medicinal chemistry perspective. *European Journal of Medicinal Chemistry*, 199. <https://doi.org/10.1016/j.ejmech.2020.112312>
- Shi, W., Chen, M., Yang, Y., Zhou, W., Chen, S., Yang, Y., Hu, Y., & Liu, B. (2020). Title: A dynamic regulatory interface on SARS-CoV-2 RNA polymerase. *BioRxiv*, 2020.07.30.229187. <https://doi.org/10.1101/2020.07.30.229187>
- Shimizu, J. F., Martins, D. O. S., McPhillie, M. J., Roberts, G. C., Zothner, C., Merits, A., Harris, M., & Jardim, A. C. G. (2020). Is the ADP ribose site of the Chikungunya virus NSP3 Macro domain a target for antiviral approaches? *Acta Tropica*, 207. <https://doi.org/10.1016/j.actatropica.2020.105490>
- Shultz M. D. (2013). The thermodynamic basis for the use of lipophilic efficiency (LipE) in enthalpic optimizations. *Bioorganic & medicinal chemistry letters*, 23(21), 5992–6000. <https://doi.org/10.1016/j.bmcl.2013.08.030>
- Simister, P., Schmitt, M., Geitmann, M., Wicht, O., Danielson, U. H., Klein, R., Bressanelli, S., & Lohmann, V. (2009). Structural and Functional Analysis of Hepatitis C Virus Strain JFH1 Polymerase. *Journal of Virology*, 83(22), 11926. <https://doi.org/10.1128/JVI.01008-09>
- Singleton, M. R., Dillingham, M. S., & Wigley, D. B. (2007). Structure and mechanism of helicases and nucleic acid translocases. *Annual Review of Biochemistry*, 76, 23–50. <https://doi.org/10.1146/ANNUREV.BIOCHEM.76.052305.115300>
- Siu, Y. L., Teoh, K. T., Lo, J., Chan, C. M., Kien, F., Escriou, N., Tsao, S. W., Nicholls, J. M., Altmeyer, R., Peiris, J. S. M., Bruzzone, R., & Nal, B. (2008). The M, E, and N Structural Proteins of the Severe Acute Respiratory Syndrome Coronavirus Are Required for Efficient Assembly, Trafficking, and Release of Virus-Like Particles. *Journal of Virology*, 82(22), 11318–11330. <https://doi.org/10.1128/jvi.01052-08>
- Skoog, D. A., Holler, F. J., & Crouch, S. R. (2007). An Introduction to Chromatographic Separations. In *Principles of Instrumental Analysis*. (Sixth ed., pp. 762–780). Thomson Brooks/Cole.
- Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L. M., Guan, Y., Rozanov, M., Spaan, W. J. M., & Gorbalenya, A. E. (2003a). Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *Journal of Molecular Biology*, 331(5), 991–1004. [https://doi.org/10.1016/S0022-2836\(03\)00865-9](https://doi.org/10.1016/S0022-2836(03)00865-9)
- Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L. M., Guan, Y., Rozanov, M., Spaan, W. J. M., & Gorbalenya, A. E. (2003b). Unique and conserved features of genome and

- proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *Journal of Molecular Biology*, 331(5), 991–1004. [https://doi.org/10.1016/S0022-2836\(03\)00865-9](https://doi.org/10.1016/S0022-2836(03)00865-9)
- Snijder, E. J., Limpens, R., de Wilde, A. H., de Jong, A., Zevenhoven-Dobbe, J. C., Maier, H. J., Faas, F., Koster, A. J., & Bárcena, M. (2020). A unifying structural and functional model of the coronavirus replication organelle: Tracking down RNA synthesis. *PLoS biology*, 18(6), e3000715. <https://doi.org/10.1371/journal.pbio.3000715>
- Snijder, E. J., van der Meer, Y., Zevenhoven-Dobbe, J., Onderwater, J. J. M., van der Meulen, J., Koerten, H. K., & Mommaas, A. M. (2006). Ultrastructure and Origin of Membrane Vesicles Associated with the Severe Acute Respiratory Syndrome Coronavirus Replication Complex. *Journal of Virology*, 80(12), 5927–5940. <https://doi.org/10.1128/jvi.02501-05>
- Sola, I., Almazán, F., Zúñiga, S., & Enjuanes, L. (2015). Continuous and Discontinuous RNA Synthesis in Coronaviruses. In Enquist, LW (Ed.), *Annual Review of Virology* (Vol. 2, pp. 265–288). <https://doi.org/10.1146/annurev-virology-100114-055218>
- Stobart, C. C., Sexton, N. R., Munjal, H., Lu, X., Molland, K. L., Tomar, S., Mesecar, A. D., & Denison, M. R. (2013). Chimeric Exchange of Coronavirus nsp5 Proteases (3CLpro) Identifies Common and Divergent Regulatory Determinants of Protease Activity. *Journal of Virology*, 87(23), 12611–12618. <https://doi.org/10.1128/jvi.02050-13>
- Stoeck, I. K., Lee, J. Y., Tabata, K., Romero-Brey, I., Paul, D., Schult, P., Lohmann, V., Kaderali, L., & Bartenschlager, R. (2017). Hepatitis C Virus Replication Depends on Endosomal Cholesterol Homeostasis. *Journal of virology*, 92(1), e01196-17. <https://doi.org/10.1128/JVI.01196-17>
- Subissi, L., Posthuma, C. C., Collet, A., Zevenhoven-Dobbe, J. C., Gorbalenya, A. E., Decroly, E., Snijder, E. J., Canard, B., & Imbert, I. (2014). One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proceedings of the National Academy of Sciences of the United States of America*, 111(37), E3900–E3909. <https://doi.org/10.1073/pnas.1323705111>
- Suzich, J. A., Tamura, J. K., Palmer-Hill, F., Warrenner, P., Grakoui, A., Rice, C. M., Feinstone, S. M., & Collett, M. S. (1993a). Hepatitis C virus NS3 protein polynucleotide-stimulated nucleoside triphosphatase and comparison with the related pestivirus and flavivirus enzymes. *Journal of Virology*, 67(10), 6152–6158. <https://doi.org/10.1128/JVI.67.10.6152-6158.1993>
- Suzich, J. A., Tamura, J. K., Palmer-Hill, F., Warrenner, P., Grakoui, A., Rice, C. M., Feinstone, S. M., & Collett, M. S. (1993b). Hepatitis C virus NS3 protein polynucleotide-stimulated nucleoside triphosphatase and comparison with the related pestivirus and flavivirus enzymes. *Journal of Virology*, 67(10), 6152–6158. <https://doi.org/10.1128/JVI.67.10.6152-6158.1993>
- Tan, J., Kusov, Y., Mutschall, D., Tech, S., Nagarajan, K., Hilgenfeld, R., & Schmidt, C. L. (2007). The "SARS-unique domain" (SUD) of SARS coronavirus is an oligo(G)-binding protein. *Biochemical and biophysical research communications*, 364(4), 877–882. <https://doi.org/10.1016/j.bbrc.2007.10.081>

- Tan, J., Vonnrhein, C., Smart, O. S., Bricogne, G., Bollati, M., Kusov, Y., Hansen, G., Mesters, J. R., Schmidt, C. L., & Hilgenfeld, R. (2009). The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes. *PLoS pathogens*, 5(5), e1000428. <https://doi.org/10.1371/journal.ppat.1000428>
- Te Velhuis, A. J. W., van den Worm, S. H. E., & Snijder, E. J. (2012). The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. *Nucleic Acids Research*, 40(4), 1737–1747. <https://doi.org/10.1093/nar/gkr893>
- Ulasli, M., Verheije, M. H., de Haan, C. A. M., & Reggiori, F. (2010). Qualitative and quantitative ultrastructural analysis of the membrane rearrangements induced by coronavirus. *Cellular Microbiology*, 12(6), 844–861. <https://doi.org/10.1111/j.1462-5822.2010.01437.x>
- Van Hemert, M. J., van den Worm, S. H., Knoop, K., Mommaas, A. M., Gorbalenya, A. E., & Snijder, E. J. (2008). SARS-coronavirus replication/transcription complexes are membrane-protected and need a host factor for activity in vitro. *PLoS pathogens*, 4(5), e1000054. <https://doi.org/10.1371/journal.ppat.1000054>
- Van Oss, C. J., Absolom, D. R., & Neumann, A. W. (1980). The “hydrophobic effect”: Essentially a van der Waals interaction. *Colloid and Polymer Science* 1980 258:4, 258(4), 424–427. <https://doi.org/10.1007/BF01480835>
- Velázquez-Campoy, A., Ohtaka, H., Nezami, A., Muzammil, S., & Freire, E. (2004). Isothermal titration calorimetry. *Current protocols in cell biology*, Chapter 17, . <https://doi.org/10.1002/0471143030.cb1708s23>
- Voet, D., Voet, J. G., & Pratt, C. W. (2016). Introduction to the chemistry of life. In *Fundamentals of Biochemistry: Life at the Molecular Level*. (5th ed., pp. 11–18). Wiley.
- Voisset, C., & Dubuisson, J. (2004). Functional hepatitis C virus envelope glycoproteins. *Biology of the Cell*, 96(6), 413. <https://doi.org/10.1016/J.BIOLCEL.2004.03.008>
- Ward, R. M., & Kearns, G. L. (2013). Proton pump inhibitors in pediatrics : mechanism of action, pharmacokinetics, pharmacogenetics, and pharmacodynamics. *Paediatric drugs*, 15(2), 119–131. <https://doi.org/10.1007/s40272-013-0012-x>
- Wazir, S., Maksimainen, M. M., Alanen, H. I., Galera-Prat, A., & Lehtiö, L. (2021). Activity-Based Screening Assay for Mono-ADP-Ribosylhydrolases. *SLAS Discovery*, 26(1), 67–76. <https://doi.org/10.1177/2472555220928911>
- Weber, R., & McCullagh, M. (2021). Role of ATP in the RNA Translocation Mechanism of SARS-CoV-2 NSP13 Helicase. *Journal of Physical Chemistry B*, 125(31), 8787–8796. <https://doi.org/10.1021/ACS.JPCB.1C04528>

- Willig, J. B., Vianna, D. R. B., Beckenkamp, A., Beckenkamp, L. R., Sévigny, J., Wink, M. R., Buffon, A., & Pilger, D. A. (2020). Imatinib mesylate affects extracellular ATP catabolism and expression of NTPDases in a chronic myeloid leukemia cell line. *Purinergic Signalling*, 16(1), 29. <https://doi.org/10.1007/S11302-019-09686-X>
- Woo, P. C. Y., Huang, Y., Lau, S. K. P., & Yuen, K. Y. (2010). Coronavirus genomics and bioinformatics analysis. *Viruses*, 2(8), 1805–1820. <https://doi.org/10.3390/v2081803>
- Woo, P. C. Y., Lau, S. K. P., Lam, C. S. F., Lau, C. C. Y., Tsang, A. K. L., Lau, J. H. N., Bai, R., Teng, J. L. L., Tsang, C. C. C., Wang, M., Zheng, B.-J., Chan, K.-H., & Yuen, K.-Y. (2012). Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene Source of Gammacoronavirus and Deltacoronavi. *Journal of Virology*, 86(7), 3995–4008. <https://doi.org/10.1128/jvi.06540-11>
- Xie, P. (2016). Dynamics of monomeric and hexameric helicases. *Biophysical Chemistry*, 211, 49–58. <https://doi.org/10.1016/J.BPC.2016.02.003>
- Xu, X., Liu, Y., Weiss, S., Arnold, E., Sarafianos, S. G., & Ding, J. (2003). Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design. *Nucleic Acids Research*, 31(24), 7117–7130. <https://doi.org/10.1093/NAR/GKG916>
- Yi, M., & Lemon, S. M. (2003). Structure-function analysis of the 3' stem-loop of hepatitis C virus genomic RNA and its role in viral RNA replication. *RNA (New York, N.Y.)*, 9(3), 331–345. <https://doi.org/10.1261/RNA.2144203>
- Zeisel, M. B., Felmlee, D. J., & Baumert, T. F. (2013). Hepatitis C virus entry. *Current Topics in Microbiology and Immunology*, 369, 87–112. https://doi.org/10.1007/978-3-642-27340-7_4
- Zhao H, Mendenhall M, Deininger MW. Imatinib is not a potent anti-SARS-CoV-2 drug. *Leukemia*. 2020;34(11):3085-3087. <https://doi:10.1038/s41375-020-01045-9>
- Ziebuhr, J., Snijder, E. J., & Gorbalenya, A. E. (2000). Virus-encoded proteinases and proteolytic processing in the Nidovirales. *Journal of General Virology*, 81(4), 853–879. <https://doi.org/10.1099/0022-1317-81-4-853>

10.CURRICULUM VITAE

Rajdeep S. Virdi.

Education

University of Wisconsin – Milwaukee

Ph.D. in Chemistry and Biochemistry

Fall 2016- Fall 2021

Areas of study: protein purification, cloning, high throughput screening (HTS), ATPase assays, binding assays for protein-DNA, protein-RNA, and protein-ligand, polarization assays, thermodynamics, kinetics, FRET, and PIFE studies.

University of Wisconsin – Milwaukee

Bachelors in Chemistry and Biochemistry

Fall 2014 – Spring 2016

Transfer student from University of Wisconsin -Madison.

Published work:

Virdi, R. S., Bavisotto, R. V., Hopper, N. C., Vuksanovic, N., Melkonian, T. R., Silvaggi, N. R., & Frick, D. N. (2020). Discovery of Drug-Like Ligands for the Mac1 Domain of SARS-CoV-2 Nsp3. *SLAS DISCOVERY: Advancing the Science of Drug Discovery*, 25(10), 1162–1170. <https://doi.org/10.1177/2472555220960428>

Frick, D. N., **Virdi, R. S.**, Vuksanovic, N., Dahal, N., & Silvaggi, N. R. (2020). Molecular Basis for ADP-Ribose Binding to the Mac1 Domain of SARS-CoV-2 nsp3. *Biochemistry*, 59(28), 2608–2615. <https://doi.org/10.1021/acs.biochem.0c00309>

Dockendorff, C., Wiedenhoef, D., Benoit, A., Porter, J., Wu, Y., **Virdi, R.**, & Shanaa, A. (2016). Design and Synthesis of Oxazoline-Based Scaffolds for Hybrid Lewis Acid/Lewis Base Catalysis of Carbon–Carbon Bond Formation. *Synthesis*, 48(15), 2413–2422. <https://doi.org/10.1055/s-0035-1560436>

Research Experience

University of Wisconsin- Milwaukee (Spring 2017- Fall 2021)

Graduate student under the supervision of Prof. David Frick.

-----2021

- Transformed, expressed, and purified (NiNTA 10 ml column and size exclusion 250 ml column)
 - SARS-CoV2 helicase in IPTG inducible NOVA blue(DE3), Rosetta(DE3) and BL21(DE3).
 - The SARS-CoV-2 helicase was assayed via SDS polyacrylamide gel and western blot.
- ATPase assay

- Malachite green assay to examine the ATPase activity of SARS-CoV2 helicase and HCV helicase in the presence and absence of DNA or RNA.
- Characterized the kinetic parameters (K_m , V_{max} and K_{cat}) of SARS-CoV-2 helicase.
- ADP-Glo assay to examine the ATPase activity of SARS-CoV2 helicase.
- *In collaboration with Prof. Hossain Group.*
 - Binding Assay:
 - HDAC glow kit was used to screen ~20 potential inhibitors of HDAC 1, 2, 3, 4, 6, and 8.
 - Upon reduced activity, additional experiments were conducted to measure IC50's with selected potential inhibitors.

2020

- Kinetic assays
 - Examine the activity of LDH from rabbit muscle and bovine heart muscle by measuring the decrease in NADH concentration at 340 nm.
- ATPase assay
 - ADP-Glo kinase assay kit was used to examine the activity of three-part CMV Helicase and examine which combination and concentration of three is most active.
 - Titer Glo assay kit was used to examine the activity of SARS-CoV2 helicase and HCV helicase.
- Expressed and purified (NiNTA 10ml column and Size exclusion 250ml column)
 - SARS-CoV2 helicase, protease and Mac1.
- Transformation
 - IPTG inducible BL21(DE3) with SARS-CoV2 helicase and protease.
- Binding Assays
 - Isothermal titration calorimetry to estimate the complete thermodynamic profile of ADP-ribose and similar nucleotides binding to Mac1 of SARS-CoV2.
 - Isothermal titration calorimetry to estimate the complete thermodynamic profile of select drugs and drug like compounds binding to Mac1 of SARS-CoV2.
 - Thermofluor assay to determine IC50 for Mac1 with ADP-ribose.
 - Thermofluor assay to screen library of 2,500 compounds with Mac1.
- Activity based screening assay
 - Examine the activity of Mac1 in the presence of alpha-NAD. (Wazir et al., 2021)
- *In collaboration with Prof. Sobolev group*
 - Designed experiments to test and quantify the antibacterial properties of UV activated surface.
 - Bacterial plaque assays

2019

- Cloning (Insert digestion, purification, Ligation), expression and transformation of Halo-tag NS3h.

- Design a plasmid that can be used in transforming BL21(DE3) cells that encodes for N-terminal Halo Tag, with TEV protease cleavage site, helicase and His tag at the C terminal (HaloTag-Tev-NS3h-Histag).
- Binding assay
 - PIFE and/or Polarization binding assay with Nud3, and F18, F38
 - PIFE and/or Polarization binding assay with Nud12 and F18
 - PIFE and/or Polarization binding assay with Nudc and F18
- Kinetic Assay
 - Examine the activity of Nud12 by measuring the change in concentration of NADH at 340 nm.
- *In collaboration with Prof. Popa group*
 - Magnetic tweezer assay
 - Examine helicase unwinding activity using partially dsDNA (42nt ssDNA with 356nt long dsDNA) in the presence of ATP.

2018

- Learned how to transfect, grow and store HEK cells with the help of another graduate student.
- Binding assay
 - PIFE assays on DNA B with 18 nucleotides long Hexachlorofluorescein labeled oligonucleotide.
 - FRET assay using with CFP-NS3h with labeled Oligonucleotide
 - Binding and FRET between CFP, YFP and MCH labeled NS3h.
- ATPase assay
 - With DnaB helicase from bacteria

2017

- Expressed and Purified (NiNTA 10ml column and Size exclusion 250ml column)
 - Arabinose induced DH5 alpha cells for Venus
 - Arabinose induced DH5 alpha cells for Sapphire
 - IPTG induced GFP in Rosetta cells.
 - IPTG induced GFP-MAVS in Rosetta cells.
 - IPTG induced YFP-NS3h in BL21 cells
 - Protein concentration was determined via Bradford assay and using Beers law.
- Binding Assay
 - FRET assays to examine binding of YFP-NS3h with labeled oligonucleotides with different fluorophores.
 - Native gel shift assay using different percentage polyacrylamide to examine binding between YFP-NS3h and oligonucleotide.
 - PIFE assay to examine the binding of NS3h with labeled oligonucleotides
 - FRET assay to examine binding of YFP-NS3h with labeled oligonucleotides under different pH conditions.
- *In collaboration with Prof. Raicu group*

- Use of multiphoton FRET microscopy to examine the binding of helicase with oligonucleotide using gel samples cut out from Native gel shift assay.

Eskra Technical Products

Summer and fall 2016

Research assistant under the supervision of Mike Eskra.

- Test, synthesize and paste various composition of graphite material for cathode to design a solid-state Lithium-Ion battery.
- Assemble lithium-ion Battery in the vacuum hood.
- Test and synthesize wet acid batteries.

University of Wisconsin- Milwaukee

Spring 2016

Research assistant under the supervision of Prof. Graham Moran.

Most experiments were repeat of experiments done by graduate students in the same lab.

- Protein prep: extraction and purification of protein from BL21(DE3) using Bio Rad BioLogic LP System with NiNTA, size exclusion and ion exchange column.
- Detect the change in the concentration of substrate using HPLC in the presence and absence of Enzyme.
- Detect the binding of LDH and MDH with 6-NADH by measuring a shift in absorbance from 340nm to 360nm upon binding.
- Also tried crystalizing proteins.

Marquette University

Summer 2014- Fall 2015

Internship under the supervision of Prof. Chris Dockendorff.

- Purpose: Synthesis of primary amine-oxazole oxazoline precatalyst that can be used to preform Syn-aldol chemistry.
- Main Techniques used:
 - Thin Layer Chromatography to monitor the progress of a chemical reaction.
 - Biotage Isolera One flash chromatograph, to separate compound of interest from impurities. Uses UV 254 and 210 nm detector SNAP 50g column 50ml/min.
 - Shimadzu LC-MS with photodiode array, and dual ESI/DUIS MS probe used to identify the purity of the compound.
 - Varian 300MHz and 400MHz NMR and Thermo/Nicolet IR spectrometer used to identify compound.

Specialized molecular biology and microbiology techniques

PCR, qRT-PCR, Western Blot, Molecular Cloning, Plasmid Extraction, Cell Transformation, RNA Extraction, Protein Purification (Sonication, Ammonium Sulfate Precipitation, Dialysis, Affinity Purification, Size-Exclusion), Chromatographic Separation (FPLC, HPLC), SDS-PAGE, Nucleic Acid Electrophoresis, Enzyme Activity Measurement and Kinetic Characterization, FRET binding assay, Protein polarization assay, Bradford Assay, Preparation of E.coli DH5 α competent cells, Fluorescence Microscopy Technology, Basic Fermentation Technique, etc.

Biochemistry relevant teaching experience

Chem 603

Spring 2021 - Fall 2021

Assist and teach students topics in biochemistry: PCR, restriction enzyme digests and ligation, E.coli transformation, protein expression and purification, in-vitro fluorescent enzyme assays, In vitro activity assays including characterization of the kinetic parameters (K_m , V_{max} and K_{cat}) for HCV and EcGR.