

May 2021

Structure & Function of Enzymes in Two Uncharacterized Gene Clusters from *Pseudomonas Brassicacearum* & *Streptomyces Griseofuscus*

Lamia Tabassum Badhon
University of Wisconsin-Milwaukee

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



Part of the [Biochemistry Commons](#), [Chemistry Commons](#), and the [Molecular Biology Commons](#)

Recommended Citation

Badhon, Lamia Tabassum, "Structure & Function of Enzymes in Two Uncharacterized Gene Clusters from *Pseudomonas Brassicacearum* & *Streptomyces Griseofuscus*" (2021). *Theses and Dissertations*. 2643.
<https://dc.uwm.edu/etd/2643>

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.

STRUCTURE & FUNCTION OF ENZYMES IN TWO UNCHARACTERIZED GENE CLUSTERS FROM

Pseudomonas brassicacearum* & *Streptomyces griseofuscus

by

Lamia Tabassum Badhon

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

Doctor of Philosophy
in Chemistry

at

The University of Wisconsin-Milwaukee

May 2021

ABSTRACT

STRUCTURE & FUNCTION OF ENZYMES IN TWO UNCHARACTERIZED GENE CLUSTERS FROM *Pseudomonas brassicacearum* & *Streptomyces griseofuscus*

by

Lamia Tabassum Badhon

The University of Wisconsin-Milwaukee, 2021
Under the Supervision of Professor Nicholas Silvaggi, PhD

Pyridoxal 5'-phosphate (PLP)-dependent enzymes harness this versatile cofactor to catalyze a variety of reactions including transamination, decarboxylation, racemization and various elimination/substitution reactions. Several years ago, a new class of PLP-dependent enzymes was discovered that uses PLP and molecular oxygen to catalyze the 4-electron oxidation of L-arginine to 4-hydroxy-2-ketoarginine. Work with the prototypical enzyme of this class, MppP from *Streptomyces wadayamensis* (SwMppP), showed that the dioxygen consumed during the reaction is reduced to hydrogen peroxide, and that the hydroxyl group installed in the product derives from water. Thus, SwMppP is an L-arginine oxidase, and not an oxygenase. This was surprising given that the hydroxylation step occurs at an un-activated methylene group of the substrate, L-arginine. SwMppP is part of the mannopeptimycin biosynthetic cluster and works together with MppR and MppQ to synthesize the non-proteinogenic amino acid L-enduracididine (L-End). A BLASTp search of the NR database using the sequence of *S. wadayamensis* MppP revealed close homologs in a number of species of pseudomonads, including *Pseudomonas brassicacearum*, *P. syringae*, *P. amygdali*, and *P. aeruginosa*.

The first section of this thesis is focused on the genomic context of the *pseudomonad* MppP homologs. None of the *pseudomonad* genomes examined contained other key mannopeptimycin biosynthetic genes like the nonribosomal peptide synthetases *mppA* and *mppB*, or the other two L-End biosynthetic genes *mppQ* and *mppR*. The lack of MppQ and MppR homologs in these organisms suggests that the arginine oxidase activity of MppP homologs is used in a distinct biochemical context. We cloned, expressed, and purified the MppP homolog from *Pseudomonas brassicacearum*, a soil-dwelling bacterium associated with the roots of plants in the mustard/cabbage family that protects the plants from several bacterial and fungal pathogens. The genomic context of the gene encoding this MppP homolog (PbrMppP) consists of a regulatory gene, the MppP homolog, and four additional open reading frames annotated as hypothetical protein (PbrHYP), dihydrodipicolinate synthase family protein (PbrDHPS), Mononuclear Fe-dependent oxygenase; (PbrOX), and 2Fe2S-binding protein ferredoxin (PbrFD). All these genes were annotated from sequence analysis. The observations that (1) the genes are overlapping, (2) that the open reading frame includes a *luxR*-type regulatory gene, and (3) 5'-untranslated region contains a putative promoter sequence suggests that these genes may constitute an operon, which we will refer to as the *P. brassicacearum* MppP containing operon (PbrMPCO). Understanding the biochemical context of PbrMppP requires assigning functions for the other 4 non-regulatory gene products. To begin, we confirmed that the structure and activity of PbrMppP are both identical to the prototypical SwMppP. Next, we analyzed the X-ray crystal structures of PbrHYP, PbrDHPS and PbrOx. We also identified condition that activates the promoter of this gene cluster. Since the structures were not sufficient to deduce the functions of these enzymes, we turned to a metabolomic strategy to determine the final product of the

operon. Knowing the final product would limit the possibilities for substrate of these putative enzymes, we show here a comparative metabolomic analysis by knocking out the entire operon and searching for “missing” metabolite(s) by multiple MS.

The second section focuses on a second MppP homolog from *Streptomyces griseofuscus*. The gene encoding this MppP homolog, SgrMppP, is flanked by 5 other genes whose products are annotated as a putative biotin carboxylase (SgrLIG), flavin-dependent monooxygenase (SgrOX), S-adenosyl methionine (SAM)-dependent methyltransferase (SgrMT), amidinohydrolase (SgrAH), and GNAT-family acetyltransferase (SgrNAT). The goal of these studies is to determine the activities of all the enzymes in this gene cluster and to identify the final product. So far I have shown, in the presence of Mg(II) and SAM, SgrMT catalyzes the transfer of a methyl group from SAM to carbon 4 of the 4-hydroxy-2-ketoarginine produced by SgrMppP. We also determined that SgrMT products are used by SgrAH as substrate and the turned over products by SgrAH are ornithine derivatives and urea. Succeeding research will involve structural and functional determination of rest of the gene products, which will eventually lead to the identification of final product of these gene-contexts.

TABLE OF CONTENTS

Abstract.....	ii
LIST OF FIGURES.....	ix
LIST OF TABLES.....	xiii
LIST OF SCHEMES.....	xv
LIST OF ABBREVIATIONS.....	xvii
ACKNOWLEDGEMENTS.....	xxi
LIST OF FIGURES.....	x
LIST OF TABLES.....	xvi
LIST OF SCHEMES.....	xviii
LIST OF ABBREVIATIONS.....	xx
ACKNOWLEDGEMENTS.....	xxiv
Chapter 1.....	1
Introduction to natural products biosynthesis.....	1
1.1 Different types of natural products.....	3
1.1.1 Alkaloids.....	4
1.1.2 Phenylpropanoids.....	6
1.1.3 Polyketides.....	8
1.1.4 Terpenoids.....	9
1.1.5 Non proteinogenic amino acids in natural products.....	11
1.1.6 Lantibiotics.....	17

1.1.7 Non ribosomal peptide and their biosynthesis	22
1.1.8 L-Enduracididine as a natural product precursor	25
1.2 A novel glycopeptide antibiotic “mannopeptimycin”	28
1.3 PLP dependent enzymes.....	32
1.3.1 PLP-dependent transaminase.....	33
1.3.2 PLP-dependent decarboxylase.....	33
1.3.3 PLP-dependent deaminase	34
1.3.4 PLP-dependent racemase	35
1.3.5 MppP: A unique PLP-dependent L-arginine oxidase	37
Chapter 2.....	41
Beyond Enduracididine Biosynthesis I: Exploring the biochemical context of MppP activity in a putative operon in <i>Pseudomonas brassicaceum</i>	41
2.1 Background	41
2.2 Methodology.....	53
2.2.1 Cloning, expression and purification of the PbrMPCO Enzymes.....	53
2.2.2 Differential scanning fluorimetry.....	55
2.2.3 Fe ^{II} loading on PbrOx	57
2.2.4 Confirmation of PbrMppP activity	57
2.2.5 NMR characterization of PbrMppP reaction products	57
2.2.6 Characterization of PbrMppP reaction products by mass spectrometry.....	58
2.2.7 Crystallization and structure determination of Pbr proteins	59
2.2.8 Steady state kinetics of PbrMppP.....	63
2.2.9 Stoichiometric analysis of PbrMppP reactions with L-Arg and L-canavanine	64
2.2.10 Pre-steady state kinetics of PbrMppP	64
2.2.11 Substrate specificity of PbrMppP	66

2.2.12 Ferredoxin NADP reductase (PbrFNR) activity test	67
2.2.13 Identifying conditions that activate the PbrMppP promoter (RS05025)	68
2.2.14 Metabolite extraction for LCMS analysis	73
2.3 Results	74
2.3.1 PbrMppP oxidizes L-Arg	74
2.3.2 H ₂ O ₂ is a product of PbrMppP catalysis	81
2.3.3 Steady state kinetics of PbrMppP	83
2.3.4 Pre-steady state kinetics of PbrMppP	85
2.3.5 PbrMppP substrate specificity	105
2.3.6 Steady state kinetics of the reaction of PbrMppP with L-canavanine	116
2.3.7 Structural characterization of PbrHYP	118
2.3.7.1 Introduction to PIN domain	118
2.3.7.2 PbrHYP is a PIN domain homolog	119
2.3.7.3 PbrHYP has no metal-binding site	121
2.3.7.4 Absence of an anti-toxin component near the PbrHYP gene	125
2.3.8 Structure of PbrDHPS	126
2.3.8.1 Type I aldolases	127
2.3.8.2 PbrDHPS is likely a Type I aldolase	128
2.3.8.3 The putative catalytic lysine residue is active in PbrDHPS	129
2.3.8.4 Lys160 forms a Schiff base with a ketoacid	132
2.3.9 Structure of PbrOX, a putative nonheme Fe(II)-dependent oxygenase	133
2.3.9.1 Introduction to mononuclear non-heme iron oxygenase	134
2.3.9.2 PbrOX is a mononuclear iron oxygenase homolog	135
2.3.9.3 Fe ^{II} Loading of PbrOX	137
2.3.9.4 Introduction to bacterial three component system	139
2.3.9.5 PbrFD is a Fe-S cluster protein	139
2.3.9.6 PbrFNR is active in the absence of any specific substrate	141

2.3.9.7 Electrons flow from NADPH towards PbrOX.....	142
2.3.9.8 Holo PbrOX is required for electron flow	144
2.3.10.1 PbrMppP operon genes are expressed only in solid culture	145
2.3.10.2 Preliminary “comparative metabolomics” of DF41 cells by LC-MS.....	148
2.4 Discussion.....	155
Chapter 3.....	156
Structure and Function of the Enzymes Expressed by the MppP-Containing Operon in <i>Streptomyces griseofuscus</i>	156
3.1 Background	156
3.2 Methodology.....	160
3.2.1 Cloning, expression, and purification of SgrMT Enzymes	160
3.2.2 Differential scanning fluorimetry	162
3.2.3 Crystallization and structure determination of SgrMPCO proteins	163
3.2.4 Mass spectrometric analysis of reaction products.....	165
3.2.5 Preparation of S- ¹³ C-labeled SAM	167
3.2.6 NMR characterization of SgrMT reaction products	167
3.2.7 Steady state kinetics determination of SgrMT	168
3.3 Results and discussions	170
3.3.1 SgrMppP is a homolog of SwMppP	170
3.3.2 SgrMT is a structural homolog of MppJ	171
3.3.3 SgrMT product identification by LCMS	175
3.3.4 Determination of SgrMT products by NMR spectroscopy	181
3.3.5 Confirmation of methyl- ¹³ C SAM synthesis by LC-MS.....	186
3.3.6 Determination of methylation position by INADEQUATE NMR	189
3.3.7 SgrMT is not specific for any divalent metal	193

3.3.8 Steady State Kinetics of SgrMT with 2KA and 4HKA.....	194
3.3.9 SgrAH is a structural homolog of ScPAH.....	197
3.3.10 SgrAH Reacts with the product(s) of SgrMT.....	200
4 Conclusion	205
5 Reference	208
Appendix A.....	226
Appendix B.....	227
Curriculum Vitae.....	229

LIST OF FIGURES

Chapter 1

Figure 1: Categories of drug approved by FDA since the last three decades (1984 to 2014). N (unaltered), ND (derivatives), and S (synthetic).	2
Figure 1.1.4: Biosynthetic route of terpenoids in plants	10
Figure 1.1.7-1: A brief mechanism of peptide growth in NRPS	23
Figure 1.1.7-2 Difference in biosynthesis of ribosomal and nonribosomal peptides.	24
Figure 1.3: Mannopectimycin open reading frame.....	30
Figure 1.4.5-1: L-enduracididine biosynthetic gene cluster in <i>Streptomyces wadayamensis</i> , is a part of mannopectimycin biosynthesis gene context.	37

Chapter 2

Figure 2.1-1: Genetic distant tree of MppP homologs constructed from PHI-BLAST	43
Figure 2.1-2: MppP-containing putative operon in <i>Pseudomonas brassicacearum</i>	45
Figure 2.1-3: The PbrMppP active site structure overlaid with SwMppP (PBD ID: 6C9B).....	48
Figure 2.1-4: Bacterial three-component oxygenase system.....	52

Figure 2.2.13-1 PbrMppP Promoter region (RS05025).....	68
Figure 2.2.13-2: Schematic of the process for cloning the putative PbrMPCO promoter region (RS05025) into the pPROBE-NT reporter plasmid. Image created with SnapGene® .	70
Figure 2.3.1-1: The structure of PbrMppP	75
Figure 2.3.1-2: UV-visible spectra of the PbrMppP reaction with L-Arg.	77
Figure 2.3.1-3: PbrMppP produces [95] and [91] in a ~2:1 ratio. The green spectrum (top) was collected for a solution of pure L-Arg in D ₂ O. The red spectrum is the PbrMppP reaction products.	79
Figure 2.3.1-4: ESI-mass spectrometric analysis of PbrMppP products in positive mode.	80
Figure 2.3.2-1: Regeneration of dioxygen after adding catalase into PbrMppP reactions at equilibrium.....	82
Figure 2.3.3-1: Michaelis Menten graph of PbrMppP catalysis. The reactions were initiated by adding PbrMppP (9 μ L of 272 μ M PbrMppP; ~5 μ M final concentration) into solutions containing 10 μ M, 20 μ M, 40 μ M, 80 μ M and 160 μ M L-arg in 50 mM BisTris Propane pH 8.5 buffer.	84
Figure 2.3.4-1: Pre-steady state kinetic data from global fitting in KinTek explorer version 10..	87

Figure 2.3.4-2: Pre-steady state kinetic data from global fitting in KinTek explorer V10, when the branch point to [91] was placed at EQ2 (A).....	92
Figure 2.3.4-3: Pre-steady state kinetic data from global fitting in KinTek Explorer when the EK branch point was placed at intermediate X1. (A).....	94
Figure 2.3.4-4: Pre-steady state kinetic data from global fitting in KinTek Explorer when the F is omitted from the model 1 (A).....	97
Figure 2.3.4-5: Pre-steady state kinetic data from global fitting in KinTek Explorer when the ER' intermediate is omitted from the model 1 (A).....	99
Figure 2.3.4-6: Pre-steady state kinetic data from global fitting in KinTek Explorer when the F is omitted, followed by omission of intermediate ER' from the model 1 (A).....	101
Figure 2.3.5-1: PbrMppP substrate specificity. Spectrometric scans of reactions with L-Arg and a panel of potential alternative substrates	109
Figure 2.3.5-2: Comparison of quinonoid-I formation by L-arg and L-canavanine in PbrMppP reaction.	113
Figure 2.3.5-3: ESI mass spectrometric analysis of MppP reaction products with L-canavanine	115

Figure-2.3.7.2-1: PbrHYP dimer structure. The yellow bond on the top indicated disulfide link between the two dimer. 120

Figure 2.3.7.3: Overlaid structures of PbrHYP (blue) and VapC (PBD ID: 4CHG; light grey)..... 125

Figure 2.3.8.4-1: Crystallographic structure of PbrDHPS with **[91]** bound..... 133

Figure 2.3.9.2-1 A) tertiary structure of PbrOX homodimer B) active site of PbrOX retains metal binding facial triad C) PbrOX (yellow) active site overlaid with UndA (red) with an RMSD value of 4.15Å over 6% sequence identity. 136

Figure 2.3.9.3-1: A) Spectra of Apo PbrOX and Fe^{II} installed PbrOX. Both Fe^{II} (FeSO₄) alone and ascorbate mixed Fe^{II} (FeSO₄) solution has a small peak near 272nm, whereas apo PbrOX doesn't have any peak other than 280 nm wavelength. B) The concentrated sample of Fe^{II} loaded PbrOX is overlaid with A. The red spectra (~250 μM PbrOX+ Fe(II)+ascorbate) is the sample concentrated after incubation with Fe and ascorbate (pink spectra: ~175 μM PbrOX with Fe(II) and ascorbate) 138

Figure 2.3.9.5-1: A) Spectra of PbrFd expressed in terrific broth supplemented with cysteine and ferric ammonium citrate (blue spectrum). The green spectrum is for PbrFd when expressed in LB media only, thus Fe-S cluster not installed. B) The elution fractions from HisTrap purification off PbrFd shows a brown color. 140

Figure 2.3.9.6-1: PbrFNR reduces dioxygen to produce H₂O₂ in the absence of any other electron acceptor. Denatured PbrFNR does not consume dioxygen..... 142

Figure 2.3.9.7-1: Electron flows from NADPH towards PbrOX 143

Figure 2.3.9.8-1: Fe(II) loading on PbrOX is required for electron flow 144

Figure 2.3.12-A: Metabolic cloud plot generated from DF41 cells expressed in solid culture (green traces) vs liquid culture (red traces) 153

Chapter 3

Figure 3.1-1: putative biosynthetic gene cluster in *Streptomyces griseofuscus*..... 156

Figure 3.1-1: A) Structure of SgrMppP homolog liganded with cofactor PLP resolved at 2.1 Å, R_{cryst}: 0.142, R_{free}: 0.184. B) SgrMppP (blue chain) active site overlaid with SwMppP (gray chain) with 51% sequence identity and RMSD value of 0.97 Å..... 171

Figure 3.3.2-1: A) structure of SgrMT liganded with Mg²⁺ and **[91]** or 2KA bound at 2.2 Å resolution. The silver sphere is the Mg²⁺ ion and the blue structure underneath the electron density map is the 2KA or **[91]**. B) Superimposed structures of SgrMT (yellow) with substrate **[91]** and Mg²⁺ bound and of MppJ (green, PDB ID: 4kig) liganded with the product β-methyl phenylpyruvate (RMSD = 1.2 Å for all Cα atoms). The red sphere is the Fe bound in the active site of MppJ..... 173

Figure 3.3.3-1: A) Extracted ion chromatograms of SgrMT reaction products B) Extracted ion chromatograms of SgrMT negative control.....	176
Figure 3.3.3-2: Mass spectra of individual chromatogram at specified retention time.....	177
Figure 3.3.4-1: determination of SgrMT reaction products by NMR spectroscopy	185
Figure 3.3.5-1 A) Chromatogram of EcMetK reaction product with S- ¹³ C labelled SAM, B) Chromatogram of naturally abundant SAM.	187
Figure 3.3.6-1: INADEQUATE (2D ¹³ C- ¹³ C coupling) spectra of SgrMT reaction products.	192
Figure 3.3.7-1: ¹ H-NMR of SgrMT reaction product in presence of Fe ^(III) and Ascorbate. The reaction substrate was pure 2KA [91]	193
Figure 3.3.8-1: Steady state kinetic data of SgrMT.....	197
Figure 3.3.9-1: SgrAH (green chain) overlaid with ScPAH (gray chain, (PDB ID: 1GQ6) with RMSD value of 0.97 Å	199
Figure 3.3.10-1: Extracted ion chromatogram of SgrAH reaction with ¹³ C- and ² D-labelled SgrMT products.	202

LIST OF TABLES

Chapter 2

Table 2-1: Gene accession codes, sequence annotations, and identity to closest characterized homolog of PbrMPCO proteins.	47
Table 2.2.1-1: Primers used for cloning Pbr proteins	53
Table 2.2.2-2: Storage buffer of PbrMPCO enzymes	56
Table 2.2.7-1: Crystallization conditions of the PbrMPCO enzymes.....	61
Table 2.2.13-1 Screening conditions used to test the putative PbrMPCO promoter.....	72
Table 2.3.2-1: Amount of dioxygen released upon the addition of catalase to PbrMppP reactions monitored in a Hansatech dioxygen electrode.....	83
Table 2.3.3-1: Steady state parameters calculated from the oxygen probe-derived initial velocity data with L-Arg.....	84
Table 2.3.4-1: Statistical analysis (observed χ^2 values) of different models explained in 2.3.4.	102
Table 2.3.6-1: Comparison of the steady state kinetic parameters of PbrMppP with L-Arg or L-canavanine as substrate.....	117

Table 2.3.12-1: list of m/z expressed highly in DF41 solid culture	154
--	-----

Chapter 3

Table 3.1-1: Gene accession codes, sequence annotations, and identity to closest characterized homolog of SgrMPCO proteins	157
--	-----

Table 3.1.1-1: Storage buffer of SgrMPCO proteins and EcMetK.....	162
---	-----

Table 3.2.3-1: Crystallization condition of Sgr enzymes	163
---	-----

Table 3.3.3-1: The major traces of m/z along with their corresponding predicted structure, analyzed from LCMS data of SgrMT reaction products. The negative control refers to SgrMppP reaction products + SAM+ Mg ²⁺ + heat denatured SgrMT.	178
--	-----

Table 3.3.5-1: Structure of major peaks determined from SAM mass spectroscopic fingerprint. The green asterisks indicates ¹³ C-isotope label	188
---	-----

Table 3.3.8-1: comparison of Michaelis–Menten kinetic fit of SgrMT with 2KA and 4HKA ...	197
--	-----

Table 3.3.10-1: Prediction of structure from ¹³ C- and ² D-labelled products from the SgrAH reaction.	203
--	-----

LIST OF SCHEMES

Chapter 1

Scheme 1.1.1: representative example of various types of alkaloids and their precursors.....	6
Scheme 1.1.3: structure of polyketides	9
Scheme 1.1.4: Structure of basic terpenoids	11
Scheme 1.1.6: non proteinogenic amino acid based natural products.	15
Scheme 1.1.6: representative lantibiotics structure.	19
Scheme 1.1.8: structure of natural products containing L-arginine derivatives, both in flexible chain form and constrained form.	28
Scheme 1.2: Structures of compounds mentioned in section 1.3	30
Scheme 1.3: Difference PLP-dependent enzymes' reaction mechanism.	36
Scheme 1.4.5-1: L-enduracididine biosynthesis by <i>Streptomyces wadayamensis</i>	38
Scheme 1.4.5-2: proposed reaction mechanism catalyzed by MppP.....	40

Chapter 2

Scheme 2.1-1: Reaction catalyzed by PbrMppP.	46
Scheme 2.1-2: Reactions catalyzed by SwMppP and SwMppR.	50

Scheme 2.3.4-1: Proposed reaction mechanism of PbrMppP	104
Scheme 2.3.8-1: Reaction catalyzed by EcDHPS.	127
Scheme 2.3.8.3-1: Structure of ShMppR product [122] when alternative aldehyde substrate 3-(2-furyl)acrolein [120] and pyruvate [122] are condensed.	130
Scheme 2.3.9-1: Reaction catalyzed by UndA	134

Chapter 3

Scheme 3.3.1-1: reaction catalyzed by SghMppJ. MppJ is part of mannopeptimycin- γ [131] biosynthetic gene cluster in <i>Streptomyces hygroscopicus</i>	Error! Bookmark not defined.
Scheme 3.3.2-1: Predicted reaction catalyzed by SgrMT.	Error! Bookmark not defined.
Scheme 3.3.4-1: Reaction catalyzed by EcMetK, where the S-connected methyl group is ^{13}C -isotope labelled	Error! Bookmark not defined.
Scheme 3.3.7-1: reaction catalyzed by SAM-methyltransferase assay kit from G-Biosciences (SAM510: SA Methyltransferase Assay, Cat. # 786-430)	Error! Bookmark not defined.
Scheme 3.3.8-1: reaction catalyzed by ScPAH	Error! Bookmark not defined.
Scheme 3.3.9-1: different form of urea in positive mode with their corresponding m/z (Ref: METLIN ID 6)	Error! Bookmark not defined.
Scheme 3.3.9-2: Predicted reaction catalyzed by SgrAH	Error! Bookmark not defined.

LIST OF ABBREVIATIONS

2KA	2-Ketoarginine
2KE	2- Ketoendiracidine
4HKA	4-Hydroxy-2-ketoarginine
AH	Amidinohydrolase
APCI	Atmospheric pressure chemical ionization
ATP	Adenosine triphosphate
BM	Basal media
COSY	Heteronuclear COrrrelation SpectroscopY
Dha	Dihydroalanine
Dhb	Dihydrobutyrine
DHPS	Dihydrodipicolinate synthase
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
DX	Double mixing
ESI-MS	Electrospray ionization mass spectrometry
FAD	Flavin adenine dinucleotide
Fd	Ferredoxin
FNR	Ferredoxin NAD(P)H reductase
Glc	Glucose

GOL	Glycerol
HSQC	Heteronuclear Single Quantum Correlation
INADEQUATE	Incredible Natural Abundance Double QUAntum Transfer Experiment
IPTG	Isopropyl β -d-1-thiogalactopyranoside
LB	Luria Broth
LCMS	Liquid Chromatography mass spectrometry
L-END	L-Enduracididine
LIG	Ligase
LS-CAT	Life science collaborate access team
MAD	Multiwavelength anomalous diffraction
MCPM	Metabolic cloud plot map
Met	Methionine
MO	Monooxygenase
Mpp	Mannopeptimycin
MRSA	Methicillin resistant <i>Staphylococcus aureus</i>
MT	Methyl ttransferase
NADH	Nicotinamide adenine dinucleotide hydrogen
NAD(P)H	Nicotinamide adenine dinucleotide phosphate hydrogen
NAT	N-Acetyl transferase
NMR	Nuclear magnetic resonance

NRPS	Non-ribosomal peptide synthase
Ox	Oxygenase
Pbr	<i>Pseudomonas brassicacearum</i>
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PIN	PiIT N-terminus
PKS	Polyketide synthase
PLP	Pyridoxal-5'-phosphate
PMP	Pyridoxamine 5'-phosphate
PP2	Proteose peptone media 2
Pyr	Pyruvate
Q1	Quinoid-1
Q2	Quinoid-2
RMSD	Root mean square deviation
SAD	Single wavelength anomalous diffraction
SAH	S-Adenosyl homocysteine
SAM	S-Adenosyl methionine
SF	Stop flow
Sgr	<i>Streptomyces griseofuscus</i>
Shg	<i>Streptomyces hygroscopicus</i>

SX	Single mixing
TEV	Tobacco etch virus
TLS	Translation-libration-screw
VRE	Vancomycin Resistant Enterococci

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my advisor Dr. Nicholas Silvaggi, for giving me the opportunity to work in the lab. I am grateful to him for mentoring me professionally as a crystallographer and enzymologist. I greatly appreciate the way he inspired me to incorporate various techniques in my research. He has always been patient to train me as a scientist. It is my pleasure to be a part of Silvaggi lab.

I am grateful to my committee members Dr. Alan Schwabacher, Dr. Sonia Bardy, Dr. David Frick & Dr. Andy Pacheco for their feedback and encouragement in my research and dissertation. I am also thankful to them for training me to use the instruments in their lab. I would also like to thank Dr. Anna Benko Dr. Frank Holger Foersterling for their enormous support through the last five years to train me in MS and NMR. I appreciate the help of Life Sciences Collaborative Access Team, Advance Photon Source in Argonne National Laboratory for data collection.

I am thankful to my lab members Lanlan Han, Trevor Melkonian, Nemanja Vuksanovic & Madeline Rettmann, for their consistent support in my research.

I want to thank my husband especially, Md Zubair Ahmed Khan, for his unconditional support and love during all these years. His endless consolation and encouragement helped me towards my graduation. My little angel Zaira Ahmed Khan also deserves a big thanks for staying calm during my graduation.

Last but not the least, I am grateful to my parents Raushan Ara Alpona & Mohammad Liaquat Ali, my brother Mohammad Asaduzzaman Shaon, my in-laws Md Abul Hashem Khan & Jesmin Akter; for understanding me and being patient during my stressful days.

Chapter 1

Introduction to natural products biosynthesis

Beyond the biochemistry of energy production, macromolecule biosynthesis, and cell division, there exists a universe of metabolites and associated chemical transformations that, while not required for cell viability, nonetheless provide distinct advantages in harsh or competitive environments. Collectively, these reactions are known as secondary metabolism. Many secondary metabolites possess biological activities like cytotoxicity, biofilm inhibition, or signaling. Collectively, these bioactive secondary metabolites are known as "natural products". Secondary metabolites can be obtained from both prokaryotes and eukaryotes. Prokaryotic secondary metabolites, in general, have defensive actions to protect the producing cells. Antibiotics are common examples of such natural products and discovery of antimicrobial agents like penicillin, streptomycin, and chloramphenicol revolutionized medicine. Eukaryotic secondary metabolites, for example alkaloids, have been used historically as anti-inflammatory, anticarcinogenic, and analgesic medicinesⁱⁱ. Over the last two decades, the pace of discovery of bioactive natural products has increased exponentially. With the aid of metabolomic profiling approaches in conjunction with computational chemistry, natural products and their derivatives has been developed as treatments for numerous diseasesⁱⁱⁱ. Since 1981, the majority of FDA-approved drugs were unaltered natural products or derivatives of natural products^{iv} (Figure1).

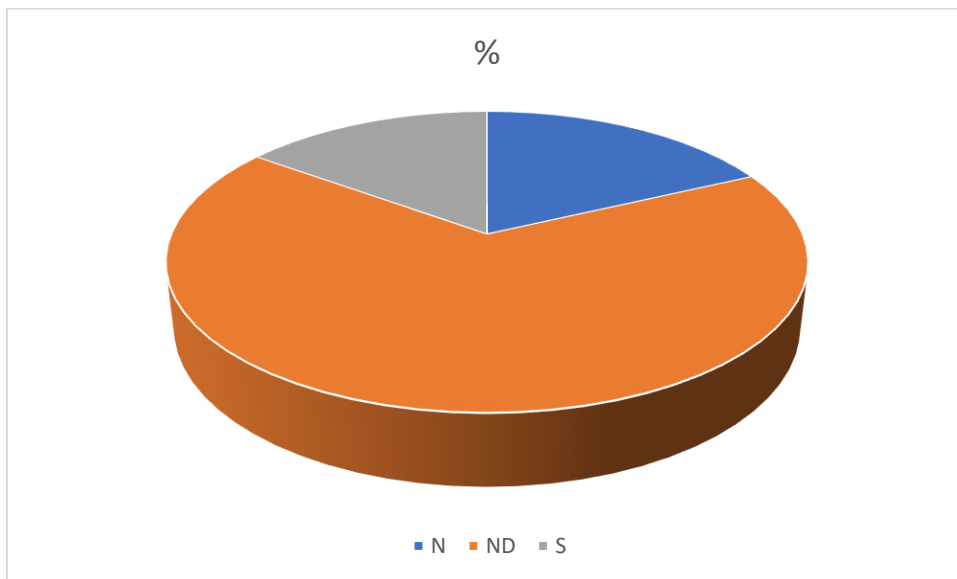


Figure 1: Categories of drug approved by FDA since the last three decades (1984 to 2014). N (unaltered), ND (derivatives), and S (synthetic).

Using natural products from direct sources threatens the natural growth of microorganism and endangers many plant species^v. A remedy to overcome this problem is systematic application and conjugation of Biotechnology and synthetic chemistry. Towards that resolution, it is indispensable to understand the biochemical pathway of the natural products.

1.1 Different types of natural products

Historically, natural products have been used as a potent source of medicinal compound. Without knowing the exact route of function, they were primarily used as therapeutic agents like cough, cold healer, and pain killer, for a long period of time. Back in early eighteenth century, the pioneering of Lavoisier principles (1789) for naming chemical elements, isolation of first recorded plant based natural product “morphine” (1804) by Friedrich Wilhelm Sertürner^{vi}, Friedrich Wohler’s breakthrough discovery for chemical synthesis of natural product (1828), and many other inventions commenced a new era for natural products and their functional characterization. For instance, the discovery of penicillin by Alexander Fleming in 1928 was facilitated by the developments mentioned above.

The emerging number of natural products discoveries require assortment for further study and modification. Natural products are classified into two major groups depending on the source: those synthesized by eukaryotes and those synthesized by prokaryotes. All three domains of life produce natural products... the bacterial and archeal systems (prokaryotes) and the Eukarya, which includes natural products from plants, fungi, and animals. Among them, plants produce highly diversified compounds with structural and functional attribution. Thus, natural products from plants are further sub-categorized into four groups depending on their biosynthetic origin^{vii} : (1) alkaloids, (2) phenylpropanoids, (3) polyketides, and (4) terpenoids. It should be noted that the production of polyketides is not limited in plants, they are also synthesized in bacteria and fungi.

Natural products from prokaryotes can be bioactive as antibiotic, many are used for cell signaling (i.e., quorum sensing) and nutrient scavengers like Siderophores. Although some produce compounds of other medicinal use, for example Bleomycin is a cytotoxic glycopeptide used for cancer treatment^{viii}. Here, for the purpose of understanding later studies in chapters 2 and 3, different classes of natural products will be discussed briefly along with the biosynthetic route of representative compounds.

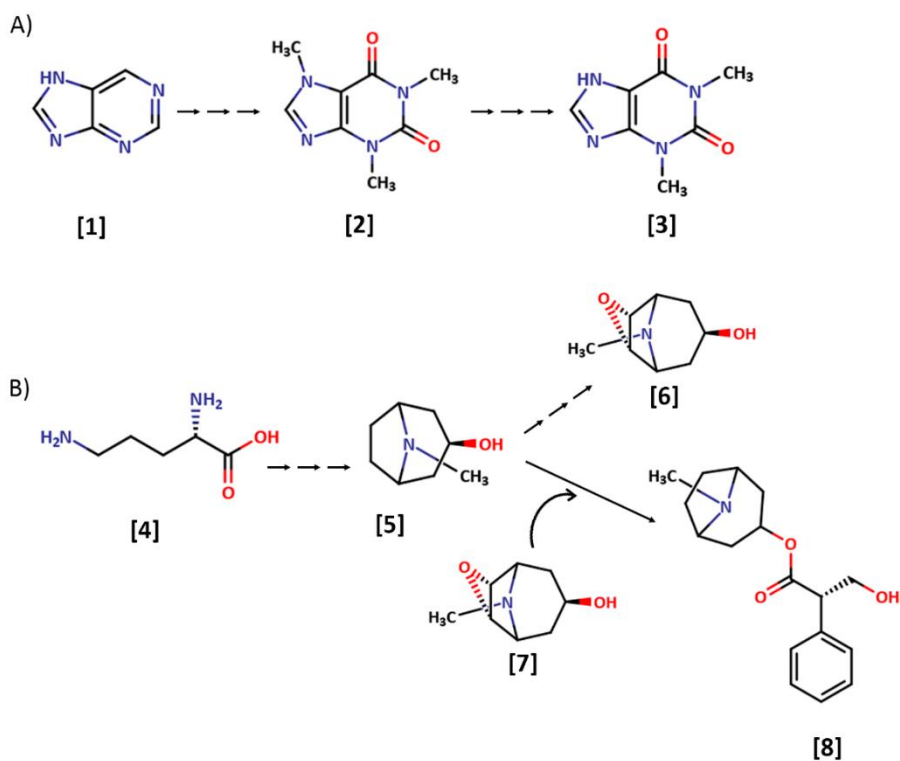
1.1.1 Alkaloids

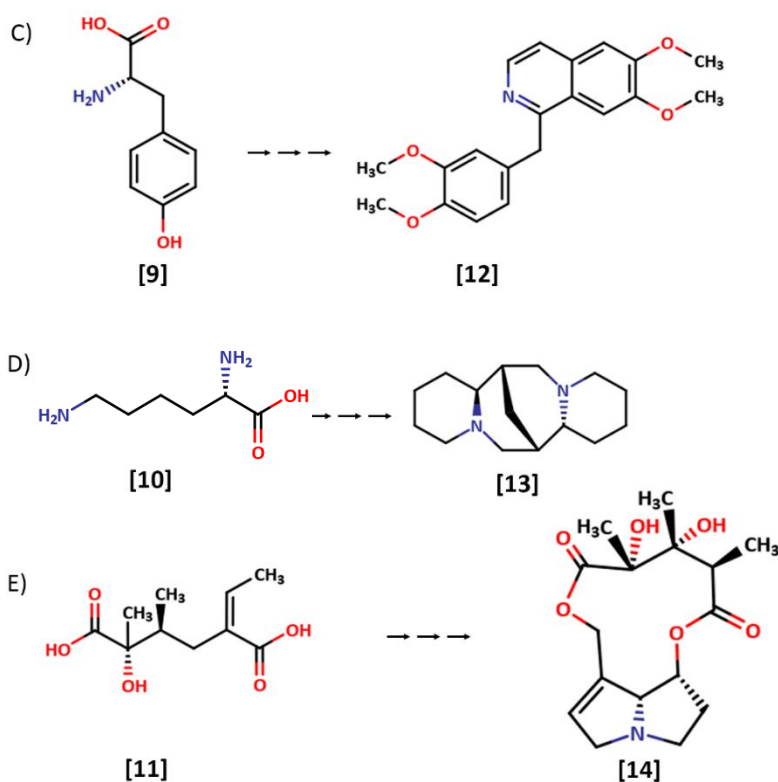
The largest group of plant based natural products are alkaloids. As Plants cannot excrete nitrogen directly, so they store it as nucleosides or amino acids. When the nucleoside pool exceeds the amount needed for primary metabolism, they are incorporated into alkaloid natural products, called nucleoside derived alkaloids. A major subcategory of nucleoside derived alkaloids are purine **[1]** alkaloids that commonly act as central nervous system stimulant and muscle relaxant^{ix}. For instance, theophylline **[3]** is a purine alkaloid derived from caffeine **[2]** (Scheme 1.1.1-A).

Similar to nucleoside derived alkaloids, when the amino acid pool exceeds what is needed for protein synthesis, these building blocks are diverted to other metabolic fates^x, which includes incorporation into alkaloid natural products, commonly known as amino acid derived alkaloids.

For example, alkaloids produced from Ornithine **[4]** precursor, are known as tropane and generally used as anti-spasmodic drug. In the biosynthesis route of tropane alkaloids, ornithine **[4]** is first converted into amino alcohol tropine **[5]**, and then branched towards scopolin **[6]**. Some alkaloids are synthesized from further esterification of tropine **[5]** and tropic acid **[7]**. For

example, (S)-hyoscyamine **[8]**, which has major use for bowel and gallstone treatment. Some alkaloids are simply derived from tyrosine **[9]**, lysin **[10]** and dicarboxylic acid necic acid or (2R,3R,Z)-5-Ethylidene-2-hydroxy-2,3-dimethylhexanedioic acid **[11]** (Scheme 1.1.1-C, D, E). Papaverine **[12]** is a benzyloisoquinoline alkaloid synthesized from tyrosine and it used to prevent vasospasm during heart disease and erectile dysfunction. Sparteine **[13]** is a quinolizidine alkaloid derived from lysine **[10]** precursor and used as medication for cardiac dysrhythmia. On the other hand, pyrrolizidine alkaloid monocrotaline **[14]** is synthesized from Necic acid **[11]** and used as feeding deterrent.



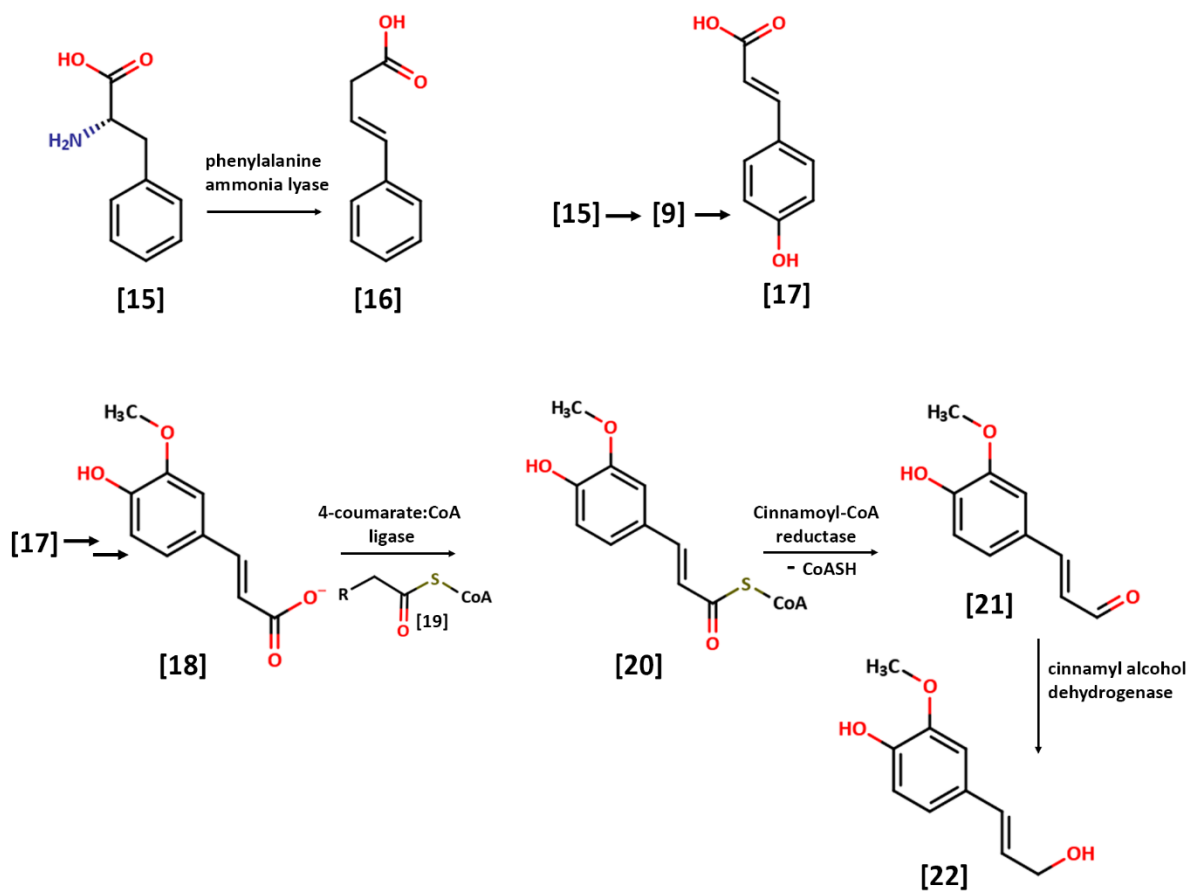


Scheme 1.1.1: representative example of various types of alkaloids and their precursors.

1.1.2 Phenylpropanoids

Naming of this group of natural products comes from their biosynthetic precursors phenyl alanine and propane. Although the actual route of synthesis is not that simple. The simplest phenylpropanoid is cinnamic acid **[16]**, derived directly from L-phenylalanine **[15]** by phenylalanine ammonia lyase enzyme. Cinnamic acid **[16]** then acts as precursor of other phenylpropanoids like lignan, lignin, phenylpropanes, benzenoids and coumarins^{xi}. Another simple phenylpropanoid would be coumaric acid **[17]**. The coumaric acid acts a a building precursor of another phenyl propanoid called ferulate **[18]**^{xii}.

A representative example of phenylpropanoid would be coniferyl alcohol **[22]**, the precursor of lignin. Coniferyl alcohol is synthesized from ferulate **[18]** in *angiosperm*^{xiii} (Figure 1.1.2B). Ferulate is condensed with an acetyl group from acetyl CoA **[19]** to form ferulyl-CoA **[20]** by the enzyme 4-coumarate:CoA ligase. Cinnamoyl-CoA reductase converts the ferulyl-CoA into coniferyl aldehyde **[21]**, which is subsequently reduced by cinnamyl alcohol dehydrogenase into coniferyl alcohol **[22]**.



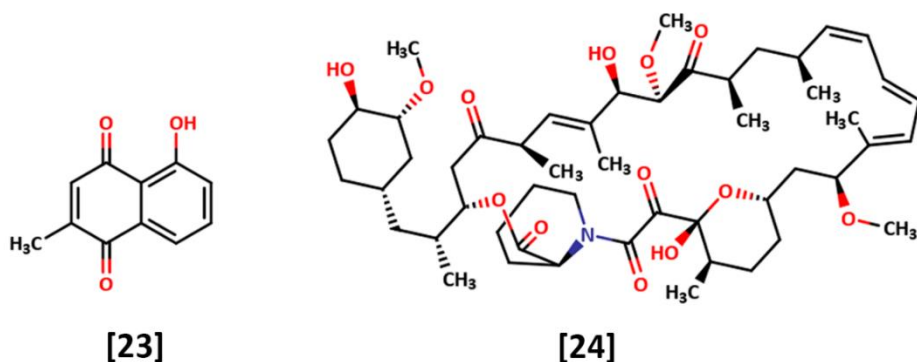
Scheme 1.1.2: Biosynthesis of phenylpropanoids

The basic medicinal use of phenylpropanoid lignans is to treat rheumatoid arthritis^{xiv}. Lignans are major antioxidants and free radical scavengers used to treat cancer^{xv}. Another major use of phenylpropanes would be as local anesthetics^{xvi}. Benzenoids, coumarins and their derivatives have medicinal uses as antinociception and anti-inflammatory agents^{xvii, xviii}.

1.1.3 Polyketides

Polyketides have medicinal attribution due to their antibiotic, antifungal, anti-cancer, and immunosuppressant properties^{xix}. Their biosynthetic source can be plant, microorganisms, fungi, or certain seawater animals. In particular, they are synthesized by enzymes expressed in module form, a concept discussed later in section 1.2.

Polyketides are natural products comprised of conjugated carbonyl group or methylene group. These molecules are incorporated from donors like acetyl- CoA or malonyl-CoA. Polyketide synthesis is similar to fatty acid synthesis. Except, unlike fatty acids, polyketides retain all their oxygen atoms in the macromolecular structure. The presence of multiple ketone groups aid for condensation reaction, thus polyketides have a variety of substitution. Some polyketides contain phenolic group forming aromatic polyketides, but their pattern is different from phenylpropanoids. For example, plumbagin (**[23]** in Scheme 1.1.3) is a plant derived polyketide used as anti-inflammatory agent. On the other hand, Rapamycin **[24]** is a polyketide synthesized in gram- positive bacteria like *Streptomyces*^{xx} and typically used as immunosuppressants. Many polyketides are glycosylated, and some of them even contain phenylpropanoid or terpenoid moieties.



Scheme 1.1.3: structure of polyketides

1.1.4 Terpenoids

Terpenes are naturally occurring hydrocarbons comprised of monomeric five-carbon unit isoprene ([25] in Scheme 1.1.4). Both isoprene [25] and methacrolein [26] are terpenes and act as terpenoid building block. Terpenoids are oxygen substituted derivative of terpenes or isoprenes. Thus, terpenoids are also known as isoprenoids. They are further subcategorized according to the number of units present in the structure^{xxi}. For example, methacrolein [26] is a Hemiterpene which is a five-carbon terpenoid natural product, Menthol [27] is a C-10 mono terpene used to treat sore throat pain. Thus, sequential incorporation of isoprene unit will be named as sesquiterpene (C-15), diterpenes (C-20), triterpenes (C-30), tetraterpenes (C-40), and polyterpenes (>C 40). For instance, labdane [28] is a diterpene and Germacrene [29] is a sesquiterpene, both having medicinal use as antimalaria, antibacterial, antifungal, and anti-inflammatory agent^{xxii}. Nowadays, Labdane is popular to inhibit replication of HIV virus^{xxiii}.

Terpenoids are exclusively derived in plants. In plants, two major subcellular organelles devoted for sugar synthesis, are plastid and cytoplasm. Two different pathways from these organelles synthesizes terpenoids (Figure 1.1.4).

One is the mevalonic acid (MVA) pathway, that occurs in the cytoplasm. The other is the desoxyxylulose phosphate (DXP) pathway, occurring in the plastid. Isoprene or isoprene derivatives are a biproducts of these pathways. Consequently, excessive sugar production in the plant is diverted into the terpenoid biosynthesis via DXP pathway. Figure 1.1.4 represents routs of terpenoids biosynthesis^{xxiv}.

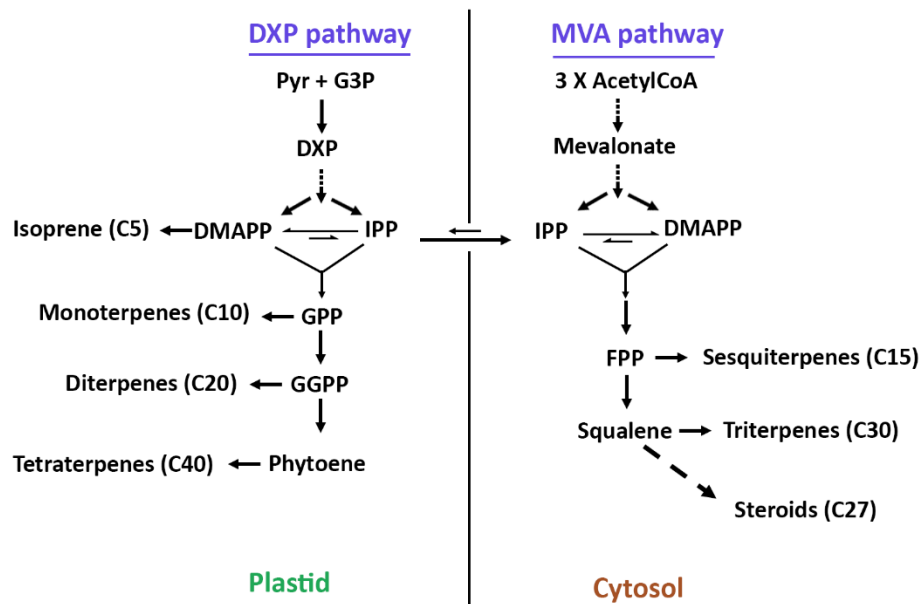
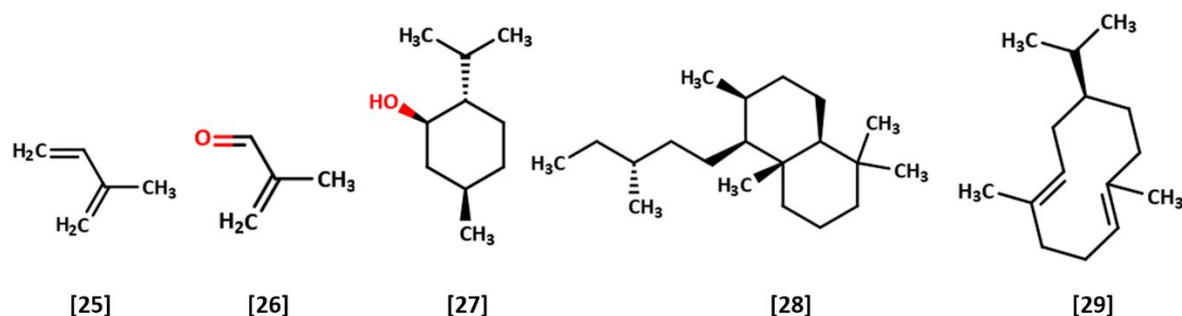


Figure 1.1.4: Biosynthetic route of terpenoids in plants. DXP, deoxyxylulose phosphate; pyr, pyruvate; G3P, glyceraldehyde-3-phosphate; DMAPP, dimethylallyl diphosphate; IPP, isopentenyl diphosphate; GPP, geranyl diphosphate; GGPP, geranylgeranyl diphosphate; MVA, mevalonate; FPP, farnesyl diphosphate.



Scheme 1.1.4: Structure of basic terpenoids

1.1.5 Non proteinogenic amino acids in natural products

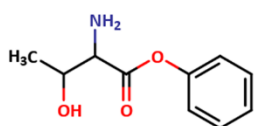
There are more than 1000 nonproteinogenic amino acids synthesized in plants, microorganisms, and other species. The naturally occurring non proteinogenic amino acids are not synthesized aimlessly, rather they serve as precursors of secondary metabolites. There is an enormous diversity of non-proteinogenic amino acid structures. For instance, β -amino acids, γ -amino acids, D-amino acids, α -hydrogen devoid amino acids, amino acid dimer or twin amino acids and so on. Here only naturally occurring non-proteinogenic amino acids will be discussed, which are direct precursors of natural products with medical significance.

β -amino acids have amino group attached to the β -carbon instead of α -carbon. There are around 100 different β -amino acid derived natural products discovered so far, which have biological significance. For example, β -amino- α -hydroxy phenyl butyrate ([30] in scheme 1.1.6) is a constituent of bestatin [31]^{xxv}, a linear dipeptide used as an angiotensin stimulator. A few more

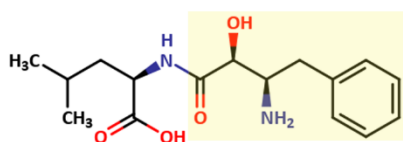
examples of β -amino acid containing natural products would be capreomycin [32], viomycin [33], bleomycin [34] and they all have antibacterial property. On the other hand, γ -amino acid analogues have clinical significance as neurotransmitter mimic. Ranunculin A [35] is a recently discovered natural product isolated from *Ranunculus ternatus*^{xxvi}, with γ -amino acid and sugar hybrids. Interestingly, Ranunculins have a trend of tail-to tail ether links between amino acid and sugar components. This extraordinary feature provides opportunity for new natural products. Structural analysis shows that compound exhibits structural similarity to GABA neurotransmitters. The extracted compound has historically been used as Chinese medicine for faucitis.

Incorporation of D-amino acids into natural products are far more complex than other groups of nonproteinogenic amino acids. Most prevalent example of D-amino acid is in the peptidoglycan layer of microorganism, which provides them resistance against peptidases^{xxvii}. Antibacterial-natural products derived from microorganisms, show versatile composition of D-amino acids in peptidyl oligomer subunit. Former in-vivo study shows that actinomycin D [36] biosynthesis in *Streptomyces parvullus* involves L-valine rather than D-valine^{xxviii, xxix}. A multifunctional enzyme, actinomycin synthetase-II (ACMS-II) catalyzes the formation of the peptide bond between threonine and valine and eventually epimerizes the L-valine into D-valine, while the residue is already in peptide bound form^{xxx}. Plant derived natural products with D-amino acid constituents, are simpler than bacterial. For example, some parasitic plants from genus *Orobanche*, can produce D-tryptophan in response to arduous environment. The D-tryptophan is simply metabolized into indole-3-pyruvate by aminotransferase activity^{xxxi}. The indole-3-pyruvate is

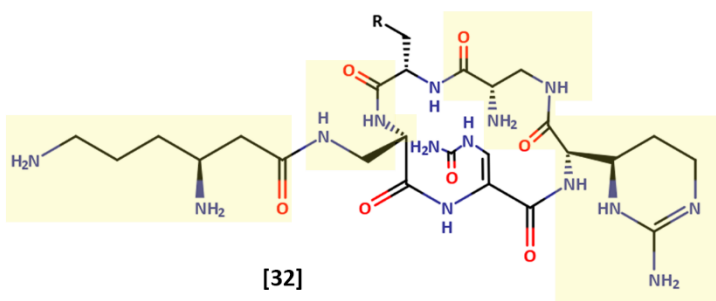
further converted into natural products like indole-3-acetic acid [37] and other hormones^{xxxii}. All these compounds have physiological aspect of cell differentiation.



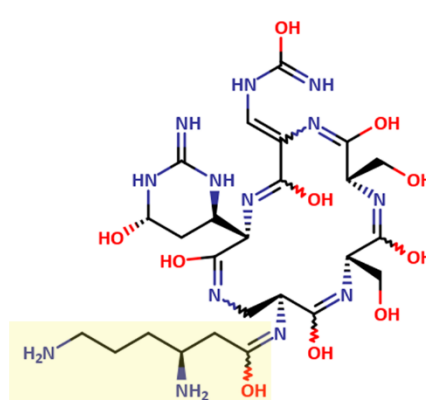
[30]



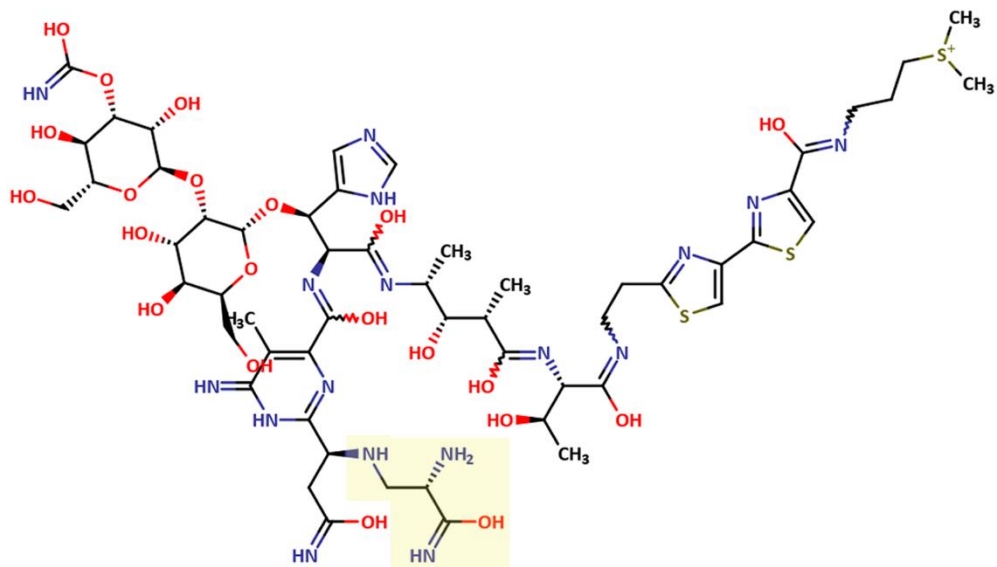
[31]



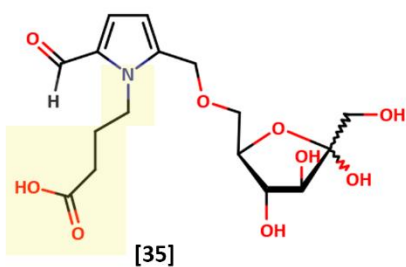
[32]



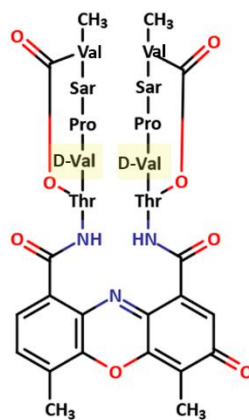
[33]



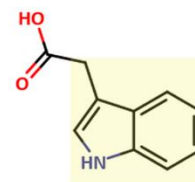
[34]



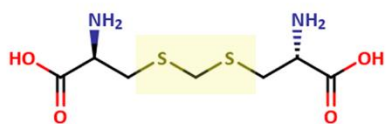
[35]



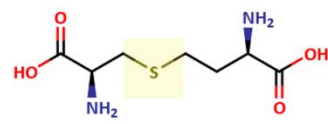
[36]



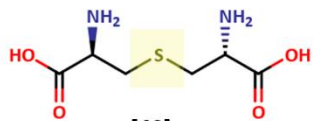
[37]



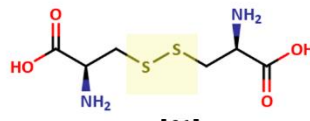
[38]



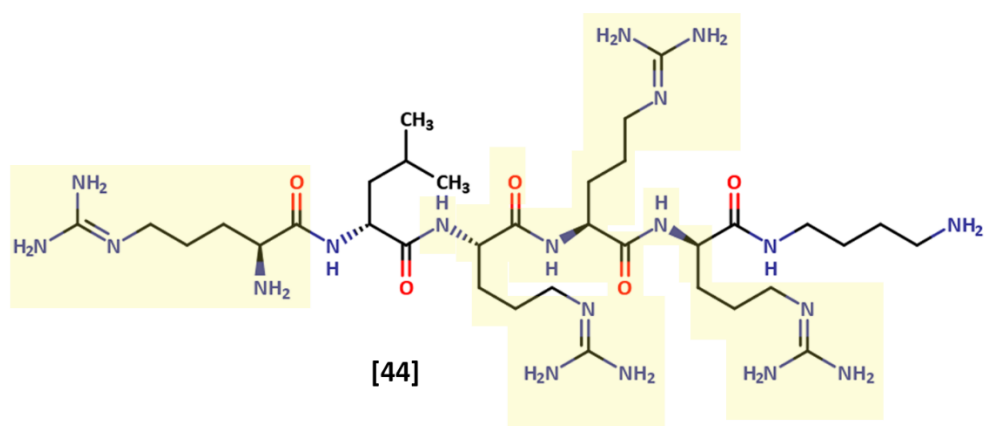
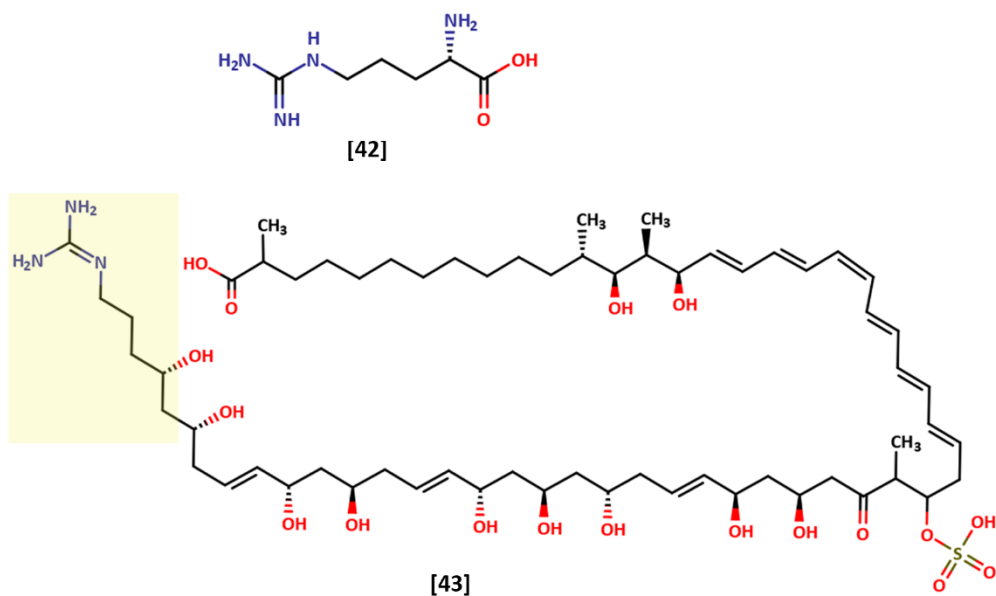
[39]



[40]



[41]



Scheme 1.1.6: non proteinogenic amino acid based natural products. Highlighted portion in the structures come from amino acids. In structure [39], [39], [40], [41], highlighted portion forms twin amino acids.

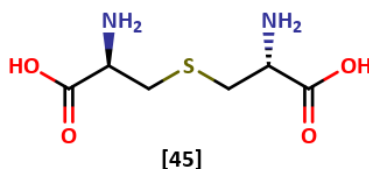
The group of non-proteinogenic amino acids where two residues are connected by bonds other than peptide bonds, are called amino acid dimers or twin amino acids. Many dimeric amino acids opt a scope of stereocenter. For example, djenkolic acid **[38]** from djenkol beans, is formed by fusion of two cysteine molecules by methylene group^{xxxiii}. Though there are yet some experimental studies required, djenkolic acid seems to produce carbonyl sulfide natural products that act as soil fumigants to naturally control plant parasitic nematodes^{xxxiv}. Cystathionine **[39]** and lanthionine **[40]** (discussed in section 1.1.7) are two other examples of twin amino acids that are natural product precursors. Cystathionine is biosynthesized from trans-sulfuration of activated serine and cysteine residues, whereas Lanthionine is formed from trans-sulfuration of two L-cysteine molecules.

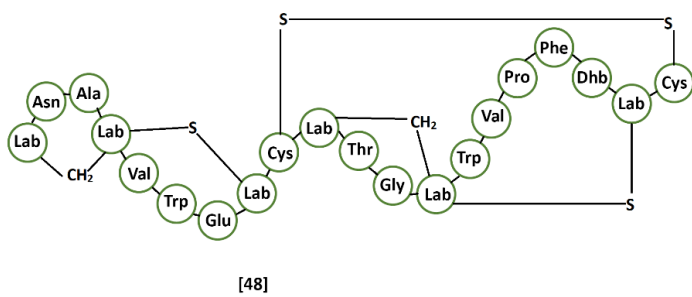
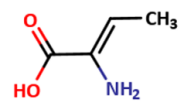
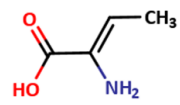
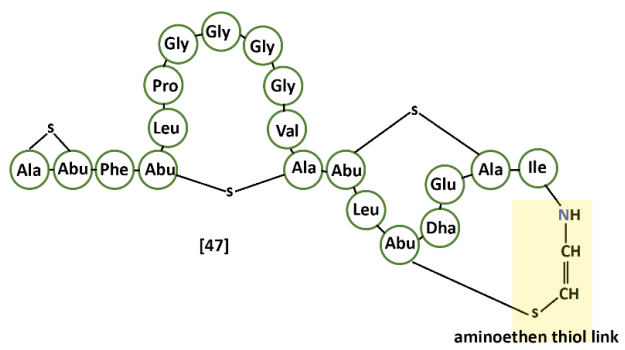
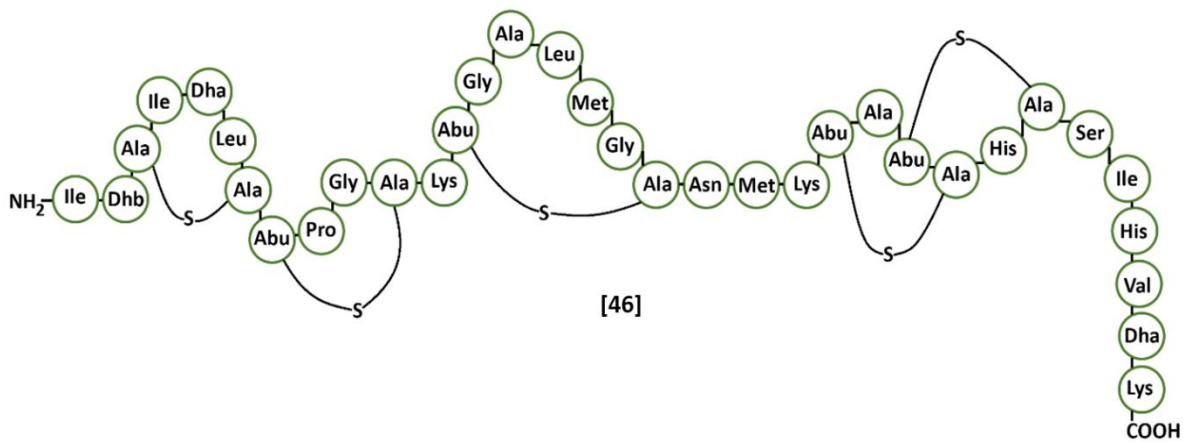
The Naturally occurring α -amino acid L-arginine (**[42]** in scheme 1.1.7) is precursor of protein biosynthesis among all the living kingdom. However, the distinguished functional feature gives it a possibility to serve as biosynthetic precursor for molecules other than proteins. The large aliphatic side chain opts flexibility for the structure. Terminal guanidinium cap provides a pK_a value of approximately 12.5, which is far higher than physiological pH, thus this functional group acts as hydrogen bond donor in proteins. Apart from that, delocalized electron in the guanidinium group allow it to participate in multiple hydrogen bonding interactions. All these features, permit L-arg to attain diverged conformation. Thus, natural products derived from L-arginine precursors, are biologically important to show different mode of action, basically of anti-bacterial and antifungal activity.

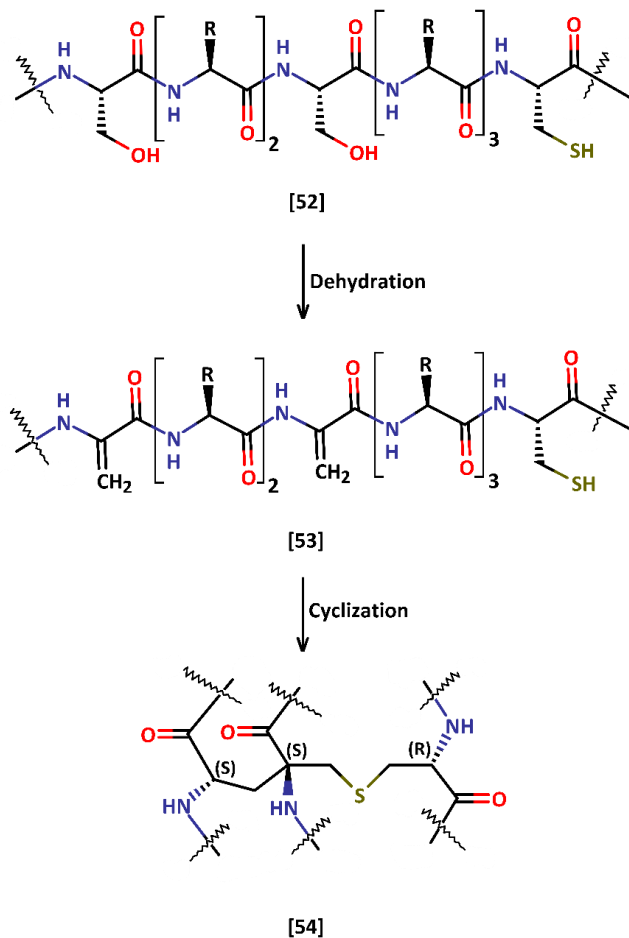
For instance, Clethramycine [43] and bicornutin [44], are naturally occurring antibiotics, containing L-arg residues. The flexible free side chain of L-arginine is retained in the final structure (highlighted).

1.1.6 Lantibiotics

Over the last three decades, the increasing prevalence of antibiotic-resistant microorganisms has highlighted the need for antibiotics with distinct modes of action. As a result, a new class of antibiotics, known as Lantibiotics has got attention. The term Lantibiotics came from its structure containing Lanthionine (Scheme 1.1.7-A), a thioether linked non-proteinogenic amino acid. The term Lanti derived from lanthionine ([45] in scheme 1.1.8), was incorporated into the name to highlight the repetitive functional thioether link into the macromolecular structure. Lantibiotics are unique for its polycyclic non-proteinogenic peptide form, interspersed with unsaturated or dehydro amino acids.







Scheme 1.1.6: representative lantibiotics structure. [58] Nisin A a class-I lantibiotic, [59] Mersacidine a class-II lantibiotic, [60] Labyrinthopterin a class-III lantibiotic.

According to the structural features and chemical characteristics, lantibiotics are grouped into three classes^{xxxv}. Class-I lantibiotics are those with elongated and flexible peptide chains^{xxxvi}. For example, Nisin A [46] isolated from *Lactococcus lactis*, is a class-I lantibiotic comprised of 34 amino acids^{xxxvii}. 80% residues in the peptide chain comes from α - amino acids, interspersed by altered amino acids like dehydroalanine (Dha) [49], dehydrobutyrine (Dhb) [50] and amino

butyric acid (Abu) [51]. These altered amino acids come from enzymatic dehydration of serine or cysteine^{xxxviii}. Furthermore, they are significant because they prevent cleavage of the polypeptide chain and studies showed that cleavage of Nisin A in the Dha or Dhb position results in inactive lantibiotic^{xxxix}. The extensible conformation enables this class of lantibiotics to interact with the cell wall. Although their actual mode of action is followed in two steps^{xi}. Primary mode of action is to cell membrane disruption by N-terminal hydrophobic residue of Nisin A, and pore formation. Pore formation occurs by Nisin A binding with lipid-II and forming a poration complex that mediates efflux of K⁺ ion^{xii}. Secondary mode of action is exerted by forming a complex with the lipids present in peptidoglycan layer, consequently inhibiting cell wall biosynthesis by depriving the substrate for transpeptidase.

Class-II lantibiotics have more globular conformations, which facilitate their translocation through the cell membrane. Class-II lantibiotics also often contain Dha, Dhb and Abu. The class-II antibiotic is different from class-I antibiotic in term of structure, as the class-I lantibiotics only thioether linkage. In contrast, class-II lantibiotics have amino-ethen thiol linkage (shown in structure of mersacidin [47]) which renders small chain chemical bonds. Thus, class-II lantibiotics attain more globular conformation comparing to class-I lantibiotics. Unlike class-I lantibiotics, class-II lantibiotics show bactericidal activity by specifically binding with phosphoethanolamine in the cell membrane, resulting in increased membrane permeability. They do not make an actual pore because of the lack of N-terminal hydrophobic residue inaccessible, which is required for cell membrane disruption. Furthermore, they also lack the flexible hinge region present in class-I lantibiotics (Scheme 1.1.7 [47]).

Class-III lantibiotics are still a small group of lantibiotics, which are usually non-antibiotic, but undergo morphological change to gain activity, thus they are pro-drugs. This class of antibiotic has special attention due to the occurrence of a rare amino acid triplet called “labionin” [54]. The biosynthetic precursors of the tripeptide are two serine and one cysteine residues. The three precursor amino acids are not adjacent to each other, rather they are two or three amino acids apart from each other [52]. Thus interestingly, this rare tripeptide is formed while present in the primary polypeptide of class-III lantibiotics. The three interspersed precursors are further modified by dehydration of serine residues and forming Dha-Dha-Cys tripeptide [53]. This tripeptide has 2S,4S,8R configuration, and consists of a central quaternary carbon atom with a lanthionine motif. Also, they possess an unusual methylene bridge, which establishes a covalent link to a further amino acid moiety. This tripeptide has to give Labionin is comprised of two Dha and one Cys residue α -amino butyric acid, D-serine, and D-cysteine, where the cysteinyl sulfur linked to D-serine is further connected with the γ -carbon of α -amino butyric acid^{xlii}. Compared with lanthionine, labionin has only one thioether bridge in peptide side chains. In addition to that, labionin facilitates an additional unusual carbacyclic ring linkage [48]^{xliii}. Thus class-III lantibiotics can form more compact structure. For instance, Labyrinthopeptin [48] synthesized in *Actinomadura namibiensis* shows antiviral activity against HIV when HIV-induced cell-cell syncytia forms in the host cell^{xliv}

1.1.7 Non ribosomal peptide and their biosynthesis

Protein synthesis relies on the ribosome (Figure 1.1.7-2) and this function is limited to the 20 proteinogenic amino acids^{xlv}. The secondary metabolites produced by bacteria and fungi often contain non-proteinogenic amino acids, which are an important source of structural diversity and often play a role in forcing such compounds into the specific conformation needed for their biological activity^{xlvi}. Since there are so many different non-proteinogenic amino acids, the ribosomal machinery is not efficient for synthesizing secondary metabolites, using non-proteinogenic amino acids as substrate. Thus, peptide natural products containing non-proteinogenic amino acids are produced by non-ribosomal peptide synthetase (NRPS) systems, where the sequence of the peptide is encoded directly in the enzymatic assembly line producing a particular non-ribosomal peptide (NRP). This enzymatic assembly line is known as module. The components of modular multidomain enzymes are transposable. Thus, they can be multi-substrate enzymes where the binding site for each substrate is programable. The product of one enzyme is carried to subsequent enzyme, thus the elongation of peptide continuous until final step of cyclization reaches. One NRPS system can be comprised of multiple modules, where each module contains enzymes for particular substrate catalysis^{xlvii}.

Figure 1.1.7-1 briefly describes route of non-ribosomal peptide. For instance, each module can have different functional domain like adenylation domain, condensation domain and tailoring domain like epimerase, hydroxylase, deaminase etc. Product from each module is carried to the subsequent module by peptidyl carried domain till the peptide chain is fully grown. Finally, a terminal enzyme i.e., thio-esterase, oxy-esterase cyclizes the chain into final product.

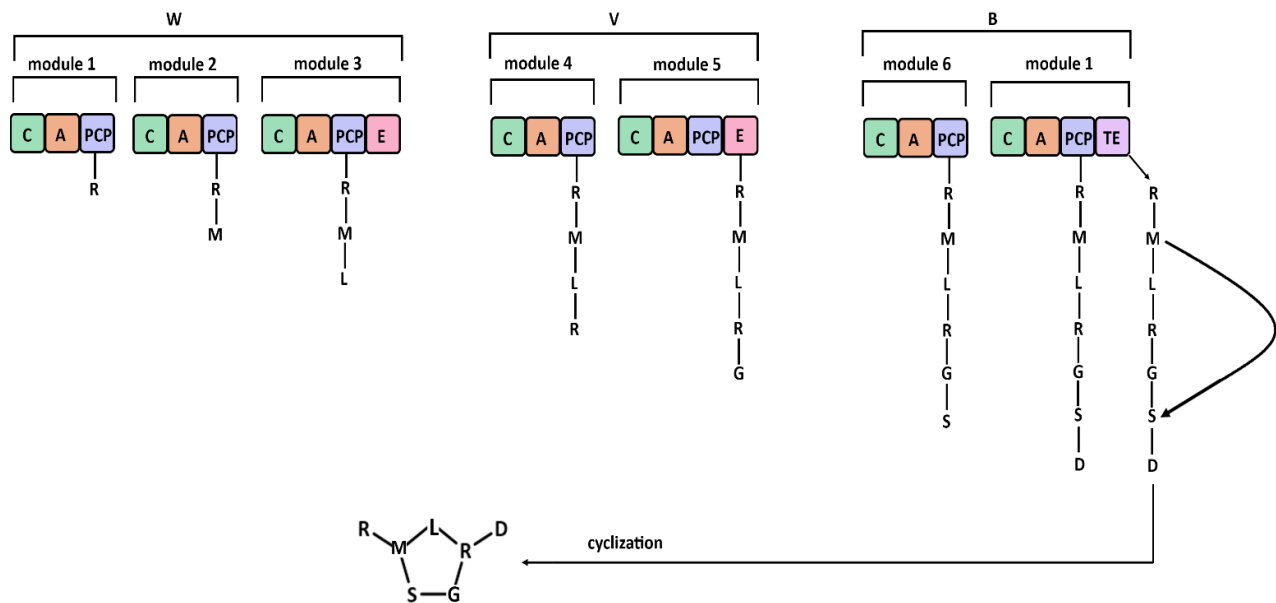


Figure 1.1.7-1: A brief mechanism of peptide growth in NRPS. C, condensation domain; A, adenylation domain; PCP, peptidyl carried domain; E, an exemplary tailoring enzyme epimerase; TE, terminal cyclization catalyzing thio-esterase. R, M, L, G, S, D are symbols used for different amino acid residues that are components of final product. Each residue is connected by peptide bond. The W, V and B indicates part of the system, that synthesizes a particular peptide.

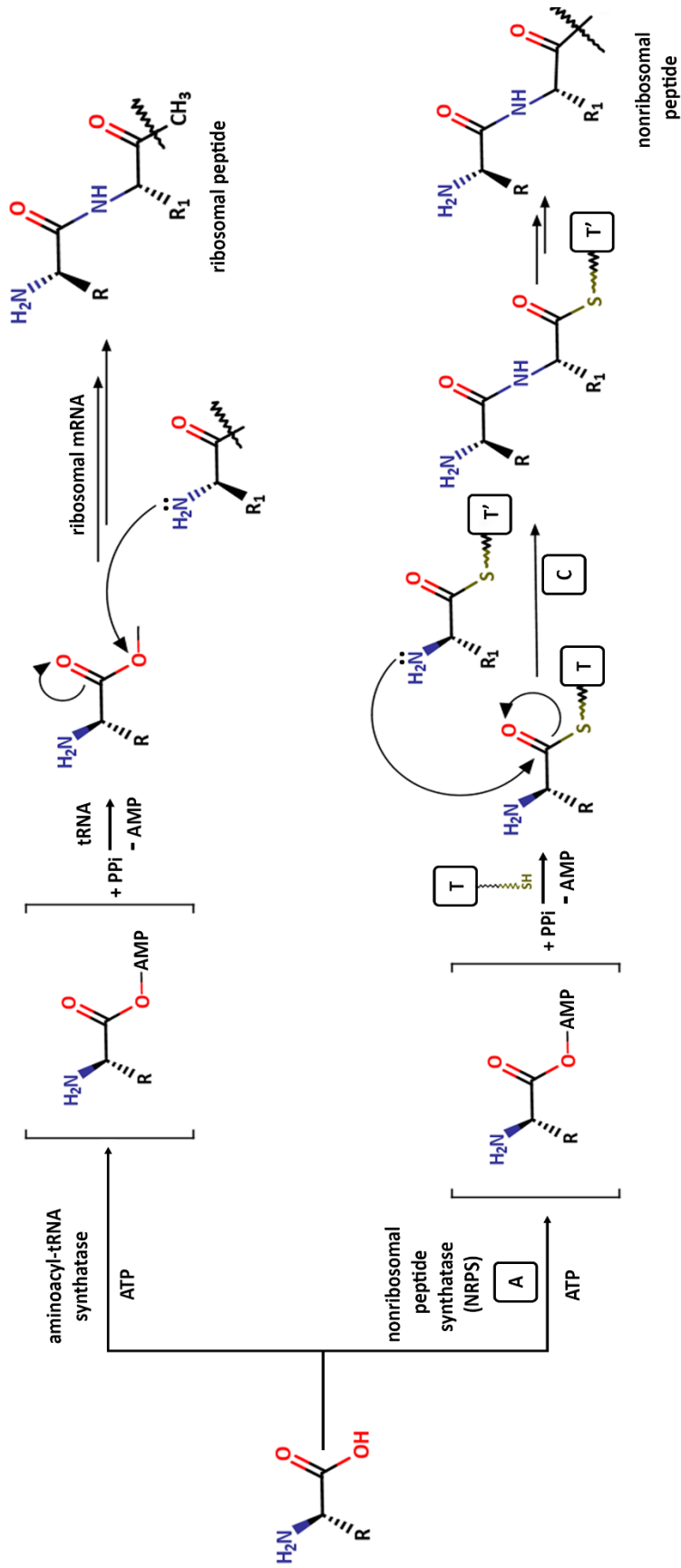
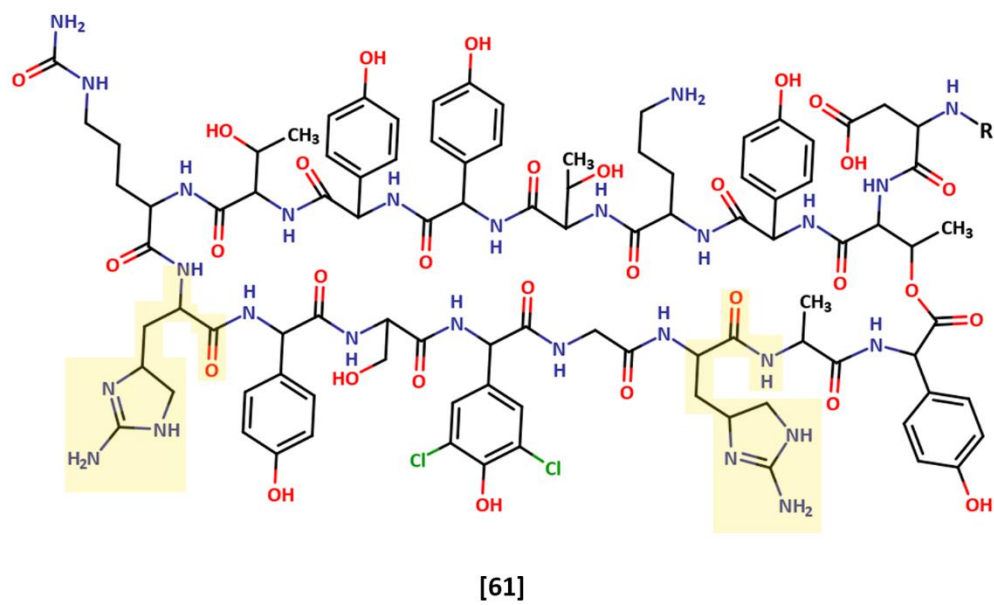
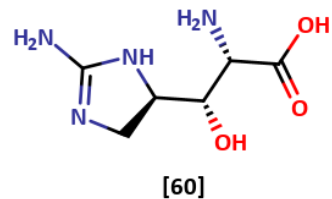
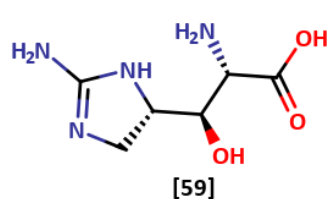
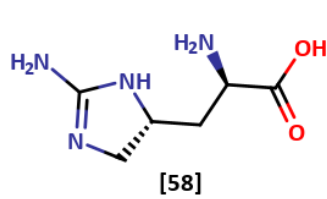
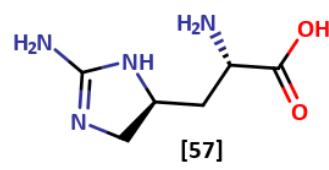
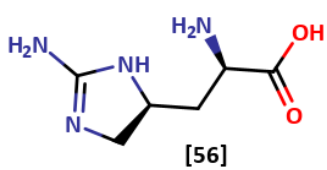
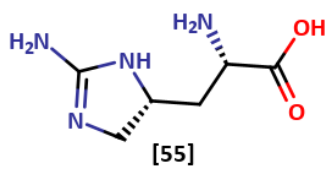


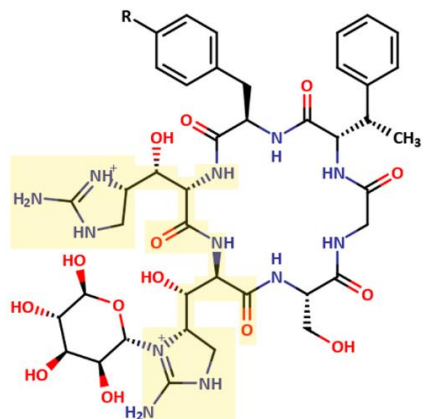
Figure 1.1.7-2 Difference in biosynthesis of ribosomal and nonribosomal peptides. ATP, adenosine triphosphate; AMP, adenosine monophosphate; PP_i = pyrophosphate; A domain, adenylation domain; T domain = thiolation (peptidyl carrier protein) domain; C domain, condensation domain.

1.1.8 L-Enduracididine as a natural product precursor

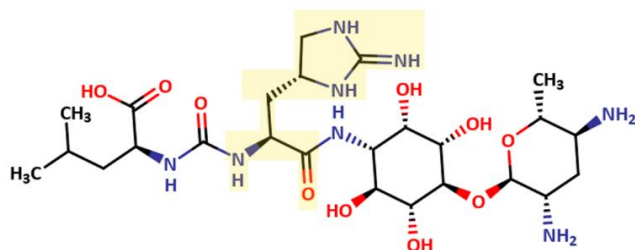
The non-proteinogenic amino acid L-enduracididine (L-End) [55] is a conformationally constrained derivative of L-arg that is found in a number of non-ribosomally produced peptide natural products. The structure of L-End is unique because it contains a five membered cyclic moiety of guanidinium group. It is a constituent of peptide antibiotics like enduracidin [61], mannopeptimycin [62], minosaminomycin [63], teixobactin [64], and enramycin [65]. There are several stereoisomers of L-enduracididine found in the nature^{xlviii}, for instance D-enduracididine [56], L-allo-enduracididine [57], D-allo-enduracididine [58]. Two more derivative of L-End, L- β -hydroxyenduracididine [59] and D- β -hydroxyenduracididine [60] are also found as natural product precursors.

Biosynthesis of L-enduracididine involves three steps, catalyzed by enzymes present in corresponding antibiotic biosynthetic gene clusters. First step is oxidation of L-arginine, followed by second step of cyclization between γ -carbon and one of the guanidinium nitrogen. Finally, transamination results in release of the products (Scheme 1.4).

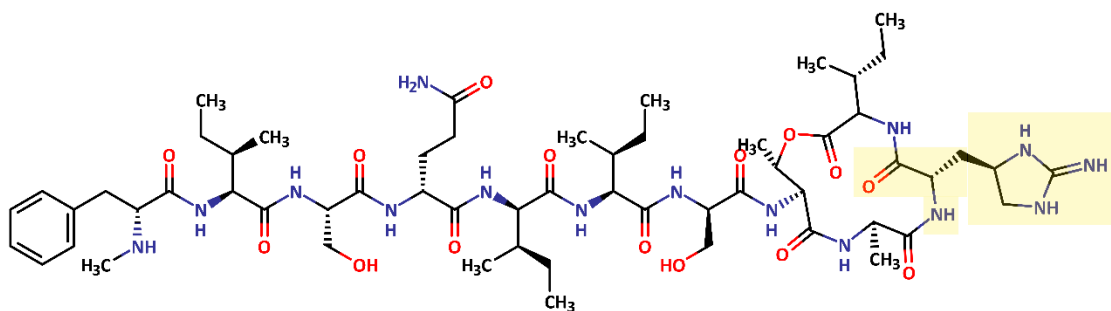




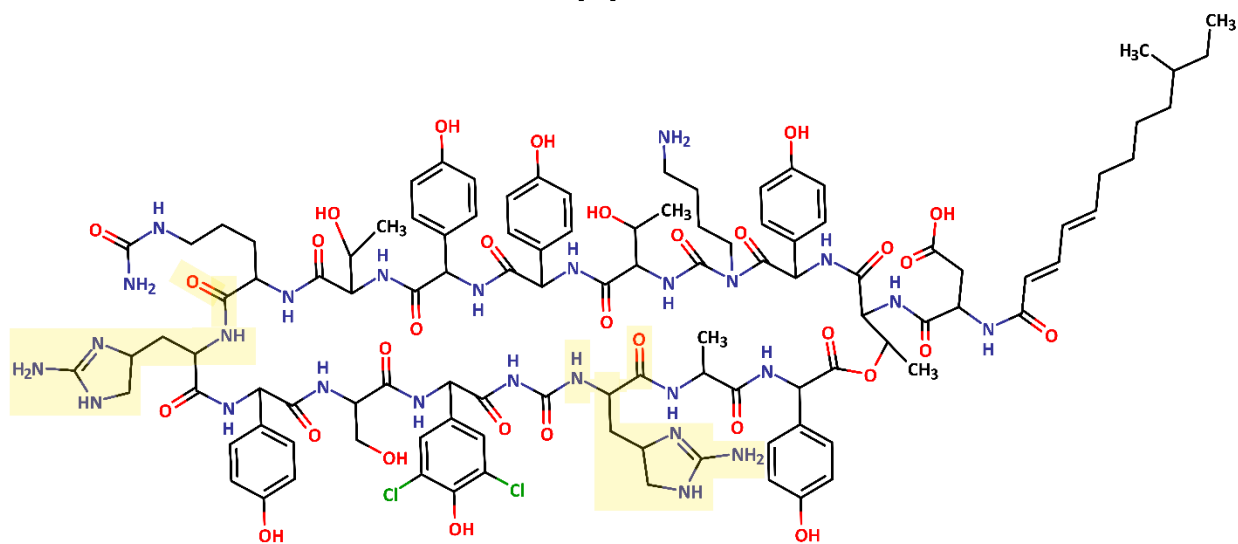
[62]



[63]



[64]

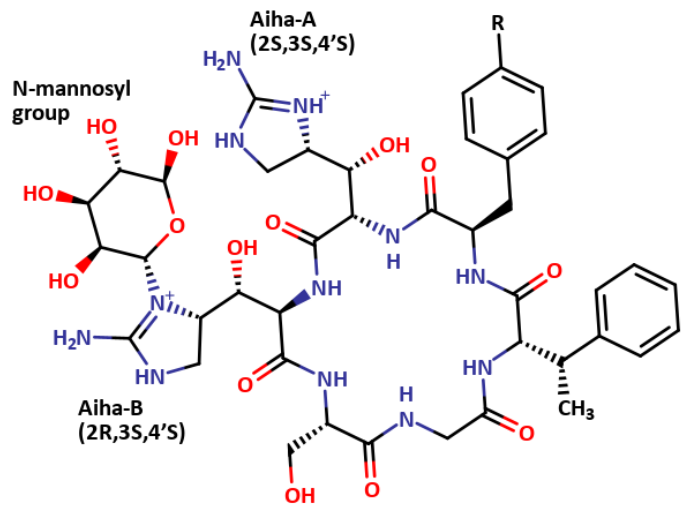


[65]

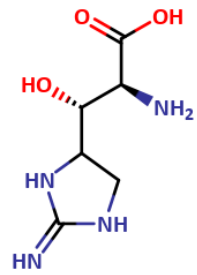
Scheme 1.1.8: structure of natural products containing L-arginine derivatives, both in flexible chain form and constrained form. Highlighted portion of the structures are derived from L-arginine, L-enduracididine or L-capreomycin. The highlighted residues are L-End (in 53-55) or L-Cap or 2S,3R-Capreomycin in Viomycin (Tubercatinomycin B or [56]).

1.2 A novel glycopeptide antibiotic “mannopeptimycin”

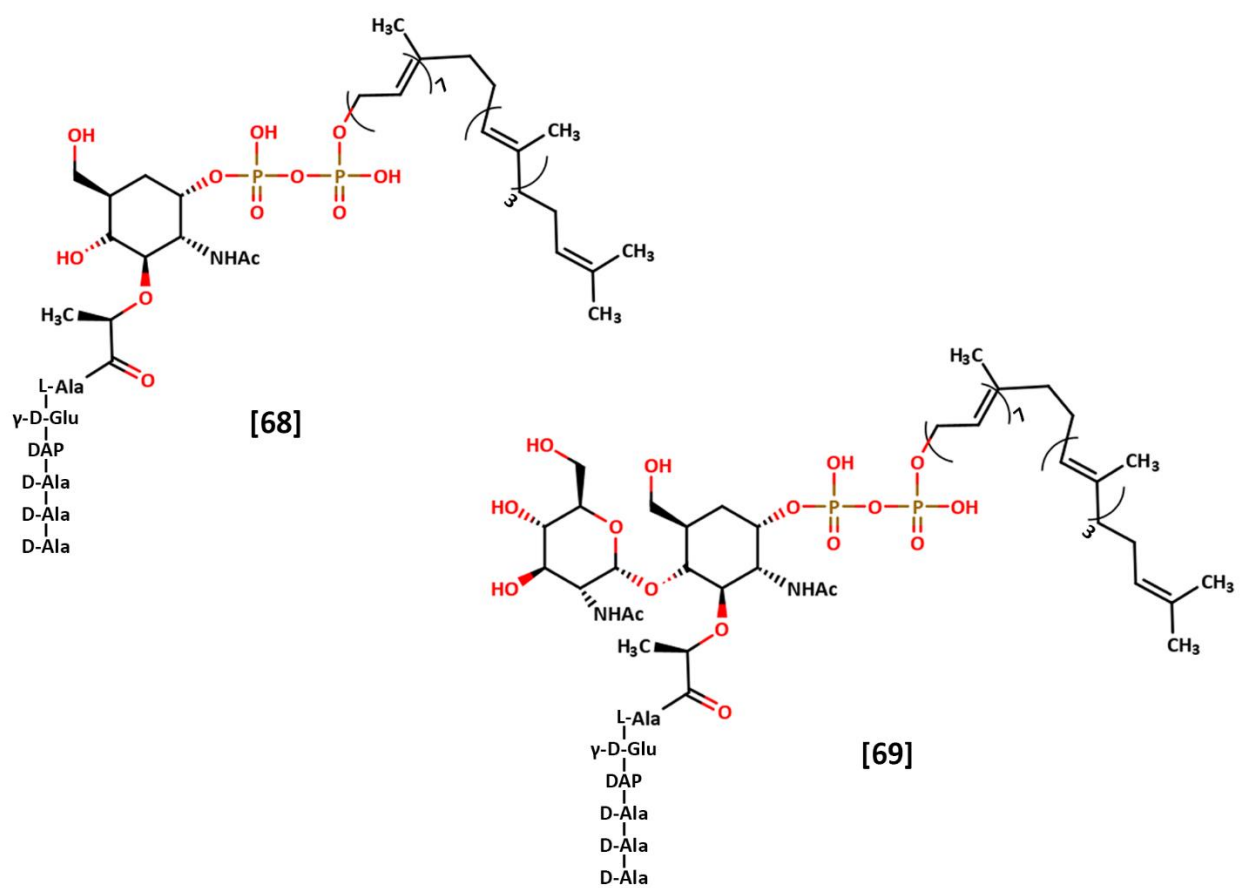
Over the last 5 decades, more than 95% of isolated *Staphylococcus aureus* worldwide have become penicillin resistant, and 60% of them are even methicillin resistant (according to WHO report)^{xlix}. As an ultimate solution to this problem, vancomycin (discovered in 1953) has had been used solution to fight infectious caused by *Staphylococcus aureus*^l. Unfortunately, a large group of enterococci, disquietingly continued to develop resistance to vancomycin^{li}. At the beginning of the 21st century (2002), a distinct class of glycopeptide antibiotics was discovered, which is effective against Gram-positive bacteria including vancomycin resistant *S. aureus* (VRSA) as well as methicillin-resistant *S. aureus* (MRSA)^{lii}. This novel glycopeptide antibiotic is called mannopeptimycin ([66] in scheme 1.2). Mannopeptimycin is a non-proteinogenic peptide antibiotic having glycosidic bond between 4-hydroxy group of D-mannose with nitrogen of cyclic guanidinium moiety present in Aiba-B (α -amino- β -[4'-(2'-iminoimidazolidinyl)]- β -hydroxypropionic acid) or β -hydroxy-L-enduracididine [67].



[66]

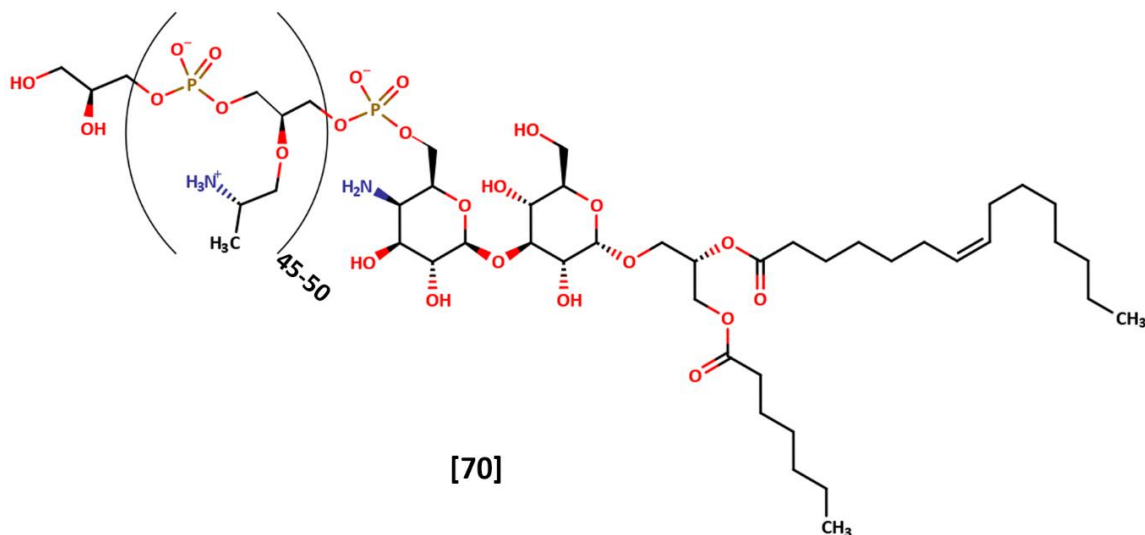


[67]



[68]

[69]



Scheme 1.2: Structures of compounds mentioned in section 1.3



Figure 1.3: Mannopeptimycin open reading frame. Black color gene represents transporter, blue for tailoring enzymes, red for non-proteinogenic amino acid formation, Green color represents NRPS (discussed previously in section 1.2), and gray stands for gene with unknown function, yellow color represents regulatory genes.

The mannopeptimycin biosynthetic gene cluster in *Streptomyces hygroscopicus*^{liii} is presented in figure 1.3. The products of the *mppJ*, *mppN*, *mppO*, and *mppP* genes are involved in non-proteinogenic amino acid biosynthesis. The *mppA* and *mppB* gene products are non-ribosomal

peptide synthetases. The rest of the genes encode tailoring, regulatory, transport proteins and resistance factors.

Similar to other glycopeptide antibiotics, mannopeptimycins also blocks bacterial cell wall synthesis via secondary inhibition of transglycosylase enzymes. It binds with the transglycosylase substrate, lipid-II [69], rather than binding with the enzyme itself^{liv}. Vancomycin inhibits cell wall synthesis by binding with the terminal D-Ala-D-Ala residue of growing polypeptide chain^{lv}. Mannopeptimycin does not compete with vancomycin and does not show any cross-resistance. As mannopeptimycin is active against vancomycin-resistant bacterial strains, it is believed that mannopeptimycin might interfere with the later stage of transglycosylation or obstructs nascent peptidoglycan maturation^{lvi}. However, it is possible that mannopeptimycin mechanism of action involves lipid-I [68] interaction rather than lipid-II [69], which would mean that mannopeptimycin acts before transglycosylase activity. Another possible target of mannopeptimycin activity is LTA (lipoteichoic acid [70]) carrier protein.

Structural analysis of mannopeptimycin [66] shows that, it has the unprecedented precursor amino acid β -hydroxy-L-enduracididine or Aiha [67]. It contains two different isomers of this non proteinogenic amino acid within the molecule, AihaA (2S, 3S, 4'S) and AihaB (2R,3S,4'S) ([66] in Scheme 1.2). The delocalized electrons in iminoimidazolidinyl ring provides resonance stability as well as scope of more electrostatic interaction. These properties might attribute to the efficiency of mannopeptimycin against vancomycin resistant bacteria.

1.3 PLP dependent enzymes

Pyridoxal 5'-phosphate (PLP) is a versatile cofactor, associated with about 4% of all known enzyme activities. Structure of PLP reveals its dynamic property to act as electron sink, enabling covalent bonding, hydrogen bonding and ionic interaction. Moreover, the pyridine ring is slightly basic, whereas phenolic group provides acidic property. Thus, PLP can easily lower the activation barrier of reactions by stabilizing intermediates^{lvii}. In the course of evolution, five different folds of PLP-dependent enzymes have emerged, where orientation of the active site residues helps the substrate to interact properly with the delocalized electrons of the cofactor. This orientation effect is responsible for reaction specificity^{lviii}. Moreover, the PLP-dependent enzymes provide stereoselectivity by optimizing transition state by maximizing the overlap of π -orbital. This extended π -orbital facilitates bond breaking by lowering activation energy. This phenomenon is also known as Dunathan's hypothesis (1966)^{lix}.

Depending on the types of reactions catalyzed, PLP-dependent enzymes are classified into 4 major groups: transaminase, decarboxylase, deaminase, racemase. During 2015, a unique class of PLP-dependent enzymes were discovered, which acts as both deaminase and oxidase^{lx} (discussed in topic 1.4.5).

1.3.1 PLP-dependent transaminase

In the transaminases, PLP forms a Schiff base with the catalytic lysine residue of the enzyme. This complex between the enzyme and PLP the cofactor is called the Internal aldimine ([71] Scheme 1.4). Upon substrate binding, a free amine on the substrate attacks the Schiff base linkage to the enzyme, displacing the catalytic lysine to form the external aldimine [72]. The α -H of the substrate is abstracted by the enzyme and the electron flows into the cofactor to form the quinoid [73]. The quinoid is re-protonated to form the ketamine [74], which is subsequently hydrolyzed to pyridoxamine phosphate (PMP) [75] and release the ketone product [76]. Then a new ketone substrate binds to the PMP-bound form of the enzyme, is attacked by the free amine of PMP to form a new external aldimine, and the entire series of steps runs in reverse to yield a new amino acid substrate. Aspartate aminotransferase (AAT) is an example of this class of enzymes^{lxi}. It takes Asp and α KG and produces oxaloacetate and Glu.

1.3.2 PLP-dependent decarboxylase

PLP-dependent decarboxylases have similar catalytic mechanism as PLP-dependent aminotransferases. The reaction mechanism of this class of enzymes also begins with enzyme-PLP complex or Internal aldimine [71]. Upon substrate binding, a schiff base forms between the PLP and substrate forming the external aldimine [77]. the substrate in the Decarboxylase is bound in a way so that the carboxylate is out of the plane of the PLP ring and thus made labile. Then the external aldimine destabilizes α -carboxylate group, resulting in

release of carbon dioxide and quinoid intermediate [78]. The rearrangement of quinoid results in release of decarboxylated product [79]. In some cases, decarboxylated product complex is further hydrolyzed into ammonia and hydrogen peroxide byproduct^{lxii}. For instance, L-DOPA decarboxylase is a PLP-dependent enzyme with akin reaction mechanism^{lxiii}.

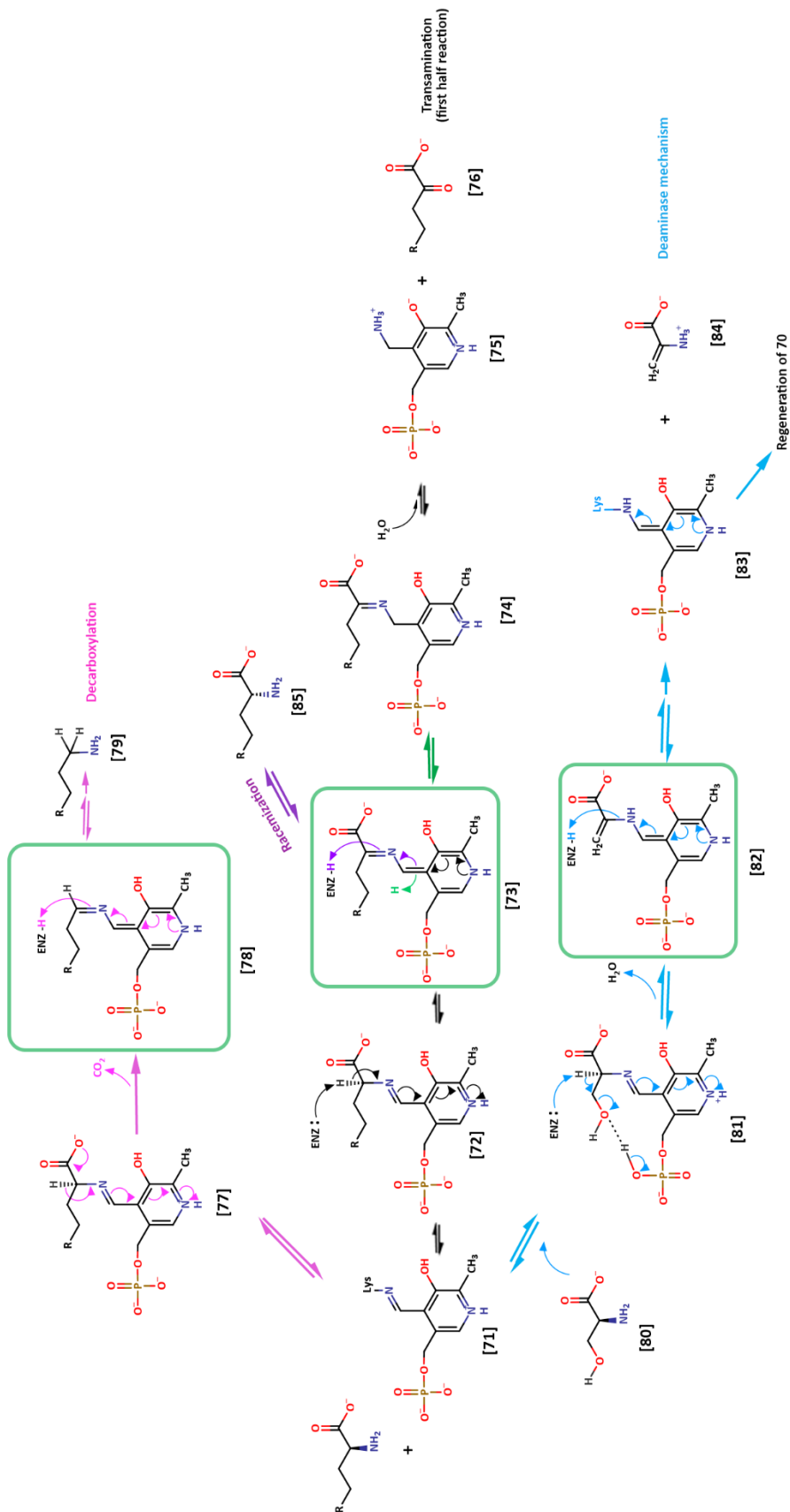
1.3.3 PLP-dependent deaminase

The mechanism of PLP-dependent deaminase are similar to that of transaminases and decarboxylases. There are very few substrates undergo this reaction. Serine dehydratase and threonine dehydratase is example of this class of PLP-dependent enzymes. Due to the presence of a beta-hydroxyl in the substrate[80], it introduces new possibilities into the standard sequence of events in the transaminase mechanism (Scheme 1.4 blue branch). Upon formation of the external aldimine [81], β -hydroxy group of the substrate forms hydrogen bond with phosphate group of PLP, destabilizing the bond, followed by release of water molecule. The following steps are similar to the transaminase reaction. The catalytic lysine residue [82] from the enzyme forms Schiff base [83], regenerating internal aldimine [71] followed by release of deaminated product [84]^{lxiv}.

1.3.4 PLP-dependent racemase

PLP-dependent racemase is prevalent to catalyze the conversion of L-amino acids into D-amino acids. Endogenous D-amino acids act as signaling molecule in both prokaryotes and eukaryotes. However, predominant enantiomer of amino acid is the L-form which occurs naturally. Thus, the amino acid racemase is an important enzyme of D-amino acid metabolism. For example, alanine racemase is such an PLP-dependent enzyme.

The mechanistic analogy of this class of enzyme involves two nucleophilic residues rather than one. The catalytic lysine residue forms the usual internal aldimine **[71]**, which eventually converts into external aldimine **[72]** upon substrate arrival. Soon after, a catalytic tyrosine residue from the B-chain (**ENZ-H** in **[73]**) forms a hydrogen bond with the α -hydrogen of the substrate in the external aldimine. Rearrangement of the hydrogen bond interaction results in protonation of the α -carboxylate group and deprotonation of α -amino group. At this point, re-orientation of the catalytic lysine results in exclusion of the B-chain tyrosine residue from the scaffold, followed by electrophilic attack by the ϵ -amino group of lysine on the α -carbanion of the substrate-PLP complex. Consequently, the deprotonated lysine residue acts as a Lewis base, donating an electron from the other side and leaving the D-amino acid product **[85]** as well as regenerating internal aldimine **[71]** ^{lxv}.



Scheme 1.3: Difference PLP-dependent enzymes' reaction mechanism.

1.3.5 MppP: A unique PLP-dependent L-arginine oxidase

L-Enduracididine biosynthesis in *Streptomyces wadayamensis* involves a unique PLP-dependent oxidase, SwMppP. The function of this enzyme is to convert L-arginine into 2-oxo-4(S)-hydroxy-5-guanidinovaleric acid ([95] in Scheme 1.4.5-2). In vitro, the enzyme produces a mixture of major product [95] and the abortive product 2-oxo-5-guanidinovaleric acid [91]. Apart from SwMppP, the L-enduracididine biosynthetic pathway involves two other enzymes, SwMppQ and SwMppR (Figure 1.4.5-1). SwMppR cyclizes the 2-oxo-4(S)-hydroxy-5-guanidinovaleric acid [95] to give the ketone form of L-enduracididine [96]. On the other hand, SwMppQ, a putative PLP-dependent aminotransferase converts the 2-Keto enduracididine into L-enduracididine [97].

The mechanism of SwMppP is interesting not only because it uses molecular oxygen to catalyze the 4-electron oxidation of L-arginine, but also it hydroxylates L-arginine at the un-activated γ -carbon. Although the molecular oxygen is reduced to H₂O₂, L-arginine is actually hydroxylated by H₂O (Scheme 1.4.5-2). These findings reveal that SwMppP is not an oxygenase, but an oxidase^{lxvi}.

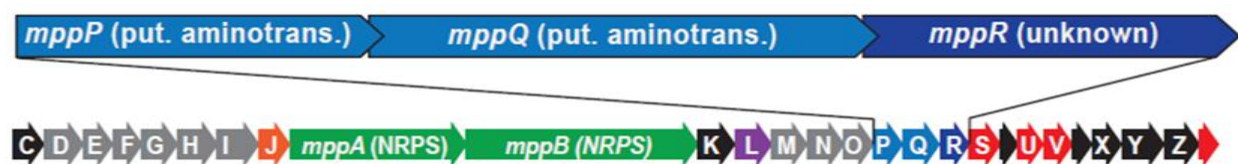
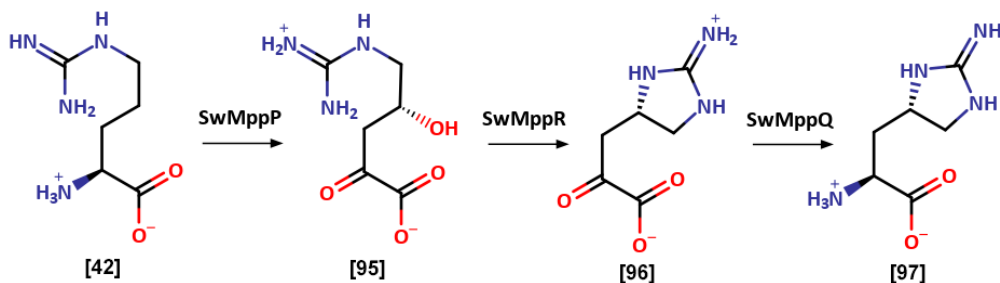


Figure 1.4.5-1: L-enduracididine biosynthetic gene cluster in *Streptomyces wadayamensis*, is a part of mannopeptimycin biosynthesis gene context.

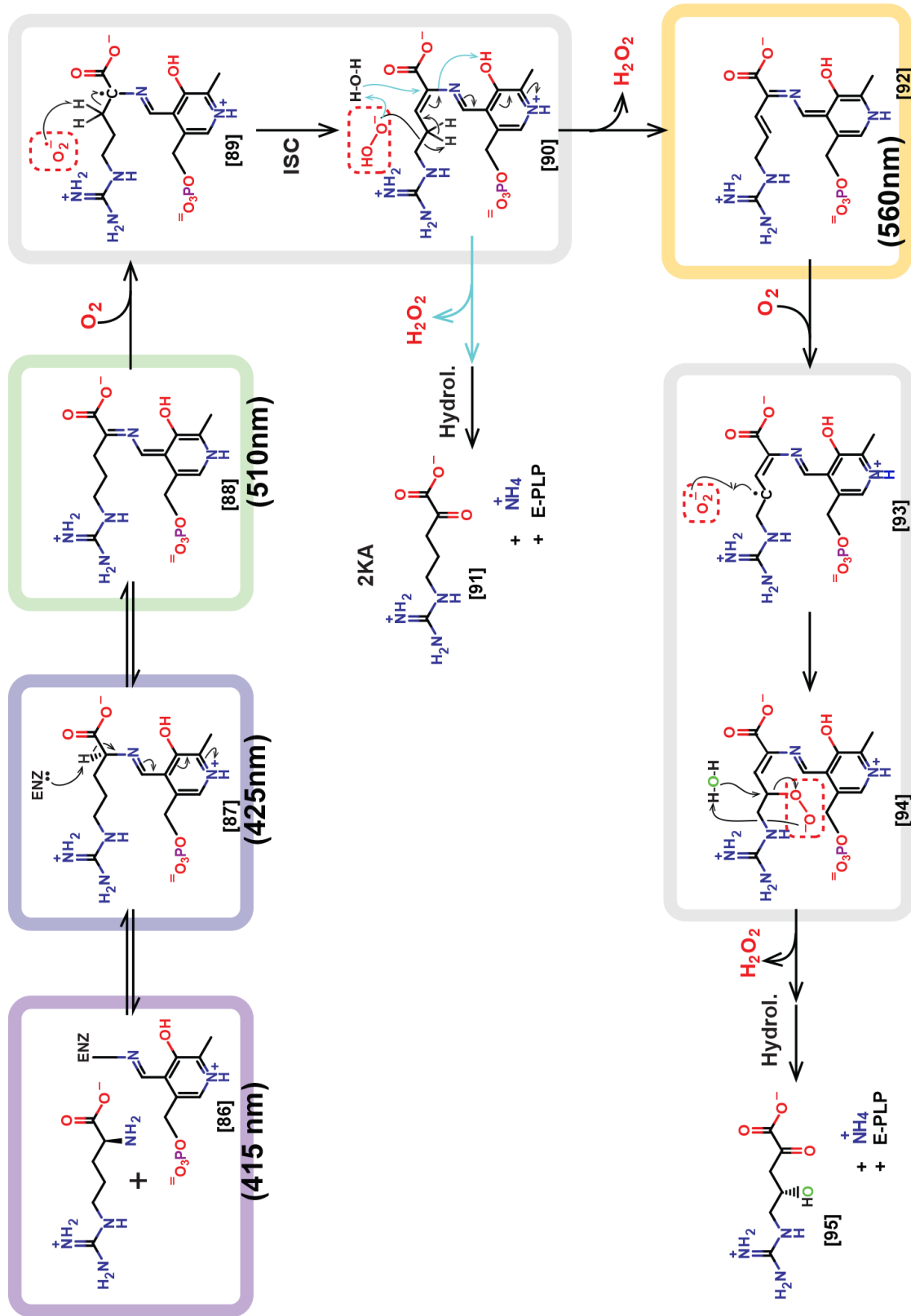


Scheme 1.4.5-1: L-enduracididine biosynthesis by Streptomyces wadayamensis

The mechanism is actually similar to other PLP-dependent enzymes, but the quinonoid is stabilized to an unusual extent (Scheme 1.4.5-2), lasting several hours in vitro. This stable quinonoid can react with O_2 when it finds its way into the active site. PLP forms Schiff base with catalytic lysine residue (lys-233) of enzyme thus PLP bound enzyme forms the internal aldimine **[86]**, giving absorbance at 415nm wavelength. Then the deamination of L-arginine followed by formation of PLP-substrate complex forms external aldimine **[87]**, giving absorbance at 425nm. At this point the MppP active site Lys-221 makes nucleophilic attack to the α -hydrogen of the L-arginine and resonance structure forms a quinoid intermediate **[88]** or Q1 that gives absorbance at 510nm. Electron transfer from Q1 to molecular oxygen could create superoxide anion that conceivably attacks the β -hydrogen **[89]** forming hydroperoxide, and eventually leaving Dehydro-arginine bound with PLP **[90]**. It could be a peroxidide or radical recombination. The exact mechanism is still unknown as the intermediates are short lived. But the first molecule of O_2 definitely results in a partial oxidation of L-arg.

From this point, MppP reaction follows two branches. The first one leads towards major product 2-oxo-4(S)-hydroxy-5-guanidinovaleric acid [95] formation, where the hydroperoxide attacks the γ -hydrogen in L-arginine, producing first molecule of H_2O_2 byproduct and quinoid-II complex [92] or Q2, that fluoresce at 560nm. A 2nd molecule of O_2 forms radical, attacking the activated γ -C [93]. Finally, water molecule likely neutralizes the reaction [94] by producing second molecule of H_2O_2 . Oxygen from the water molecule incorporates ultimately with the substrate to form 2-oxo-4(S)-hydroxy-5-guanidinovaleric acid [95].

On the contrary, the reactive peroxide anion complex follows a second pathway using a water molecule, where the peroxy-anion attacks the water molecule's H to produce H_2O_2 and MppP minor product 2-oxo-5-guanidinovaleric acid [91].



Scheme 1.4.5-2: proposed reaction mechanism catalyzed by Mppp

Chapter 2

Beyond Enduracididine Biosynthesis I: Exploring the biochemical context of MppP activity in a putative operon in *Pseudomonas brassicaceum*

2.1 Background

As discussed in topic 1.1.7, the conformationally restricted form of arginine L-enduracididine (L-End) is present in a number of antibiotics, and the first step in the biosynthesis of this nonproteinogenic amino acid is the oxidation of L-arginine by MppP. Work on an MppP homolog by the Ryan laboratory^{lxvii} showed that the L-arginine oxidase activity of MppP operates in at least one other biochemical context beyond L-End biosynthesis. We were interested to know how diverse the biochemical contexts of the MppP activity are. To find out, we did a PHI-BLAST search^{lxviii} using a sequence motif specific for MppP homologs (T-x-x-E-x(10,20)-H-x(150,170)-D-x(2)-F-x(20,30)-[ED]-D-[TS]-G-K). This motif was identified by comparing the sequences of three known MppP homologs with those of a small number of fold type I PLP-dependent enzymes (*e.g.* aspartate aminotransferase, kynurenine aminotransferase, and LL-diaminopimelate aminotransferase). The PHI-BLAST search returned 198 MppP homologs, which were analyzed using MEGA-X^{lxix} to generate a distance tree (Figure 2.1-1). Manual analysis of the gene contexts of all 198 identified MppP homologs shows that they group into a three-lobed family of proteins. For the most part, the clusters, or clades, on the distance tree each have a consistent genomic context, but there is considerable variation in genomic contexts from clade to clade.

For example, starting in the lower-left lobe on the distance tree, the MppP homologs in the red clade are flanked by the genes for β -hydroxy-enduracididine production, MppQ, MppR and MppO (discussed in topic 1.3 and 1.4.5) and are likely all mannopeptimycin biosynthetic clusters. SwMppP belongs to this clade (the yellow circle in the red clade). The sequences in the neighboring clade (maroon) have a similar genomic context where the MppP gene is flanked by MppQ and MppR genes, indicating that these MppP homologs are involved in L-End production in the context of enduracidin biosynthesis. The MppP homolog from *Verrucosispora maris* (VmMppP; 62 % identical to SwMppP, highlighted yellow) is an example. The MppP-like sequences in the orange clade are all associated with genes annotated as “agmatinase family”, SAM-dependent methyltransferase, N-acetyltransferase, and two hypothetical proteins. The MppP homolog from *Streptomyces grecciofuscus* (SgMppP; 50 % identical to SwMppP) is an example of this group. Finally, the MppP homolog from *Micromonospora chokoriensis* (McMppP; 44 % identical to SwMppP) belongs to the green branch, where the MppP homolog is flanked by genes similar to *vioC* and *vioD* of the viomycin biosynthesis pathway.

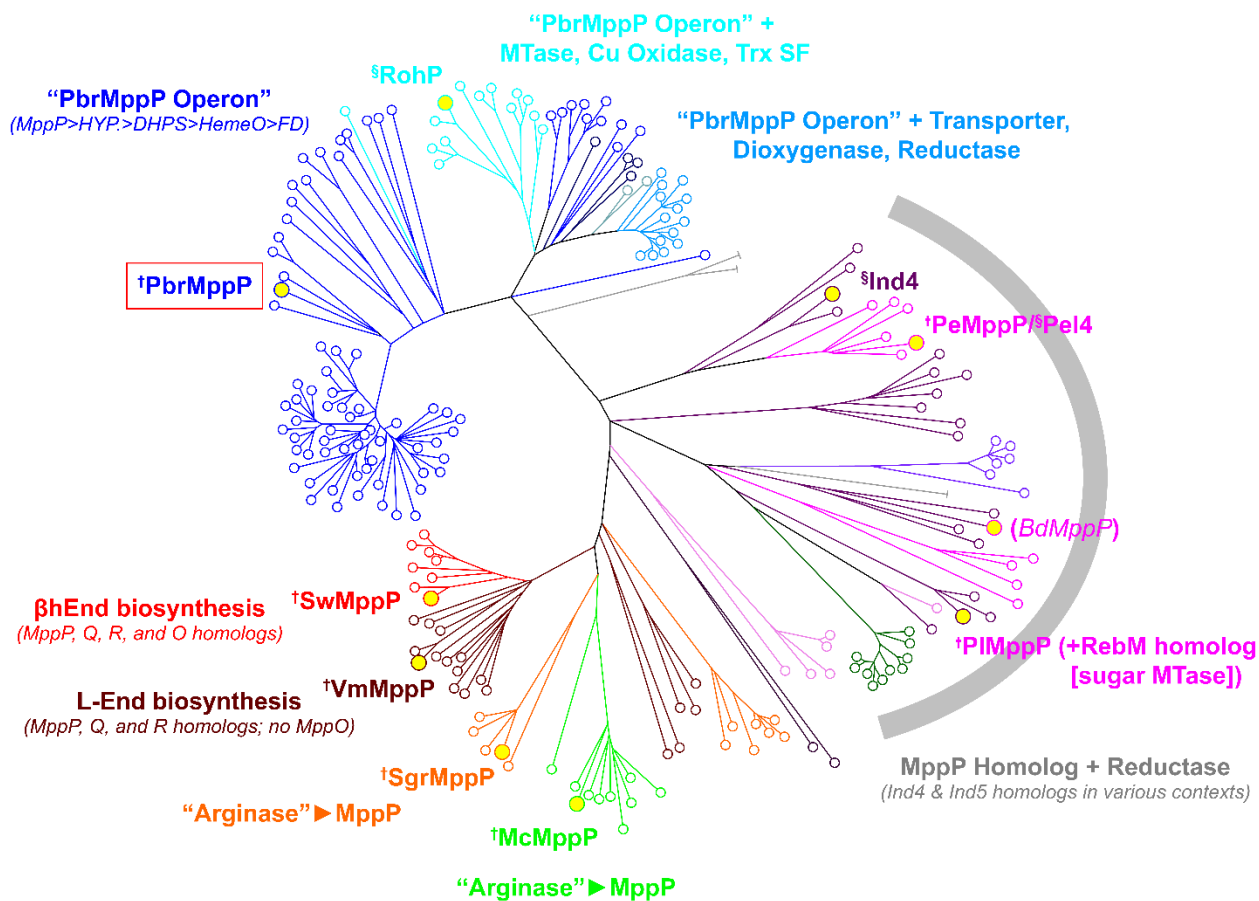


Figure 2.1-1: Genetic distant tree of *MppP* homologs constructed from PHI-BLAST

The *MppP* homologs in the lower-right lobe (marked with a gray semicircle) are all associated with an NAD(P)H-dependent reductase and are similar to *Ind4*, which is involved in indolmycin biosynthesis^{lxx, lxxi}. The *MppP* homologs from *Pseudomonas elgii* (*PeMppP*) and *Bartonella doshiae* (*BdMppP*) are *Ind4* homologs. In contrast, the homolog from *Pseudoalteromonas luteoviolacea* (*PIMppP*) is associated with a slightly different genomic context, which also includes a gene homologous to the rebeccamycin biosynthetic gene *rebM*, the product of which is a sugar O-methyltransferase^{lxxii}. As rebeccamycin contains no obvious guanidine-derived

moiety, these homologs are likely involved in the biosynthesis of a different, as-yet-uncharacterized natural product.

The sequences in the top lobe (blue and cyan clades) is dominated by a single genomic context, where the MppP homolog is flanked by 6 open reading frames encoding a LuxR-type regulatory protein, 4 putative enzymes, and an accessory protein (Figure 2.1-2). The *Pseudomonas brassicacearum* MppP homolog (PbrMppP; highlighted in Figure 2.1-1) is the prototype of this group. Interestingly, all these genes have 4 to 6 bp overlapping region at the end, which is a compelling evidence to say that they are co-expressed under same promoter as well as they can be exalted as “Operon”^{lxxiii}. The 500bp region between the genes encoding the LuxR and MppP homolog, contains a putative promoter sequence. On the opposite strand, in the region of putative promoter, is an acylhomoserinelactone synthase (AHLS) which has quorum-sensing ability in response to fluctuations in cell-population/density^{lxxiv}. The additional open reading frames encode 5 non regulatory genes, predicted as PbrMppP, a hypothetical protein (PbrHYP), a dihydrodipicolinate synthase family protein (PbrDHPS), a mononuclear Fe(II)- and α -keto glutarate-dependent oxygenase (PbrOx), and a 2Fe2S-binding protein similar to ferredoxin (PbrFD).

The cyan clade contains MppP homologs similar to the enzyme RohP, which has been characterized by the Ryan group^{lxxv} and flanked by methyltransferase, Cu-oxidase and thioredoxin genes. RohP is an enzyme involved in polyketide biosynthesis, which converts rhodomycin-D into 10-carboxy-13-deoxy caminomycin^{lxxvi}. To conclude, the presence of MppP

homolog in genomic context lacking MppQ and MppR homologs, suggests that the oxidase activity of MppP is more broadly employed than in simply making L-End.

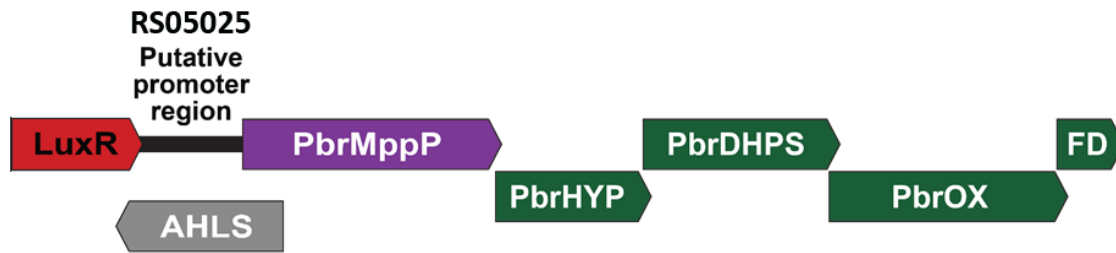
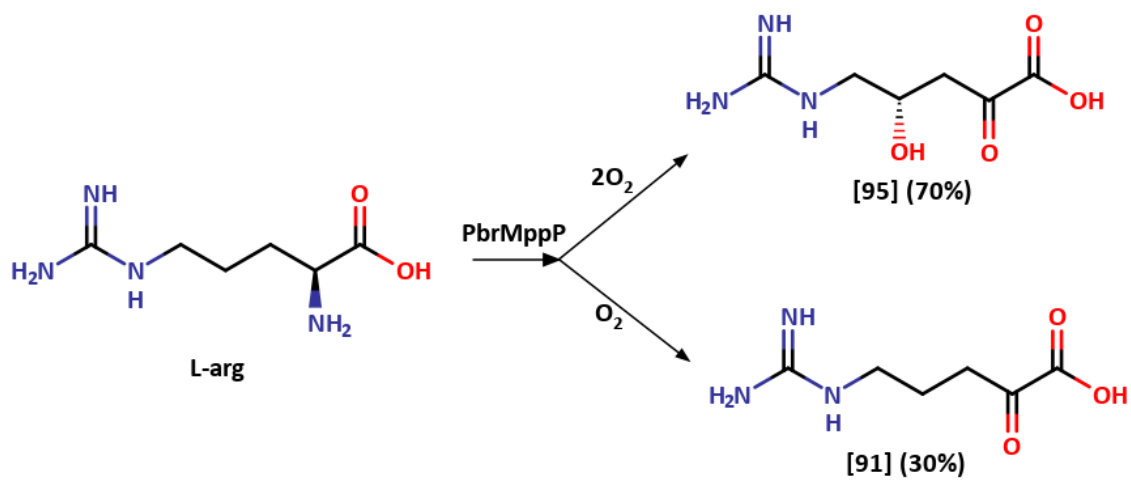


Figure 2.1-2: *MppP*-containing putative operon in *Pseudomonas brassicacearum*. The *LuxR* gene (red) encodes a regulatory protein, followed by a putative promoter region. The gene labeled “AHLS” (acylhomoserinelactone synthase) is encoded by the opposite stand. Downstream of the putative promoter region, is the *MppP* homolog gene, followed by four genes predicted as hypothetical protein (*PbrHYP*), dihydrodipicolinate synthase family protein (*PbrDHPS*), mononuclear Fe(II)- and α -ketoglutarate-dependent oxygenase (*PbrOx*), and a 2Fe2S-binding protein (similar to ferredoxin, *PbrFD*).



Scheme 2.1-1: Reaction catalyzed by PbrMppP. The major product of MppP reaction is 2-oxo-4(S)-hydroxy-5-guanidinovaleric acid [95], the minor product is 2-oxo-5-guanidinovaleric acid [91].

Table 2-1: Gene accession codes, sequence annotations, and identity to closest characterized homolog of PbrMPCO proteins.

Gene Accession	Protein	Functional prediction	Sequence Identity(%)
CD58_RS05015	LuxR	LuxR type TX regulator ^{lxxvii, lxxviii} in <i>Pseudomonas chlororaphis</i> O6	72 %
RS05020 (Opposite strand)	AHLS	<i>Pseudomonas chlororaphis</i> PCL1606 Acylhomoserinelactone Synthase ^{lxxix}	53 %
CD58_RS05025	PbrMppP	<i>Streptomyces wadayamensis</i> ^{lxxx} PLP-dependent L-arginine Oxygenase	33%
CD58_RS05030	PbrHYP	Hypothetical protein, closest homology with VapC, the toxin component of the virulence-associated protein complex from <i>Mycobacterium tuberculosis</i> ^{lxxxi}	9.8%
CD58_RS05035	PbrDHPS	<i>E. coli</i> Dihydrodipicolinate synthase ^{lxxxii}	26%
CD58_RS05040	PbrOX	<i>Pseudomonas aeruginosa</i> PAO1 Mononuclear Fe(II)- and α -ketoglutarate-dependent oxygenase ^{lxxxiii}	56%
CD58_RS05045	PbrFD	Rieske Fe-S domain-containing reiske protein in <i>Pseudomonads</i> ^{lxxxiv}	90%

PbrMppP and SwMppP are similar in both their structures and mechanisms. PbrMppP has 33% sequence identity with SwMppP. However, the superimposed structures of these two homologs has a small RMSD value of 1.2 Å for all C α atoms. This similarity in the structures extends from the overall fold to the detailed active site architecture. Figure 2.1-3 shows the overlay of SwMppP and PbrMppP and highlights the similarities in the orientations of all of the key active site residues as well as the bound product [95].

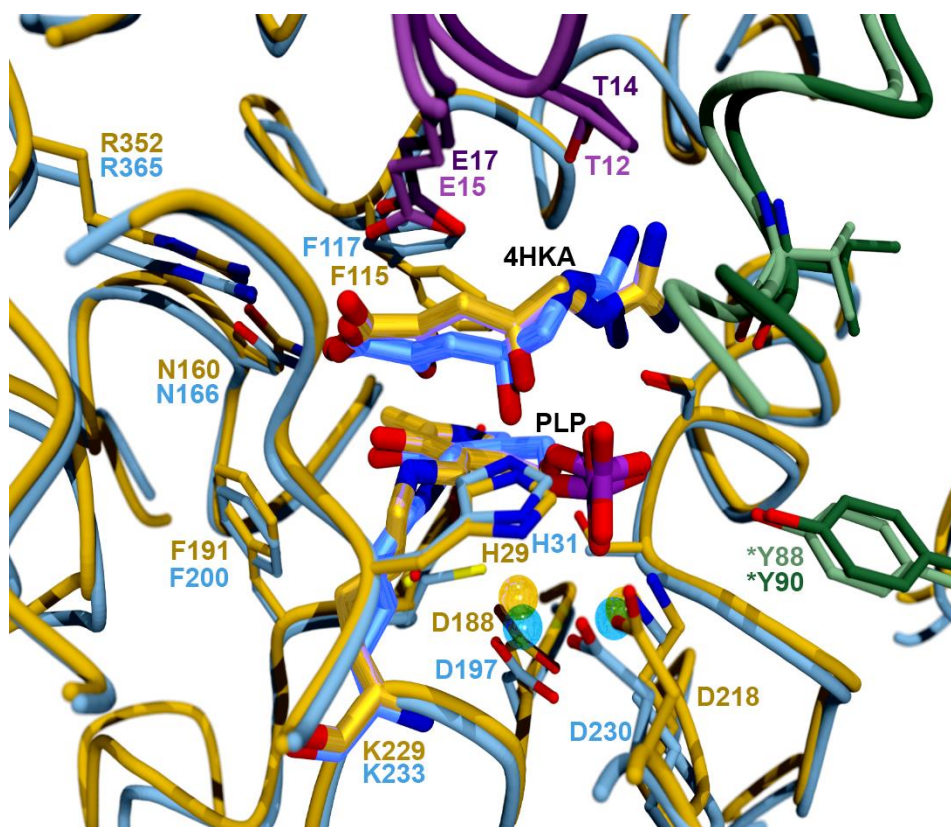


Figure 2.1-3: The PbrMppP active site structure overlaid with SwMppP (PDB ID: 6C9B). The dark purple and light purple section is coming from N-terminus of PbrMppP and SwMppP, respectively.

The main chain trace and side chains of PbrMppP are colored blue (carbon atoms only), while SwMppP is colored yellow. The labels correspond to the structure colors. The dark green and light green residues belong to the other chain of PbrMppP and SwMppP dimers, respectively.

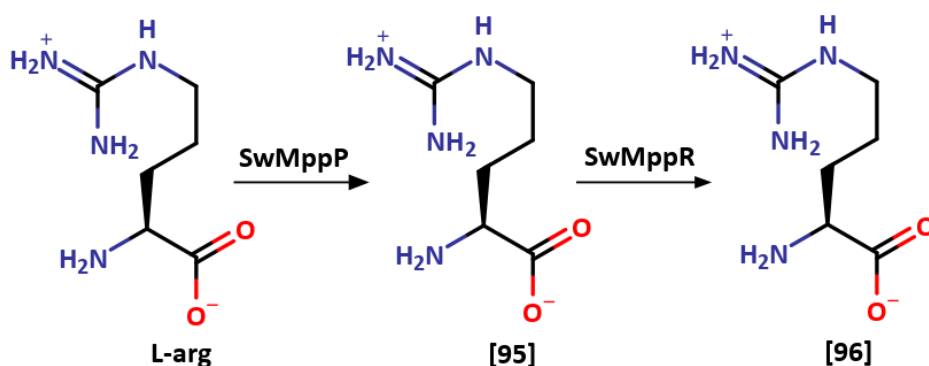
The N-terminal residue of Threonine-12 and glutamate-15 is essential for hydroxylation and catalytic efficiency of the substrate^{lix}. In SwMppP, when these N-terminal residues are mutated to eliminate hydrogen bonds to the substrate, or the N-terminal 22 residues are deleted, the SwMppP variants can only produce 2-ketoarginine. Asn160 and Arg352 make hydrogen bonding and electrostatic interactions, respectively, with the carboxylate group of [95] (Figure 2.1-3). The side chain carboxylate group of Asp218 forms a hydrogen bond with the catalytic Lys229 amino group in the external aldimine forms of the enzymes (Scheme 1.4.5-2). Moreover, His29 is crucial as acid-base catalyst^{lixv}.

The next gene in the PbrMppP gene context is PbrHYP, with 51% sequence identity with VapC, the toxin half of the virulence-associated protein toxin-antitoxin system in *Mycobacterium tuberculosis*

(Vap-BC complex)^{lixxv}. VapC possess a (PIN) domain fold, which, in VapC, contains a divalent metal center and is associated with ribonuclease activity^{lixxvi}. We show here that PbrHYP shares the PIN domain fold, but the loss of several acidic residues in the putative active site of PbrHYP has eliminated the metal-binding site, and presumably, the ribonuclease activity.

PbrDHPS shares 26% sequence identity with the *bona fide* *E. coli* DHPS. The function of DHPS in *E. coli* is to condense pyruvate with aspartate semi-aldehyde (ASA) to form 4-hydroxy-tetrahydro-

picolinic acid (HTPA), an intermediate that dehydrates non-enzymatically to yield 2,3-dihydrodipicolinate. The product of *E. coli* DHPS is used in the L-lysine and *meso*-diaminopimelate synthesis pathways in plants and bacteria^{lxxxvii}. Interestingly, SwMppR, one of the 3 enzymes that, together with SwMppP generates L-End in *S. wadayamensis*, catalyzes an intramolecular cyclization of [95] to give [96]. The chemistry uses a catalytic lysine residue and a Schiff base intermediate and is thus like the type I aldolases. This observation leads to the possibility that, while PbrDHPS may be a true aldolase like EcDHPS, it might also catalyze an intramolecular modification of some intermediate similar to SwMppR.



Scheme 2.1-2: Reactions catalyzed by SwMppP and SwMppR. SwMppP catalyzes conversion of L-arg [42] to 2-oxo-4(S)-hydroxy-5-guanidinovaleric acid [95], SwMppP catalyzes conversion of [95] to 2-ketoendutacididine [96]

The next enzyme in the putative operon, PbrOX, is a mononuclear Fe^{2+} - and α -ketoglutarate-dependent oxygenase homolog. Its sequence identity is 33% with the non-heme iron oxygenase UndA from *Pseudomonas aeruginosa* PAO1^{lxxxviii}. We show here that Fe(II) loading onto PbrOX is necessary for catalytic activity. Finally, PbrFD is a putative Fe-S cluster protein 90% identical to Toluene 1,2-dioxygenase system ferredoxin in *Pseudomonas putida*^{lxxxix}. The co-existence of PbrOX and PbrFD in the same putative operon suggests they belong to a possible three-component oxygenase system. The function of three component oxygenase systems^{xc} in bacteria is to catalyze redox reactions (figure 2.1-4) where the first component is a NAD(P)H-dependent ferredoxin-reductase (FNR)^{xcj}. FNR oxidizes NAD(P)H to NAD(P)⁺ and transfers the electron to the second component in the system, the ferredoxin (FD). Finally, the reduced FD (hypothetically, PbrFD in the PbrMPCO gene context), delivers the electron to the metal center of the third component, a terminal oxygenase (perhaps PbrOX here), to initiate catalysis.

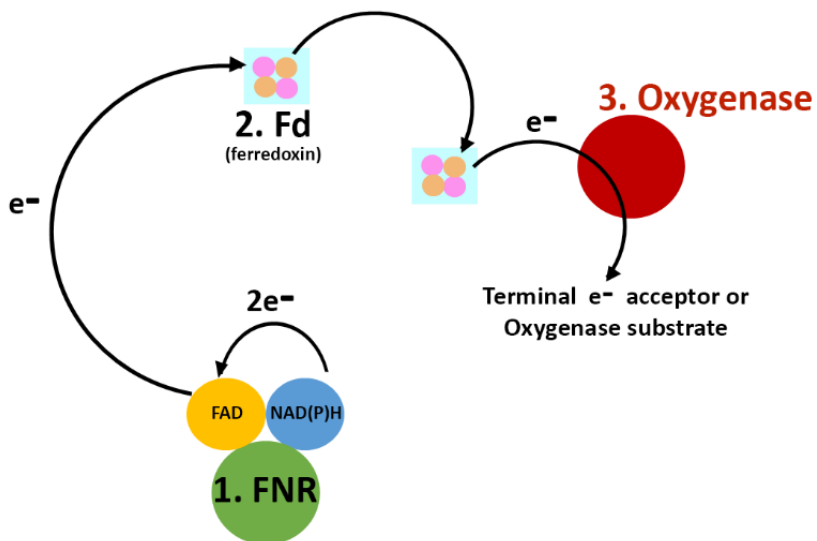


Figure 2.1-4: Bacterial three-component oxygenase system. FNR, the first component that oxidizes NAD(P)H into NAD(P)⁺ and transfers the electron to the second component, ferredoxin. The FD then transfers the electron to the terminal oxygenase, where the electron is utilized to reduce the metal center of the oxygenase, readying the system for catalysis.

Interestingly, there are only two annotated NAD(P)H-dependent ferredoxin reductases in the *Pseudomonas brassicacaerum* DF41 genome. Either or both of these FNR enzymes could work with PbrFD and PbrOX to catalyze the oxidation of an as-yet-unknown substrate.

In this thesis, we show the structural characterization of PbrMppP, PbrHYP, PbrDHPS, and PbrOX, along with kinetic studies and substrate specificity determinations of PbrMppP. At the end of this chapter, we will also characterize the functions of PbrOX and PbrFD to show their interdependence as part of three-component oxygenase system.

2.2 Methodology

2.2.1 Cloning, expression and purification of the PbrMPCO Enzymes

The coding sequences of PbrMppP, PbrHYP, PbrDHPS, PbrOX and PbrFD were optimized for expression in *E. coli* and synthesized by GenScript Inc. (Piscataway, NJ). The synthetic genes were cloned into the pE-SUMO_{kan} expression vector. PbrMppP and PbrFNR were cloned into the pET-TEV_{amp} expression vector by GenScript Inc.

Table 2.2.1-1: Primers used for cloning Pbr proteins

Names	Primer sequences
PbrHYP_F	5'-GGTCTCAAGGTATGGCGGAGGCGAGCATC-3'
PbrHYP_R	5'-GCTCTAGATCATTAGCCAACGTGTTGAAAGTCTTCTTTA-3'
PbrDHPS_F	5'-GGTCTCAAGGTATGAGCGGCAAGCACATCT-3'
PbrDHPS_R	5'-GCTCTAGATCATTACGCCATAACTTCTTGCTCCG-3'
PbrOX_F	5'-GGTCTAAGGTATGAGCGACCTGCAGAAGGC-3'
PbrOX_R	5'-GCTCTAGATCATTAATCGCACGTTTCGATCATACGC-3'
PbrFD_F	5'-GGTCTCAAGGTATGAACGAGCCGAGCGAATTCT-3'
PbrFD_R	5'-GCTCTAGATCATTAGCTCAGACGTTTCACTTCCAGAC-3'
RS05025_pProbeNT_HindIII_F	5'-GGTGGTAAGCTTCTTGAAGAAGTTTCGAAACGGAC-3'
RS05025_pProbeNT_EcoRI_R	5'-GTTGTTGAATTCATTGTTATTCCTATAGGTGCTAAAGATATTT-3'
GFP_F (for sequencing)	5'-GGTCCTTCTTGAGTTTGTAAC-3'
GFP_R (for sequencing)	5'-CCATCTAATTCAACAAGAATTGGGACAAC-3'

All proteins were expressed using similar protocols. The His₆-SUMO fusion proteins were expressed from *E. coli* BL21 Star (DE3) cells (Invitrogen Inc., Carlsbad, CA) carrying one of the various pE-SUMO plasmids. Cultures were grown in Luria-Bertani medium with 50 µg/mL kanamycin at 37 °C. Cells containing plasmid pE-SUMO-PbrFD were cultured in Terrific Broth (TB) containing 0.1 mg/ml ferric ammonium citrate, 1 mM cysteine, and 50µg/ml Kanamycin at 37°C. When the cultures reached an OD_{600nm} of ~0.7, protein expression was induced with 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The temperature was reduced to 25°C, and the cultures were grown overnight with shaking at 250 rpm. Cells were harvested by centrifugation at 6000 rpm for 20 minutes. The cell pellets were collected and resuspended in 5 mL/g cell paste of buffer A (25 mM TRIS pH 8.0, 300 mM NaCl, 10 mM imidazole) supplemented with 0.1 mg/mL DNase I (Worthington Biochemical Corp., Lakewood, NJ). In the case of PbrMppP, buffer A included 20 µM PLP to stabilize the enzyme. The cells were lysed using a Branson Sonifier S-450 cell disruptor (Branson Ultrasonics Corp, Danbury, CT) for a total of 10 min at 60% amplitude with 30 s pulses separated by 30s rest periods. The temperature was maintained at or below 4 °C by suspending the steel beaker in an ice bath directly over a spinning stir bar. The lysates were clarified by centrifugation at 18,000 rpm for 45 min and then applied to a 5 mL HisTrap column (GE Lifesciences) at a flow rate of 5 mL/min to isolate the His₆-SUMO fusion proteins. The proteins were eluted by a four-step gradient of buffer B (25 mM TRIS pH 8.0, 300 mM NaCl, and 250 mM imidazole; 5, 15, 50, and 100%). The His₆-SUMO fusion proteins were eluted in the third and fourth steps and were all >90% pure, as judged by Coomassie-stained sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE). Peak fractions were pooled and dialyzed overnight at 4°C against 4 L of 25 mM TRIS pH 8.0, 150 mM NaCl, and ~3 µM SUMO protease

(LifeSensors Inc). In the case of PbrMppP, the dialysis buffer contained 100 μ M PLP to keep the protein active. The dialysates were passed through the HisTrap column a second time to remove the cleaved His₆-SUMO tag as well as the protease. The resulting protein preparations were >95% pure. The proteins were desalted using a HiTrap desalting column (GE Lifesciences) into 10.0 mM MES pH 6.7 and 20 μ M PLP and stored at -80°C . All other enzymes were purified by the same protocol, with the omission of 20 μ M PLP from the buffers. Selenomethionine-labeled PbrHYP and PbrDHPS were expressed from T7 Express Crystal cells (New England Biolabs, Ipswich, MA) carrying the pE-SUMO-PbrHYP or pE-SUMO-PbrDHPS plasmids. Cells were grown in Selenomethionine Medium Complete (Molecular Dimensions, Newmarket, Suffolk, U.K.) with 50 μ g/mL kanamycin. All other aspects of the expression and purification were identical to those described for the unlabeled proteins.

2.2.2 Differential scanning fluorimetry

The purified enzymes were used for differential scanning fluorimetry (DSF), or “thermofluor assay”. For the thermofluor assay, 200 μ L of Spyro Orange dye dilute solution (50X) were prepared by adding 2 μ L of 5000x Spyro Orange to 198 μ L of deionized water and 25 μ L were portioned into each vial of 8 PCR tubes. Similarly, dilute protein solutions (1 mg/mL) were prepared and 25 μ L were portioned into each vial of 8 separate PCR tubes. Using an 8-channel pipet, 11 μ L of deionized water were placed into each of the wells in a 96 well Eppendorf TwinTec plate. Then 2 μ L of the prepared dilute protein solution was added, followed by 5 μ L of the Slice pH kit (Hampton Research, CA). Finally, 2 μ L of the 50X Spyro orange dye solution were

added and the samples were mixed by carefully pipetting up and down twice. The plates were immediately sealed with an optically clear sealing film and centrifuged for 30s at 1000 rpm to collect the samples in the bottoms of the wells. The sealed plates were placed in the RT-PCR instrument and equilibrated to 25°C for 5 minutes, in order for the temperature of the plate and reagents to equilibrate with the block as well as allow the reagents to diffuse and mix. Finally, the thermofluor assay was performed from 25°C to 95°C with a temperature ramp of 1°C per minute. Based on the results of differential scanning fluorimetry experiments with the Slice pH screen (Hampton Research,CA), buffers with the highest T_m values were selected for the storage buffers and purified enzymes were stored accordingly (refer to Table 2.2.2-2).

Table 2.2.2-2: Storage buffer of PbrMPCO enzymes

Protein	Storage buffer
PbrMppP	10mM MES pH 6.7, 20 μM PLP
PbrHYP	10mM BisTris pH 7.5
PbrDHPS	10mM Succinate pH 6.0
PbrOX	10mM HEPES pH 7.4
PbrFD	10mM BisTris pH 6.3
PbrFNR	10mM BisTris pH 6.3

2.2.3 Fe^{II} loading on PbrOX

After cleavage of the His₆-Tag, PbrOX was diluted with storage buffer (10mM HEPES pH 7.4) to ~2 μM. Small amounts of FeSO₄ (~1mg) and ascorbate (~1mg) were taken and mixed together dry by crushing with the tips of two spatulas. The crushed mixture was added into the dilute enzyme and incubated at 4°C for 4 hours^{xcii}. After the incubation period, the enzyme mixture was centrifuged at 4000 rpm for 20 minutes to remove any insoluble material. The supernatant was used to check the spectrum of Fe(II)-bound PbrOX and also used in steady state assays on the Clark electrode (O₂ probe) with PbrFD and PbrFNR.

2.2.4 Confirmation of PbrMppP activity

A Hewlett-Packard 8453 diode array spectrophotometer was used to collect spectra from a 500 μL reaction containing ~20 μM PbrMppP with 1 mM L-Arg and oxygen-saturated buffer (20 mM BisTris propane pH 9.0). The scans were taken every 0.5 s from 300 nm to 600 nm.

2.2.5 NMR characterization of PbrMppP reaction products

To perform reactions for NMR analysis, PbrMppP was desalted into 20mM sodium phosphate buffer pH 8.4 using a HiTrap desalting column (GE Lifesciences). Similarly, 1 mL of 1mg/mL catalase from bovine liver (Sigma-Aldrich, St. Louis, MO) was desalted into deionized water to remove the trehalose stabilizer from the catalase preparation. Catalase was added to remove the

hydrogen peroxide that is produced as a by-product of MppP catalysis, since it leads to non-enzymatic oxidative decarboxylation of the products [95] and [91]. Reactions were assembled in 50mL corning tubes by adding 724 μL of 277 μM desalted PbrMppP, 2 μL of 500 $\mu\text{g}/\text{mL}$ desalted catalase, and 5mL of 4mM L-Arg stock solution prepared in 20mM sodium phosphate buffer pH 8.4. The pH of all solutions was adjusted using a Fisherbrand™ accumet™ AE150 Benchtop pH meter. The volume of each reaction mixture was adjusted to 10mL by adding 4.7mL of phosphate buffer to reach final concentrations of approximately 20 μM PbrMppP, 2mM L-Arg and 1 $\mu\text{g}/\text{mL}$ catalase. Reaction tubes were shaken at 25°C, 200 rpm for 4 hours and protected from light. At the end of the reaction time, samples were evaporated to dryness using a Centrifan PE Rotary Evaporator (KD Scientific, Holliston, MA) at 50°C overnight. The resulting residue was taken up in 600 μL D₂O and centrifuged at 17,000 x g for 10 min to remove any remaining particulates. The supernatants (500 μL) were loaded into NMR tubes for ¹H NMR analysis to determine the structure(s) of the reaction product(s).

2.2.6 Characterization of PbrMppP reaction products by mass spectrometry

Reactions were performed in 1.5 mL Eppendorf tubes containing 1 mL of total volume. From a 4mM L-Arg stock in Phosphate buffer (titrated down to 8.0, 250 μL were added for a final concentration of 1mM L-Arg. From a catalase (Sigma-Aldrich, St. Louis, MO) stock solution of 500 $\mu\text{g}/\text{mL}$ (desalted into deionized water as described in section 2.2.3), 1 μL was added to each reaction for a final catalase concentration of 0.5 $\mu\text{g}/\text{mL}$. Finally, 72 μL of PbrMppP enzyme stock

(277 μ M) was added for a final PbrMppP concentration of approximately 20 μ M. Reaction volumes were adjusted to 1mL with mqH₂O. The negative control was performed in same way, except the PbrMppP enzyme was denatured by heating in boiling water for 10 minutes. Reactions were run in the dark at 25 °C for 4 hours with shaking at 200 rpm. After this period, all samples were boiled for 10 min and centrifuged at 30,000 x g for 10 min. The supernatants were directly injected into a Shimadzu LCMS-2020 to analyze by mass spectrometry. Samples were separated with a binary gradient to 80% organic solvent across 2 minutes. Molecules were ionized using the ESI method with a scan speed of 2500 unit/second.

2.2.7 Crystallization and structure determination of Pbr proteins

Initial crystallization conditions of PbrMppP, PbrHYP, PbrDHPS, and PbrOX were all pursued in the same way. Briefly, ~10 mg/mL of protein were screened against the Index HT screen (Hampton Research) and LMB Screen (Molecular Dimensions). After optimization, diffraction-quality crystals were obtained by the hanging drop vapor diffusion method. Final crystallization conditions are listed in Table 2.2.7-1. Crystals were cryo-protected by sequential soaks in the holding solutions listed in Table 2.2.7-1 and flash-cooled by plunging into liquid nitrogen. The structures of PbrMPCO proteins in complexes with various ligands were obtained by transferring crystals into 50 μ L drops of the cryo-protectant solution supplemented with ~20 mM ligand.

X-ray diffraction data were collected at beamline 21-ID-D of the Life Science Collaborative Access Team (LS-CAT) at the Advanced Photon Source (APS) using a MAR 300 CCD detector. A total of

360 frames were collected from 0° to 180° with an oscillation range of 0.5°. All data were processed with the HKL2000 software package^{xciii} or MOSFLM^{xciv} and AIMLESS^{xcv} of the CCP4 software suite^{xcvi}.

As SwMppP and PbrMppP share 33% sequence identity, the PbrMppP structure was determined by molecular replacement using SwMppP as the reference model. The diffraction data were collected from a crystal using a wavelength of 0.97921 Å. PbrMppP crystallized in space group $P2_1 2_1 2_1$ with two chains in the asymmetric unit. The molecular replacement solution was subjected to iterative cycles of manual model building in COOT⁸ ^{xcvii} and maximum likelihood-based refinement using the PHENIX package^{xcviii}. Ordered solvent molecules were added automatically in phenix.refine⁴ ^{xcix}. Hydrogen atoms were added to the model using phenix.reduce^c and were included in the later stages of refinement to improve the stereochemistry of the model. The positions of H atoms were refined using the riding model with a global B factor. Regions of the model for translation-libration-screw (TLS) refinement were identified using phenix.find_tls_groups (P. V. Afonine, unpublished work), and the TLS parameters were refined in phenix.refine⁶. Once the refinement converged ($R_{\text{cryst}} = 0.137$, and $R_{\text{free}} = 0.166$), the model was validated using the tools implemented in COOT^{ci}. Structure of PbrMppP with 4HKA bound was determined by difference Fourier.

Table 2.2.7-1: Crystallization conditions of the PbrMPCO enzymes.

Protein	Crystallization condition	Cryo-protectant solution
PbrMppP (12mg/mL)	100 mM Ammonium sulfate+ 25% PEG+ 100mM BisTris pH 5.5	20% Glycerol
PbrHYP (10mg/mL)	25% PEG 3350, 0.2M NaCl, 0.1M BIS-TRIS pH 5.5, 0.25M LiCl	Paratone N
PbrDHPS (15 mg/mL)	25% PEG 3350, 0.2M MgCl ₂ , 0.1M TRIS pH 8.5	Paratone N
PbrOx (10mg/mL)	30% Tacsimate, pH 7.0, 10mM Bis-TRIS pH 6.3	30% PEG 400

For PbrHYP, the nearly cubic crystals formed after 3-5 days. Crystals were cryoprotected and flash-cooled by plunging into liquid nitrogen. The PbrHYP structure was determined by single wavelength anomalous diffraction (SAD) using data collected from a crystals at 0.97849 Å, 62.0 eV below the tabulated K-edge wavelength for Se (0.97950 Å). PbrHyp crystallized in space group C2 with two chains in the asymmetric unit. AutoSHARP^{xxxi} was used to determine the Se-substructure and calculate density-modified electron density maps. The initial model was built automatically using the PHENIX package (phenix.autobuild)^{cii}. The solution was subjected to iterative cycles of manual model building in COOT^{8xcvii} and maximum likelihood-based refinement using the PHENIX package. Solvent molecule, H-atoms, and TLS refinements were performed

similarly described above. Once the refinement converged ($R_{\text{cryst}} = 0.183$, and $R_{\text{free}} = 0.231$), the model was validated using the tools implemented in COOT^{xc}. The final refined model was used to determine the structure of the PbrHYP.

PbrDHPS crystals formed after 10 days and grew to maximum dimensions of $\sim 10 \mu\text{m} \times \sim 10 \mu\text{m} \times \sim 5 \mu\text{m}$. The structure of PbrDHPS with **[91]** bound was obtained by soaking a crystal in crystallization solution supplemented with $\sim 12 \text{ mM}$ **[91]** and a small amount of cryoprotectant solution for 4 hours. PbrDHPS structure was also determined by SAD. Data collected from a crystal at 0.97849 \AA , 61.0 eV below the tabulated K-edge wavelength for Se (0.97950 \AA). PbrDHPS crystallized in space group $C 2 2 2$ with two chain in one asymmetric unit. AutoSHARP^{xxxi} was used to determine the Se substructure and calculate density-modified electron density maps. The initial model building and refinement strategy was similar as described above. Once the refinement converged ($R_{\text{cryst}} = 0.186$, and $R_{\text{free}} = 0.234$), the model was validated using the tools implemented in COOT^{xc}. The final refined model was used to determine the structure of the PbrDHPS. Structure of PbrDHPS with **[91]** bound was determined by difference Fourier.

The SeMet-substituted PbrOX crystals formed needle shaped after 3 days and grew to maximum dimensions of $\sim 160 \mu\text{m} \times \sim 50 \mu\text{m} \times \sim 2 \mu\text{m}$. Crystals were directly cryo-protected by soaking in holding solutions listed in table 2.2.7-1. The structure of SeMet-substituted PbrOX was determined by single wavelength anomalous diffraction (SAD) using data collected from crystals at 0.97853 \AA , 61.0 eV below the tabulated K-edge wavelength for Se (0.97950 \AA). PbrOX crystallized in space group $P42$ with a single chain in the asymmetric unit, though a symmetry-

related chain forms an extensive interface where each chain buries $\sim 1300 \text{ \AA}^2$. Thus, PbrOX is likely a dimer in solution. AutoSHARP^{5 xxxi} was used to determine the Se substructure and calculate density-modified electron density maps. The initial model building and refinement strategy was similar as described for PbrHYP. Once the refinement converged ($R_{\text{cryst}} = 0.159$, and $R_{\text{free}} = 0.202$), the model was validated using the tools implemented in COOT^{xc}. The final refined model was used to determine the structure of the PbrOx.

Data collection and model refinement statistics are listed in Appendix A.

2.2.8 Steady state kinetics of PbrMppP

All steady state kinetic assays (500 μ L) were conducted in triplicate at 25 °C in 20 mM BisTris propane, pH 9.0. The initial velocity of PbrMppP-catalyzed hydroxylation of L-Arg was measured directly by monitoring the decrease in dioxygen concentration as measured using a Clark-type electrode (Hansatech Instruments, Norfolk, UK). The concentration of L-Arg was varied from 5 to 6400 μ M. The k_{cat} and K_{M} values were determined by fitting the initial velocity data using equation 1:

$$V_0 = V_m[A]/(K_M + [A])$$

where, V_0 is the initial velocity, V_m is the maximal velocity, K_M is the Michaelis constant and $[A]$ is the concentration of L-Arg. The same procedure was followed for the reaction of PbrMppP with L-canavanine.

2.2.9 Stoichiometric analysis of PbrMppP reactions with L-Arg and L-canavanine

The stoichiometry of the reaction was estimated by measuring the quantity of dioxygen consumed in reactions with a fixed enzyme concentration (5 μ M PbrMppP) and varied L-Arg concentrations. Reactions were monitored using the Clark electrode. The concentration of dioxygen remaining when the reaction reached equilibrium was subtracted from the starting concentration to obtain the concentration of dioxygen consumed. The stoichiometry was computed using Equation 2:

$$\text{O}_2 \text{ consumed per eq. of L-Arg added} = ([\text{O}_2]_o - [\text{O}_2]_f) / [\text{L-Arg}]_t$$

where $[\text{O}_2]_o$ is the initial concentration of O_2 , $[\text{O}_2]_f$ is the final concentration of O_2 and $[\text{L-Arg}]_t$ is the total concentration of L-arg added.

The same procedure was repeated with L-canavanine.

2.2.10 Pre-steady state kinetics of PbrMppP

Pre-steady state kinetics data were collected on a Model SF-61DX double-mixing stopped flow spectrophotometer at 25 $^{\circ}$ C. The instrument was prepared for anaerobic work by overnight incubation with a solution of 1 mM glucose, 10 U/mL glucose oxidase, which had been flushed through the entire system. Solutions for anaerobic stopped flow experiments were prepared in

a tonometer. Briefly, 7 ml of 26 μM PbrMppP in 10 mM MES pH 6.7 and 100 μM glucose (with 1 $\mu\text{g}/\text{mL}$ catalase) were placed in the main body of the tonometer and 25 μL (1 U/ μl) glucose oxidase were placed in the side arm. The tonometer was attached to a Schleck line and dioxygen was removed from the solution using 45 cycles of vacuum (~ 5 s) and flushing with argon (~ 30 s). Because the PbrMppP protein is light-sensitive and somewhat unstable at room temperature, this procedure was conducted on ice and the tonometer was covered with foil to protect it from light. Different concentrations of L-arg in 100 mM BIS-TRIS propane pH 9.0 with 200 μM glucose were placed in glass syringes and sparged with argon gas for 5 min to remove dioxygen. Single-mixing experiments were conducted with 10 μM PbrMppP and 63 to 1000 μM L-Arg (all concentrations quoted are the concentrations in the cuvette after mixing 1:1).

To measure L-Arg binding, anaerobic PbrMppP (13 μM) was mixed with anaerobic L-Arg and changes at the PLP center were monitored by exciting the internal aldimine at 415 nm and measuring the total fluorescence emission for 20 s after initiating the reaction. The rate of formation of the first quinonoid intermediate (Q1) was measured by monitoring the absorbance at 510 nm for 20 s after mixing anaerobic enzyme and L-Arg solutions. To measure the rates of the dioxygen-dependent steps, double-mixing experiments were performed where anaerobic enzyme was mixed 1:1 with a saturating concentration of anaerobic L-Arg. This was then mixed 1:1 with buffer containing concentrations of dioxygen ranging from 82 to 136 μM . The first tonometer contained 40 μM enzyme and 1 $\mu\text{g}/\text{mL}$ catalase in 10 mM MES pH 6.7 under anaerobic conditions (100 μM glucose and 25 μL or 1 U/ μl glucose oxidase). The second tonometer contained 2,000 μM L-Arg in 100 mM BIS-TRIS propane pH 9.0. The third tonometer contained 100 mM BIS-TRIS propane pH 9.0 with concentrations of dioxygen of 30, 82, 166, 252, 402, and

523 μM . Defined dioxygen concentrations were obtained using a Maxtec Maxblend gas blender to mix 90% pure O_2 gas with gaseous N_2 . The concentrations of dioxygen produced at different settings of the mixing valve were determined using the oxygen electrode. Samples of oxygenated buffer were prepared by sparging the 100 mM BIS-TRIS propane pH 9.0 buffer with a set concentrations of dioxygen for 5 min. After the first mixing step (enzyme plus L-Arg), the reactions were aged for 20 s to give all of the enzyme time to convert to the first quinonoid intermediate prior to the second mixing step. The second step mixed the anaerobic reaction mixture with oxygenated buffer. The reactions were monitored for 100s at 510 nm or 560 nm for 100s.

2.2.11 Substrate specificity of PbrMppP

Reactions containing 500 μl of 30 μM PbrMppP in 20 mM BIS-TRIS propane pH 9.0 were initiated by adding 10 μl of different L-Arg analogs, including D-Arg, L-citrulline, D-ornithine, L-ornithine, L-Lys, N-methyl-L-Arg or L-canavanine, to a final concentration of 10mM. The reactions were monitored by recording spectra from 200-700 nm (0.5 s between scans) for 300 s in a Hewlett-Packard 8453 diode array spectrophotometer. The products of these reactions were analyzed by ESI-MS as discussed above.

The steady state kinetics of L-canavanine was determined using the steady state dioxygen consumption assay described above. L-Canavanine concentrations ranging from 100 μM to 10,000 μM were added to 20mM BIS-TRIS propane pH 9.0. Reactions were initiated by the addition of PbrMppP to a final concentration of 30 μM and initial velocities were recorded.

These initial velocities were plotted against the L-canavanine concentration and fitted to Equation 1, where [A] is the concentration of L-canavanine.

2.2.12 Ferredoxin NADP reductase (PbrFNR) activity test

Reactions containing 1 μM PbrFNR with concentrations of NADPH ranging from 1 μM to 40 μM were monitored by recording spectra from 200-600 nm (0.5s intervals) for 300 s in a Hewlett-Packard 8453 diode array spectrophotometer. NADPH was diluted with buffer (10mM HEPES pH 7.0) and incubated at least 10 min at RT before performing the assay. Reactions were initiated by adding PbrFNR with a final concentration of 1 μM .

To inspect the interactivity of PbrFNR, PbrFD and PbrOX, the decrease in absorbance at 340 nm was monitored to observe the oxidation rate of NADPH. In the cuvette containing different concentration of NADPH and 1 μM PbrFNR, PbrFD (final concentration of 1 μM) was added to observe the NADPH oxidation rate by monitoring decrease in absorbance at 340 nm. When the decrease in 340 nm almost reached plateau, PbrOX (1 μM) was added to see if the NADPH oxidation rate changes.

To measure the quantity of O_2 consumed in reactions with 1 μM PbrFNR, 10 μM PbrOX, 10 μM PbrFD, and 10 μM NADPH, reactions were monitored using the Hansatech dioxygen electrode. The concentration of dioxygen remaining when the reaction reached equilibrium was subtracted from the starting concentration to obtain the concentration of dioxygen consumed.

2.2.13 Identifying conditions that activate the PbrMppP promoter

(RS05025)

Pseudomonas brassicacearum DF41

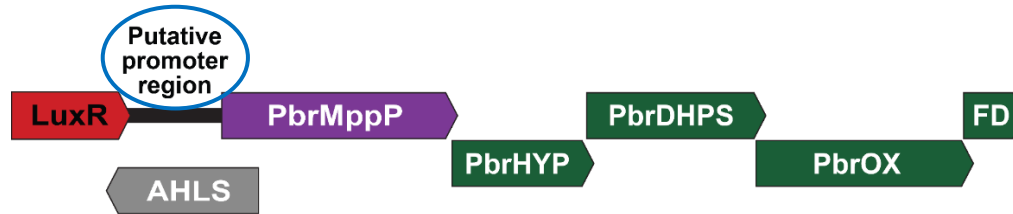


Figure 2.2.13-1 *PbrMppP* Promoter region (RS05025)

To identify the conditions under which the putative promoter of the *PbrMppP*-containing operon is activated, the untranslated sequence between the *luxR* and *mppP* genes was cloned into the reporter plasmid pPROBE-NT. The putative promoter region of the *PbrMPCO* was amplified from *Pseudomonas brassicacaerum* DF41 genomic DNA using Phusion[®] High-Fidelity PCR Master Mix with HF Buffer (M0531S, New England Biolabs, Ipswich, MA) and primers “RS05025_pProbeNT_HindIII_F” and “RS05025_pProbeNT_EcoRI_R” from Table 2.2.1-1, which added the HindIII and EcoRI restriction sites to the 5' and 3' ends of the amplified product, respectively. The amplified DNA was separated by 0.8% agarose gel, isolated band was excised, and purified from the gel using the QIAquick gel extraction kit (28115, Qiagen, Inc). The purified DNA was digested with HindIII (HindIII-HF[®] R3104S) and EcoRI (EcoRI-HF[®] R3101S) restriction enzymes (New England Biolabs). The empty pPROBE-NT (kanamycin-resistant) was digested using the same restriction enzymes. The digested insert and vector were ligated using T4 DNA ligase

(M0202S, New England Biolabs) in 1:3 (vector:insert) ratio. Ligation reactions (10 μ L) containing ~20ng DNA were used to transform NEB5 α high efficiency competent *E. coli* cells. The transformations (~50 μ L) were streaked into LB-agar plates containing 50 μ g/ml kanamycin and incubated at 37°C for 16 hours. Transformants from these plates were used to inoculate 10 mL of LB medium for isolation of plasmid DNA using the Miniprep kit (Qiagen, Inc). Isolated plasmids were used for subsequent PCR to confirm the presence of insert. The purified plasmids were screened for successful inclusion of the PbrMppP promoter region (RS05025 insert) by GENEWIZ (South Plainfield, NJ) using the sequencing primers GFP_F and GFP_R (Table 2.2.1-1).

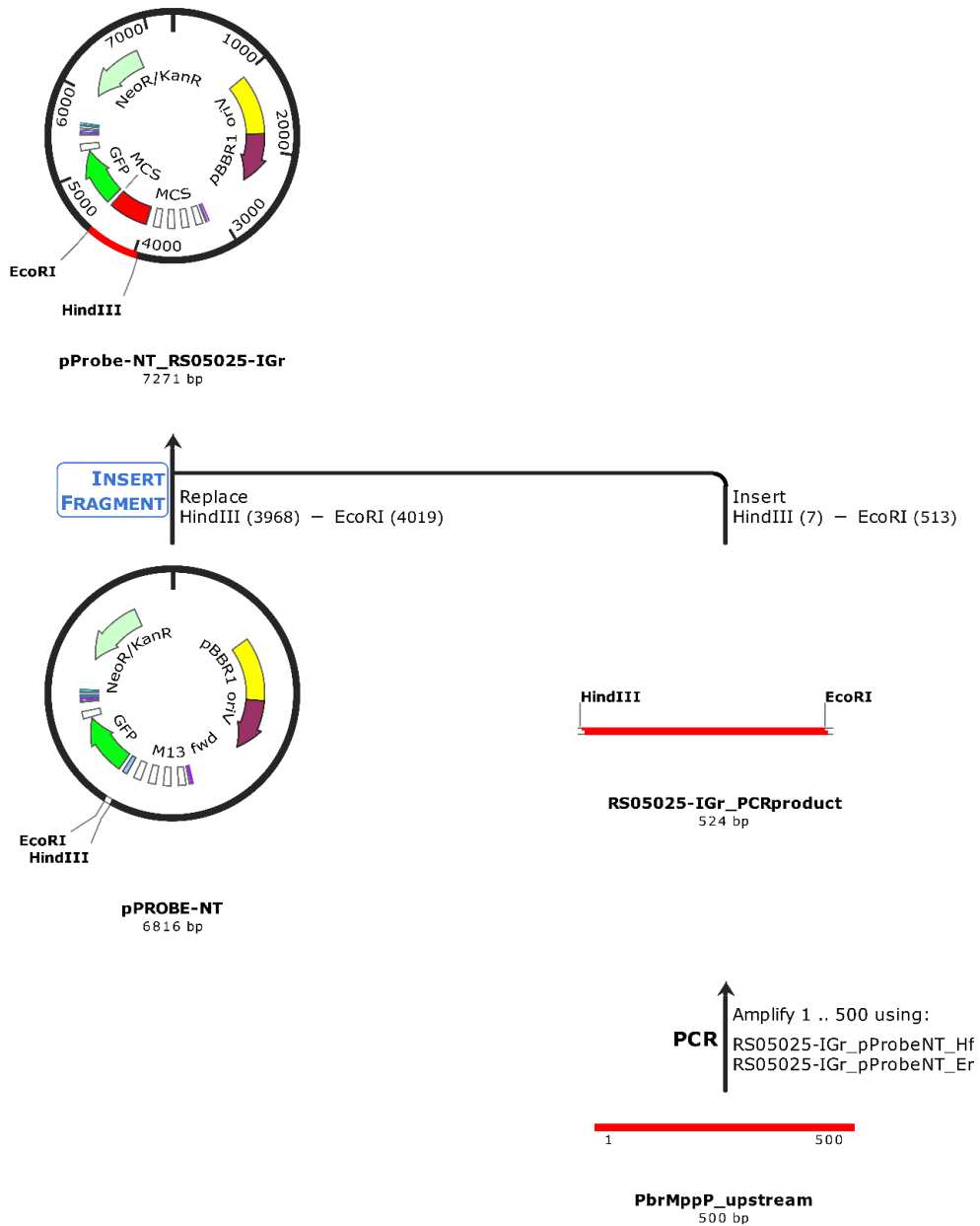


Figure 2.2.13-2: Schematic of the process for cloning the putative *PbrMPCO* promoter region (*RS05025*) into the *pPROBE-NT* reporter plasmid. Image created with SnapGene® *ciii*.

Pseudomonas brassicacaerum (DF41) cells were made competent by washing 3 times with ice-cold 10% glycerol. Cells were resuspended in ice-cold 10% glycerol and centrifuged at 12,000 rpm for 2 minutes between washes. This procedure was repeated three times to make the cells

electrocompetent. The putative PbrMPCO promoter cloned into the pPROBE-NT plasmid was used to electroporate DF41 cells. Approximately 74 ng of plasmid DNA (50 μ L) was added with 50 μ L of competent cells and transferred into a 2mm electroporation cuvette. The DF41 cells were electroporated by applying 1600 V using a BTX electroporation system. Immediately after electroporation, 1 mL SOC media was added quickly and then the cells were incubated for 2 hours at 30°C. After this incubation period, transformed DF41 cells were centrifuged at 12,000 rpm for 2 minutes. From the supernatant ~50 μ L was discarded and the pelleted cells were resuspended in the remaining 50 μ L of SOC media. The concentrated cells were streaked onto an LB-agar plate containing 50 μ g/ml kanamycin. DF41 cells were transformed with the empty pPROBE-NT vector as a negative control.

DF41 cells harboring the GFP (green fluorescence protein) reporter plasmid (pPBROBE-NT) were grown under a series of conditions to identify which one(s) caused expression of the GFP reporter gene under control of the putative PbrMPCO promoter. Media compositions are listed in Table 2.2.13-1.

Table 2.2.13-1 Screening conditions used to test the putative PbrMPCO promoter.

Media	Composition (Final concentrations in media)
1. Basal media 2	C source: 0.40% Glucose/ 20 mM succinate/20 mM gluconate/ 20 mM pyruvate. 7 mM (NH ₄) ₂ SO ₄ , 40 mM K ₂ HPO ₄ , 22 mM KH ₂ PO ₄ , 0.5 mM MgSO ₄ , 0.1 mM FeSO ₄ . *PH adjusted to 7.0 during ammonium-phosphate base (BM2) 10X preparation with (NH ₄) ₂ SO ₄ , K ₂ HPO ₄ , KH ₂ PO ₄ stock solutions.
2. PP2 media	Proteose peptone (10g/L)
3. Phosphate sufficient media	100mM HEPES pH 7.0, 80 mM NaCl, 20 mM KCl, 20mM NH ₄ Cl, 4.26 μM FeCl ₂ , 3 mM Na ₂ SO ₄ , 200 μM CaCl ₂ , 1mM MgCl ₂ , 0.4% (w/v) glucose, 1.0mg/ml thiamine, 660 μM K ₂ HPO ₄
4. Phosphate deficient media:	100mM HEPES pH 7.0, 80 mM NaCl, 20 mM KCl, 20mM NH ₄ Cl, 4.26 μM FeCl ₂ , 3 mM Na ₂ SO ₄ , 200 μM CaCl ₂ , 1mM MgCl ₂ , 0.4% (w/v) glucose, 1.0mg/ml thiamine, 42μM K ₂ HPO ₄
5. Defined minimal media:	93.5 mM NH ₄ Cl, 25 mM K ₂ HPO ₄ , 5 mM MgSO ₄ , 18 μM FeSO ₄ , 2.5% glycerol, 51mM monosodium glutamate, 110 mM Disodium succinate, 10 mM Nitritotriacetic acid (Ca, Mg chelator)
6. LB	Luria borth (25g/L)
7. No salt media	Tryptone (10g/L), yeast extract (1g/L)
8. M9 minimal media	42 mM Na ₂ HPO ₄ , 22 mM K ₂ HPO ₄ , 8.6 mM NaCl, 18.7 mM MgSO ₄ , 0.2% glucose. *PH adjusted to 7.0 during 10X M9 salt preparation with Na ₂ HPO ₄ , K ₂ HPO ₄ , NaCl, MgSO ₄ stock solutions.

All the carbon sources were used in individual media

*condition screened in both liquid and Agar (Solid) culture, No CaCl₂ used for DF41 cells

2.2.14 Metabolite extraction for LCMS analysis

P. brassicacearum DF41 cells expressing GFP (exhibiting a green glow) were scooped from LB-agar plates and re-suspended in 10 mL solvent comprising 10% methanol in mqH₂O , , . Metabolites from *P. brassicacearum* DF41 cells carrying the empty pPROBE-NT vector were extracted in the way to use as the negative control. Resuspended cells were lysed by sonication at 60% amplitude, 30 s pulses for a total of 10 minutes of sonication. The lysate was clarified by centrifugation at 18,000 rpm for 45 min. The supernatant was tested to confirm GFP expression by measuring spectra from 250 to 600 nm. These samples were used for comparative metabolomics studies by LC-MS. A C18 reversed phase column (Phenomenex, 30 mm X 2 mm; particle size 2.5 Å) was used to separate the ions. The mobile phase consisted of 0.1% formic acid in deionized water (A) and 100% LCMS-grade methanol (Milipore Sigma; B). Ions were separated with a binary gradient, increasing from 5% B to 80% B over 6 minutes. To regenerate the column after the ion separation, a gradual step down to 5% B over the next 5 minutes. The interface voltage was set to 4.5 kV and the temperature of column oven was set to 35oC. The column pressure was set to a maximum of 4000 psi. Molecules were ionized using ESI or APCI, separately, with a scan speed of 7500 units/second.

2.3 Results

2.3.1 PbrMppP oxidizes L-Arg

Since PbrMppP is a homolog of SwMppP with a sequence identity of 33%, it is perhaps not surprising that the activity of PbrMppP would be the same as SwMppP. The X-ray crystal structure of PbrMppP shows that its tertiary and quaternary structures are almost identical to those of SwMppP (RMSD value 1.2 Å, Figure 2.1-3). Likewise, the active sites of the two homologs are nearly identical as well. The substrate and products bind in very similar orientations in both enzymes and make the same interactions with the same active site residues.

PbrMppP crystallized in space group $P2_12_12_1$ with unit cell dimensions of $a = 85.7 \text{ \AA}$, $b = 108.3 \text{ \AA}$, and $c = 195.4 \text{ \AA}$ with 4 molecules in the asymmetric unit arranged as 2 independent homodimers (Figure 2.3.1-1A). The domain architecture is identical to SwMppP, comprising a large domain (Asp35-Asp284), a small domain (Phe25-Gln35 and Asp285-Asn388) and an N-terminal extension (Met1-Arg24). The large domain contains a seven β -sheet flanked by ten α -helices forming an α - β - α sandwich, while the small domain has a topology of mixed α -helices and β -sheets.

In the absence of substrate, PLP is covalently attached to Lys233 by an aldimine linkage (Figure 2.3.1-1B). The negative charge of Asp197 stabilizes the pyridinium form of the cofactor by making a salt bridge to the pyridine N atom. This interaction has been shown to be critical for catalysis in a number of PLP-dependent enzymes like aspartate aminotransferase^{civ}. The cationic pyridinium allows the cofactor to accept electrons from the substrate, stabilizing the carbanion formed during catalysis.

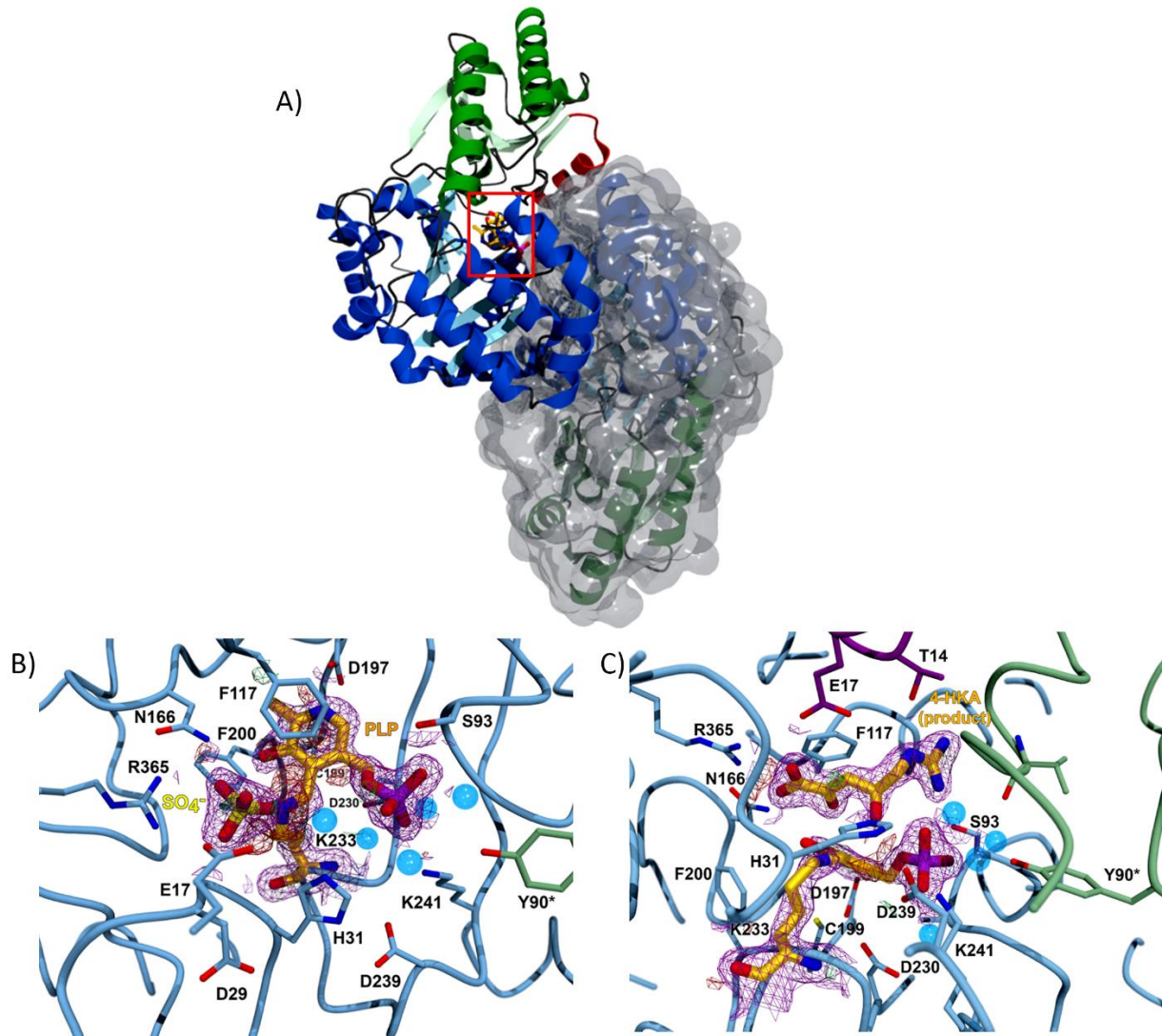


Figure 2.3.1-1: The structure of PbrMppP. A) quaternary structure of PbrMppP. The dimer structure shows one chain represented by a ribbon and the other chain as a transparent surface. The blue ribbon represents the large domain, whereas the green ribbon represents small domain. The red α -helix on the top represents N-terminus. Catalytic K233-PLP internal aldimine in the active sites is shown in yellow stick (highlighted by red square). B) PbrMppP active site showing internal aldimine formed by PLP and K-233. C) PbrMppP structure with 4HKA bound in the active site.

The structure of the PbrMppP•4HKA complex (Figure 2.3.1-1C) was obtained by soaking PbrMppP crystals in a solution containing 10 mM L-Arg as described above (see Section 2.2.6). PbrMppP is active in the crystalline form, thus after X-ray diffraction, the product **[95]** was observed bound in the active site. Once the substrate binds in the active site, the substrate α -amino group attacks C4' of the internal aldimine to form the external aldimine, displacing Lys233. The sulfate ion bound to Arg365 is replaced by the α -carboxylate of L-Arg. Thr14 and Glu17 in the N-terminal helix form hydrogen bonds with the guanidinium and carboxylate groups of L-Arg. When the product is bound in the active site, the side chains of His31, Asp197 and Asp239 come closer to C4' of the external aldimine. The ring of Phe117 rotates away from the PLP and moves closer to the bound product, **[95]**.

The PbrMppP reaction involves a number of spectroscopically active intermediates. Initiation of the reaction by adding PbrMppP to a cuvette with 1mM L-Arg under aerobic conditions results in an immediate shift of the 415nm peak to 425nm (Figure 2.3.1-2) and a slow increase in absorbance at 510nm as the reaction consumes the dioxygen in the cuvette. When the dioxygen concentration approaches zero, the 510nm peak reaches a maximum. Soon after the 510nm peak appears, a second peak at 560nm appears and then decays to zero as the 510nm peak plateaus. The 415nm peak is due to the internal aldimine, while the 425nm peak corresponds to the external aldimine. The longer wavelength peaks at 510nm and 560nm are due to the first **[88]** and second **[92]** quinoid intermediates, respectively. The 510nm persists after the 560nm peak has decayed away, because quinonoid I is the first intermediate that requires dioxygen for the reaction to proceed.

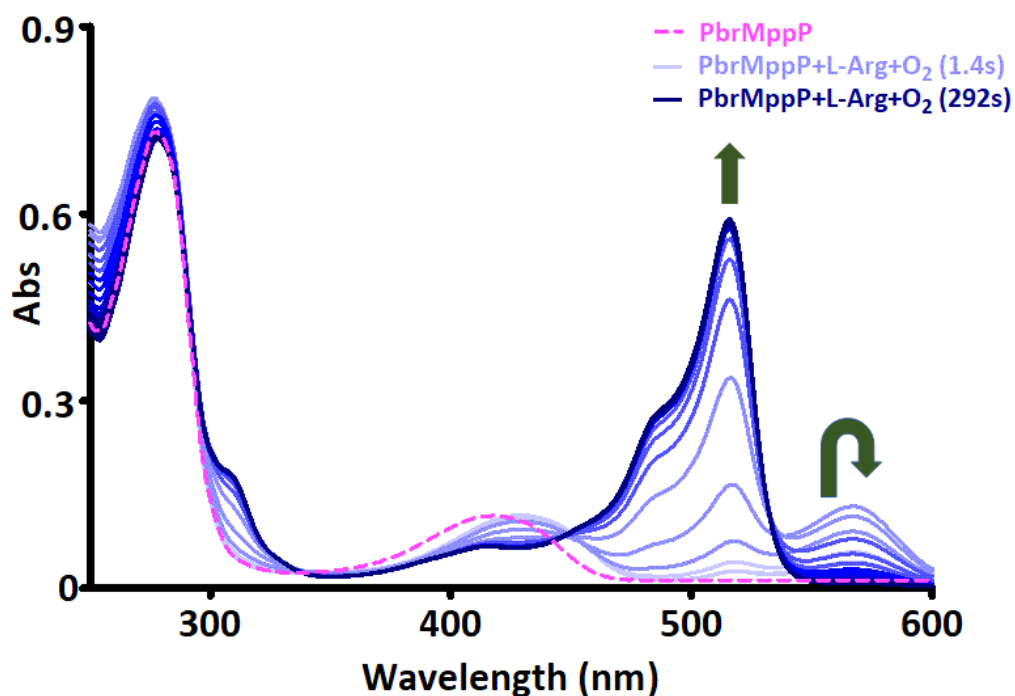


Figure 2.3.1-2: UV-visible spectra of the PbrMppP reaction with L-Arg. The pink spectrum is for PbrMppP, where the 415nm peak shifts to 425nm upon L-Arg addition. The spectra are colored according to time, with the light blue spectra representing the early stage of the reaction, beginning at 1.4s, and transitioning to dark blue by the end of the reaction at 292s. The increase in absorbance at 510nm is for quinoid I [88] formation, followed by a peak at 560nm for quinoid II [92] formation. After a short span of time, the absorbance at 560nm decays due to the reaction proceeding towards 4H2KA [95].

NMR analysis of PbrMppP reaction products (Figure 2.3.1-3) shows that, like SwMppP^{lix}, PbrMppP produces a mixture of **[95]** and **[91]** in a ratio of ~2:1. Reaction conditions were as described in Section 2.2.4. The mass spectrometric analysis confirms the NMR results. The major ions observed in the positive mode (Figure 2.3.1-4A) analysis of the reaction mixtures in the presence of catalase were $m/z = 190$, corresponding to **[95]** (190.1 Da) and $m/z = 174$, corresponding to **[91]** (174.1 Da). In reactions without catalase (Figure 2.3.1-4B), the major ions in the positive mode analysis are $m/z = 162$ and 146 , corresponding to the decarboxylated forms of **[95]** and **[91]**, respectively. It was observed in previous studies of SwMppP that H₂O₂ produced during the reaction catalyzes non-enzymatic decarboxylation of the ketoacids^{lix}.

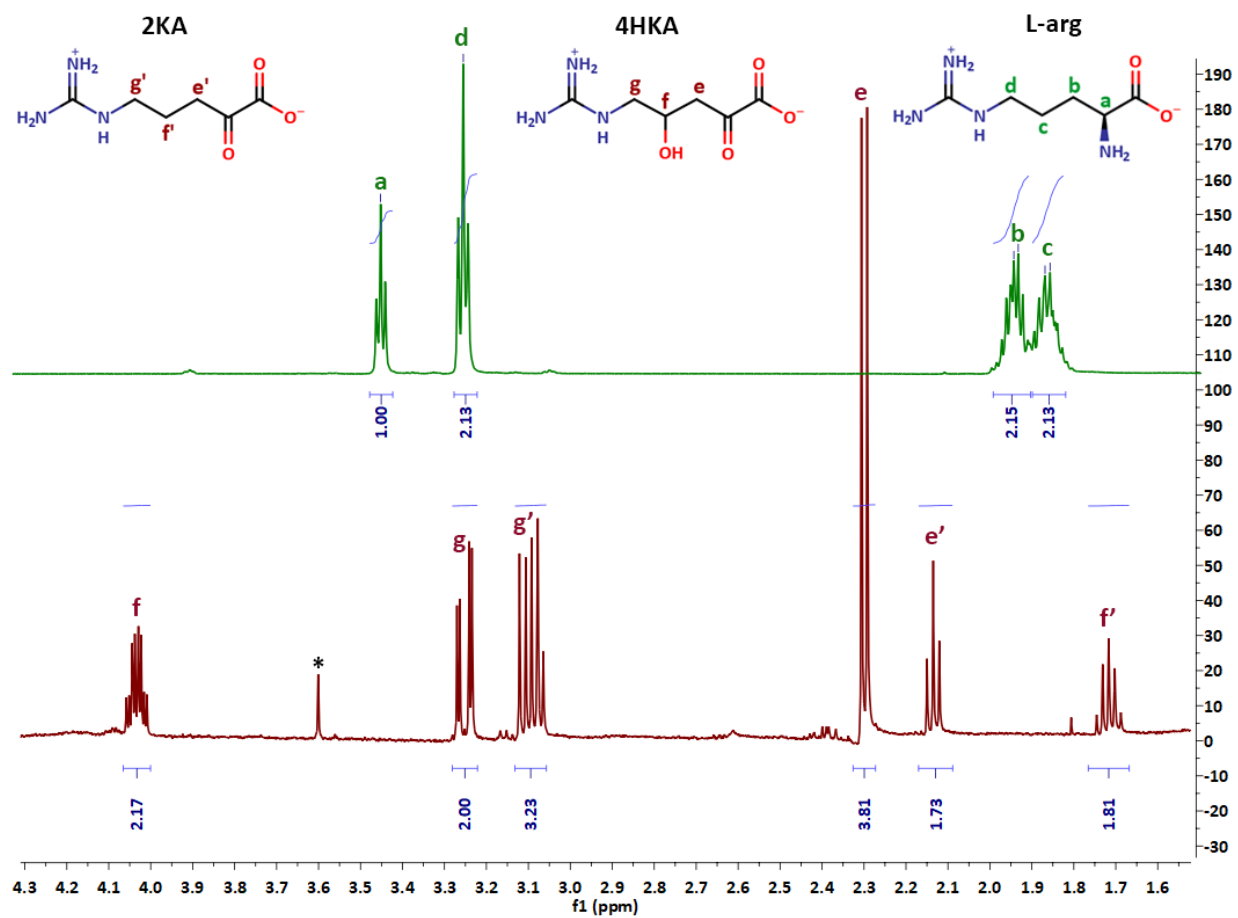


Figure 2.3.1-3: *PbrMppP* produces [95] and [91] in a ~2:1 ratio. The green spectrum (top) was collected for a solution of pure L-Arg in D₂O. The red spectrum is the *PbrMppP* reaction products.

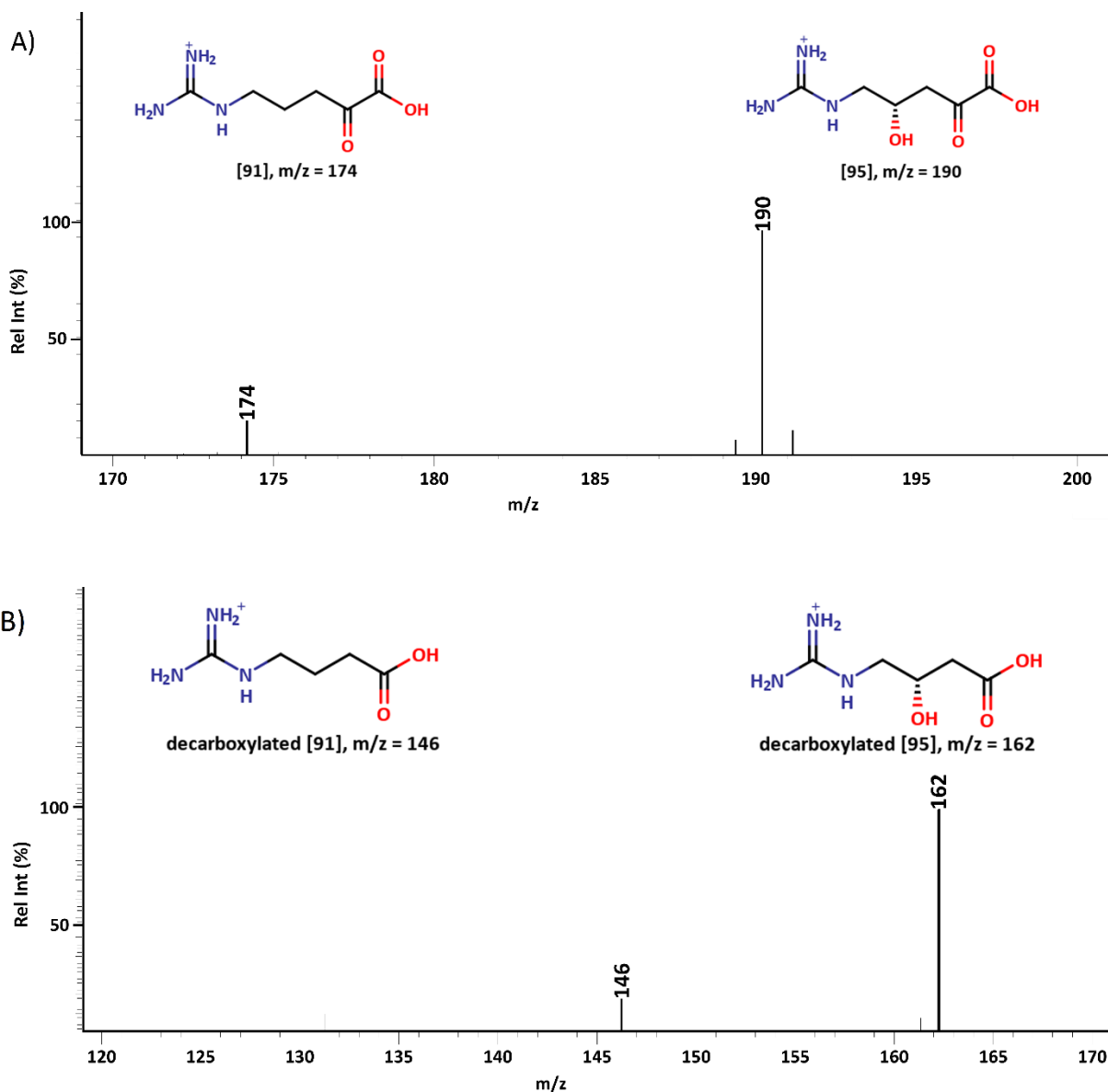


Figure 2.3.1-4: ESI-mass spectrometric analysis of PbrMppP products in positive mode. Labelled m/z peaks show the major reaction products vs relative intensity (%). (A) m/z of PbrMppP products [95] ($m/z = 190.1$) and [91] ($m/z = 174.1$). There was no trace of L-arginine ($m/z = 175.1$) after 4 hours, indicating that the reaction progressed to completion. (B) The ESI-MS spectra of the catalase-free (decarboxylated) PbrMppP products. The $m/z=146$ peak is due to the decarboxylated [91] and the $m/z=162$ peak is due to decarboxylated [95].

2.3.2 H₂O₂ is a product of PbrMppP catalysis

PbrMppP reactions with concentrations of L-Arg ranging from 10 μM to 160 μM were monitored using a Clark electrode (Figure 2.3.2-1). Catalase (1 μL of 5mg/mL) was added when the 500μL reactions reached equilibrium. The oxygen probe traces show that, in all cases, the addition of catalase resulted in the generation of an amount of dioxygen equal to one half of the total dioxygen consumed by PbrMppP. This observation shows that PbrMppP is producing hydrogen peroxide, which is then being disproportionated by catalase. The ratios of dioxygen consumed by PbrMppP vs the amount of L-Arg in the reaction are listed in Table 2.3.2-1. The average ratio of dioxygen consumed to L-Arg added is ~1.3. Assuming that all of the L-Arg was consumed in these reactions, as it was in the reactions for NMR analysis, a ratio between 1 and 2 suggests the operation of two branches of the mechanism (Scheme 2.1.1) where the production of the fully oxidized product **[95]** consumes two equivalents of dioxygen and the production of the abortive product **[91]** consumes only one equivalent.

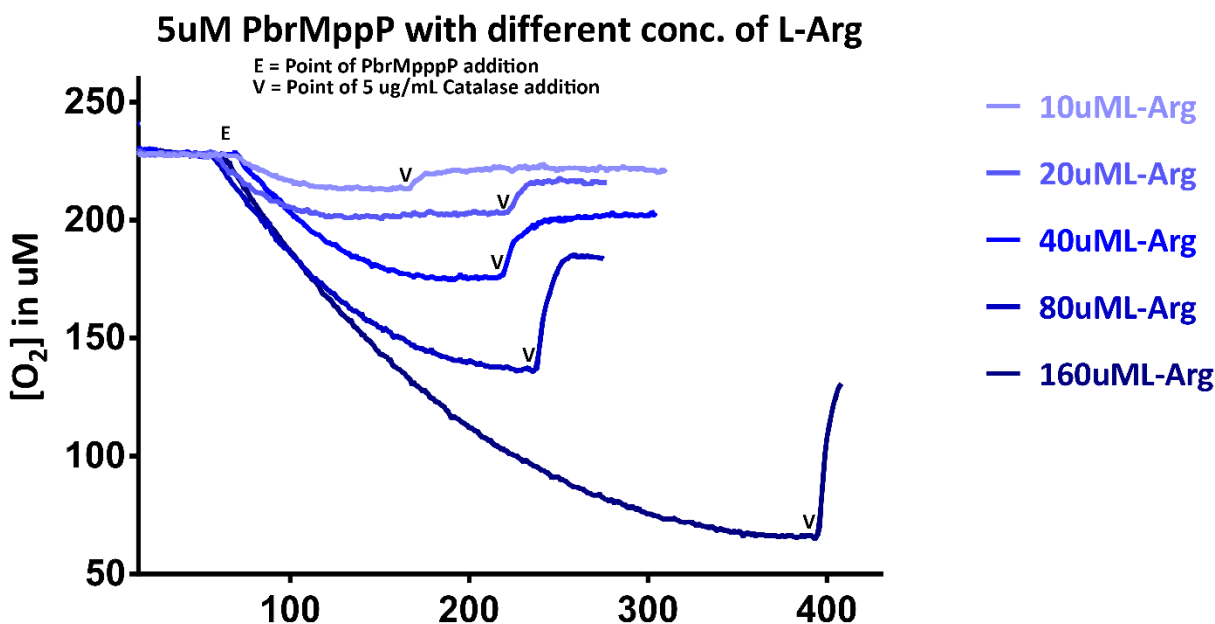
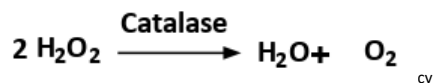


Figure 2.3.2-1: Regeneration of dioxygen after adding catalase into PbrMppP reactions at equilibrium. The concentration of dioxygen was monitored in a Hansatech dioxygen electrode. The reactions were initiated by adding PbrMppP (9 μL of 272 μM PbrMppP; $\sim 5 \mu\text{M}$ final concentration) into solutions containing 10 μM , 20 μM , 40 μM , 80 μM and 160 μM L-arg in 50 mM BisTris Propane pH 8.5 buffer. When the reaction reached equilibrium (i.e. plateaued), 1 μL of 5mg/mL catalase were added into the reactions for a final concentration of 1 $\mu\text{g}/\mu\text{L}$. The symbol “E” in the figure is used to mark the event of PbrMppP addition and the symbol “V” is used to mark the event of catalase addition. The regeneration of O_2 suggests that H_2O_2 is a byproduct of the reaction.

Tab 2.3.2-1: Amount of dioxygen released upon the addition of catalase to PbrMppP reactions monitored in a Hansatech dioxygen electrode.

[L-arg]	[O ₂] used	[O ₂]/ [L-arg]	[O ₂] retrieved by catalase
10 uM	14.0 μM	1.40:1	8.8 μM
20 uM	25.5 μM	1.27:1	13.2 μM
40 uM	52.7 μM	1.32:1	27.2 μM
80 uM	107.1 μM	1.34:1	56.2 μM
160uM	161.6 μM	1.01:1	85.2 μM

2.3.3 Steady state kinetics of PbrMppP

The initial velocity data from dioxygen probe experiments were used to determine the steady state kinetics of the reaction of PbrMppP with L-Arg. The kinetic parameters determined from these data (K_M , k_{cat} , and the specificity constant) are listed in Table 2.3.3-1.

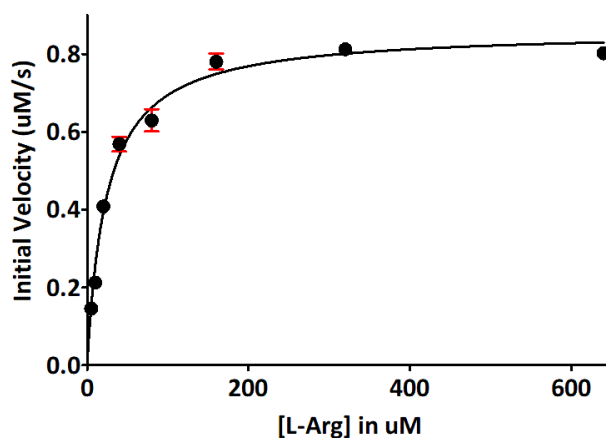


Figure 2.3.3-1: Michaelis Menten graph of PbrMppP catalysis. The reactions were initiated by adding PbrMppP (9 μL of 272 μM PbrMppP; $\sim 5 \mu\text{M}$ final concentration) into solutions containing 10 μM , 20 μM , 40 μM , 80 μM and 160 μM L-arg in 50 mM BisTris Propane pH 8.5 buffer.

Table 2.3.3-1: Steady state parameters calculated from the oxygen probe-derived initial velocity data with L-Arg.

Michaelis-Menten constant, K_M	$24 \pm 1.5 \mu\text{M}$
Turnover number, k_{cat}	$0.16 \pm 0.02 \text{ s}^{-1}$
Specificity constant, k_{cat}/K_M	$(6.8 \pm 0.1) \times 10^3 \text{ M}^{-1}\text{s}^{-1}$

2.3.4 Pre-steady state kinetics of PbrMppP

Changes at the PLP center of PbrMppP, upon binding of L-Arg were monitored using stopped flow spectroscopy under anaerobic conditions (SX) by exciting the internal aldimine at 415 nm. All concentrations mentioned here are after mixing (final concentrations in the cuvette). PbrMppP (13 μM) was mixed with L-Arg (125, 250, 500, and 1000 μM) under anaerobic conditions and the internal aldimine was excited at 415 nm. Changes at the PLP center were monitored by the total fluorescence. Figure 2.3.4-1A represents the kinetic model of the fittings.

In the absence of dioxygen, the fluorescence signal from the internal aldimine ($\lambda_{\text{ex}} = 415\text{nm}$) exhibits complex behavior. First, there is increase in fluorescence that peaks at 0.1-0.2 s (Figure 2.3.4-1B). The amplitude of this fluorescence increases with the L-Arg concentration. Following the increase in fluorescence, there is a decrease almost down to the initial values. We interpret this signal, tentatively, to represent binding of L-Arg to give the Michaelis complex (ER), which refers to the PLP-enzyme complex with N-terminus open conformation where the substrate L-arg is just hanging at the active site. The decrease in fluorescence is consistent with conversion of the external aldimine to the first quinonoid intermediate (EQ1). These data can be fitted to a triple-exponential function (Equation 3):

$$Y = f_0 + A_1 * \exp^{-k_1*x} + A_2 * \exp^{-k_2*x} + A_3 * \exp^{-k_3*x}$$

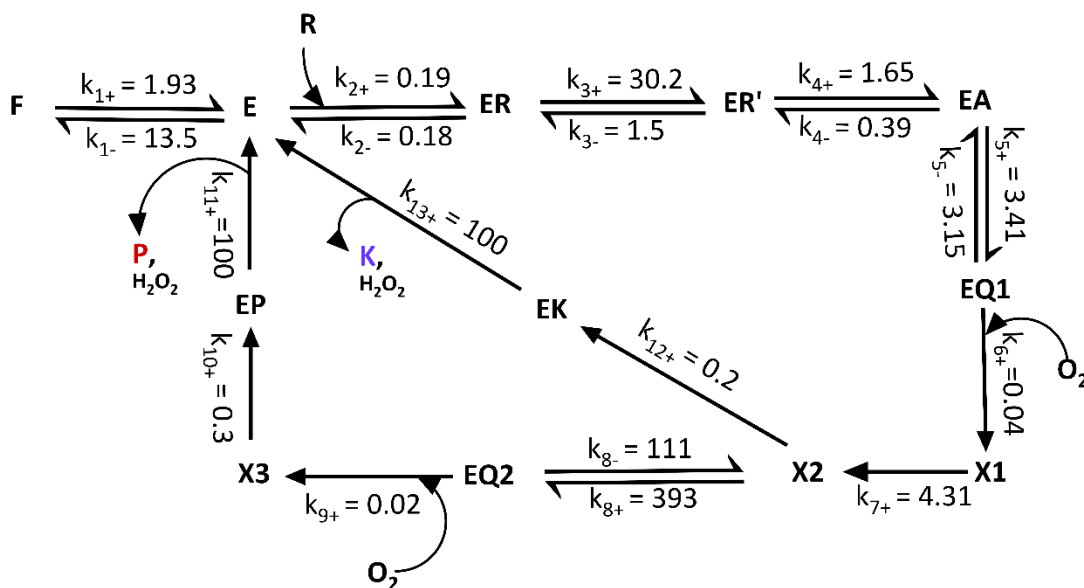
where f_0 is a constant, A_{1-3} are the amplitudes of the three phases, k_{1-3} are the associated rate constants, and x is time. These data, however, were not analyzed using the traditional re-plotting

of amplitudes and rate constants derived from such non-linear least squares fitting. Rather, the data from all experiments were fitted simultaneously by simulation in KinTek Explorer (version 10.0.200514). The results of fitting by simulation in KinTek Explorer are described in detail below, after the descriptions of all of the experiments performed.

The model of the fitting along with observed rate constants are represented in Figure 2.3.4-1A.

A)

Model 1



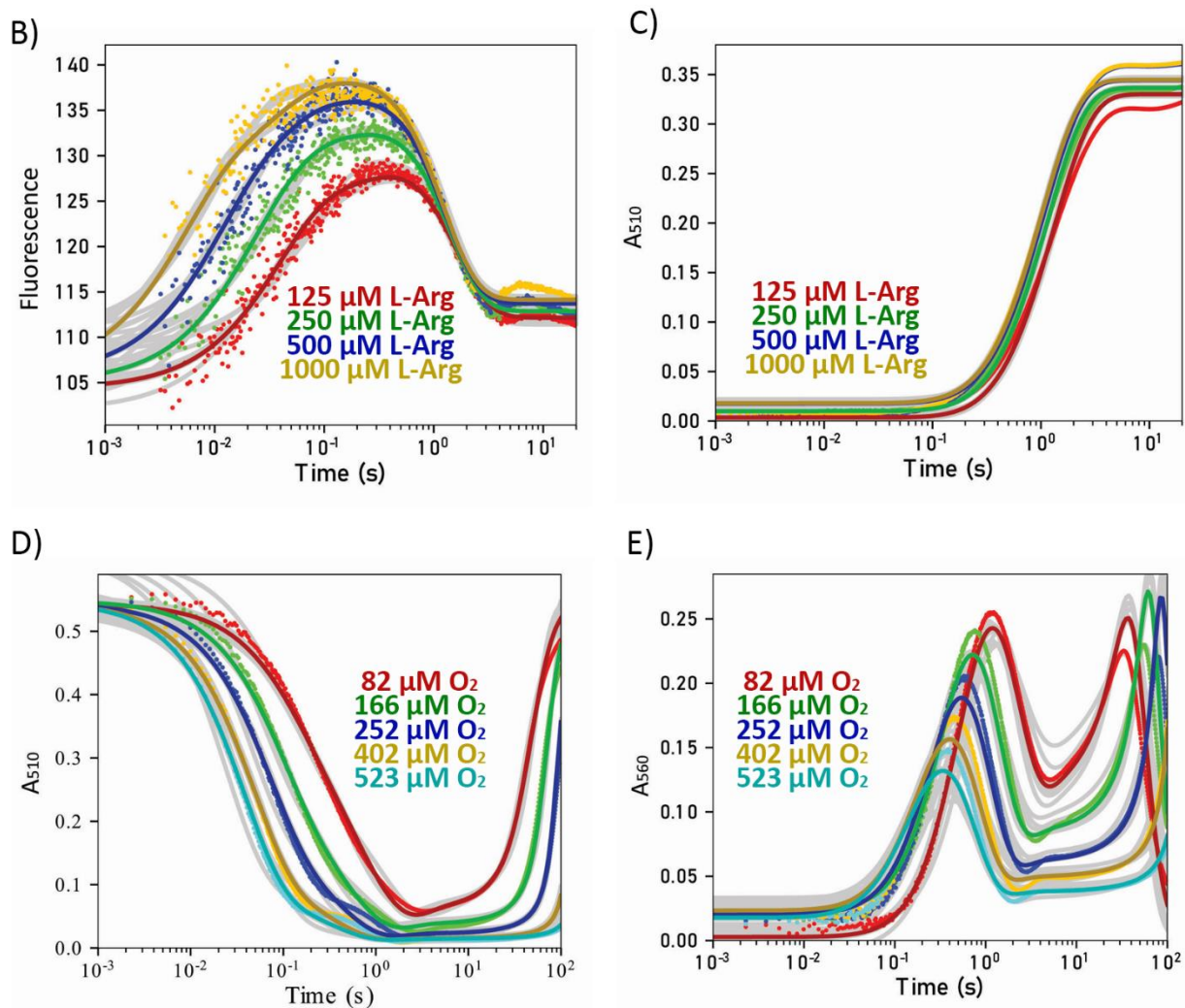


Figure 2.3.4-1: Pre-steady state kinetic data from global fitting in KinTek explorer version 10. A) Model 1, the hypothetical model with preliminary microscopic rate constants. The steps with unidirectional arrows were forced to be unidirectional by manually applying $k_{-x} = 0$. The purple “K” refers to the minor product of the PbrMppP reaction, [91]. The red “P” refers to the major product, [95]. B) the change in fluorescence after mixing 13 μM anaerobic PbrMppP with various concentrations of anaerobic L-Arg. C) The same experiment as described in (B) monitored at 510 nm. D) Double mixing experiments where 40 μM (before the first mix) PbrMppP was mixed with

2000 μM L-Arg in the first mixing step and aged for 10 seconds. In the second mixing step, 100 mM Bis-Tris propane buffer at pH 8.5 containing various concentrations of dioxygen was mixed and the absorbance at 510 nm was monitored for 100 s. E) The same reactions as (D), monitored at 560 nm, the absorbance maximum of the second quinonoid intermediate. The gray shading associated with the spectra refers to the fit through the entire ranges of all the parameters derived from fitting.

The accumulation of the first quinonoid intermediate was observed at the absorption maximum (510 nm) under anaerobic conditions (Figure 2.3.4-1C). The experiment was set up as described for the fluorescence experiment, except that the reactions were monitored by absorbance rather than fluorescence. The results are shown in Figure 2.3.4-1C. EQ1 forms in a single exponential process with a ~ 100 ms lag that corresponds to the formation and ultimate decay of the external aldimine. The signal saturates when all of the enzyme has been converted to EQ1, as evidenced by the dependence on the amplitudes of the transients on the substrate concentration.

The reaction of EQ1 with dioxygen were monitored by double-mixing stopped-flow spectrophotometry. The prepared anaerobic PbrMppP (40 μM) in 10 mM MES pH 6.7, 20 μM PLP was first mixed with 2000 μM L-Arg in 100 mM Bis-Tris Propane pH 8.5 and aged for 10 s prior to the second mix of different defined concentrations of dissolved O_2 in 100 mM Bis-Tris propane pH 8.5 buffer. The dissolved O_2 concentration in the buffer was defined by sparging an inverted syringe containing 100 mM BisTris propane pH 8.5 buffer for 5 min with blended N_2 and O_2 of known partial pressures supplied by a Maxtec Maxblend gas blender. The

concentration of dissolved O_2 was determined by first sparging the reaction chamber of a Hansatech dioxygen electrode filled with same buffer to define the equilibrium concentration of dissolved O_2 . The EQ1 complex was observed at 510 nm (Figure 2.3.4-1D) and the EQ2 complex was observed at 560 nm (Figure 2.3.4-1E) upon the second mixing with O_2 .

The absorbance at 510 nm was monitored for 100 s and the curve first decays until the dioxygen was completely consumed, at which point EQ1 began to accumulate again. The EQ1 decay happens before the second quinonoid forms. Hence, EQ2 only accumulates after EQ1 has decayed.

The microscopic rate constants resulting from fitting by simulation in KinTek Explorer, together with the model used for fitting, are shown in Figure 2.3.4-1A. In the hypothesized model, F is an inactive conformation of the enzyme, which attains the active conformation E, either upon binding with the cofactor PLP, or undergoing some conformational change. Thus, E is the internal aldimine form of PbrMppP. Binding of the substrate L-Arg gives the Michaelis complex (ER). To satisfy the fluorescence data, a hypothetical ER' step was added to the model that could represent a conformational change of the substrate or enzyme that is required to progress to the external aldimine. For example, the ER conformation could have the N-terminus disordered, and ER' represents the species where the N-terminal helix orders and forms its interactions with the substrate. In the ER' conformation, all of the active site residues are positioned to react with the substrate to give the external aldimine, which is subsequently converted to EQ1. In presence of dioxygen, EQ1 reacts with O_2 to form intermediate X1, the presence of which is required to satisfy the data and could be interpreted as a complex with superoxide that progresses to a peroxo intermediate, X2, most likely at the L-Arg C α . X2 then

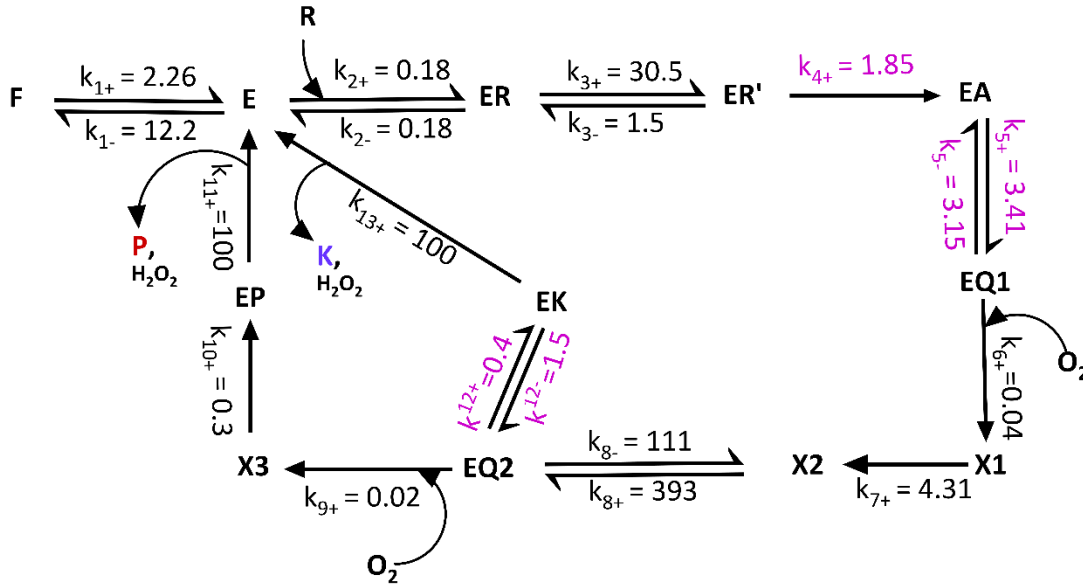
proceeds to the second quinonoid intermediate. EQ2 reacts with a second molecule of dioxygen to form intermediate X3, that breaks down to an enzyme-product complex (EP). Finally, EP is hydrolyzed to give the major product **[95]** (the red “P” in Figure 2.3.4-1A) and regenerating the free enzyme, E. A branch of the mechanism goes from intermediate X2 to form the enzyme-2KA complex, EK, which hydrolyzes to give the abortive product 2KA (the purple “K” in Figure 2.3.4-1E) and regenerating the free enzyme.

We tested this model by moving the branch point of the bifurcated mechanism and excluding the various hypothetical intermediates (e.g. F, EA) one at a time.

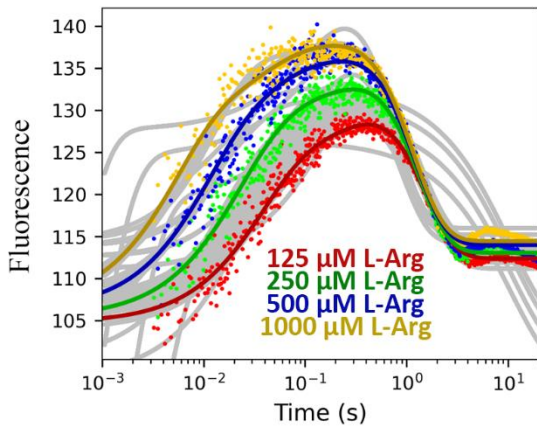
When the branch point to the abortive product **[91]** was moved from intermediate X2 to EQ2 (Figure 2.3.4-2A and the simulation was re-run, the quality of the fit was severely degraded (Figure 2.3.4-2B-E). For instance, in the anaerobic single mixing experiment monitoring EQ1 formation at 510nm (Figure 2.3.4-2C), accumulation of the quinonoid is completely abolished in the fit. Likewise, in the experiments looking at the effect of dioxygen concentration on the decay of EQ1 (Figure 2.3.4-2D) and on the accumulation and decay of EQ2 (Figure 2.3.4-2E) are also impacted. The fits (solid-colored lines) fit the data less well, and the parameter ranges (denoted by the grey lines) have become unacceptably wide, meaning that a large range of rate constants would fit the data equally poorly. The conversion of the external aldimine ER' to intermediate EA was unidirectional (the reverse rate refined to 0 s⁻¹) in the case of Model 2 (Figure 2.3.4-2A), which is not physically realistic. The formation of EK from EQ2 was also two times slower than the equivalent step in Model 1, conversion of X2 to EK.

A)

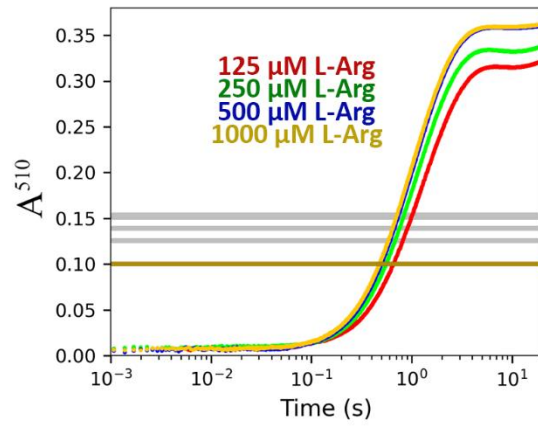
Model 2



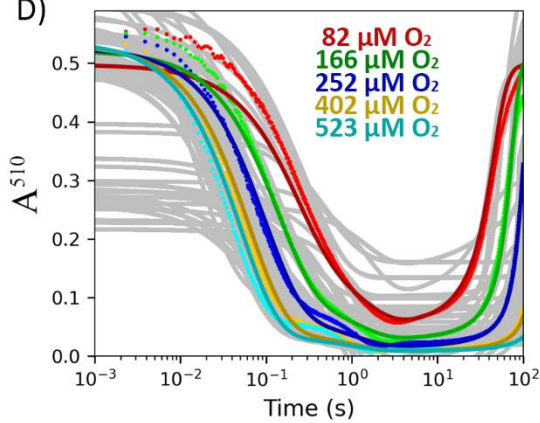
B)



C)



D)



E)

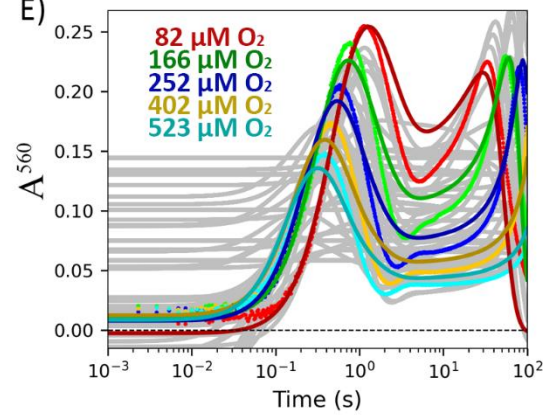


Figure 2.3.4-2: Pre-steady state kinetic data from global fitting in KinTek explorer V10, when the branch point to **[91]** was placed at EQ2 (A). B) the change in fluorescence after mixing 13 μM anaerobic PbrMppP with various concentrations of L-Arg under anaerobic conditions. C) The absorbance at 510 nm monitored in the same condition as for B. D) Double mixing experiment where 40 μM PbrMppP was mixed with 2000 μM L-Arg in the first step and aged for 10 s. In the second step, 100 mM BisTris propane buffer pH 8.5 containing various concentrations of dioxygen was mixed and absorbance at 510 nm was monitored. E) The same reaction as for D, but monitored at 560 nm. The gray shading underneath the data refers to the range of parameters (rate constants and extinction coefficients) that would be consistent with the fit. The steps with unidirectional arrows were forced to be unidirectional by manually applying $K^- = 0$. The K labeled with purple color refers to the minor product of MppP reaction, **[91]**. The P with red color refers to the major product of MppP reaction, **[95]**. The rate constants with pink labeling refer to deviations from Model 1.

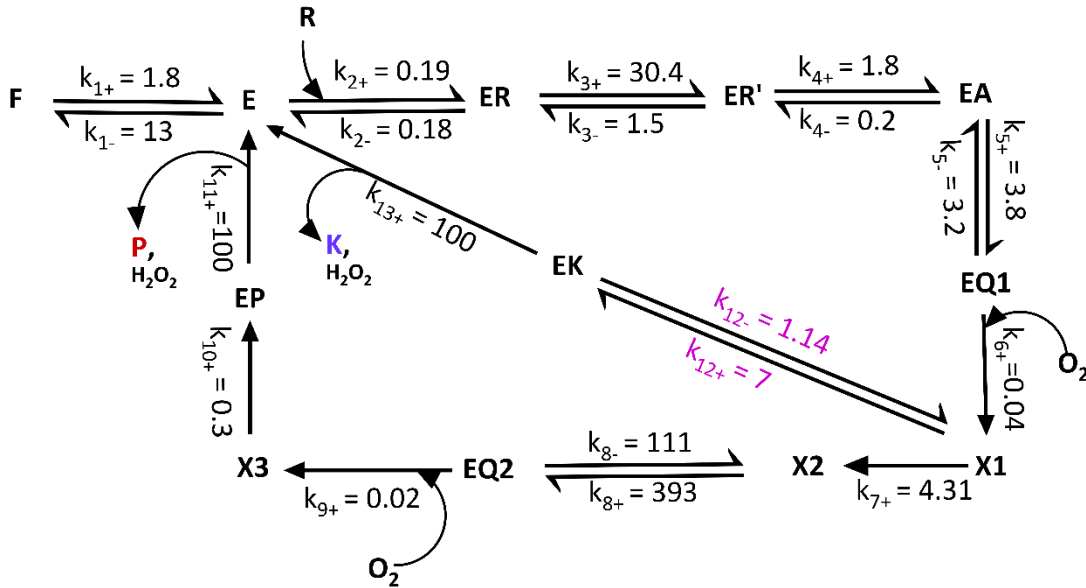
When the branch point to the abortive product **[91]** was moved from intermediate X2 to X1 (Figure 2.3.4-3A and the simulation was re-run, the quality of the fit was also degraded (Figure 2.3.4-3B-E). Although the anaerobic single mixing experiments (Figure 2.3.4-2B,C) didn't have much impact, but the double mixing experiments (Figure 2.3.4-2D,E) are largely impacted. The unidirectional step of EK conversion (the reverse rate refined to 0 s^{-1}) has also become four times slower in the case of Model 3 (Figure 2.3.4-3A). We use here the χ^2 value of each experiment the test fittings, as a quantitative parameter to evaluate how good the fittings are (Table 2.3.4-1). The χ^2 values refers to the placement of the best-fit lines relative to the data, whereas, the grey

parameter ranges shows how well the model reflects the data. The EQ1 decay experiment of model 3 has four times χ^2 value and the EQ2 accumulation experiment of model 3 has ten times χ^2 value, comparing to that of model 1.

Thus the analysis of model 2 and model 3 shows that the best fit of abortive product [91] branch point would be X1.

A)

Model 3



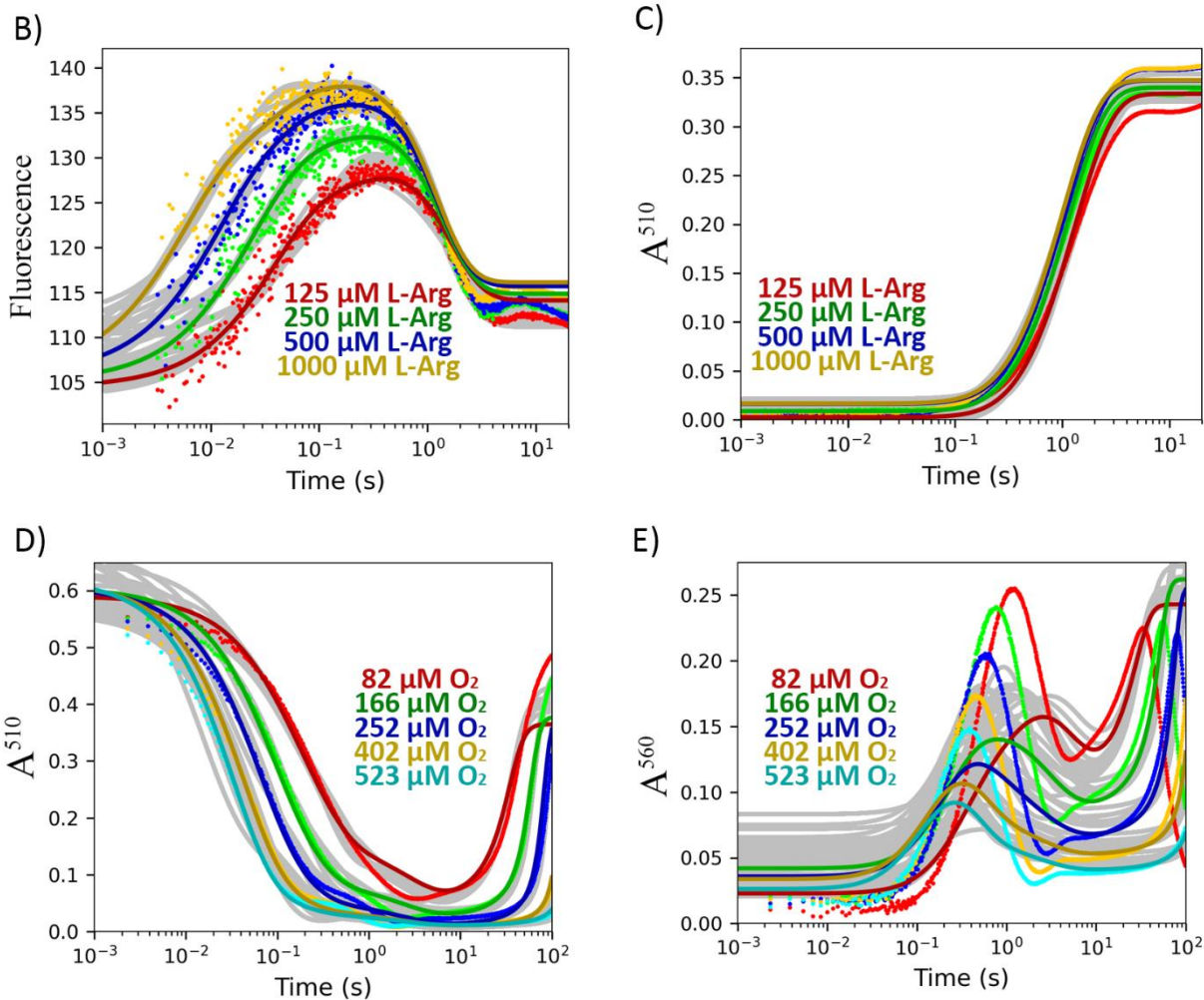


Figure 2.3.4-3: Pre-steady state kinetic data from global fitting in KinTek Explorer when the EK branch point was placed at intermediate X1. (A). B) the change in fluorescence after mixing 13 μM anaerobic PbrMppP with various concentrations of L-Arg under anaerobic conditions. C) The absorbance at 510 nm monitored in the same condition as for B. D) Double mixing experiment where 40 μM PbrMppP was mixed with 2000 μM L-Arg in the first step and aged for 10 s. In the second step, 100 mM BisTris propane buffer pH 8.5 containing various concentrations of dioxygen was mixed and absorbance at 510 nm was monitored. E) The same reaction as for D, but

monitored at 560 nm. The gray shading underneath the data refers to the range of parameters (rate constants and extinction coefficients) that would be consistent with the fit. The steps with unidirectional arrows were forced to be unidirectional by manually applying $K^- = 0$. The K labeled with purple color refers to the minor product of MppP reaction, [91]. The P with red color refers to the major product of MppP reaction, [95]. The rate constants with pink labeling refer to deviations from Model 1.

When the inactive form of the enzyme, F was omitted from Model 1 (Figure 2.3.4-4A), the aerobic experiments were more affected than the anaerobic experiments (Figure 2.3.4-4B-E). The O_2 dependency of Q2 (Figure 2.3.4-2E) is expanded far from all parameter ranges. The impact on rate constant was also deviated. The conversion of external aldimine to intermediate EA was 5 times slower than model 1. Whereas, the formation of EA intermediate from external aldimine ER' is almost unidirectional. Finally, EQ1 formation was also slower in case of model 4 (Figure 2.3.4-4E)

Model 4

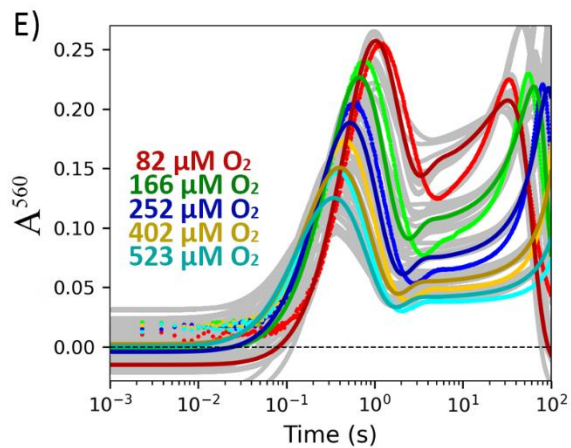
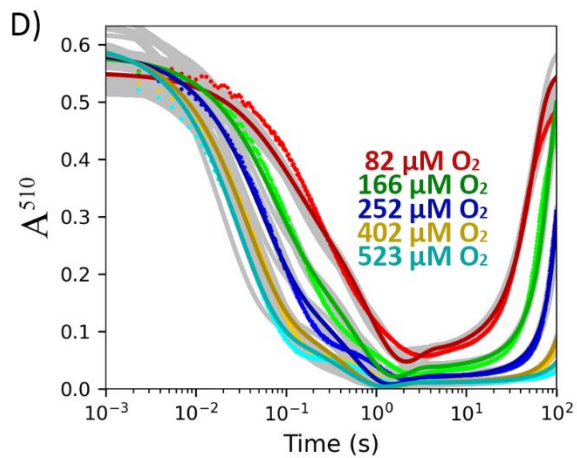
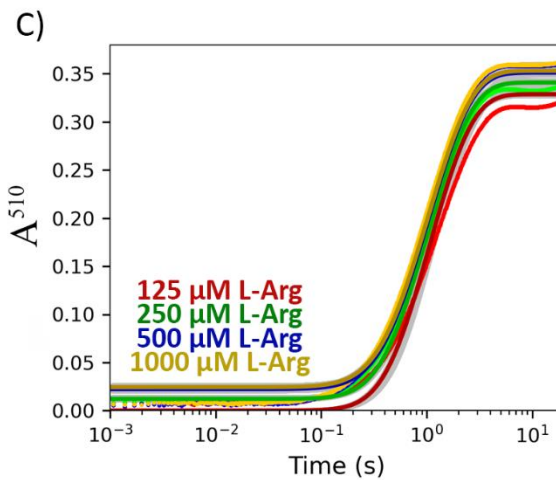
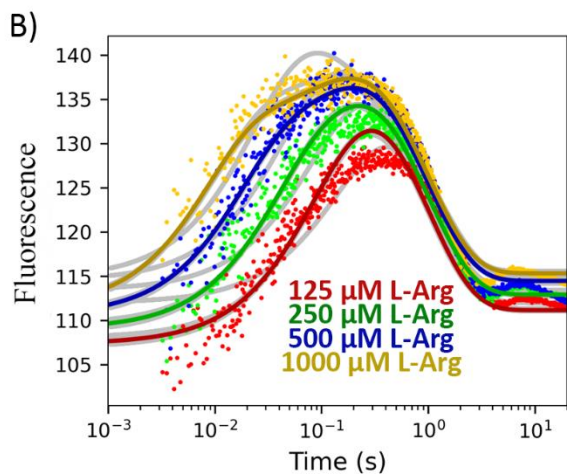
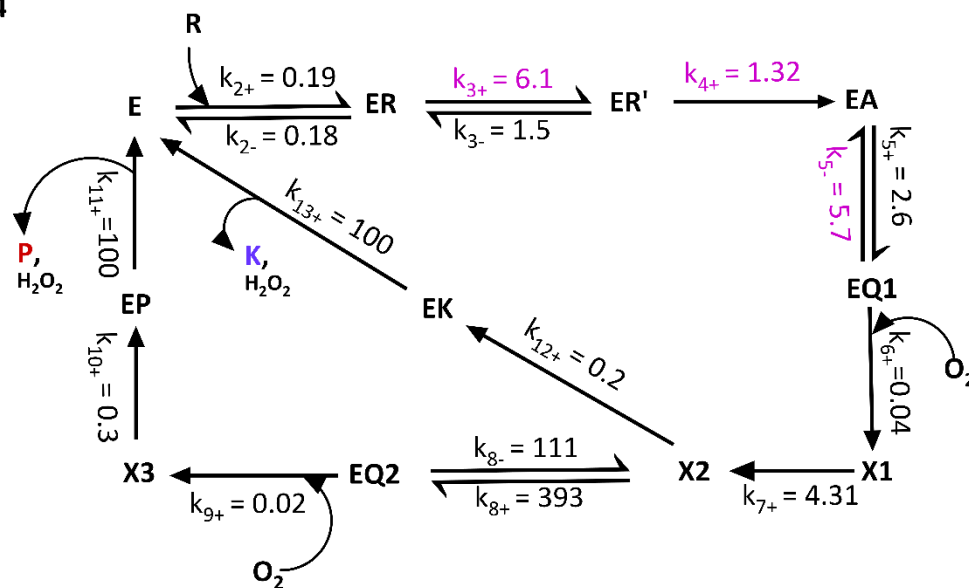
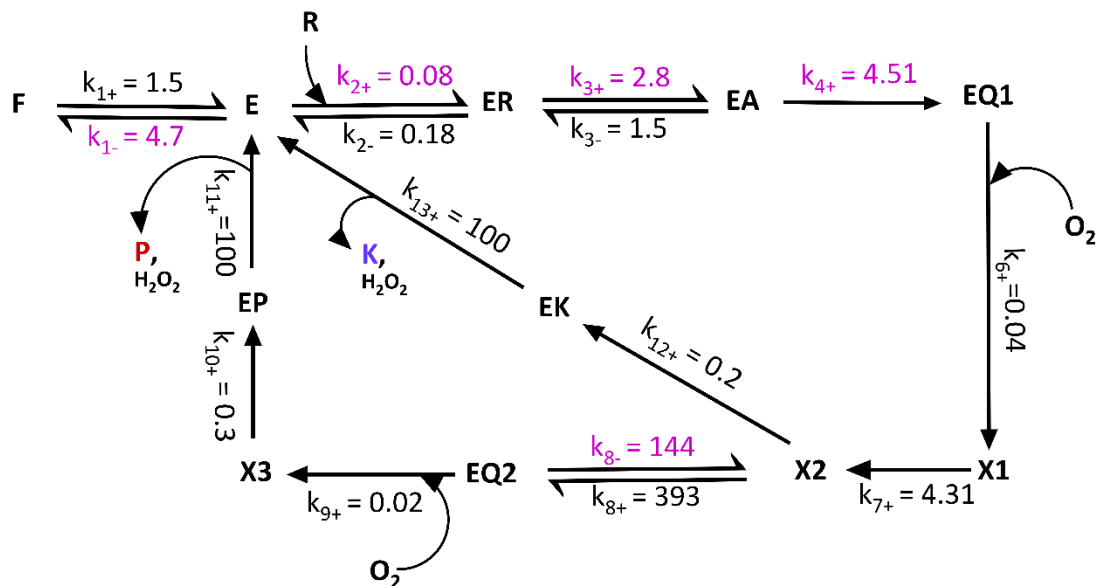


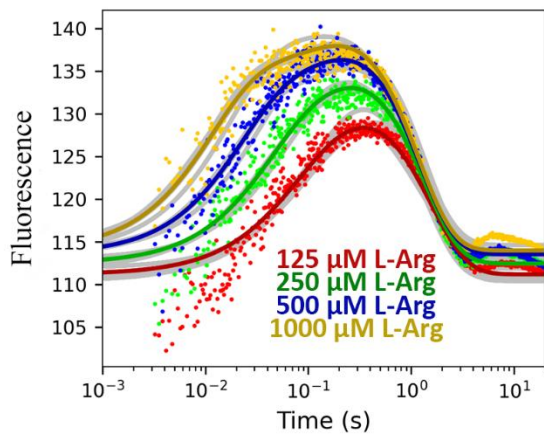
Figure 2.3.4-4: Pre-steady state kinetic data from global fitting in KinTek Explorer when the F is omitted from the model 1 (A). B) the change in fluorescence after mixing 13 μM anaerobic PbrMppP with various concentrations of L-Arg under anaerobic conditions. C) The absorbance at 510 nm monitored in the same condition as for B. D) Double mixing experiment where 40 μM PbrMppP was mixed with 2000 μM L-Arg in the first step and aged for 10 s. In the second step, 100 mM BisTris propane buffer pH 8.5 containing various concentrations of dioxygen was mixed and absorbance at 510 nm was monitored. E) The same reaction as for D, but monitored at 560 nm. The gray shading underneath the data refers to the range of parameters (rate constants and extinction coefficients) that would be consistent with the fit. The steps with unidirectional arrows were forced to be unidirectional by manually applying $K = 0$. The K labeled with purple color refers to the minor product of MppP reaction, [91]. The P with red color refers to the major product of MppP reaction, [95]. The rate constants with pink labeling refer to deviations from Model 1.

When intermediate ER' was omitted from model 1 (Figure 2.3.4-5A), the most impact was on Q2 accumulation phase (Figure 2.3.4-5E), as the model fitting expanded the parameter ranges far away from actual data points. Although the conversion step of inactive enzyme F to active enzyme E was slightly faster than that one in model 1, the consecutive step of internal aldimine formation was slightly slower. Finally, the EQ1 formation step was forced to be unidirectional.

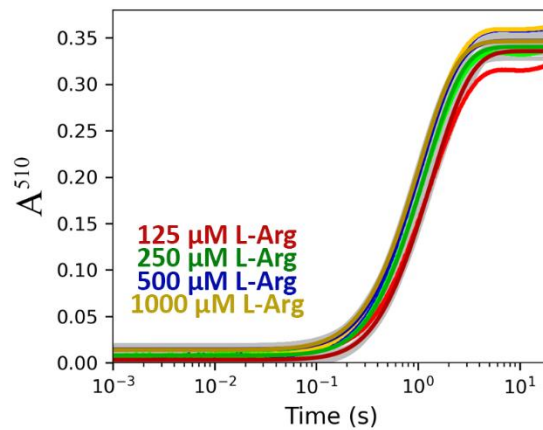
Model 5



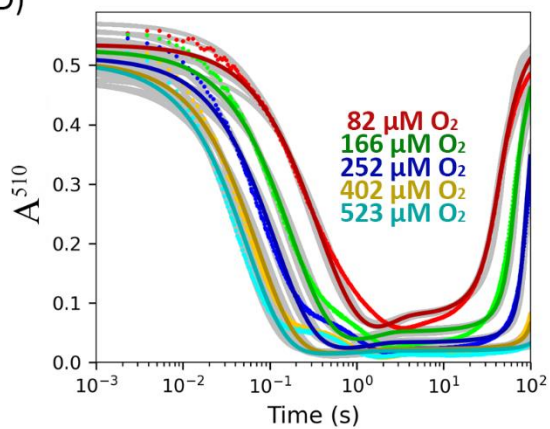
B)



C)



D)



E)

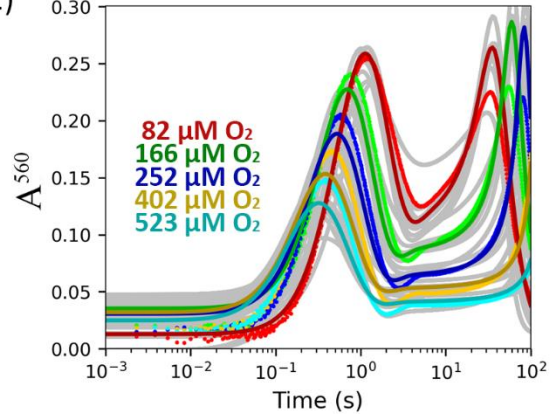
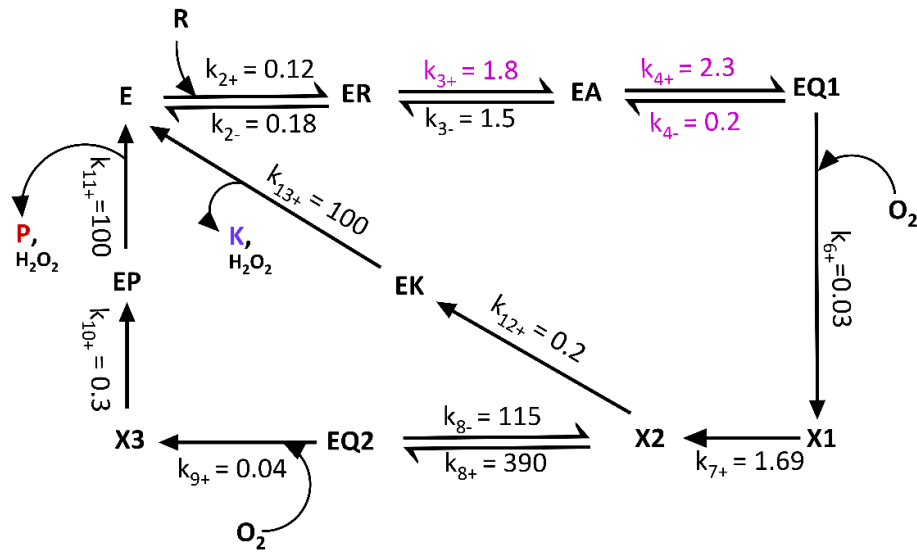


Figure 2.3.4-5: Pre-steady state kinetic data from global fitting in KinTek Explorer when the ER' intermediate is omitted from the model 1 (A). B) the change in fluorescence after mixing 13 μM anaerobic PbrMppP with various concentrations of L-Arg under anaerobic conditions. C) The absorbance at 510 nm monitored in the same condition as for B. D) Double mixing experiment where 40 μM PbrMppP was mixed with 2000 μM L-Arg in the first step and aged for 10 s. In the second step, 100 mM BisTris propane buffer pH 8.5 containing various concentrations of dioxygen was mixed and absorbance at 510 nm was monitored. E) The same reaction as for D, but monitored at 560 nm. The gray shading underneath the data refers to the range of parameters (rate constants and extinction coefficients) that would be consistent with the fit. The steps with unidirectional arrows were forced to be unidirectional by manually applying $K^- = 0$. The K labeled with purple color refers to the minor product of MppP reaction, [91]. The P with red color refers to the major product of MppP reaction, [95]. The rate constants with pink labeling refer to deviations from Model 1.

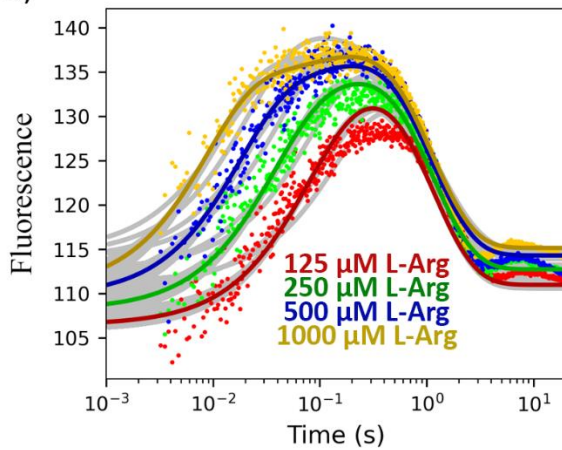
When F and ER' were both omitted from Model 1 to give Model 6, the effect was similar to Model 5. It is obvious that the formation of ER and EA is a major component of the mechanism, thus we omitted F followed by omission of ER'. This optimization disrupted the fitting of anaerobic experiment with fluorescent excitation (Internal aldimine monitoring, figure 2.3.4-6B), aerobic experiment of Q1 accumulation (Figure 2.3.4-6D) and Q2 accumulation (Figure 2.3.4-6E). All these model fitting expanded the parameter ranges far away from actual data points. Moreover, the conversion step of ER to ER' was 15 times slower than that one in model 1.

A)

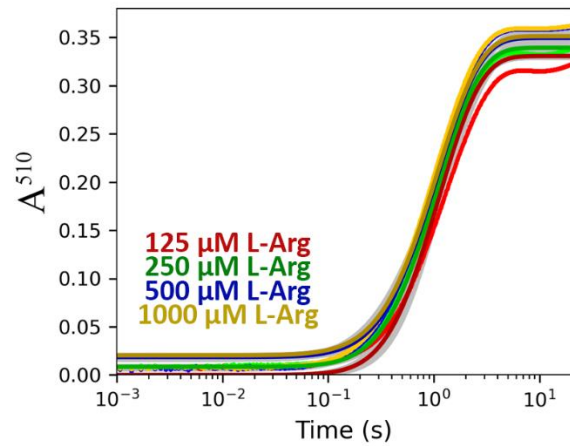
Model 6



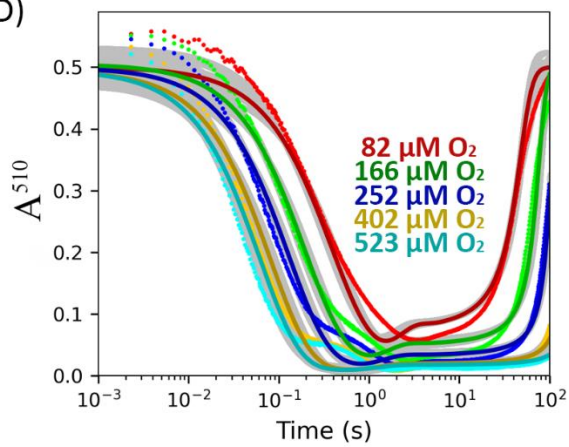
B)



C)



D)



E)

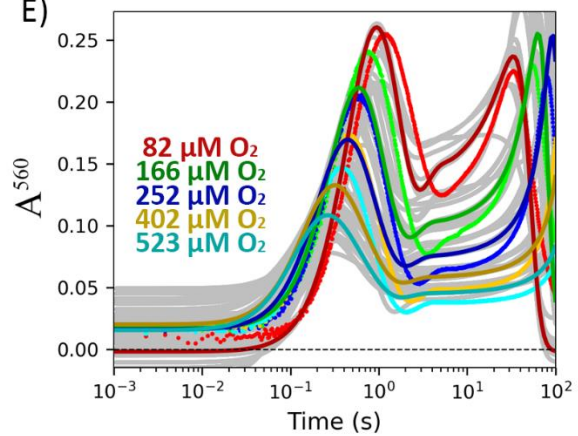


Figure 2.3.4-6: Pre-steady state kinetic data from global fitting in KinTek Explorer when the F is omitted, followed by omission of intermediate ER' from the model 1 (A). B) the change in fluorescence after mixing 13 μM anaerobic PbrMppP with various concentrations of L-Arg under anaerobic conditions. C) The absorbance at 510 nm monitored in the same condition as for B. D) Double mixing experiment where 40 μM PbrMppP was mixed with 2000 μM L-Arg in the first step and aged for 10 s. In the second step, 100 mM BisTris propane buffer pH 8.5 containing various concentrations of dioxygen was mixed and absorbance at 510 nm was monitored. E) The same reaction as for D, but monitored at 560 nm. The gray shading underneath the data refers to the range of parameters (rate constants and extinction coefficients) that would be consistent with the fit. The steps with unidirectional arrows were forced to be unidirectional by manually applying $K = 0$. The K labeled with purple color refers to the minor product of MppP reaction, [91]. The P with red color refers to the major product of MppP reaction, [95]. The rate constants with pink labeling refer to deviations from Model 1.

The statistical analysis of all the alternative models is recorded in Table 2.3.4-1. The primary model, Model 1, was the best global fit to the entire data set with the lowest all-experiment chi squared (χ^2) value. The statistics for Models 2 and 3 show that both alternative placements of the branch point in the mechanism fit the data poorly. Similar, more subtle impacts were observed for Models 4-6, testing the value of including the equilibrium between active and inactive protein (having F in the model) and of including the conformational change from ER to ER'. While not as devastating to the fit as placing the branch point in the wrong place in the mechanism, omitting F and/or ER' significantly degrades the fit of Models 4-6 to the data. Thus,

we consider Model 1 as being the best model for this set of experimental data. From this validated model, we proposed a reaction mechanism for PbrMppP catalysis (Scheme 2.3.4-1).

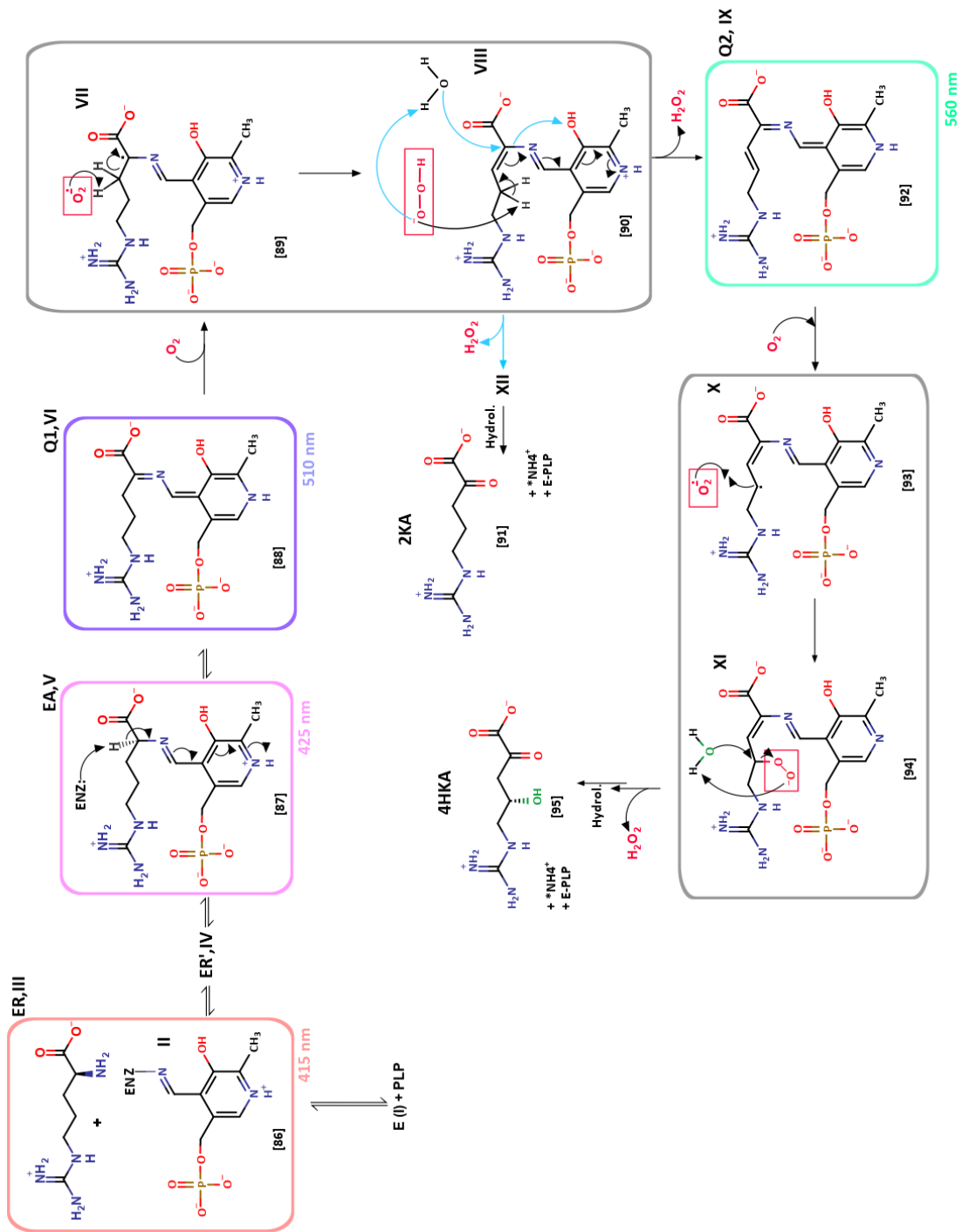
Table 2.3.4-1: Statistical analysis (observed χ^2 values) of different models explained in 2.3.4.

	Model 1	Model 2 (EQ2 to EK)	Model 3 (X1 to EK)	Model 4 (Omit F)	Model 5 (Omit ER')	Model 6 (Omit F & ER')
Exp 1 (415fl anaerobic)	3182	3048	6370	5931	4568	5129
Exp 2 (510 anaerobic)	2390	748663	2368	3360	2067	2862
Exp 3 (510+O₂)	1192	2950	4942	2894	3731	6377
Exp 4 (560+O₂)	618	1180	6900	916	862	1929
All-Exp χ^2	7.4 x 10³	7.6 x 10⁵	2.1 x 10⁴	1.3 x 10⁴	1.1 x 10⁴	1.1 x 10⁴

Starting from the active, internal aldimine form of PbrMppP (Scheme 2.3.4-1, II; the $E \rightleftharpoons F$ equilibrium has been omitted from Scheme 2.3.4-1), it could be binding of free PLP by enzyme with no cofactor, or it could be some conformational change that results in the two different

form of enzyme. Probably the N-terminus-open conformation of the enzyme allows the substrate L-Arg to bind. At this point, the internal aldimine (**ER,III**) form of enzyme absorbs light at 415 nm. Upon closing of the N-terminus, the active site residues are oriented properly for nucleophilic attack of L-Arg on C4' of PLP, resulting in the intermediate **ER' (IV)** formation, which ultimately forms the external aldimine (**EA, V**). The external aldimine absorbs light at 425 nm. Then the external aldimine eventually forms quinoid-I intermediate (**Q1, VI**). In the presence of O_2 , Q1 form an excited state dynamic (**VII**) where molecular oxygen forms reactive oxo-anion to attack the beta hydrogen of L-arg in PLP bound form or Q1 that has absorption maxima at 510 nm, and also originates hydroperoxide-anion. The other beta hydrogen rearranges to form double bond between alpha and beta carbon. In the next excited state dynamic (**VIII**), the mechanism follows two branches of reaction. If the hydroperoxide -anion attacks water molecule and leaves as hydrogen peroxide, then the aldimine carbon attacks the hydrogen from hydroxyl anion (**XII**) followed by incorporation of oxygen the alpha position to give the minor product 2-oxo-5-guanidinovaleric acid [91].

On the other hand, if hydroperoxide anion attacks the γ -hydrogen of L-arg in PLP bound form (**VIII**), then second molecule of H_2O_2 byproduct is produced along with quinoid-II complex (**Q2, IX**) that has absorption maxima at 560nm. A 2nd molecule of O_2 forms radical, attacking the activated γ -C (**X**). Finally, water molecule likely neutralizes the reaction producing intermediate **XI** and another molecule of H_2O_2 . Oxygen from the water molecule incorporates ultimately with the substrate to form 2-oxo-4(S)-hydroxy-5-guanidinovaleric acid [95].



Scheme 2.3.4-1: Proposed reaction mechanism of PbrMppp

2.3.5 PbrMppP substrate specificity

To explore the substrate specificity of PbrMppP, a series of L-Arg isosteres including D-Arg, L-Lys, L-ornithine, L-citrulline, N-methyl-L-Arg, and L-canavanine were tested for activity with PbrMppP. All reactions were performed in a total volume of 500 μ L containing 10 μ M PbrMppP and 10mM of each potential substrate. Reactions were monitored using a diode array spectrophotometer to detect the presence and rate of accumulation of the quinonoid species. In cases where turnover was observed (as build-up of Q1 due to dioxygen depletion), the product(s) were identified by ESI-mass spectrometry.

D-Arg, L-Lys, L-ornithine, and L-citrulline were not substrates for PbrMppP. When PbrMppP reacted with D-Arg [98], no reaction was observed, as compared with the reaction with L-Arg (Figure 2.3.5-1A,B). The structure of SwMppP with D-Arg bound suggests that the α -proton of D-Arg is oriented away from the catalytic Lys233 residue and is thus not in a position to progress past the external aldimine^{lix}. The reactions with L-Lys [99] reaction shows an absorption maximum at 432 nm suggestive of the external aldimine [100]. It is less clear why this peak decreases so markedly over time, but the shoulder at \sim 320 nm leads us to propose that either the external aldimine is slowly hydrolyzed, leaving the active site with pyridoxamine 5-phosphate (PMP) bound, or a ketimine intermediate is formed (Figure 2.3.5-1C and Scheme 2.3.5-1B). The reaction with L-ornithine [101] showed similar absorbance maxima at \sim 320 nm and 425 nm (Figure 2.3.5-1D). Like L-Lys, ornithine [101] is able to form the external aldimine [102] (Scheme 2.3.5-1C). As for lysine, the new peak at \sim 320 nm might be due to PMP or to ketimine [103] formation. When PbrMppP reacts with L-citrulline [104], the absorption maximum immediately

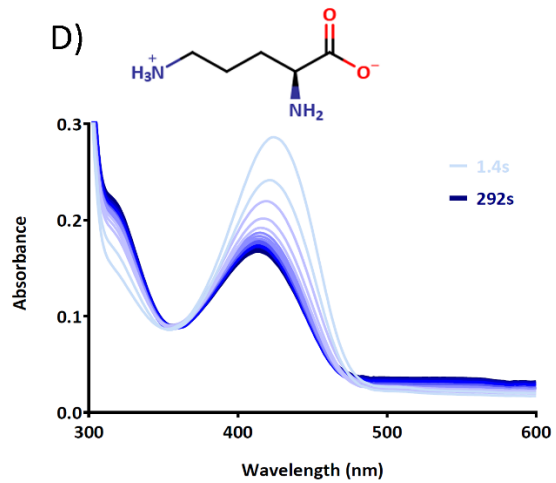
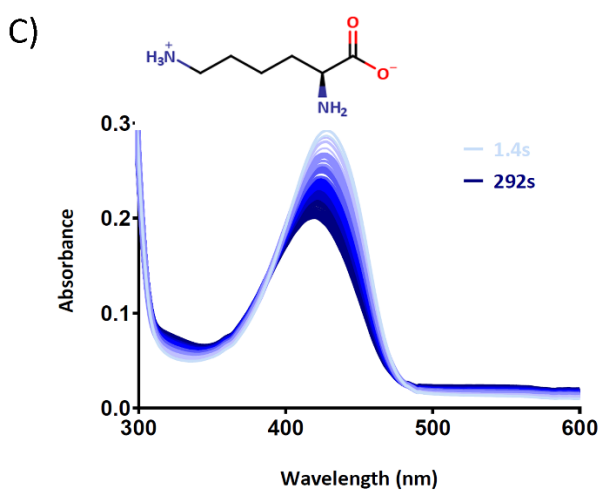
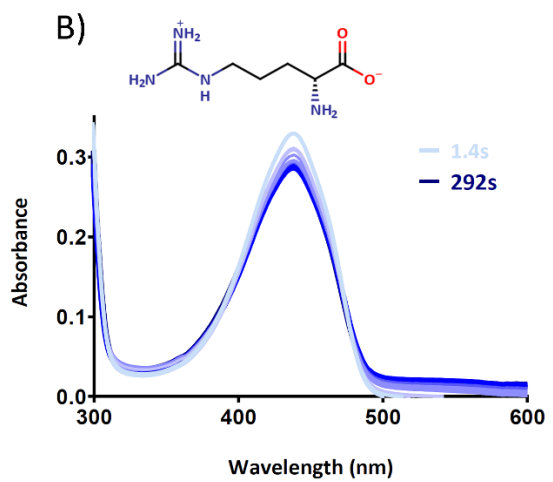
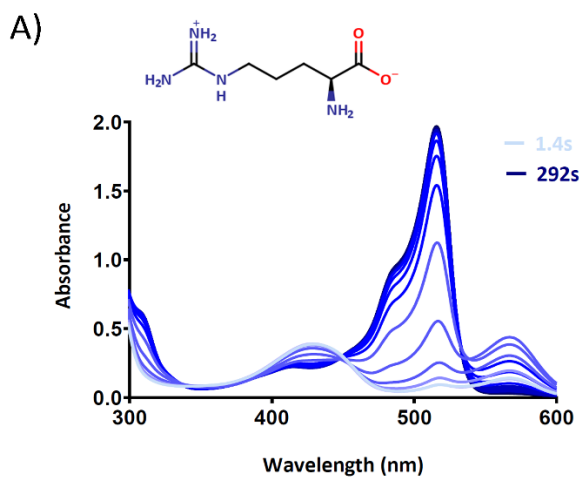
after mixing the enzyme and citrulline is at 425 nm, consistent with external aldimine formation [105]. Curiously, the absorbance at 425 nm decreases over time, but it is not replaced by any peak at ~320 nm (Figure 2.3.5-1E). While it is difficult to explain this observation, it is clear that L-citrulline is not a substrate for PbrMppP.

With N-methyl-L-Arg [106] as a substrate, a small accumulation of Q1 [108] is observed at 510 nm (Figure 2.3.5-1F). The observation of Q1 indicates that it is a substrate, but the fact that it does not accumulate much past its initial value suggests that the reaction is much slower than that with L-Arg so that the cuvette is not depleted of oxygen during the assay period. This slow reaction may be explained by the bulky addition of the methyl group to the L-Arg guanidinium nitrogen misorienting the substrate in the active site. It is also possible that, although the addition of the methyl group does not significantly alter the side chain pK_a of arginine, it may affect the charge distribution across the guanidinium group^{CVI}. If, as we hypothesize, the guanidinium group of the substrate provides the positive charge in the active site to encourage the formation of superoxide, then any deviation of this charge from its natural position might adversely affect the activation of dioxygen for electron transfer from Q1.

The L-canavanine [109] reaction shows Q1 [111] formation (Figure 2.3.5-1G), but no accumulation of Q2 (560 nm). This result suggested that L-canavanine [109] supports only two-electron oxidation, forming only 2-keto-L-canavanine [114]; the L-canavanine substrate is not converted to the fully oxidized product “4-hydroxy-2-ketocanavanine”. A proposed mechanism for the reaction of PbrMppP with L-canavanine, is presented in Scheme 2.3.5-1F. Beginning with the enzyme as the internal aldimine [86] form of enzyme that absorbs light at 415 nm, upon

closing of the N-terminus, the active site residues are oriented properly for nucleophilic attack of L-canavanine on C4' of PLP. Thus, the internal aldimine **[86]** is converted to external aldimine **[110]** due to proper orientation of the active site residues. The external aldimine **[110]** is eventually converted into quinoid-1 or Q1 **[111]** intermediate that gives absorbance at 510nm. Afterward, O₂ molecule forms reactive oxygen species that attacks the β-H forming hydroperoxide, leaving Dehydro-L-canavanine bound with PLP. The hydroperoxide then attacks the γ-H on the PLP bound L-canavanine and produces H₂O₂ byproduct. Finally, the product 2-keto-L-canavanine **[114]** leaves the active site, regenerating the enzyme-PLP complex.

Due to the presence of an oxygen atom at the δ-position, the pKa of the γ-carbon is shifted by the oxygen, making it more difficult to abstract a proton. Consequently, possible hydroperoxide radical can't attack the γ-H. As a result, PbrMppP catalyzes only 2 electron oxidation of L-canavanine.



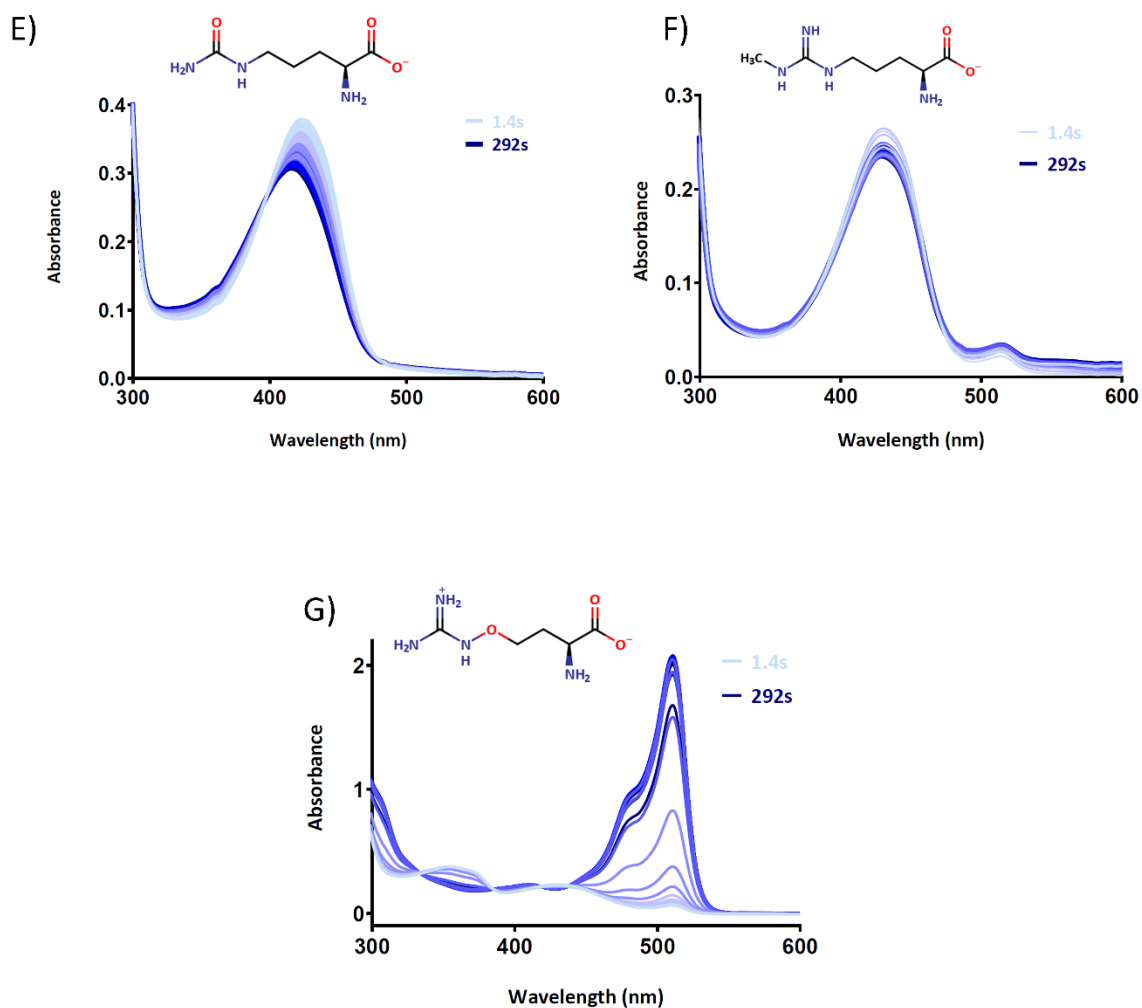
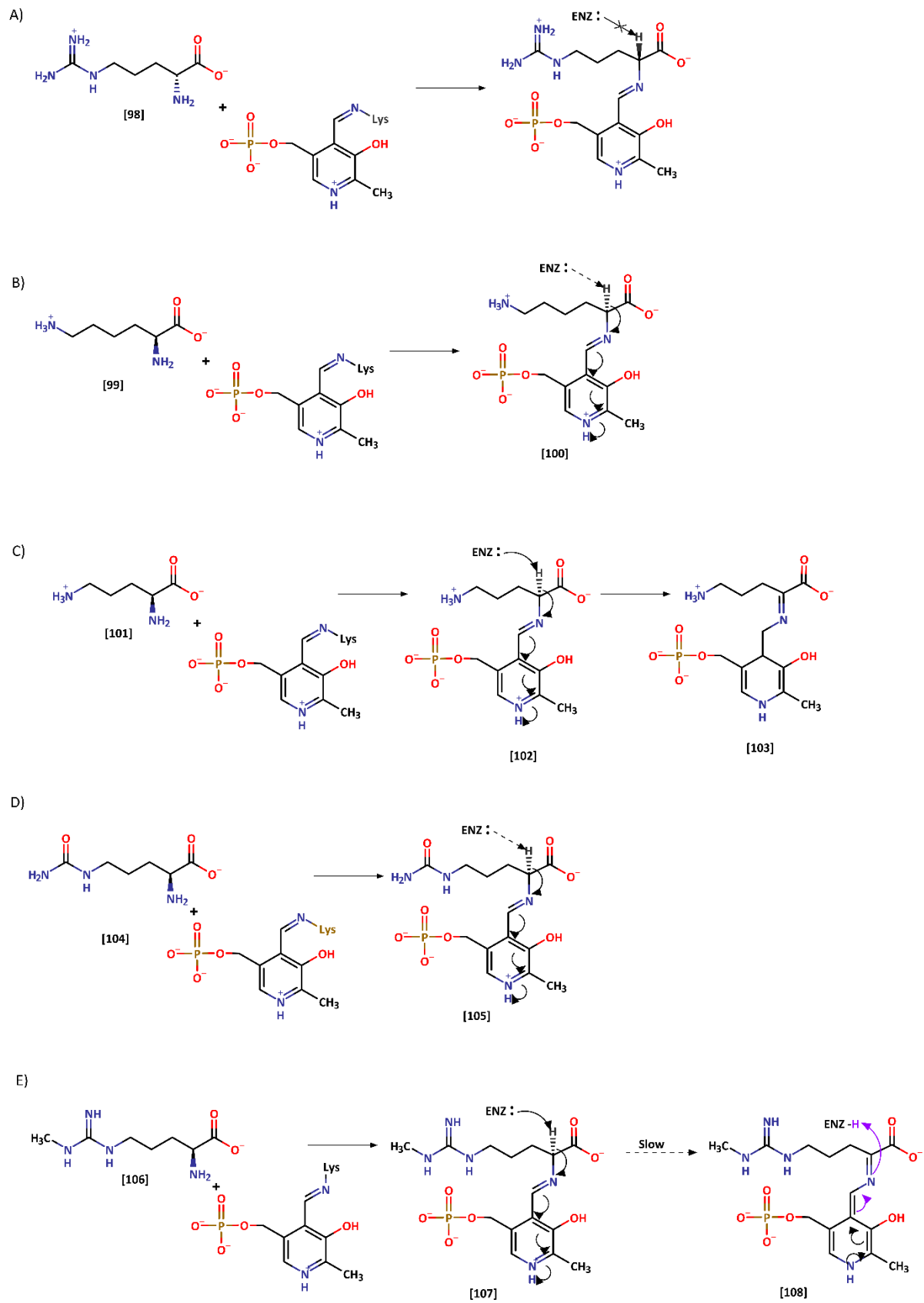
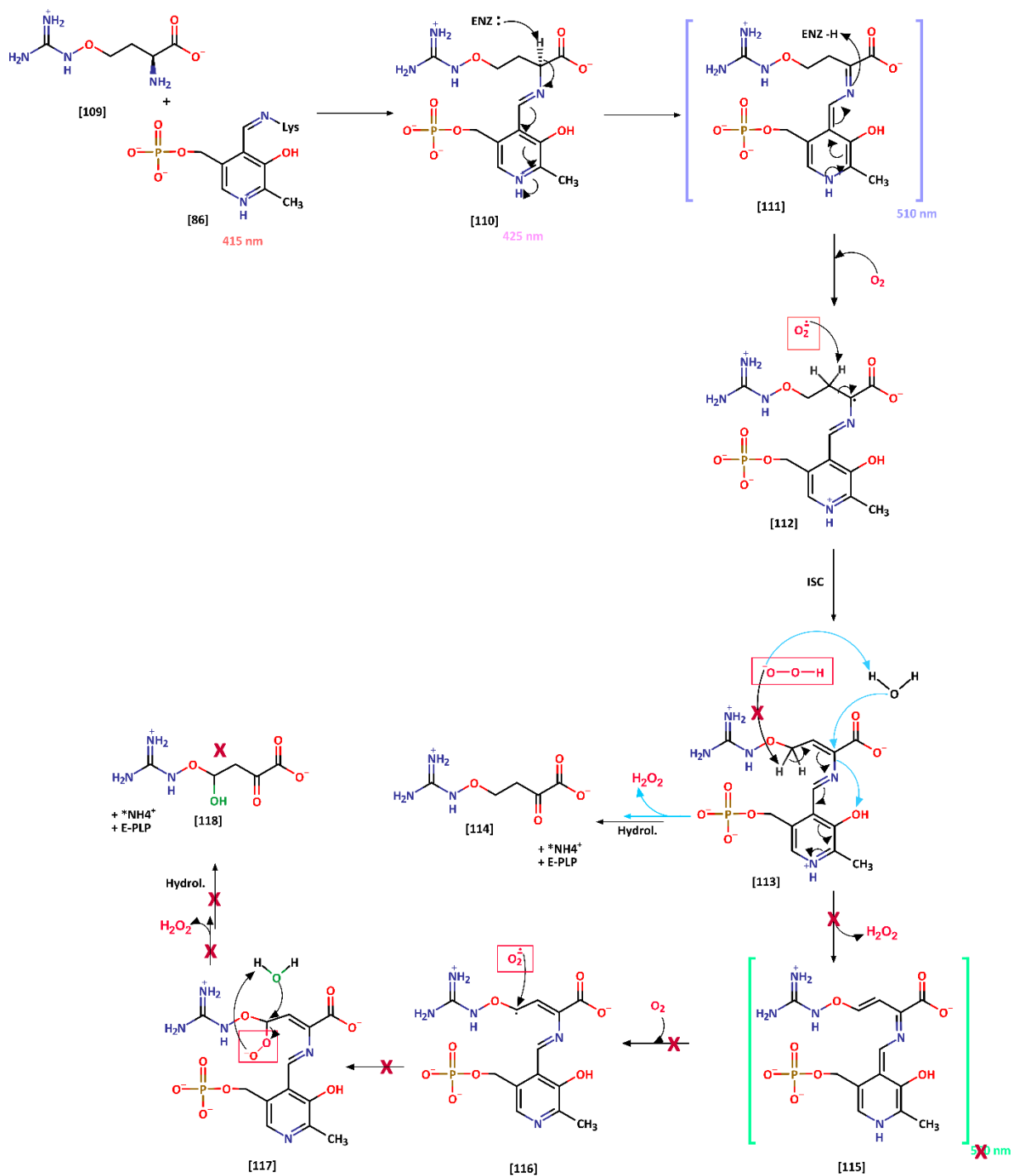


Figure 2.3.5-1: *PbrMppP* substrate specificity. Spectrometric scans of reactions with L-Arg and a panel of potential alternative substrates: (A) L-Arg, (B) D-Arg, (C) L-Lys, (D) L-ornithine, (E) L-citrulline, (F) N-methyl-L-Arg, (G) L-canavanine. The structure in the subset of each panel shows corresponding substrate structure. The light blue spectrum represents the first measurable point of the reaction (1.4s). Subsequent spectra (every 25 s pictured) grade from light blue to dark blue, which is the last spectrum collected (292 s).



F)



Scheme 2.3.5-1: Possible reaction mechanisms of PbrMppP with different substrate analogs. (A) D-Arg, (B) L-Lys, (C) L-ornithine, (D) L-citrulline, (E) N-methyl-L-Arg, (F) L-canavanine. A) D-Arg [98] barely forms external aldimine due to orientation of α H away from catalytic K-233; B) L-lysine; [99] forms a possible external aldimine [100] which shows absorption maxima at 432 nm; (C) L-ornithine [101] forms external aldimine [102] followed by ketamine [103]; D) L-citrulline [104] probably gives a slow reaction to form external aldimine [105] which eventually diffuses out of active site; E) N-methyl-L-arg [106] forms external aldimine [107] followed by very slow formation of Quinoid-I [108] which gives absorbance at 510nm; F) Proposed reaction mechanism of PbrMppP with L-canavanine [109]. L-canavanine [109] is a true substrate of PbrMppP following one equivalent of oxygen reduction. The hypothetical reaction forms expected external aldimine [110] ($\lambda_{max} = 425$ nm) followed by Quinoid-I formation [111] ($\lambda_{max} = 510$ nm), consequently the oxidized product 2-keto-L-canavanine [114] is released regenerating PLP-PbrMppP complex of internal aldimine ($\lambda_{max} = 415$ nm). Due to the presence of electronegative oxygen atom at δ -position, bonding electron from γ -C is attracted towards the oxygen atom, making the γ -H less likely an electrophile. Consequently, possible hydroperoxide can't attack to the γ -H. As a result, PbrMppP catalyzes only 2 electron oxidation of L-canavanine. Otherwise, it is also possible that the fully oxidized form of L-canavanine is converted so quickly that it is not tractable via spectrophotometry.

Comparison of the rate of Q1 formation with L-Arg and L-canavanine shows that PbrMppP accumulates Q1 faster with L-Arg, as would be expected for the natural substrate. What is interesting here, is that the canavanine reacts quickly enough to deplete the cuvette of dioxygen. It is not clear at this time what species is absorbing light at 350 nm. This species decays slowly, but that is all that can be said at this point.

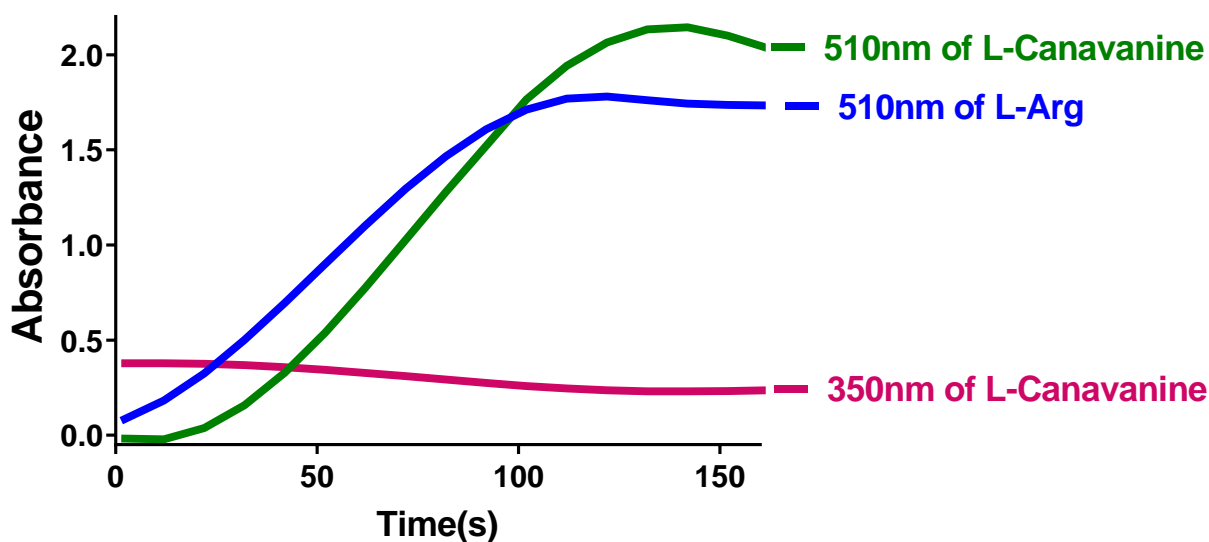
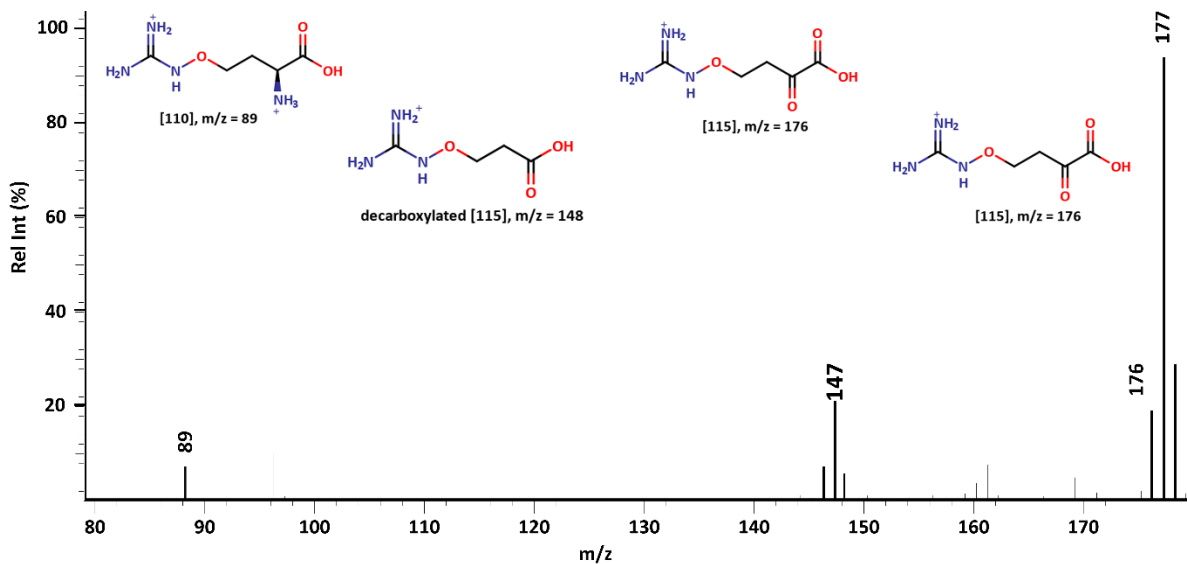


Figure 2.3.5-2: Comparison of quinonoid-I formation by L-arg and L-canavanine in PbrMppP reaction. L-canavanine conversion is slower (green) than L-arg conversion (Blue) thus L-arg reaction reached equilibrium early. On the other hand, slow decay of 350nm species (pink) is proportional to the formation of quinoid-I in L-canavanine reaction.

The reaction product of PbrMppP and L-canavanine was also analyzed by ESI-Mass spectrometry (Figure 2.3.5-3). In the positive mode, m/z 89 is consistent with the doubly charged cationic form of L-canavanine, whereas m/z 177 is consistent with the singly charged cationic form **[109]**. The $m/z=176$ is the MppP 2-electron oxidized product 2-keto-L-canavanine **[114]** and $m/z=148$ reveals the presence of decarboxylated **[114]**. In the negative mode, $m/z=175$ stands for the singly charged anionic form of L-canavanine. The PbrMppP product **[114]** appears at $m/z=174$, whereas the decarboxylated **[114]** appears at $m/z=146$ in negative mode. The mass difference between the L-canavanine starting material and the 2-electron oxidized product of the PbrMppP reaction is small. The presence of the decarboxylated product (m/z 148 in positive mode and m/z 146 in negative mode) is an important indicator that L-canavanine was truly modified by PbrMppP, and it, thus, a substrate for the enzyme.

A)



B)

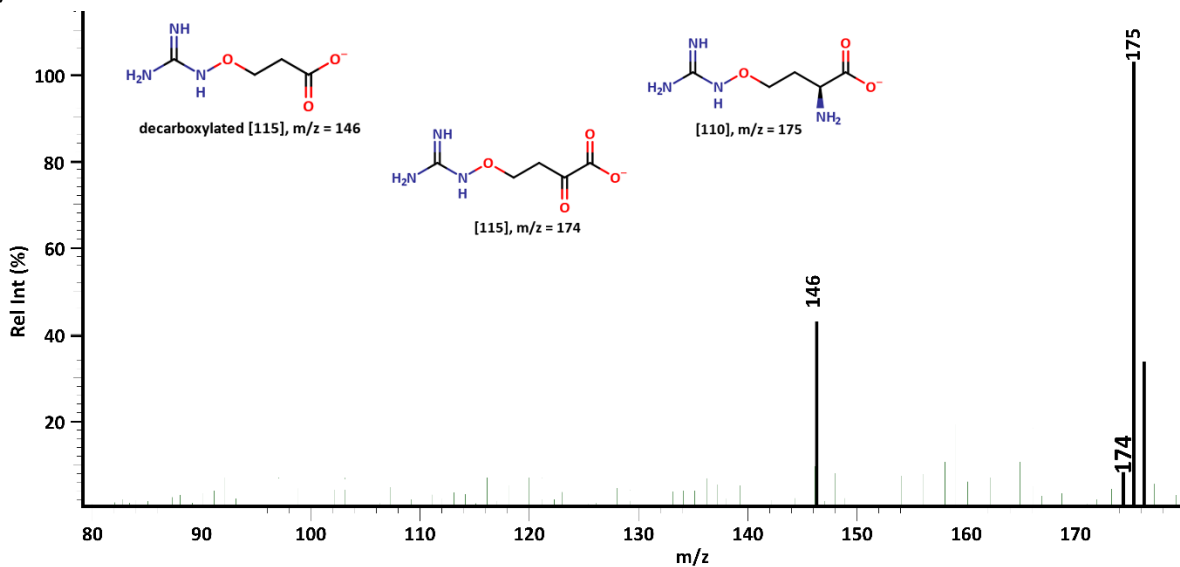


Figure 2.3.5-3: ESI mass spectrometric analysis of MppP reaction products with L-canavanine; (A) positive mode, (B) Negative mode. Reactions were performed in 500 μ L volume, with 50 μ M PbrMppP and 5 mM L-Canavanine in 10 mM BisTris Propane pH 8.5.

2.3.6 Steady state kinetics of the reaction of PbrMppP with L-canavanine

The steady state kinetic parameters (K_M , k_{cat} , specificity constant) and Michaelis-Menten plot in Figure 3.7-1 for the reaction of PbrMppP with L-canavanine were determined by calculating the initial rates of dioxygen consumption in oxygen probe experiments as was done for PbrMppP kinetics with L-Arg. The kinetic constants are listed in Table 2.3.6-1 (L-canavanine Kinetic parameters highlighted green). As L-canavanine is not capable of 4-electron oxidation, there is no branch in the mechanism with this substrate. The Turnover number with L-canavanine of $7.9 \pm 0.5 \text{ s}^{-1}$ is ~ 50 -fold higher than the turnover number with L-Arg ($0.16 \pm 0.03 \text{ s}^{-1}$). The proposed mechanism of the reaction of PbrMppP with L-canavanine is presented in Figure 2.3.5-1F.

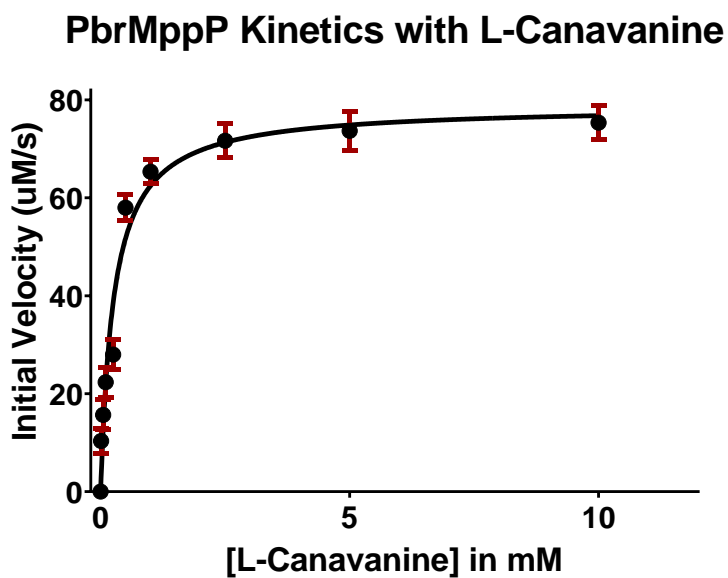


Figure 2.3.6-1: PbrMppP Michaelis-Menten plot with L-canavanine as substrate.

Table 2.3.6-1: Comparison of the steady state kinetic parameters of PbrMppP with L-Arg or L-canavanine as substrate.

PbrMppP Kinetics	L-Arg	L-canavanine
Michaelis-Menten constant, K_M	$24 \pm 1.5 \mu\text{M}$	$262 \pm 17.8 \mu\text{M}$
Turnover number, k_{cat}	$0.16 \pm 0.02 \text{ s}^{-1}$	$7.9 \pm 0.5 \text{ s}^{-1}$
Specificity Constant, k_{cat}/K_M	$(6.8 \pm 0.1) \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$	$(30 \pm 0.002) \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$

PbrMppP seems to have lower affinity for L-canavanine than L-Arg, which is somewhat surprising given that the two are almost identical in size, shape, and placement of hydrogen bond donors and acceptors. It should be noted, however, that as the δ -oxo analog of L-Arg, canavanine has an oxyguanidinium group, which has a significantly lower pKa (7.0) than the guanidinium group of L-Arg (~ 12.5)^{cvi}. These reactions were performed at pH 9.0, so the side chain of L-canavanine would be almost completely deprotonated. Without the charge on the oxoguanidine, L-canavanine may not interact with the PLP phosphate group in exactly the same way as L-Arg. Of course, if L-canavanine is not positively charged under the assay conditions, and the substrate cation really is required for activation of dioxygen, it is very surprising that the turnover number with L-canavanine is so much higher than the turnover number with L-Arg. While work on this question is ongoing, we feel it is very clear that PbrMppP is a homolog of the prototypical SwMppP that catalyzes exactly the same reaction with very similar kinetic properties. In order to find out how this activity fits into the putative PbrMPCO, we set out to determine the structures of the rest of the PbrMPCO enzymes in the hope of getting clues to their respective functions.

2.3.7 Structural characterization of PbrHYP

Primary structure analysis of PbrHYP reveals that it is most similar to sequences having the PilT N-terminus (PIN) domain fold. PIN domains are most commonly associated with endoribonuclease activity that is involved in the pathogenicity in of Gram-negative bacteria. *Pseudomonas brassicacaerum* (DF41) is a soil-dwelling, non-pathogenic Gram-negative bacterium associated with the roots of plants in the mustard/cabbage family that protects the plants from several bacterial and fungal pathogens^{cviii}.

2.3.7.1 Introduction to PIN domain

The PIN domain superfamily contains small proteins (~120-130 amino acids) with distinctive structures comprised of a core of at least four parallel beta strands flanked by at least three alpha helices^{cxix} (Figure 2A). The superfamily is named after the *pilT* gene product from numerous Gram-negative bacteria. A distinctive feature of this fold is that aspartate and glutamate residues that form a binding site for divalent metal ions^{cx}, especially Mg²⁺ or Mn²⁺^{cxii}. The active site contains a metal-binding tetrad (DEDD motif) comprised of three highly conserved acidic residue (DED) and one poorly conserved residue where the aspartate is sometimes replaced by asparagine^{cxii,cxiii}.

PIN domain proteins occur in archaea, prokaryotes, and eukaryotes. Interestingly, PIN domain proteins evolved activities in processes ranging from cell motility to cellular stress responses. In

archaea^{cxiv} and bacteria^{cxv}, PIN domain-like proteins act as part of a type-II toxin-antitoxin system (TA system), also familiar as virulence associated proteins.

The antitoxin protein binds with the toxin enzyme to suppress the activity under normal conditions. Degradation of the antitoxin protein is triggered when the cell encounters a stressful environment (*e.g.* competition with other species). In the absence of the antitoxin protein, the endoribonuclease activity of the toxin is activated^{cxvi}. Bacterial PIN domain-like proteins act as 5'-3' endonuclease in type-II TA systems, cleaving own RNA to undergo into static state. This feature is held responsible for prolonged and recurring infection in host^{cxvii}. On the other hand, eukaryotic PIN domain-like proteins function as exonucleases in various RNA surveillance mechanisms like RNA interference (RNAi), degrading nonsense mRNA^{cxviii, cxix}.

2.3.7.2 PbrHYP is a PIN domain homolog

PbrHYP crystallized in space group P4₁2₁2 and the crystals of PbrHYP diffracted to 2.0 Å resolution. There were 2 molecules in one asymmetric unit. From the crystallographic structure (Figure 2.3.7.2-1), PbrHYP appears to be a disulfide-linked homodimer. The overall fold is similar to the PIN domain,ⁱⁱⁱ with a central six-stranded beta sheet surrounded by seven alpha helices. This homo-dimeric quaternary structure is prevalent in the PIN domain-like superfamily^{cxx}. A PDBefold^{xxi} search shows that the structure of PbrHYP has highest structural similarity to VapC (PDB ID 4CHG). Even though the sequence identity is only 10%, the root mean square deviation for the superposition of the PbrHYP and VapC structures is 2.46 Å. VapC is the toxic endoribonuclease from the *Mycobacterium tuberculosis* type-I TA system. Under favorable

growth conditions, VapC activity is suppressed by binding with the antitoxin protein VapB. Under adverse conditions, VapB is degraded, activating VapC's endoribonuclease activity^{xxii}. The *vapB* and *vapC* genes are adjacent in *M. tuberculosis* chromosome, which is pivotal for bacterial TA system function^{cxxi, cxxii, cxxiii}.

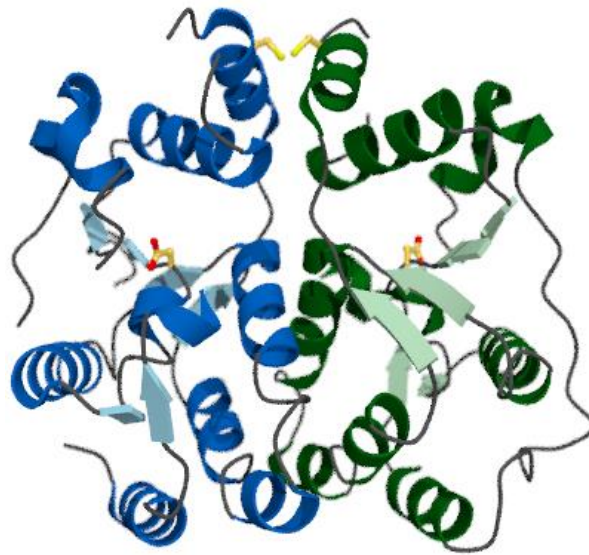


Figure-2.3.7.2-1: PbrHYP dimer structure. The yellow bond on the top indicated disulfide link between the two dimer.

2.3.7.3 PbrHYP has no metal-binding site

PbrHYP possesses a low degree of sequence identity (<10%) with the PIN domain protein VapC, it is perhaps not surprising that important functional features of the latter protein are missing from PbrHYP. The superimposed structures of the PbrHYP and VapC active sites (Figure 2.3.7.3-1) show that in VapC, three aspartate residues, Asp10, Asp122, and Asp96, coordinate two Mg^{2+} ions (purple spheres). Two of the aspartate residues, D10 and D122, are retained in the PbrHYP active site, but the other metal-binding residues of VapC, Asp96, and Asp116, are replaced with lysine (Lys104) and serine (Ser124), respectively, in PbrHYP. The lysine-for-aspartate substitution effectively eliminates any potential metal-binding character in PbrHYP. Glu42 of VapC, which is part of the second shell of residues around the dimetal center, is replaced by isoleucine (Ile49) in PbrHYP. Thus, while PbrHYP retains the gross features of the PIN domain fold, it has lost the functional residues central to the endoribonuclease activity of other PIN domain superfamily members and likely represents an unknown group of proteins within the superfamily with an as-yet-uncharacterized function.

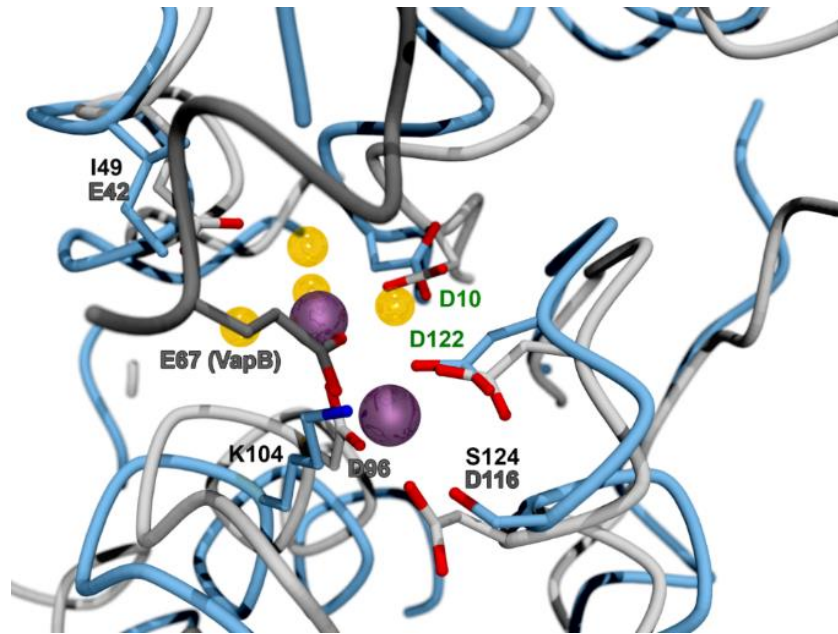


Figure 2.3.7.3: Overlaid structures of PbrHYP (blue) and VapC (PBD ID: 4CHG; light grey). The dark grey ribbon is a part of VapB that interfaces with the dimetal center of VapC to prevent endoribonuclease activity.

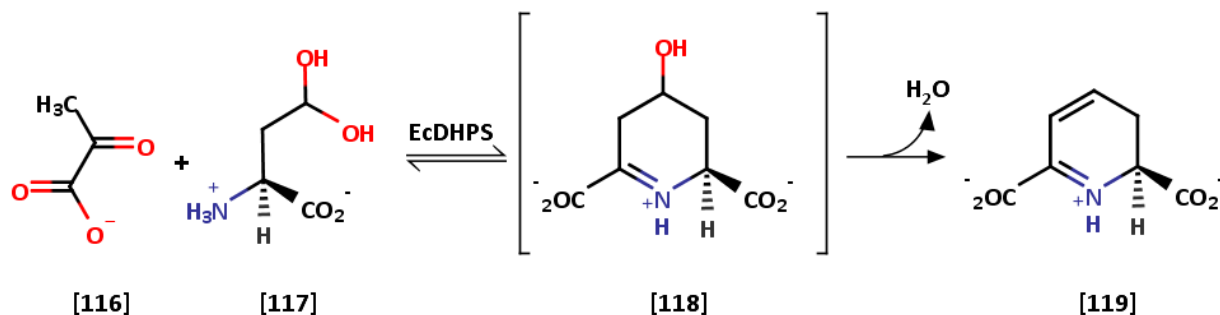
2.3.7.4 Absence of an anti-toxin component near the PbrHYP gene

In prokaryotic TA systems, the toxin and antitoxin components are expressed together under same promoter, normally from adjacent genes^{xxv, xxvi, xxvii}. Additionally, at the structural level, the interaction of the toxin and antitoxin parts plays an imperative role. Furthermore, the antitoxin actually does not dissociate completely from the toxin part. Rather, a glutamate residue consistently interacts with the toxin at the binding interface, preferably by an arginine residue^{cxxiv}. Regardless of the fact that antitoxin removal activates the toxin enzyme, a weak ribonuclease activity was observed *in vitro* with the toxin-antitoxin complex^{cxxv}.

Curiously, there is no antitoxin protein near PbrHYP in DF41 genome. None of the other proteins in PbrMppP gene cluster are homologous to antitoxins. As putative antitoxins are small protein comprise of 80-85 amino acids, we made a genome wide search for DF41 and interestingly only one annotated antitoxin (CD58_RS00980) was found in *pseudomonas brassicacaerum* (DF41) genome^{cxvii}, which is not adjacent to the PbrHYP gene.

2.3.8 Structure of PbrDHPS

The next enzyme in the PbrMPCO is a dihydrodipicolinate synthase family protein we are tentatively calling PbrDHPS until its actual activity can be identified. PbrDHPS is predicted to be a Type I aldolase. Type I aldolases use a catalytic lysine to form a Schiff base intermediate with the substrate, as opposed to the metal-dependent Type II aldolases. We determined the X-ray crystal structure of PbrDHPS at 2.0 Å resolution. The enzyme was annotated as DHPS because of its homolgy (26% sequence identity) with dihydrodipicolinate synthase from *E.coli* (EcDHPS)^{cxviii}. A PDBefold^{cxviii} search also shows highest structural similarity to EcDHPS (RMSD 1.8Å) The function of EcDHPS is to condense pyruvate **[116]** and 2-(S)-aspartate semi-aldehyde **[117]** into 4-hydroxy-tetrahydropicolinic acid **[118]**, which is then non-enzymatically dehydrated into 2,3-dihydrodipicolinate **[119]** (Scheme 2.3.8-1). This compound is incorporated into the L-Lys and *meso*-L-diaminopimelate biosynthesis pathways in plants and bacteria^{cxix}.



Scheme 2.3.8-1: Reaction catalyzed by EcDHPS.

2.3.8.1 Type I aldolases

The Type I aldolases^{cxxx} are a ubiquitous enzyme group with members found in all domains of life. From a structural perspective, most type I aldolases possess the TIM barrel fold with eight parallel β -strands forming a barrel in the center of the protein (Figure 2.3.8.2-1A). Each strand is connected to the next one by an α -helix. These 8 helices pack against the outside of the central barrel.^{cxxxii} The TIM barrel is so far most abundant protein fold known in the Protein Data Bank (PDB), and in all cases, TIM barrel proteins have proven to be enzymes. The catalytic site is invariably located at the C-terminal end of the barrel.

As mentioned earlier, Type I aldolase mechanisms are characterized by the involvement of a catalytic lysine residue that forms an iminium complex (*i.e.* Schiff base) with the substrate^{cxxxii} (Scheme 2.3.8-1). The catalytic lysine residue makes electrophilic attack at the carbonyl group of the substrate, forming the iminium ion. It can catalyze both the aldol condensation of a ketone and an aldehyde substrate, or the retroaldol cleavage of a ketone substrate. In the aldol condensation reaction, the binding of substrates is typically sequential ordered with the ketone

substrate binding first to form animinium, that then facilitates binding of the aldehyde substrate^{CXXXIII}.

2.3.8.2 PbrDHPS is likely a Type I aldolase

The quaternary structure of PbrDHPS (Figure 2.3.8.2-1A) reveals that it is a 2-fold symmetrical homodimer. The active site structure (Figure 2.3.8.2-1B) shows that there is a lysine residue (Lys160) adjacent to a tyrosine (Tyr132) that resemble the catalytic dyad found in many Type I aldolases^{CXI}.

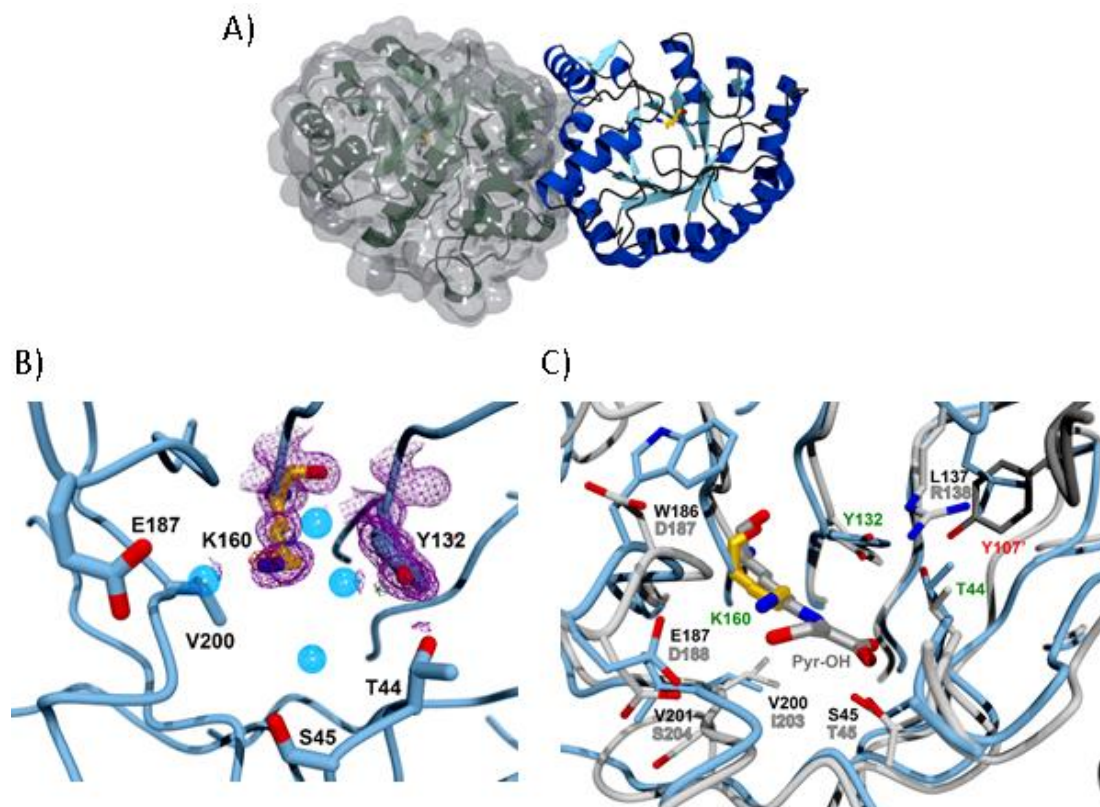


Figure 2.3.8.2-1: (A) PbrDHPS homodimer with core alpha-beta barrel fold. (B) PbrDHPS active site residues with the Lys-160 and Tyr-132 shown with electron density map contoured at +1.0

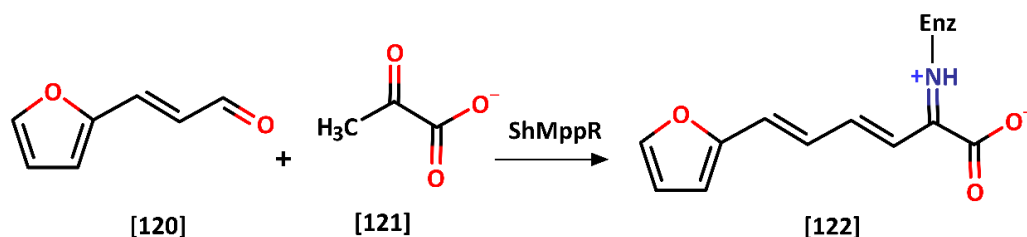
sigma. (C) Overlay of PbrDHPS (blue) and EcdHPS (gray, PDB ID: 1DHP)^{cxxxiv}. The root mean square deviation for matching C α atoms is 1.8Å. EcdHPS is shown with pyruvate covalently bound as the enamine with Lys-161. The Tyr107' (dark grey) is contributed by the other protomer of the dimer.

The superimposed structures of PbrDHPS and EcdHPS (Figure 1C) have an RMSD value of 1.8Å for all matched C α atoms (SSM algorithm)^{cxxxv}. The Lys161/Tyr133 catalytic dyad of EcdHPS (Lys160/Tyr132) is conserved in PbrDHPS. Most other active site residues are either conserved or are changed to chemically similar amino acids. For example, Thr44 is conserved and Thr45 of EcdHPS is replaced by Ser45 in PbrDHPS, and Ile203 is replaced by Val200. Less conservative changes are observed at Arg138 and Asp187 in EcdHPS, which are substituted by Leu137 and Trp186, respectively, in PbrDHPS. Overall, the PbrDHPS appears to be slightly more hydrophobic than the EcdHPS active site. Asp187 is a catalytically important residue in the EcdHPS reaction to anchor the substrate in the active site^{cxxxvi}. Its substitution by tryptophan in PbrDHPS likely reflects a different substrate preference in PbrDHPS^{cxxxvi}.

2.3.8.3 The putative catalytic lysine residue is active in PbrDHPS

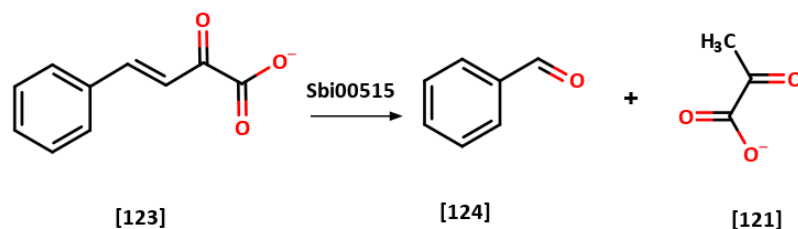
Previous studies showed that the enzyme MppR from *Streptomyces hygroscopicus* (ShMppR) catalyzes the cyclization of the MppP product [95] into the ketone form of enduracididine [96] (scheme 1.4.5-1)^{cxxxvii}. Although its fold bears no relationship to the TIM barrel fold of the Type I aldolases, MppR uses a Schiff base-forming lysine residue and is capable of catalyzing aldol condensation reactions, albeit poorly. At high concentrations of ShMppR (~100 μ M) condensation

of the aldehyde substrate 3-(2-furyl) acrolein **[120]** with pyruvate **[121]** (Scheme-3) was observed as a bright yellow color ($\lambda_{\max} = 360 \text{ nm}$) resulting from the condensation product **[122]**^{cxxxviii}.



Scheme 2.3.8.3-1: Structure of ShMppR product **[122]** when alternative aldehyde substrate 3-(2-furyl)acrolein **[120]** and pyruvate **[122]** are condensed. Compound **[120]** has $\lambda_{\max} = 325 \text{ nm}$ and compound **[122]** has $\lambda_{\max} = 360 \text{ nm}$.

A structural homolog of MppR from *Streptomyces bingchenggensis*, Sbi00515, is a *bona fide* aldolase, catalyzing the retro-aldol cleavage of benzylidene pyruvate **[123]** (Scheme-4) into benzaldehyde **[124]** and pyruvate^{cxxxix}, and aldol condensations of pyruvate with a wide variety of aldehyde substrates^{cxl}.



Scheme 2.3.8.3-2: Sbi00515 acts as an aldolase. Retro-aldol cleavage of benzylidene pyruvate **[123]** by Sbi00515 produces benzaldehyde **[124]** and pyruvate **[121]**.

To determine whether PbrDHPS might have similar aldolase activity, 10 μ M PbrDHPS was mixed with 3-(2-furyl) acrolein [120] and pyruvate [121] (Scheme 2.3.8.3-1) and incubated for 1 hr at 25 $^{\circ}$ C, after which the reaction was passed through a 10 kDa MWCO concentrator. The retentate was a yellowish orange color, indicating that 3-(2-furyl)acrolein was bound to the enzyme, most likely as the condensation product with pyruvate (Figure 2.3.8.3-1A). PbrDHPS, thus possesses aldolase activity, and Lys160 is likely the nucleophile for the reaction.

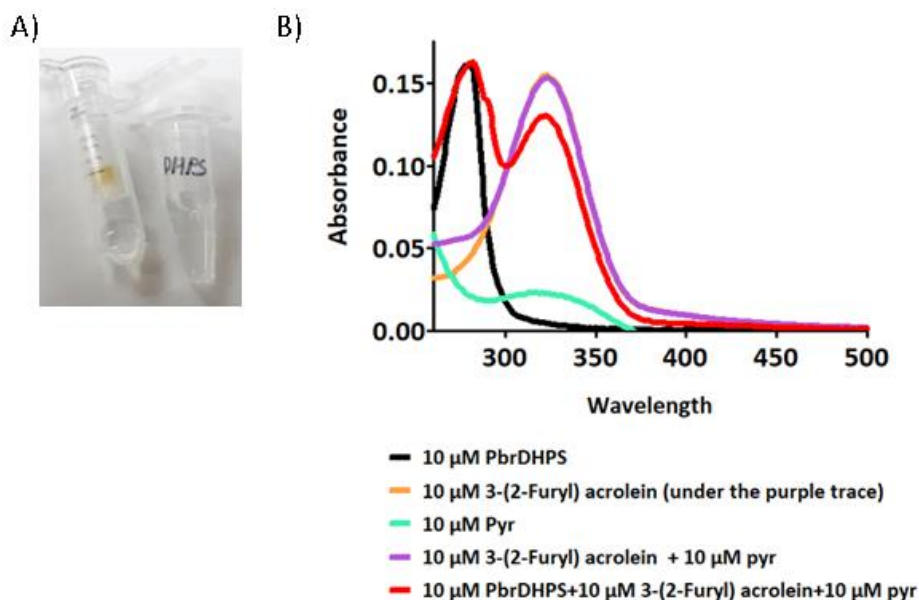


Figure 2.3.8.3-1: A) 3-(2-Furyl) acrolein and pyruvate condensed product stays in the membrane by binding covalently with PbrDHPS. The orange color section in the mini concentrator indicates the covalently bound fraction. B) UV-vis spectra of PbrDHPS, possible substrate analog and products.

The retentate from the concentrator was analyzed by UV-Vis spectrophotometry (Figure 2.3.8.3-1B, red trace). There is no peak at 360 nm corresponding to pure **[122]**, but there is a double-peak at 380-420nm that may be due to the Schiff base of **[122]** with the enzyme, which would have more extended conjugation and would conceivably absorb at a longer wavelength. It is worth noting that there is no peak at 325 nm for pure 3-(2-Furyl) acrolein.

2.3.8.4 Lys160 forms a Schiff base with a ketoacid

The abortive product **[91]** (Scheme 1.4.5-2) was tested for binding to PbrDHPS with the idea that the true substrate for the enzyme is likely to be some modified form of arginine. PbrDHPS crystals were soaked with 12mM **[91]** for 2 hours, and the resulting X-ray crystal structure reveals **[91]** covalently bound to the catalytic Lys160 residue in the active site (Figure 2.3.8.4-1A). The guanidinium group of **[91]** makes a salt bridge with the carboxylate group of Glu187. The superimposed structures of PbrDHPS and EcdHPS (Figure 2.3.8.4-1B) with ligands bound show that most of the active site residues align similarly, as do the ligands.

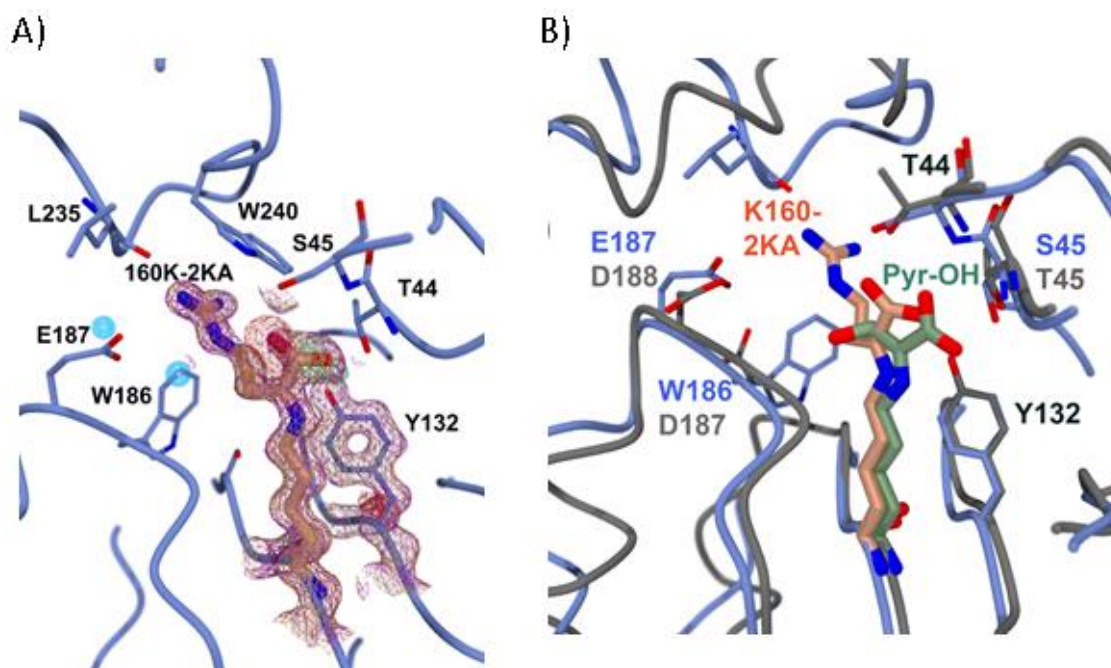
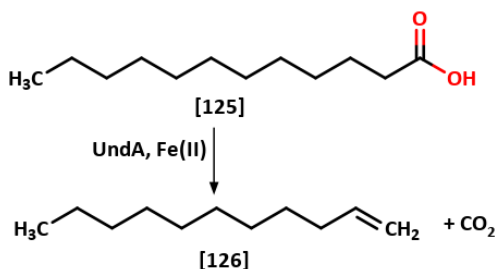


Figure 2.3.8.4-1: Crystallographic structure of PbrDHPS with [91] bound. (A) [91] is covalently bound to Lys160, shown in the electron density map of difference flourier with a contour level of +1.0. (B) Superimposed structure of PbrDHPS (blue chain) and EcdHPS (gray chain) where PbrDHPS is bound with 2KA and EcdHPS is bound with its substrate analog hydroxy pyruvate. Both ligands have similar orientations in the overlaid structures.

2.3.9 Structure of PbrOX, a putative nonheme Fe(II)-dependent oxygenase

The next gene in the PbrMPCO encodes a putative nonheme iron-dependent oxygenase we have labeled PbrOX. We determined the structure of PbrOX at 2.1 Å resolution. A PDBefold^{cxli} search shows highest structural identity of PbrOX with the nonheme iron oxygenase UndA (RMSD 4.15 Å) from *Pseudomonas protegens* Pf-5. UndA (PDB ID: 4WX0^{cxlii}) is involved in biosynthesis of the

medium chain 1-alkene 1-undecene [126] from lauric acid [125] (Scheme 2.3.9-1)^{cxliii}. While it is not possible at this point to test potential substrates of PbrOX, we did test ways to produce iron-loaded holoenzyme, since no metal was observed in the likely iron-binding site in the PbrOX active site. During the course of experiment, we identified using spectrometry that the Fe was loaded onto PbrOx by incubating the enzyme with small amounts of FeSO₄ (~1mg) and ascorbate (~1mg) at 4°C for 4 hours . After the incubation period, the enzyme mixture was centrifuged at 4000 rpm for 20 minutes to remove any insoluble material. The supernatant was used to check the spectrum of Fe(II)-bound PbrOX.



Scheme 2.3.9-1: Reaction catalyzed by UndA

2.3.9.1 Introduction to mononuclear non-heme iron oxygenase

Mononuclear nonheme iron oxygenases are versatile enzymes that contain a reduced iron center to catalyze transfer of one or both oxygen atoms from molecular oxygen to organic substrates. Oxygenases incorporating one oxygen atom into the product (and the other into water) are known as monooxygenases^{cxliv}. Enzymes that incorporate both atoms of dioxygen into their products are called dioxygenases^{cxlv,cxlvi}. One of the features of nonheme iron oxygenases that

distinguishes them from heme oxygenases is that, they have catalytic iron atom coordinated with amino acids from active site residues rather than the organic cofactor heme. In most cases the so-called “facial triad” of residues that coordinate the iron atom in the non-heme enzymes is comprised of two histidine residues and one aspartate or glutamate^{cxlvii}. The metal, thus coordinated, is not only able to act as an electron sink, but also to facilitate the binding of multiple substrates. In many cases, the electrons required for reduction of the iron come from an α -ketoacid co-substrate, such as α -ketoglutarate. These are α -ketoacid-dependent mononuclear iron oxygenases^{cxlviii}. Other subfamilies include aromatic ring-cleaving dioxygenases^{cxlix}, tetrahydrobiopterin-dependent mononuclear dioxygenases^{cl} and Rieske oxygenases^{cli, clii} which have an Fe-S cluster domain.

2.3.9.2 PbrOX is a mononuclear iron oxygenase homolog

PbrOX crystallized in space group $P4_2$ with a single chain in the asymmetric unit, though a symmetry-related chain forms an extensive interface where each chain buries $\sim 1300 \text{ \AA}^2$. Thus, PbrOX is likely a dimer in solution (Figure 2.3.9.3-1A). The active site of PbrOX is particularly rich in tyrosine residues and contains a Glu-His-His triad (E124, H134 and H218) that may constitute a metal-binding site (Figure 2.3.9.3-1B). Alignment of the PbrOX and UndA structures (Figure 2.3.9.3-1C) suggests that the mononuclear iron-binding site of UndA is likely retained in PbrOX. The much more polar binding pocket of PbrOX suggests that its substrate is more polar than that of UndA.

The active site structure of PbrOX shows that the His134 adopts two alternate conformations (Figure 2.3.9.3-1B). Presumably, the orientation would be stabilized by metal coordination in the

holoenzyme. The Glu98 forms a hydrogen bonding interaction with Tyr214. Tyr138, Tyr215 and Tyr200 are also interacting through two water molecules (blue spheres in Figure 2.3.9.3-1B).

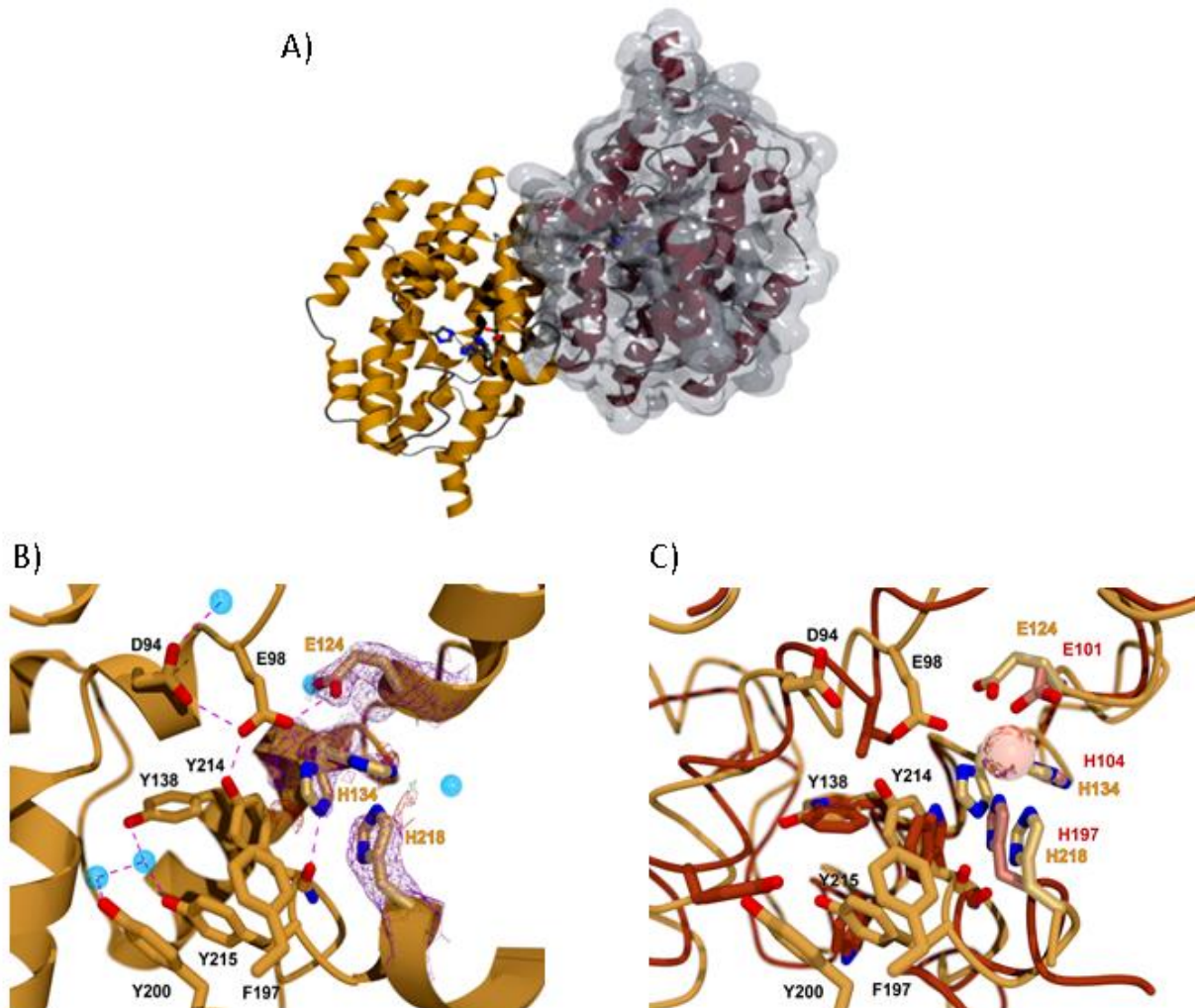
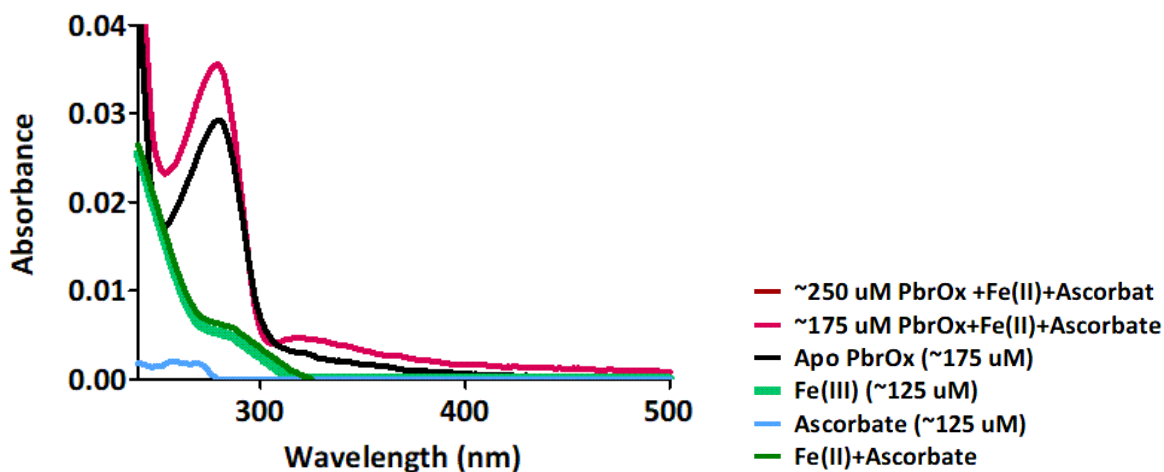


Figure 2.3.9.2-1 A) tertiary structure of PbrOX homodimer B) active site of PbrOX retains metal binding facial triad C) PbrOX (yellow) active site overlaid with UndA (red) with an RMSD value of 4.15Å over 6% sequence identity.

2.3.9.3 Fe^{II} Loading of PbrOX

The holoenzyme (iron(II)-bound form) for PbrOX was reconstituted by adding a small amount of iron sulfate crystals crushed together with dry ascorbic acid^{xxi}. After incubation with Fe^{II} and ascorbate, the UV-Vis spectrum of PbrOX shows a broad feature centered at 325nm (Figure 2.3.9.4-1) that seems to be associated with the Fe(II)-bound (holo) enzyme, since free Fe(II) shows a peak at 272 nm. Apo PbrOX, Fe(III), and ascorbate have no feature at 325nm. We propose that ascorbate acts here as a potential one-electron reducing agent for Fe^{III}, which can then bind to the enzyme. Upon concentrating the dilute protein solution (~175 μM before concentrating up to 250 μM) using a 10kDa MWCO concentrator, the 325 nm peak becomes more intense, supporting the suggestion that this feature is associated with enzyme-bound Fe(II).

A)



B)

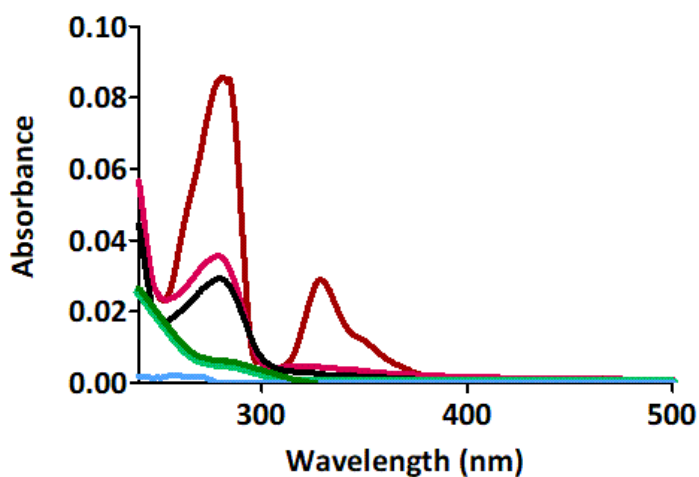


Figure 2.3.9.3-1: A) Spectra of Apo PbrOX and Fe^{II} installed PbrOX. Both Fe^{II} (FeSO₄) alone and ascorbate mixed Fe^{II} (FeSO₄) solution has a small peak near 272nm, whereas apo PbrOX doesn't have any peak other than 280 nm wavelength. B) The concentrated sample of Fe^{II} loaded PbrOX is overlaid with A. The red spectra (~250 μM PbrOX+ Fe(II)+ascorbate) is the sample concentrated after incubation with Fe and ascorbate (pink spectra: ~175 μM PbrOX with Fe(II) and ascorbate)

2.3.9.4 Introduction to bacterial three component system

Bacterial three-component oxygenase systems are most commonly associated with the biodegradation of environmentally hazardous compounds^{cliv}. This multienzyme systems consist of three soluble proteins (Figure 2.1-4) that interact to form an electron-transport chain to transfer electrons from NAD(P)H via flavin and [2Fe-2S] redox centers to a terminal dioxygenase^{clv}. The first component is a ferredoxin-NAD(P)H reductase (FNR) that oxidizes NAD(P)H and transfers two electrons to flavin adenine dinucleotide (FAD). FNR transfers these electrons, one electron at a time, from the flavin to the second component, a [2Fe-2S] cluster-containing protein called ferredoxin. Finally, the electrons are transported by the ferredoxin to the iron center of a terminal oxygenase, where they are used to reduce molecular oxygen to oxidize a substrate. In almost all cases, the proteins involved in bacterial three-component oxygenase systems are co-expressed^{clvi, clvii, clviii}.

Interestingly, the co-existence of PbrOX with PbrFd suggests a possible three component electron transport system in this gene cluster^{clix}. Although there is no presence of FNR in the pbr gene context, we used one of the only two annotated FNR enzyme from the *P. brassicacearum* genome to test this hypothesis.

2.3.9.5 PbrFD is a Fe-S cluster protein

The PbrFD is a [2Fe-2S] cluster protein homolog with 90% sequence identity to the ferredoxin from *Pseudomonads*^{clx}. The iron-sulfur cluster is not installed properly in protein from cells grown

on plain LB medium (Figure 2.3.9.5-1A, green spectrum). When BL21* cells containing the pE-SUMO_{kan}-PbrFD expression vector were grown in Terrific Broth (TB) rich medium supplemented with 1 mM L-cysteine 0.1 mg/ml ferric ammonium citrate, the protein produced had the [2Fe-2S] cluster properly installed (Figure 2.3.9.5-1A, blue spectrum). The HisTrap eluate during PbrFD purification had a brown color characteristic of [2Fe-2S] clusters (Figure 2.3.9.5-1B, fractions 28 and 29). The UV-Vis spectrum (Figure 2.3.9.5-1A, blue spectrum) shows a characteristic Fe-S charge transfer band^{clxi} around 320nm when reduced and 450 nm when it is oxidized. The blue spectrum in Figure 2.3.9.5-1A is thus a mixture of reduced and oxidized PbrFD.

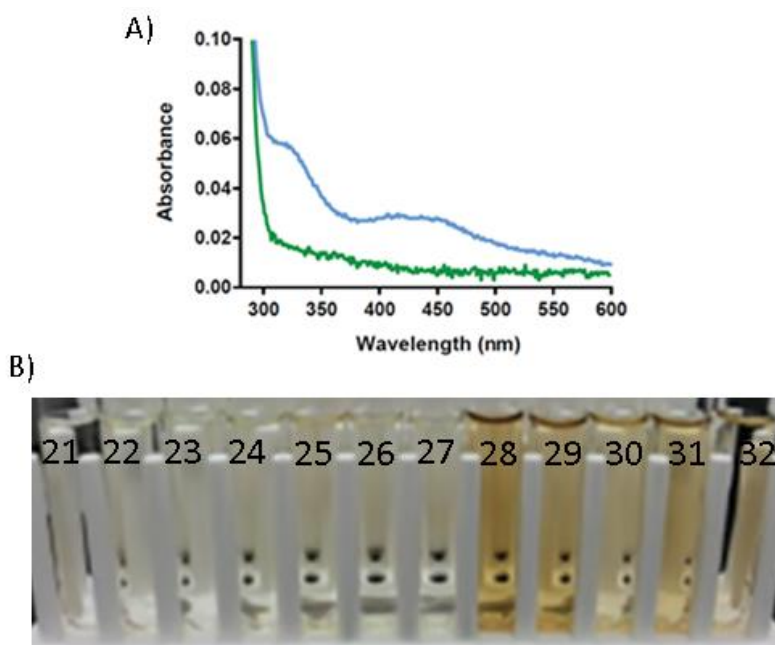


Figure 2.3.9.5-1: A) Spectra of PbrFd expressed in terrific broth supplemented with cysteine and ferric ammonium citrate (blue spectrum). The green spectrum is for PbrFd when expressed in LB media only, thus Fe-S cluster not installed. B) The elution fractions from HisTrap purification of PbrFd shows a brown color.

2.3.9.6 PbrFNR is active in the absence of any specific substrate

When examining FNR in activity tests, we found that the enzyme oxidizes NADPH in the absence of any other electron acceptor (data not shown). This observation was independent of the buffer, with the same phenomenon being observed in 50 mM Tris pH 8.5, 50 mM BisTris propane pH 9.0 and 50mM sodium phosphate pH 8.5. To test whether the electrons from NADPH are reducing dioxygen in this reaction, we performed a similar experiment in the oxygen electrode (Figure 2.3.9.6-1). We observed that addition of 1 μM PbrFNR to a solution with 20 μM NADPH in 50 mM Tris pH 8.5 resulted in consumption of ~ 8 μM dioxygen. Once the rate of dioxygen consumption began to change, the addition of 1 $\mu\text{g}/\text{ml}$ catalase (final concentration in 500 μL) resulted in an increase in the dioxygen concentration of ~ 4 μM . This observation suggests that PbrFNR is reducing dioxygen to produce H_2O_2 . Heat-denatured PbrFNR was used as a control (Figure 2.3.9.6-1, grey curve).

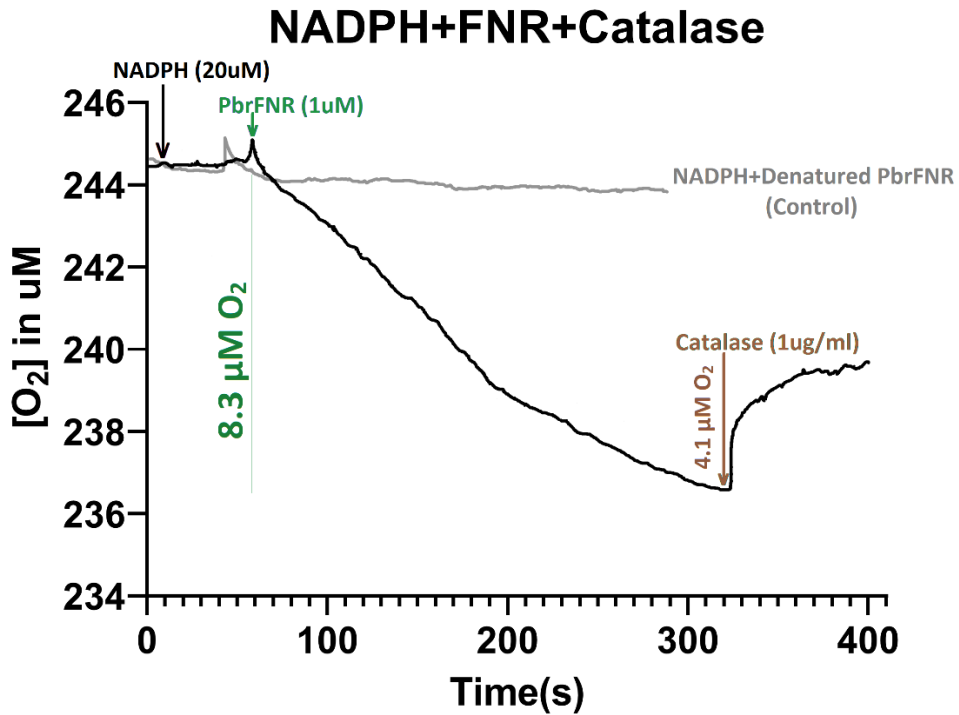


Figure 2.3.9.6-1: PbrFNR reduces dioxygen to produce H₂O₂ in the absence of any other electron acceptor. Denatured PbrFNR does not consume dioxygen.

2.3.9.7 Electrons flow from NADPH towards PbrOX

The oxygen probe experiment described in the previous section was continued with additions of PbrFD and PbrOX (Figure 2.3.9.7-1). When O₂ consumption by FNR slowed down, addition of approximately 1 μM PbrFD resulted in more O₂ consumption. Adding PbrOX in the absence of PbrFD does not alter the rate of dioxygen reduction by PbrFNR (Figure 2.3.9.7-1, grey curve), but upon addition of PbrFD, the rate increases significantly, showing that PbrFD is likely an

intermediate transporter of electron. The observation leads to the conclusion that electrons flow from NADPH via PbrFD to the metal center of PbrOX.

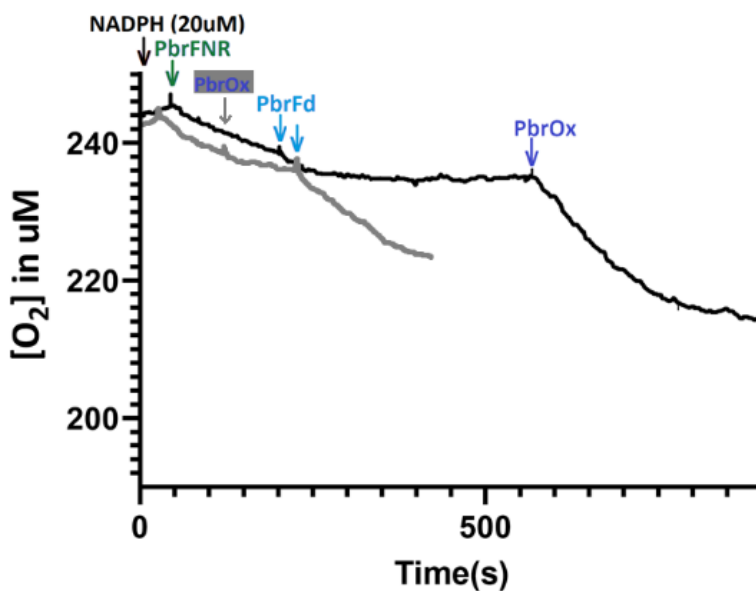


Figure 2.3.9.7-1: Electron flows from NADPH towards PbrOX. The black curve indicates rate of dioxygen reduction in the NADPH solution (in 50mM Tris pH 8.5) due to addition of PbrFNR, followed by PbrFD and PbrOX. The gray curve is the control where rate of dioxygen reduction in the NADPH solution (in 50mM Tris pH 8.5) is shown due to addition of PbrFNR, followed by heat denatured PbrOX. No significant dioxygen reduction was observed upon addition of this by heat denatured PbrOX (control). But, addition of PbrFD resumed dioxygen reduction.

2.3.9.8 Holo PbrOX is required for electron flow

The oxygen probe experiment was continued to determine if the holo PbrOX prepared as described earlier has Fe^{II} bound in a catalytically competent manner, or if there is simply free Fe^{II} ions in solution. The apo PbrOX and heat-denatured PbrOX were used as negative controls (Figure 2.3.9.8-1, grey curve). Neither of these inactive forms of PbrOX altered the rate of uncoupled dioxygen reduction by PbrFNR. Adding 500 nM FeCl₂ and ascorbate resulted in a dramatic consumption of dioxygen.

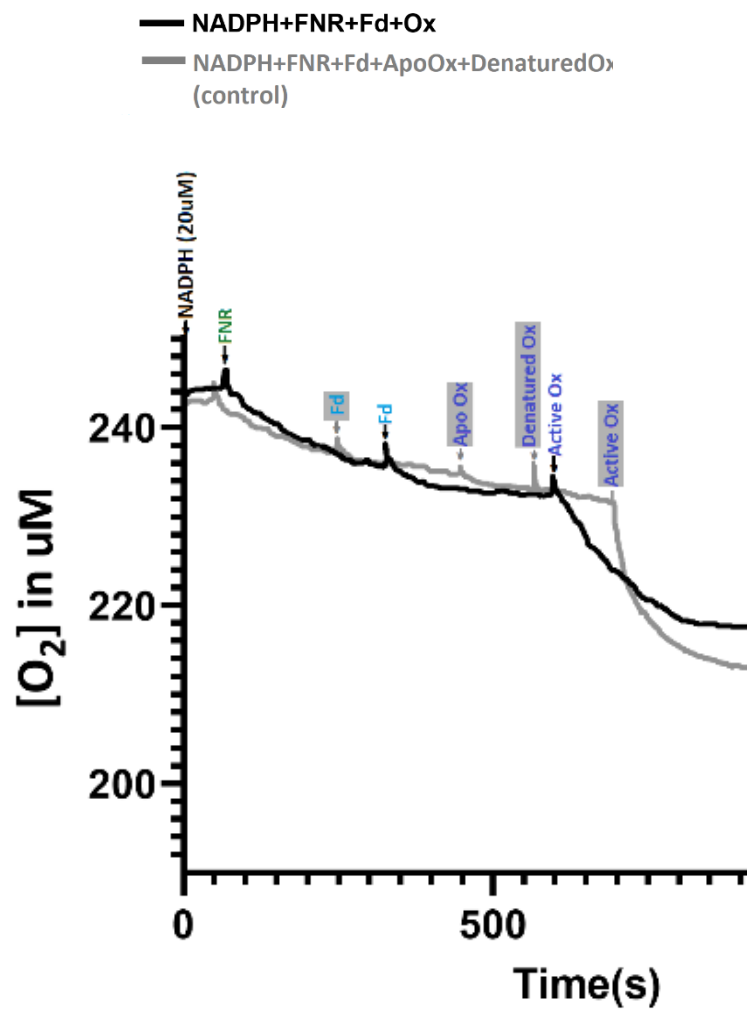


Figure 2.3.9.8-1: Fe(II) loading on PbrOX is required for electron flow

2.3.10.1 PbrMppP operon genes are expressed only in solid culture

Since the structures described above provide few clues about possible functions of the enzymes in PbrMPCO, we decided to use a metabolomics approach to determine the final product of the pathway. This information would dramatically reduce the possibilities for reactions and substrates involved. With this idea, the putative promoter region of the PbrMPCO (Figure 2.2.13-1) was cloned into a promoterless reporter vector, pPROBE-NT, upstream of the green fluorescent protein (GFP) reporter gene (Figure 2.2.13-1). Thus, if the promoter is activated, GFP will be expressed and the cells will glow when illuminated by UV light. The conditions listed in Table 2.2.13-1 were tested. The PbrMPCO promoter is activated only when *P. brassicacearum* (DF41) cells are grown on solid media (Figure 2.3.10.1-1A,B). To check if the cells are glowing due to expression of GFP, *P. brassicacearum* (DF41) cells were scooped from the agar plates and resuspended in 10 mL of solvent (see Section 2.2.13). The suspension was sonicated to lyse the cells and the suspensions were centrifuged to remove cell debris. The collected supernatants were scanned in a UV-Vis spectrophotometer to confirm the presence of GFP (Figure 2.3.10.1-1C,D). From these observations, it seems that GFP/PbrMPCO expression is activated when cells reach a sufficient density (quorum sensing). *P. brassicacearum* (DF41) cells can form biofilm, and this regulates the expression of a variety of genes. The presence of an acylhomoserine lactone (AHL) synthase on the opposite strand of the putative promoter region (Figure 2.2.12-1) supports the involvement of the PbrMPCO proteins in activities related to biofilm formation or maintenance, since AHL is an important quorum sensing messenger^{clxii}. It is possible that we are not observing quorum sensing behavior, but rather changes in gene expression due to oxygen

depletion. Additional tests will need to be performed in defined atmospheres to rule this possibility out.

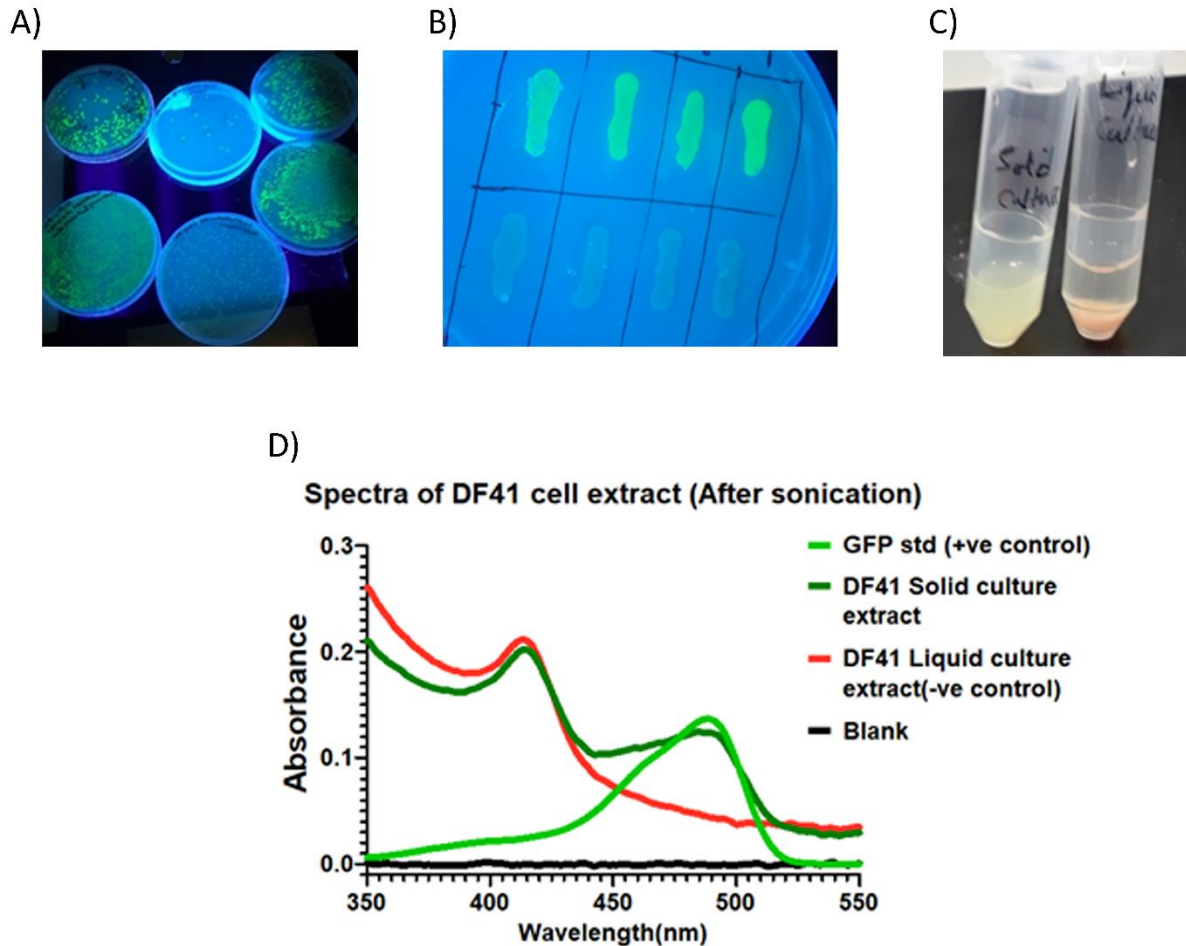


Figure 2.3.10.1-1: (A) DF41 cells electroporated with *PbrMppP* promoter: *pPROBE-NT* plasmid construct. (B) *PbrMppP* operon in DF41 cells are expressed in solid culture, thus transformed cells are glowing under UV (top row of cell smear on B). The bottom row of cell smear is the negative control. i.e empty vector transformed DF41 cells. (C) Sonicated cell extract DF41 cells grown in LB-agar plates (greenish cell extract on the left). Cells from liquid LB media, was used as negative

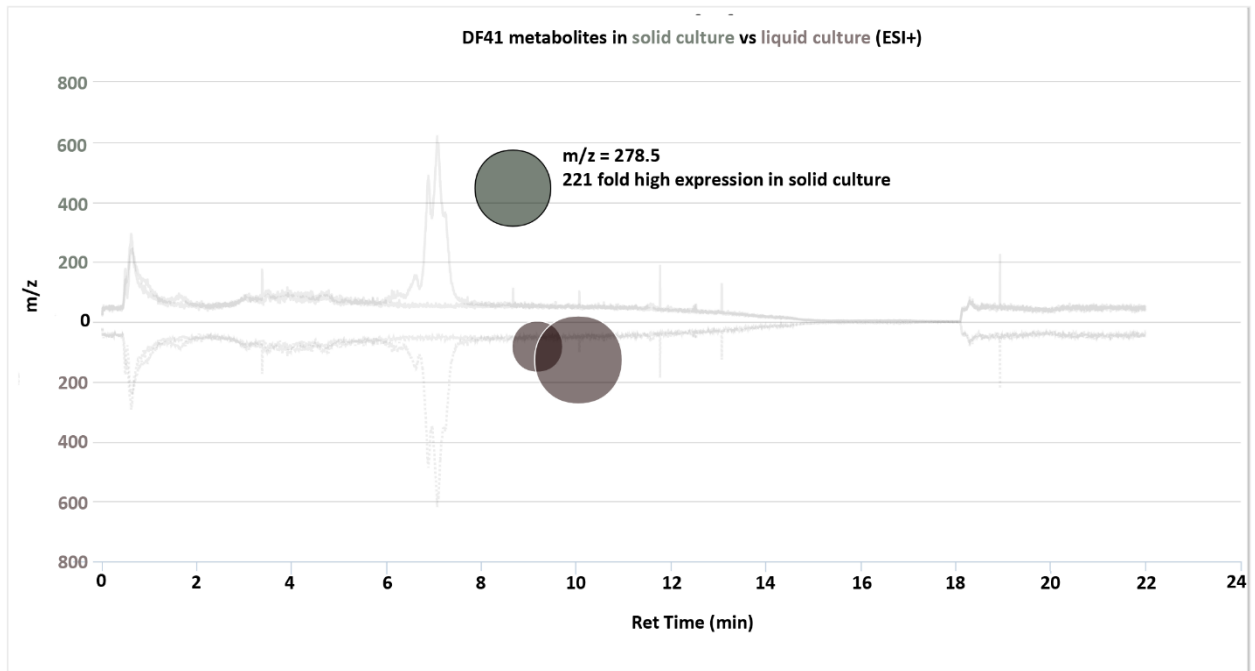
control as PbrMppP operon knocked out strain was not available and also, liquid culture doesn't turn-on PbrMppP operon (pinkish cell extract on the right). (D) Supernatant of the cell extract check via spectroscopy.

The DF41 cells electroporated to transform with PbrMppP promoter: pPROBE-NT plasmid construct. PbrMppP operon in DF41 cells are expressed in solid culture, thus the transformed cells are glowing under UV (Figure 2.3.10.1-1 B). The bottom row of cell smear (Figure 2.3.10.1-1 B) is the negative control i.e DF41 cells transformed with empty vector, thus not glowing under UV. The transformed cells were sonicated and cell extract (greenish cell extract on the left of Figure 2.3.10.1-1C) was used to check the expression of GFP. The UV-vis spectra of DF41 cell extract was compared with standard GFP (Figure 2.3.10.1-D). Both the standard GFP and electroporated DF41 cell extract have the absorption maxima of approximately 479 nm^{clxiii,clxiv}. The same construct in DF41 cells grown on liquid LB media, was used as negative control as PbrMppP operon knocked out strain was not available yet. Moreover, liquid culture of transformed DF41 cells doesn't turn-on PbrMppP operon as well as GFP expression (pinkish cell extract on the right of Figure 2.3.10.1-1C). Thus, Liquid culture is persuasive negative control in absence of knocked out strain.

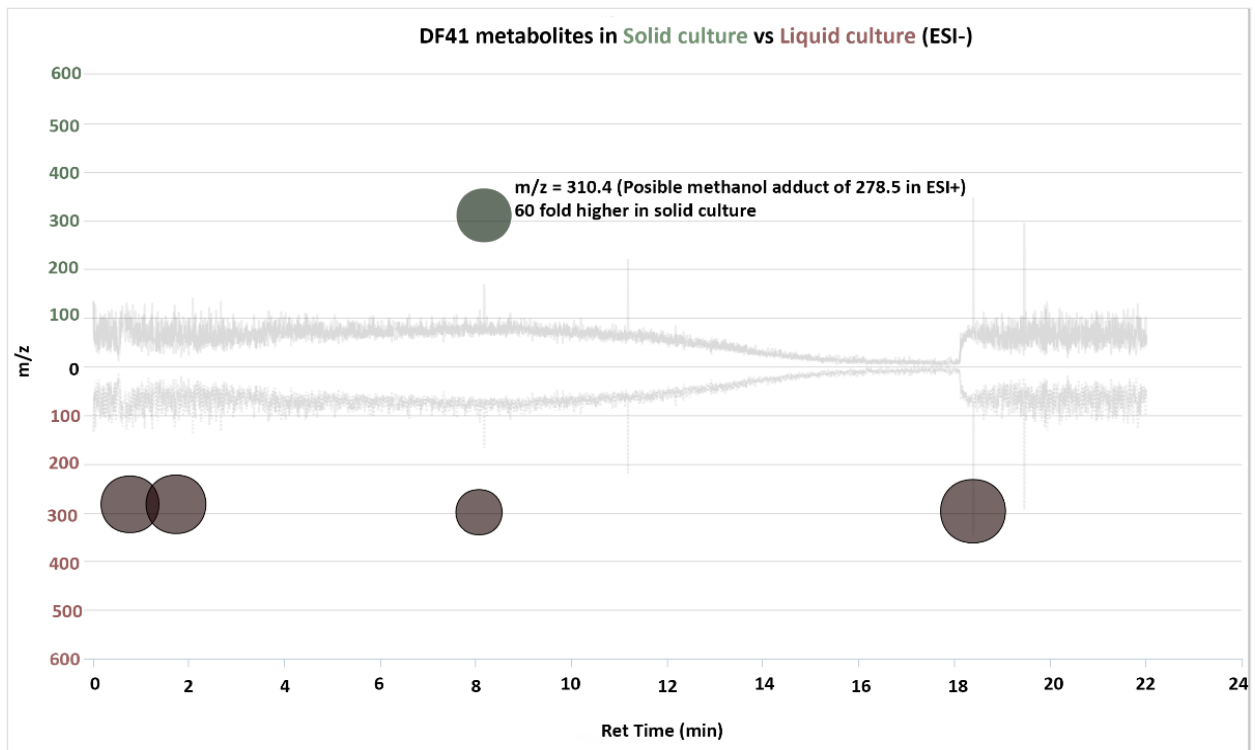
2.3.10.2 Preliminary “comparative metabolomics” of DF41 cells by LC-MS

With the hypothesis that very high cell densities activate the PbrMPCO, our next goal was to knock out the entire PbrMPCO from the *P. brassicacearum* (DF41) genome using single-strand overhang extension PCR (SOE PCR)^{clxv, clxvi}. Once we have the knockout strain, the metabolites of the knockout strain can be compared with the wild type DF41 metabolites using LC-MS/MS^{clxvii, clxviii}. Therefore, we can proceed towards finding the final product of the pathway. So far, we have the PCR products prepared for knockout process. The gene knockout and following analysis of the knockout strain will be done by the successor to this project. In the absence of a validated knock out strain, we took advantage of the fact that the promoter is only turned on under solid phase growth. Cells collected from solid cultures (PbrMPCO active) were compared to *P. brassicacearum* (DF41) cells from liquid cultures (PbrMPCO inactive). While these results are not definitive, some changes in the metabolite ions were observed between the two culture conditions. The LC-MS data from solid culture and liquid culture were analyzed with XCMS^{clxix} and aligned pairwise to prepare a metabolite could plot map (Figure 2.3.10.2-1). The summary of the preliminary comparative metabolomics results is presented in Table 2.3.10.2-1.

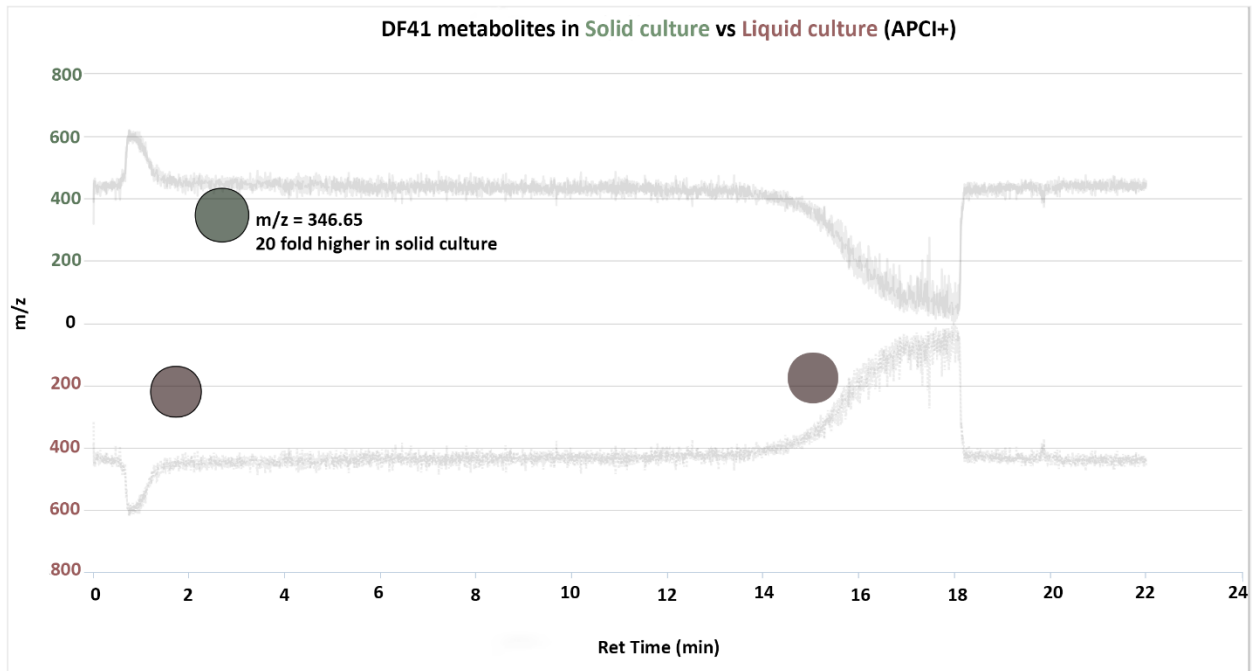
A)



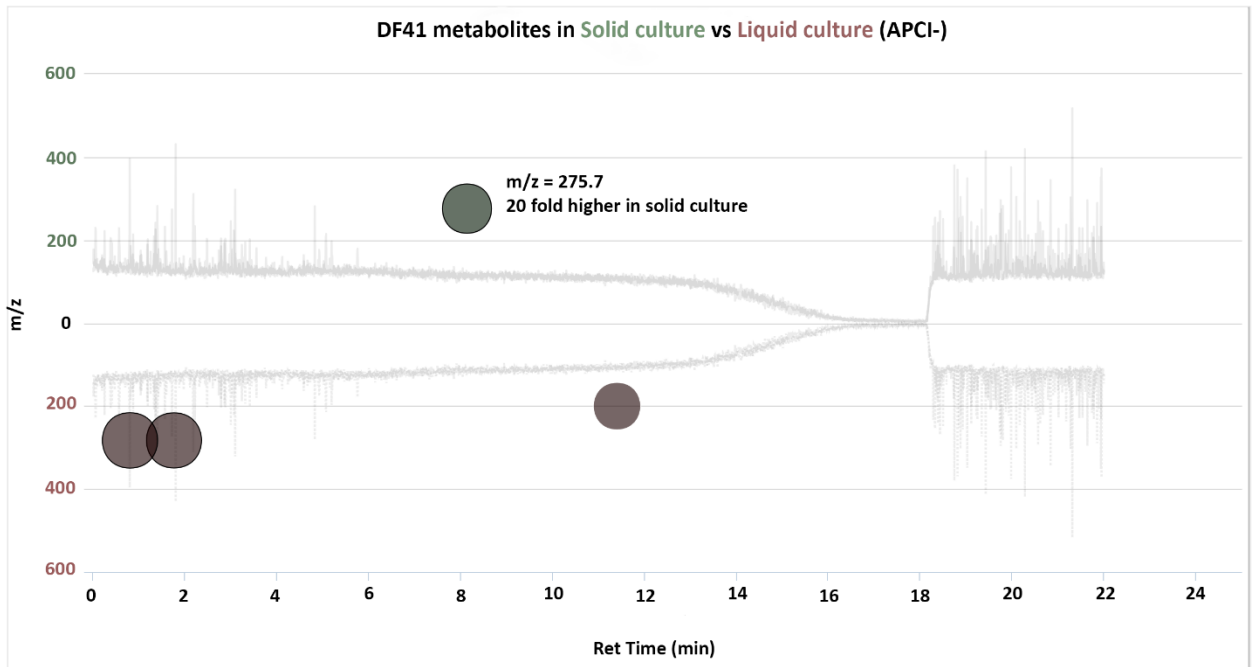
B)



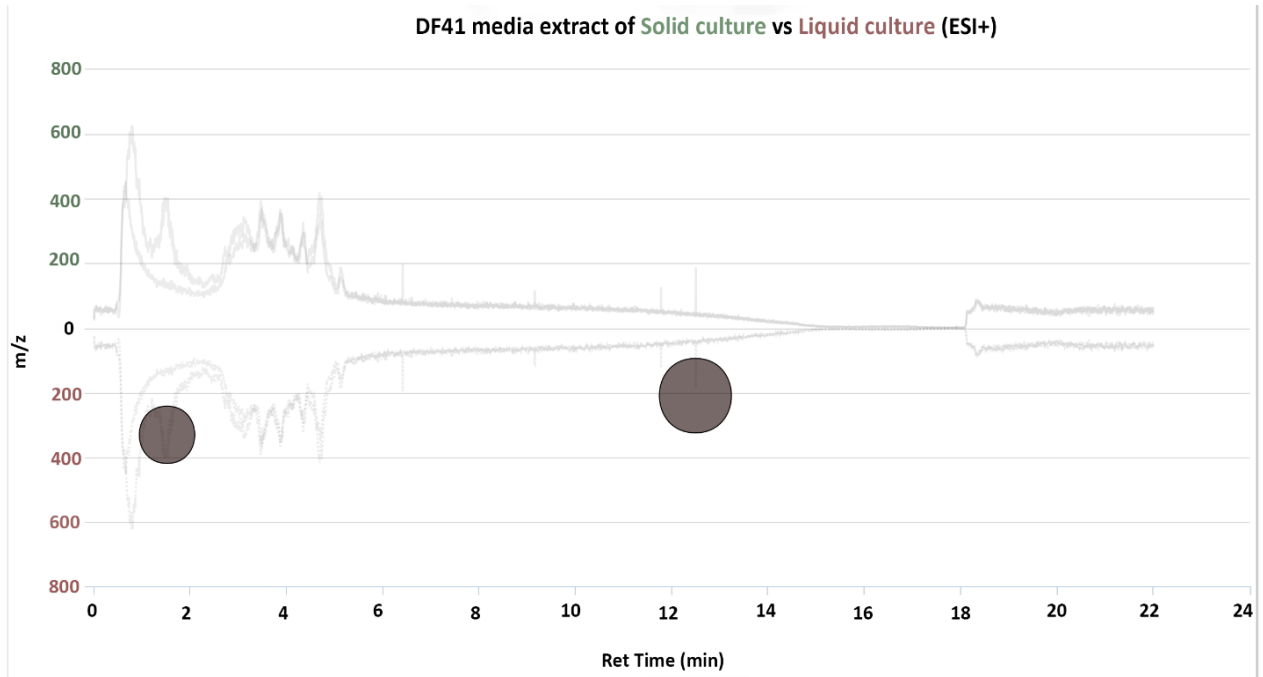
C)



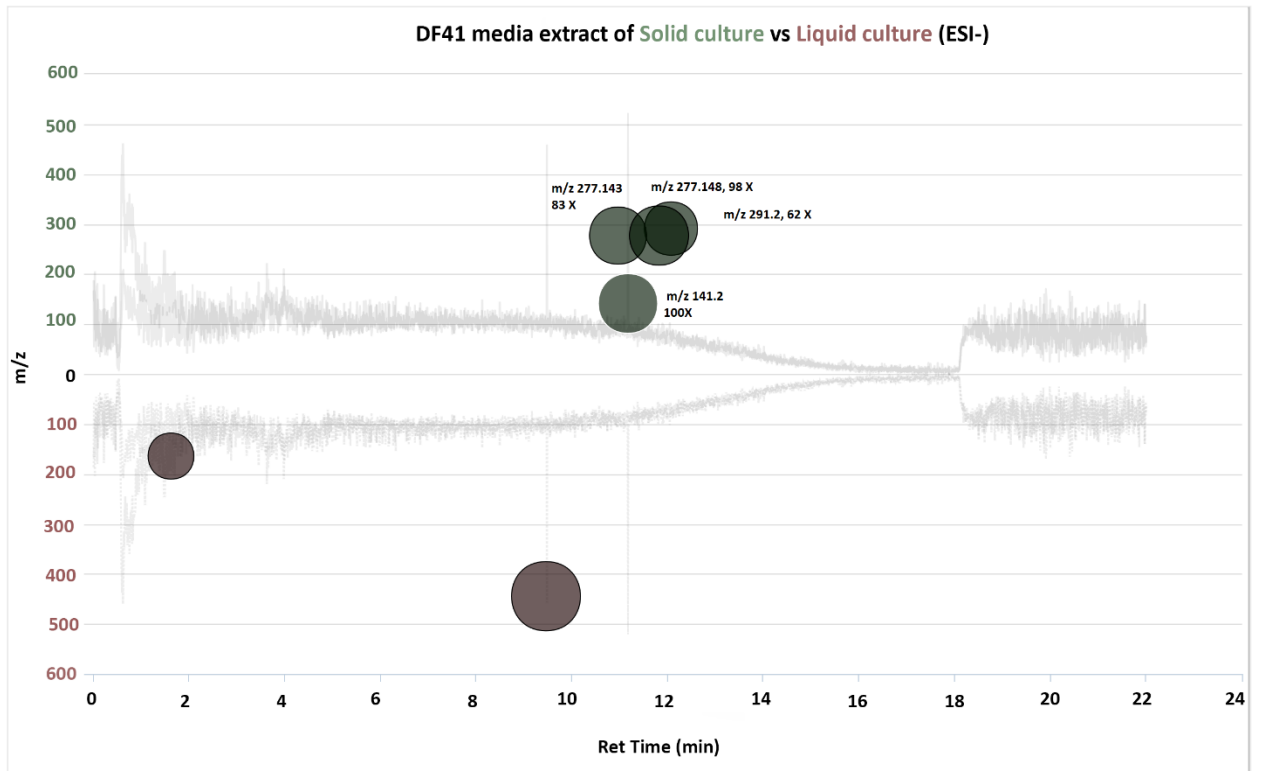
D)



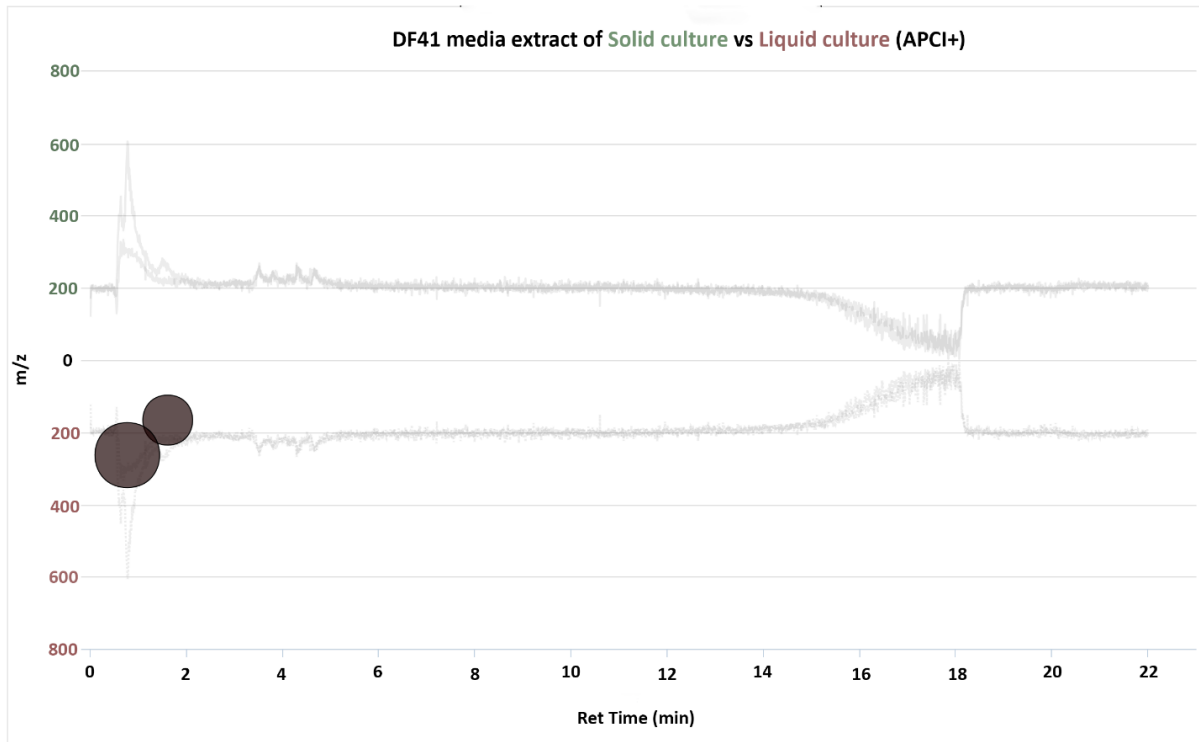
E)



F)



G)



H)

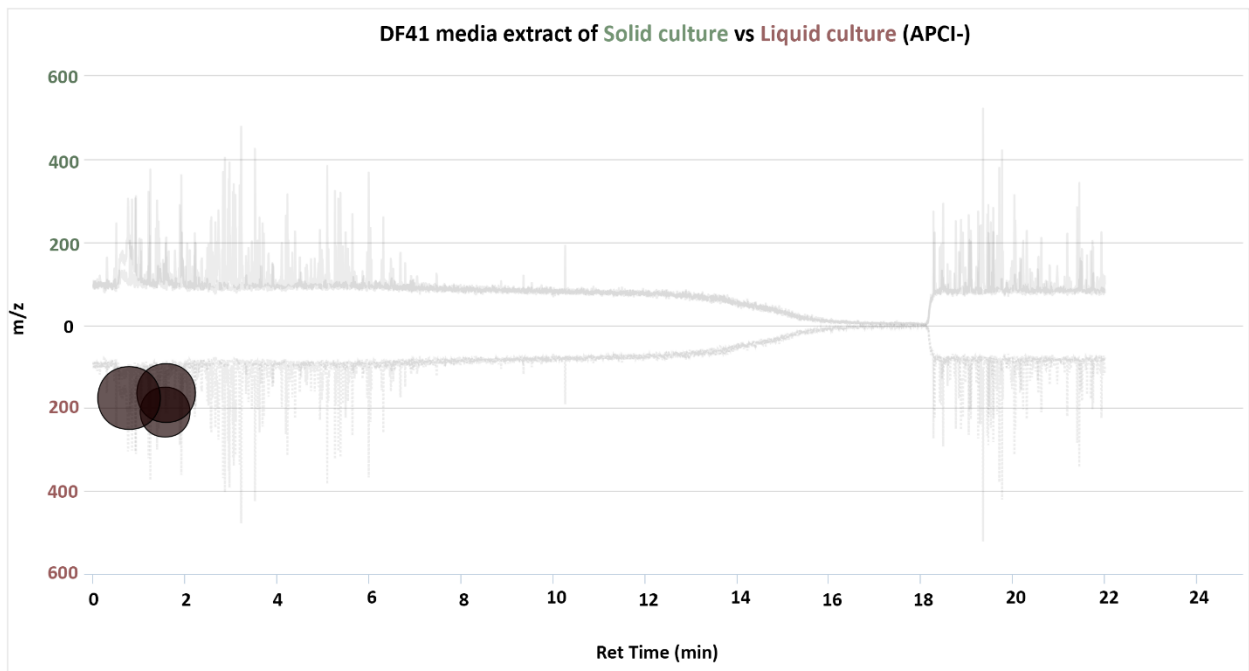


Figure 2.3.12-A: Metabolic cloud plot generated from DF41 cells expressed in solid culture (green traces) vs liquid culture (red traces). X-axis represents the retention time in minutes, Y-axis represents the m/z. The green labeled m/z on the upper half represents m/z associated in solid culture metabolites, whereas the red labeled m/z on the lower half represents m/z associated in liquid culture metabolites. All the data screened here was with minimum absolute intensity of 25000, at least 20-fold higher expressed in any solid culture than liquid culture. Finally, the bubbles in cloud plot correspond to their relative intensity. Metabolites of DF41 cells scooped from solid culture were analyzed using (A) ESI in positive mode, (B) ESI in negative mode, (C) APCI in positive mode, (D) APCI in negative mode, media extract analyzed using (E) ESI in positive mode, (F) ESI in negative mode, (G)) APCI in positive mode, (H) APCI in negative mode.

Table 2.3.12-1: list of m/z expressed highly in DF41 solid culture

Ion source	m/z	Ionization method	Relative intensity in solid media vs liquid media
Solid culture	278.5	ESI (+)	221 X
Solid culture	310.4	ESI (-)	60 X
Solid culture	346.65	APCI (+)	20 X
Solid culture	275.7	APCI (-)	20 X
Media extract of DF41 solid culture	Nothing found	ESI (+)	N/A
Media extract of DF41 solid culture	277.143	ESI (-)	83 X
	277.148	ESI (-)	98 X
	291.2	ESI (-)	62 X
	141.2	ESI (-)	100 X
Media extract of DF41 solid culture	Nothing found	APCI (+)	N/A
Media extract of DF41 solid culture	Nothing found	APCI (-)	N/A

We did a bioinformatics analysis of the putative promoter region using TFsitiescan^{clxx}. The promoter analysis shows that the PbrMppP promoter (Figure 2.2.13-1) might interact with an RpoS type σ -factor^{clxxi}. Such transcription regulators, in combination with histone-like nucleoid protein, is expressed under stress conditions. The accordance of an acylhomoserine lactone synthase (AHLS) at the opposite strand, is suggesting the possible involvement of such σ -factor in response to cell density^{clxxii}.

2.4 Discussion

The MppP homolog of *Pseudomonas brassicacaerum* is quite like the first characterized MppP enzyme, SwMppP. The surprising feature of the PbrMPCO is the incorporation of the L-Arg oxidase activity of MppP into a new biochemical context completely distinct from L-enduracididine biosynthesis.

PbrMppP is basically identical to SwMppP, but the biochemical context of the L-Arg oxidase activity is different. The structures of the various proteins showed us proteins with similar structures, but revealed no tangible clues to the exact functions of any of the other proteins, or even about their substrate specificities. Finally, the PbrMppP operon is activated under conditions of high-density cell growth. From the XCMS analysis of DF41 cell metabolites, we were able to narrow down some precursor ions which might be of interest for the next step. Thus structure determination of the product in PbrMCPO pathway can be determined. We are looking forward to the fragmentation of some of the ions that are most enhanced in the solid culture samples in an effort to determine the structures of those metabolites.

Chapter 3

Structure and Function of the Enzymes Expressed by the MppP-Containing Operon in *Streptomyces griseofuscus*

3.1 Background

Since the activities of the PbrMPCO enzymes proved too cryptic for rapid identification, we looked for an alternative MppP homolog having a unique genomic context, but one where the associated genes had more definitive function annotations. Just such a gene cluster was found in *Streptomyces griseofuscus*. This MppP containing operon in *Streptomyces griseofuscus* or SgrMPCO has MppP homolog flanked by five other gene products (Figure 3.1-1). All these genes appear to be co-transcribed as they have 4 to 5 bp overlapping gene sequences. The gene products are annotated as a putative biotin carboxylase (SgrLIG), flavin-dependent monooxygenase (SgrMO), SAM-dependent methyltransferase (SgrMT), amidinohydrolase (SgrAH), PLP-dependent L-arginine oxidase (SgrMppP), and GNAT-family acetyltransferase (SgrNAT). These proteins are more similar to their closest homolog comparing to that of PbrMppP gene cluster. Thus with the hope that it will be more tractable, to get the final product of this pathway, I worked on the second project.



Figure 3.1-1: putative biosynthetic gene cluster in *Streptomyces griseofuscus*

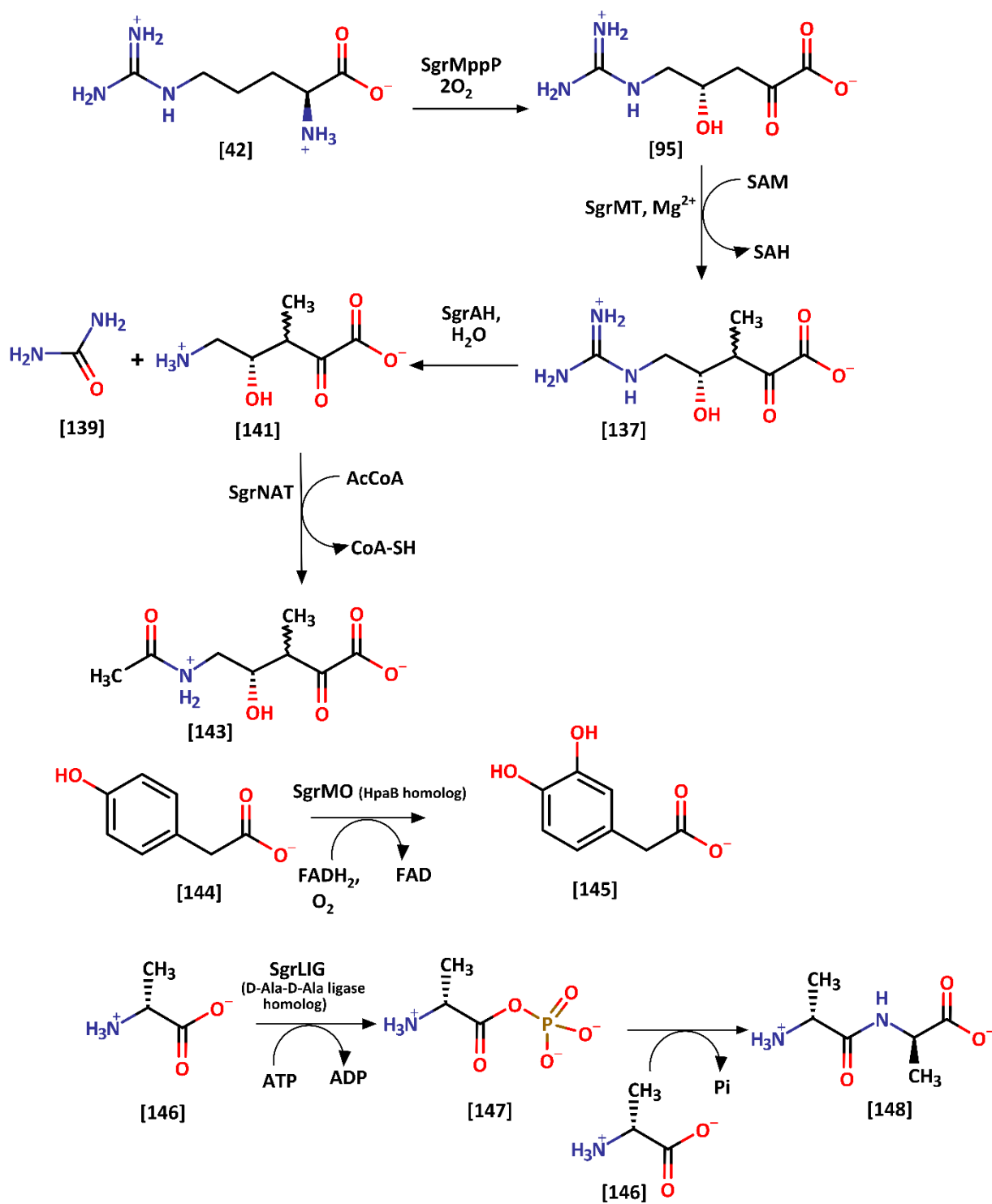
Table 3.1-1: Gene accession codes, sequence annotations, and identity to closest characterized homolog of SgrMPCO proteins

Annotation from Genebank	Functional prediction	Sequence Identity(%)
SgrLIG	D-Ala-D-Ala ligase	54% identical to biotin carboxylase superfamily ^{clxxiii}
SgrMO	Fe(II) and Flavin dependent monooxygenase	33 % identical to the oxygenase component of 4-HPA-3-monooxygenase (HpaB) ^{clxxiv}
SgrMT	SAM and Mg ²⁺ dependent methyltransferase	20% identical to norcoclaurine-6-O-methyltransferase ^{clxxv}
SgrAH	Amidino hydrolase	43% identical to procalvamate amidino hydrolase (PAH) ^{clxxvi}
SgrMppP	L-Arginine oxidase	51% identical to SwMppP ^{clxxvii}
SgrNAT	GNAT family N-acetyltransferase	no homolog in PDB

Similar to the PbrMPCO project, the bioinformatics analysis shows an MppP operon in *Streptomyces griseofuscus* where the MppP homolog behaves in exactly the same enzymatic way as PbrMppP and SwMppP. The Superimposed image of SgrMppP with SwMppP active site shows that it retains most of the active site residues with exact same conformation (Figure 3.3-1)

The sequence annotations of the PbrMPCO enzymes do not give sufficient clues to their activities, and none seem to react with the products of PbrMppP. In contrast, the higher degree of confidence in the functional annotations of the SgrMPCO proteins led us to pursue this gene cluster. We did build a plausible pathway (Scheme 3.1.-1) using the sequence analysis of hypothetical enzymatic activities.

The goal of these studies is to determine the activities of all of the enzymes in this gene cluster and to identify the product of their concerted action. Our preliminary characterization has demonstrated that, in the presence of Mg(II) and SAM, SgrMT catalyzes the transfer of a methyl group from SAM (S-Adenosyl-methionine) to carbon 3 of the fully oxidized product [95] of SgrMppP. Crystal structures of SgrMT in the unliganded state, and with 2-ketoarginine bound, show that it is similar to the Fe(II)- and SAM-dependent methyltransferase MppJ^{clxxviii}. We also determined the structure of the amidinohydrolase homolog, SgrAH, at 2.1 Å resolution. The closest homolog of SgrAH is the procalvaminic acid hydrolase from *Streptomyces clavuligerus* (ScPAH, PDB ID: 1GQ6) with RMSD value of 1.0 Å for all C α atoms and 51% sequence identity. Mass spectrometric analysis of reaction products with the SgrMT product as the substrate, shows that SgrAH produces urea from the product of SgrMT. Although the major product of SgrAH has yet to be determined, it may be 4-hydroxy-3-methyl-2-ketoornithine or a cyclized derivative thereof. Structural characterization of the other proteins in this putative biosynthetic pathway is ongoing.



Scheme 3.1-1: Hypothetical pathway of SgrMPCO

3.2 Methodology

3.2.1 Cloning, expression, and purification of SgrMT Enzymes

The coding sequences of SgrMppP, SgrMT, SgrAH and EcMetK were optimized for expression in *E. coli*, synthesized, and cloned into the pET-TEV_{amp} expression vector by GenScript Inc. (Piscataway, NJ). All proteins were expressed using similar protocols. The His₆-TEV-fusion proteins were expressed from *E. coli* BL21 Star (DE3) cells (Invitrogen Inc., Carlsbad, CA) carrying one of the pET-TEV-X plasmids, where “X” is SgrMppP, SgrMT, SgrAH, or EcMetK. Cultures were grown in Luria-Bertani medium with 100 µg/mL ampicillin at 37 °C. When the cultures reached an OD_{600nm} of ~0.7, protein expression was induced with 0.4 mM IPTG. The temperature was reduced to 25°C, and the cultures were grown overnight while being shaken at 250 rpm. Cells were harvested by centrifugation at 6000 rpm for 20 minutes. The cell pellets were collected and resuspended in 5 mL/g of buffer A (25 mM TRIS pH 8.0, 300 mM NaCl, 10 mM imidazole) supplemented with 0.1 mg/mL DNase I (Worthington Biochemical Corp., Lakewood, NJ). In the case of SgrMppP, buffer A included 20 µM PLP to stabilize the enzyme. The cells were lysed using a Branson Sonifier S-450 cell disruptor (Branson Ultrasonics Corp., Danbury, CT) for a total of 10 min at 60% amplitude with 30 s pulses separated by 30 s rest periods. The temperature was maintained at or below 4 °C by suspending the steel beaker in an ice bath directly over a spinning stir bar. The lysates were clarified by centrifugation at 18,000 rpm for 45 min and then applied to a 5 mL HisTrap column (GE Lifesciences, Piscataway, NJ) at a flow rate of 5 mL/min to isolate the His₆-SUMO fusion proteins. The proteins were eluted by a four-step gradient of buffer B (25

mM TRIS pH 8.0, 300 mM NaCl, and 250 mM imidazole; 5, 15, 50, and 100%). The His₆-TEV-fusion proteins were eluted in the third and fourth steps and were ~90% pure, as judged by Coomassie-stained sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE). Peak fractions were pooled and dialyzed overnight at 4 °C against 4 L of 25 mM TRIS pH 8.0, 150 mM NaCl, 5mM DTT and ~3 μM TEV protease^{clxxxix}. The dialysates were passed through the HisTrap column a second time to remove the cleaved His₆-TEV tag as well as the protease. The resulting protein preparations were >95% pure.

The purified proteins were used in differential scanning fluorimetry (DSF, aka Thermofluor) assays (described in topic 3.2.2) to identify the storage buffer with the greatest increase in melting temperature over the control. Each protein was desalted into its optimal buffer (10 mM) using a HiTrap desalting column (GE Lifesciences) and stored at –80°C. Storage condition for each of the SgrMPCO proteins are listed in table 3.2.1-1. Selenomethionine-substituted SgrMT was produced in the same way, except that T7 Express Crystal cells (New England Biolabs, Ipswich, MA) carrying the pET-TEV-SgrMT plasmid were used for expression. These cells were grown in Selenomethionine Medium Complete (Molecular Dimensions, Newmarket, Suffolk, U.K.) with 100 μg/mL ampicillin.

Table 3.1.1-1: Storage buffer of SgrMPCO proteins and EcMetK

Protein	Storage buffer
SgrMppP	10mM MES pH 6.7, 20 μ M PLP
SgrMT	10mM BisTris pH 7.5 + 20 μ M MgCl ₂
SgrAH	10mM Succinate pH 6.0 + 10 μ M MnCl ₂
EcMEtK	10mM HEPES pH 7.4

3.2.2 Differential scanning fluorimetry

For the DSF assay, 200 μ L of Spyro Orange dye solution (50X) were prepared by adding 2 μ L of 5000x Sypro Orange concentrate to 198 μ L of deionized water. This dye solution was distributed into 8 PCR tubes (25 μ L each). Similarly, dilute protein solutions at 1 mg/mL were prepared and 25 μ L were aliquoted into 8 fresh PCR tubes. Using an 8-channel micropipet, 11 μ L of deionized water were transferred to each well of 96-well Eppendorf TwinTec plates (one for each protein under study). Next, 2 μ L of dilute protein solution were added, followed by 5 μ L of the 96-condition Slice pH screen (Hampton Research, Aliso Viejo, CA). Finally, 2 μ L of the 50X Spyro Orange dye solution were added. Each well was carefully mixed by pipetting up and down twice. The plates were immediately sealed with an optically clear sealing film and centrifuged for 30 s at 1000 rpm to ensure that the solution was at the bottoms of all the wells. The sealed plates were placed in a RealPlex RT-PCR instrument (Eppendorf) and equilibrated to 25°C for 5 minutes, in order for the temperature of the plate and reagents to equalize with the heating block. Finally, the DSF assay was performed from 25°C to 95°C with a temperature ramp of 1°C per minute.

3.2.3 Crystallization and structure determination of SgrMPCO proteins

Each of the purified SgrMPCO proteins was screened for initial crystallization conditions at 8-12 mg/mL against the Index HT (Hampton Research) and LMB (Molecular Dimension). After optimization, diffraction-quality crystals were obtained by the hanging drop vapor diffusion method. Final crystallization conditions are listed in table 3.2.3-1.

Table 3.2.3-1: Crystallization condition of Sgr enzymes

Protein	Crystallization condition	Cryo-protectant solution
SgrMppP (12mg/mL)	100 mM Ca-acetate+ 50 mM CaCl ₂ + 25% PEG+ 100mM BisTris pH 5.5	20% Glycerol+ Crystallization solution
SgrMT(8mg/mL)	0.2M MgCl ₂ , 025% PEG 3350, 0.1M TRIS pH 8.5	20% Glycerol+ Crystallization solution
SgrAH (9 mg/mL)	0.16M Li ₂ SO ₄ , 100mM MnCl ₂ , 30% PEG 3350, 0.1M BisTris propane pH 7.	ParatoneN

X-ray diffraction data were collected at beamlines 21-ID-F of the Life Science Collaborative Access Team (LS-CAT) at the Advanced Photon Source (APS). All data were processed with HKL2000 version 715.

As SwMppP and SgrMppP share 51% sequence identity, the PbrMppP structure was determined by molecular replacement using SwMppP as the reference model. The diffraction data were collected from a crystal using a wavelength of 0.97921 Å. PbrMppP crystallized in space group $P2_1 2_1 2_1$ with two chains in the asymmetric unit. The molecular replacement solution was

subjected to iterative cycles of manual model building in COOT^{8 clxxx} and maximum likelihood-based refinement using the PHENIX package^{clxxxi}. Ordered solvent molecules were added automatically in phenix.refine^{4clxxxii}. Hydrogen atoms were added to the model using phenix.reduce^{clxxxiii} and were included in the later stages of refinement to improve the stereochemistry of the model. The positions of H atoms were refined using the riding model with a global B factor. Regions of the model for translation-libration-screw (TLS) refinement were identified using phenix.find_tls_groups (P. V. Afonine, unpublished work), and the TLS parameters were refined in phenix.refine^{clxxii}. Once the refinement converged ($R_{\text{cryst}} = 0.142$, and $R_{\text{free}} = 0.184$), the model was validated using the tools implemented in COOT^{clxxxiv}.

SgrMT formed rod-shaped crystals after 3–4 days and grew to maximum dimensions of $\sim 150 \times \sim 60 \times < 5 \mu\text{m}$. Crystals were cryo-protected using 20% glycerol in the holding solution and were flash cooled by plunging into liquid nitrogen. The structure of SgrMT with **[91]** bound was obtained by soaking crystals in crystallization solution with 10mM **[91]**. After soaking for 2 hours at RT, crystals were cryo-protected by sequential soaks in holding solutions containing 5, 10, or 20% glycerol with 10mM **[91]** in crystallization solution listed in table 3.2.3-1, and flash cooled by plunging into liquid nitrogen.

The SgrMT structure was determined by the single wavelength anomalous diffraction (SAD) method using data collected at 0.97856 Å. SgrMT crystalized in space group $P2_12_12_1$ with two molecules in the asymmetric unit. The initial model was built automatically using the PHENIX package (phenix.autobuild)^{clxxxv}. The solution was subjected to iterative cycles of manual model building in COOT^{8 xcvi} and maximum likelihood-based refinement using the PHENIX package. Ordered solvent molecules were added automatically in phenix.refine^{clxxii}. Solvent molecule, H-

atoms, and TLS refinements were performed similarly described above. Once the refinement converged ($R_{\text{cryst}} = 0.22$, and $R_{\text{free}} = 0.31$), the model was validated using the tools implemented in COOT^{c/xxiv}. The final refined model stripped of water molecules and H atoms and with all B factors optimized, was used to determine the final. Structure of SgrMT liganded with 2KA was determined by difference Fourier.

The SgrAH formed rod-shaped crystals after 5 days and grew to maximum dimensions of $\sim 140 \times \sim 40 \times \sim 5 \mu\text{m}$. These crystals were cryo-protected using Paratone N and flash-cooled by plunging into liquid nitrogen. SgrAH SAD using similar procedure described for SgrMT. The initial model building; solvent molecule, H-atoms, and TLS refinements were also performed similarly described above, till the refinement converged ($R_{\text{cryst}} = 0.142$, and $R_{\text{free}} = 0.184$), the model was validated using the tools implemented in COOT^{c/xxiv}. Data collection and model refinement statistics are listed in Appendix B.

3.2.4 Mass spectrometric analysis of reaction products

To perform reactions for LCMS analysis, SgrMppP and SgrMT enzymes were separately desalted into 20mM sodium phosphate buffer pH 8.4 using a HiTrap desalting column (GE Lifesciences). Similarly, 1 mL of 1mg/mL catalase from bovine liver (Sigma-Aldrich, St. Louis, MO) was desalted deionized water to get rid of the trehalose stabilizer from the catalase preparation. Catalase was added to remove the hydrogen peroxide that is produced as a by-product of SgrMppP catalysis, since it leads to non-enzymatic oxidative decarboxylation of [91] and [95].

Reactions were assembled into a 1.5mL PCR tube by adding 36 μL of 277 μM SgrMppP desalted enzyme, 1 μL of 50 $\mu\text{g}/\text{mL}$ desalted catalase and 250 μL of 4mM L-arginine stock solution prepared in 20mM sodium phosphate buffer pH 8.4. pH of all solutions was adjusted using a Fisherbrand™ accumet™ AE150 Benchtop pH Meter. The reaction mixture was adjusted to 500 μL volume by adding 213 μL of phosphate buffer to reach the final concentration of approximately 20 μM SgrMppP, 2mM L-arginine and 0.1 $\mu\text{g}/\text{mL}$ catalase. Reaction tubes were shaken at 25°C, 200 rpm for 4 hours and protected from light. At the end of the SgrMppP reaction, 54 μL of 445 μM SgrMT, 84 μL of 15 mM S-adenosyl methionine (SAM) and 2.8 μL of 5.0 mM MgCl_2 was added to reach the final concentration of 20 μM SgrMT, 1.8 mM SAM and 20 μM MgCl_2 , considering final volume of 700 μL . The negative control contains all the same ingredients, except SgrMT was denatured by heating on boiling water bath.

The products of SgrMT reaction was analyzed using triple quad LCMS-8040 (UPLC) from Shimadzu Scientific Instruments in the Shimadzu Laboratory for Advanced and Analytical Chemistry, UWM. Compounds in the reaction samples and control, were ionized using ESI method with an applied current of 4.4 volt. A ZIC® HILIC column from EMD Millipore corporation, MA was used to separate the ions. Stationary phase was zwitterionic particles comprised of silica, amide and sulfobetaine. For aqueous mobile phase, 2mM NH_4COOH pH 7.4 was used, and LCMS grade acetonitrile (From Fischer chemical) was used as organic mobile phase. A binary gradient was applied starting with 90% organic phase and going down to 50% across 8 minutes, followed by increment up to 90% from 8-12 minutes. The 90% organic phase were continued to 15th minute to store the HILIC column. Enzymatically synthesized S-¹³C labelled SAM and SgrAH reaction products were also identified using similar method.

3.2.5 Preparation of S-¹³C-labeled SAM

Reactions were assembled in 50mL corning tubes by adding 3.3 mL of 300 μ M EcMetK, 2 mL of 10 mM ATP and 2 mL of 10 mM methyl-¹³C L-methionine (Cambridge Isotope Laboratories, Inc.; catalog number CLM-206-PK). The reaction mixture was adjusted to 10 mL by adding 2.7 mL of 20 mM sodium phosphate buffer, pH 8.4, to reach final concentrations of 100 μ M EcMetK, 2 mM methyl-¹³C L-Met, and 2 mM ATP. The reactions were incubated at 25°C with shaking at 200 rpm for 6 hours. After 6 h, the reactions were evaporated to dryness using a Centrifan PE Rotary Evaporator (KD Scientific, Holliston, MA) at 50 °C overnight. The resulting residue was taken up in 600 μ L H₂O and centrifuged at 17,000 x g for 10 min to remove any remaining particulates. The supernatants were analyzed by LCMS to confirm the presence of methyl-¹³C labelled SAM as described in Section 3.2.4.

3.2.6 NMR characterization of SgrMT reaction products

Reactions were assembled in 50mL corning tubes by adding 724 μ L of 277 μ M SgrMppP in 20 μ M phosphate buffer pH 8.4 , 2 μ L of 500 μ g/mL catalase (desalted into H₂O) and 5 mL of 4 mM L-Arg stock solution prepared in 20mM sodium phosphate buffer pH 8.4. The reaction mixture was adjusted to 10mL with 20 mM sodium phosphate buffer to reach final concentrations of 20 μ M SgrMppP , 2mM L-Arg and 1 μ g/mL catalase. The reactions were incubated at 25 °C with shaking at 200 rpm for 4 hours in the dark. After the SgrMppP reactions were complete, 539 μ L of 445 μ M SgrMT, 1413 μ L of 15 mM S-adenosyl methionine (SAM) and 48 μ L of 5.0 mM MgCl₂ were

added to reach final concentrations of 20 μM SgrMT, 1.8 mM SAM and 20 μM MgCl_2 in a total volume of 12 mL. To determine the promiscuousness of SgrMT for divalent metals, the same procedure was followed, except that the MgCl_2 solution was replaced with either FeCl_2 or MnCl_2 . These reactions were incubated for 4 hr at 25 $^\circ\text{C}$ with shaking at 200 rpm. The reactions were then evaporated to dryness in the rotary evaporator at 50 $^\circ\text{C}$ Overnight. The resulting residue was taken up in 600 μL D_2O and centrifuged at 17,000 x g for 10 min to remove any remaining particulates. The supernatants (500 μL) were loaded into NMR tubes. Samples were shimmed in the Bruker Avance III HD 500 MHz spectrometer for taking ^1H NMR, ^1H - ^1H COSY (correlated spectroscopy) and HSQC (heteronuclear single quantum transfer correlation spectroscopy) spectra. For the Incredible Natural Abundance Double QUAntum Transfer Experiment (INADEQUATE) analysis, a similar reaction was performed using ^{13}C -L-Arg and methyl- ^{13}C labeled SAM (preparation described in 3.2.5).

3.2.7 Steady state kinetics determination of SgrMT

A SAM-dependent methyltransferase assay kit from G-Biosciences (SAM510: SA Methyltransferase Assay, Cat. # 786-430) was used to study the steady state kinetics of SgrMT. Assays (115 μL) were conducted in triplicate at 37 $^\circ\text{C}$ in 100 mM TRIS pH 8.0. The initial velocity of SgrMT-catalyzed methylation of **[91]** was measured using a Hewlett-Packard 8453 diode array spectrophotometer. For each trial, 100 μL of the SAM assay master mix, 5 μL of $\sim 8 \mu\text{M}$ SgrMT and 10 μL of various concentrations of **[91]** ranging from 10 μM to 300 μM (methyl group acceptor substrate) were mixed in a cuvette and the absorbance at 510 nm was measured immediately.

The concentration of substrate **[91]** was varied from 10 to 300 μM . The initial velocity for each reaction was calculated from linear regression of each curve, using equation 4:

$$\Delta\text{Abs}/\text{min} = \frac{(\text{Abs at Time 2}) - (\text{Abs at Time 1})}{\text{Time 2} - \text{Time 1}}$$

Activity of methyltransferase was obtained using equation 5:

$$\text{Methyltransferase Activity } (\mu\text{M}/\text{s}) = \frac{\Delta\text{Abs}/\text{min}}{0.026 / (\mu\text{M})} * \frac{115 \mu\text{L}}{5 \mu\text{L}} * \frac{1}{60 \text{ s}}$$

The k_{cat} and K_{M} values were determined by fitting the initial velocity data using the equation 1.

The same procedure was followed for analysis of SgrMT kinetics with **[95]**. Kinetic parameters are listed in table 3.3.5-1.

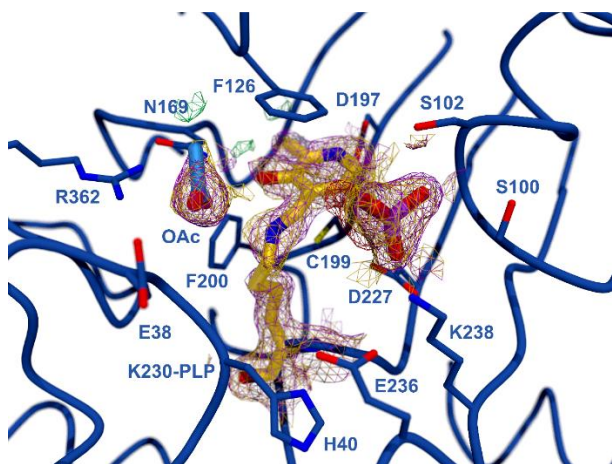
3.3 Results and discussions

3.3.1 SgrMppP is a homolog of SwMppP

SgrMppP is a homolog of SwMppP with a sequence identity of 54%, thus the activity of SgrMppP is same as SwMppP (data not shown here). The X-ray crystal structure of PbrMppP shows that its tertiary and quaternary structures are almost identical to those of SwMppP (RMSD value 0.9 Å, Figure 2.1-3). Likewise, the active sites of the two homologs are nearly identical as well.

SgrMppP crystallized in space group $P2_12_12_1$ with unit cell dimensions of $a = 85.7 \text{ \AA}$, $b = 108.3 \text{ \AA}$, and $c = 195.4 \text{ \AA}$ with 4 molecules in the asymmetric unit arranged as 2 independent homodimers (Figure not shown here). The domain architecture is identical to SwMppP. In the absence of substrate, PLP is covalently attached to Lys230 by an aldimine linkage (Figure 3.3.1-1B). The negative charge of Asp197 stabilizes the pyridinium form of the cofactor by making a salt bridge to the pyridine N atom. This interaction has been shown to be critical for catalysis in a number of PLP-dependent enzymes like aspartate aminotransferase^{clxxxvi}. The cationic pyridinium allows the cofactor to accept electrons from the substrate, stabilizing the carbanion formed during catalysisⁱ.

A)



B)

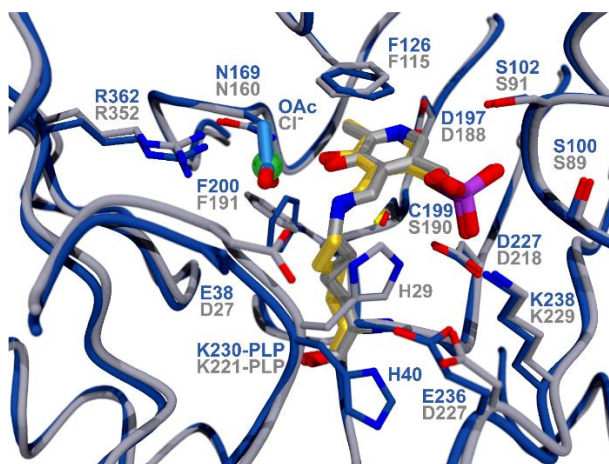
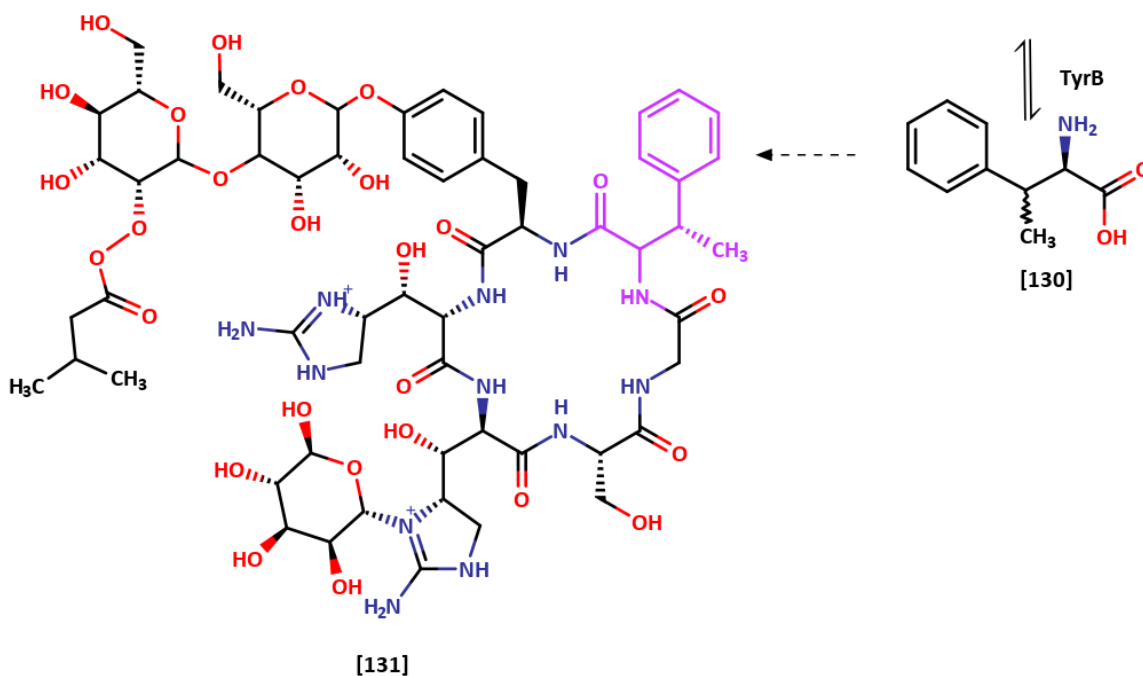
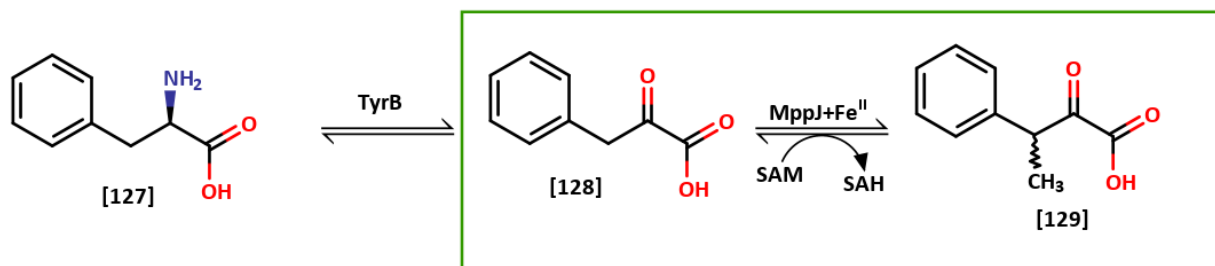


Figure 3.1-1: A) Structure of *SgrMppP* homolog liganded with cofactor PLP resolved at 2.1 Å, R_{cryst} : 0.142, R_{free} : 0.184. B) *SgrMppP* (blue chain) active site overlaid with *SwMppP* (gray chain) with 51% sequence identity and RMSD value of 0.97 Å.

3.3.2 *SgrMT* is a structural homolog of *MppJ*

SgrMT crystallized in space group $P2_12_12_1$ with two molecules in the asymmetric unit. The crystallographic structure of *SgrMT* (Figure 3.3.2-1A) reveals that it has a potential divalent metal-binding triad comprised of His232, Glu233 and His290. *SgrMT* shares 21% sequence identity with the enzyme *MppJ* (PDB ID: 4kig), a methyltransferase from *Streptomyces hygroscopicus*. In this organism, *MppJ* is a part of the mannopeptimycin- γ biosynthetic cluster where it functions in a SAM- and Fe(II)-dependent manner to methylate phenylpyruvate into β -methyl phenylpyruvate (Scheme 3.3.2-1).

Superimposition of the structure of SgrMT with Mg^{2+} and **[91]** bound on that of MppJ with β -methyl phenylpyruvate bound (Figure 3.3.2-1B) shows that the orientations of most active site residues as well as the metal-binding triad are similar. The carboxylate group of **[91]** is coordinated with the Mg ion. The guanidinium group of **[91]** interacts with the carboxylate group of Asp136.



Scheme 3.3.2-1: reaction catalyzed by SghMppJ. MppJ is part of mannopeptimycin- γ [131] biosynthetic gene cluster in *Streptomyces hygroscopicus*. TyrB oxidizes phenylalanine [127] into phenylpyruvate [128]. MppJ methylates [128] in Fe^{II} and SAM dependent manner to produce β -methyl phenylpyruvate [129], which finally incorporates into mannopeptimycin- γ [131]. The purple highlighted structure in [131] is the coming from MppJ product derivative.

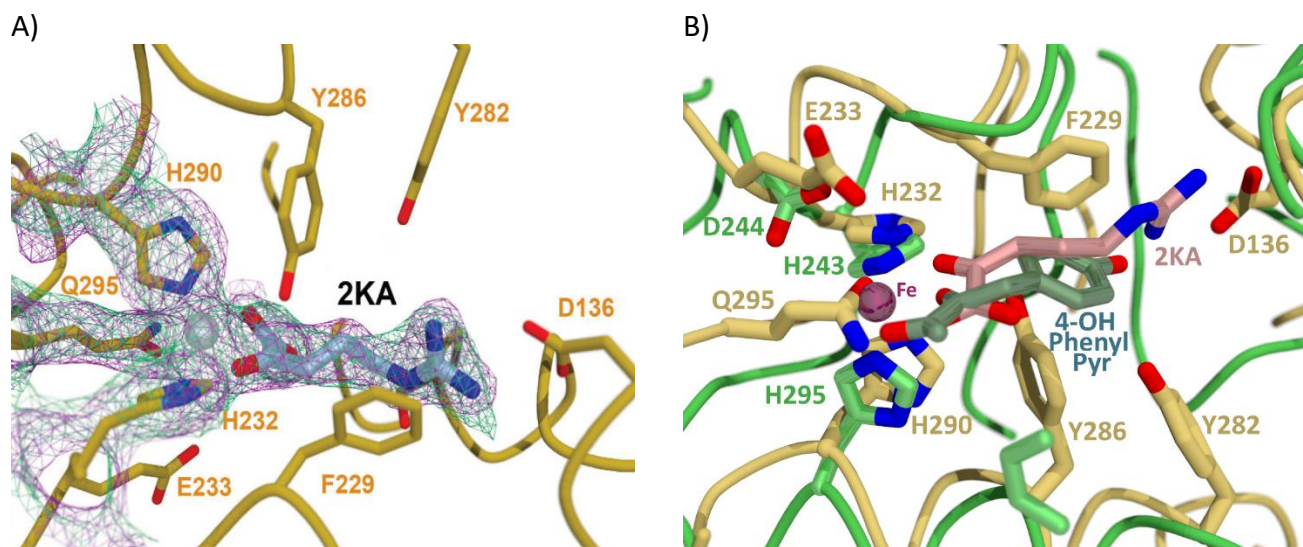
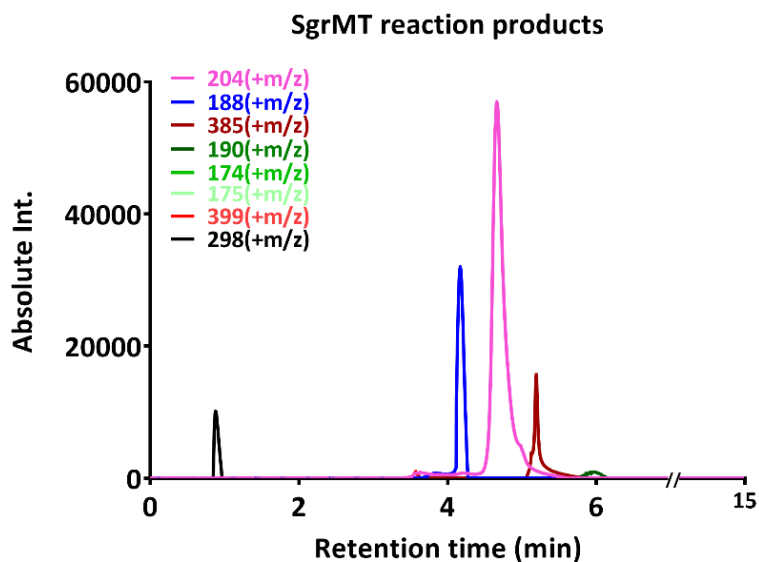


Figure 3.3.2-1: A) structure of SgrMT liganded with Mg²⁺ and [91] or 2KA bound at 2.2 Å resolution. The silver sphere is the Mg²⁺ ion and the blue structure underneath the electron density map is the 2KA or [91]. B) Superimposed structures of SgrMT (yellow) with substrate [91] and Mg²⁺ bound and of MppJ (green, PDB ID: 4kig) liganded with the product β -methyl phenylpyruvate (RMSD = 1.2 Å for all C α atoms). The red sphere is the Fe bound in the active site of MppJ

3.3.3 SgrMT product identification by LCMS

The products of the SgrMT reaction were analyzed by LC-MS using ESI ionization. Analytes were separated by hydrophilic interaction liquid chromatography using a ZIC[®] HILIC column. The extracted ion chromatogram (in positive mode) from LC-MS analysis is presented in Figure 3.3.3-1A. The SgrMT negative control (Figure 3.3.3-1B) shows base peaks of $m/z = 190$ and $m/z = 174$, which correspond to the SgrMppP products **[95]** and **[91]**. There is a trace of L-Arg starting material ($m/z = 175$), but it is very slight. The vast majority of the starting material was turned over by SgrMppP. The peak at $m/z = 399$ is for intact SAM (Scheme 3.2.2-1) **[134]** and $m/z = 298$ is a fragment of SAM that appears early in the separation method (ref METLIN ID: 3289). Notably, the 298 species only comes from SAM. The SAM byproduct after methyl donation, S-adenosyl homocysteine (SAH, Scheme 3.3.2-1) **[135]** does not have any similar fragment. The SgrMT reaction mixture shows a peak at $m/z = 204$ (Figure 3.3.3-1A), which corresponds to the methylated 4-electron oxidized product of SgrMppP **[136]** (Scheme 3.3.3-1). Interestingly, another peak is observed at $m/z = 188$, which refers to methylated 2-electron-oxidized SgrMppP product **[137]** (Scheme 3.3.3-1). Thus, SgrMT turns over both the major, **[95]**, and minor, **[91]**, products. Also, in the SgrMT reaction sample, a peak at $m/z = 385$ is found due to the SAM byproduct SAH (ref METLIN ID: 296). These peaks (m/z 204, 188 and 385) are absent in negative control reactions lacking Mg^{2+} and SAM. Thus, it can be clearly stated that SgrMT is a SAM- and Mg^{2+} -dependent methyltransferase. Individual mass spectra for each chromatogram are presented in Figure 3.3.3-2. All the data show m/z species above a 5% relative abundance threshold. Table 3.3.3-1 represents all the major peaks along with their corresponding predicted structures.

A)



B)

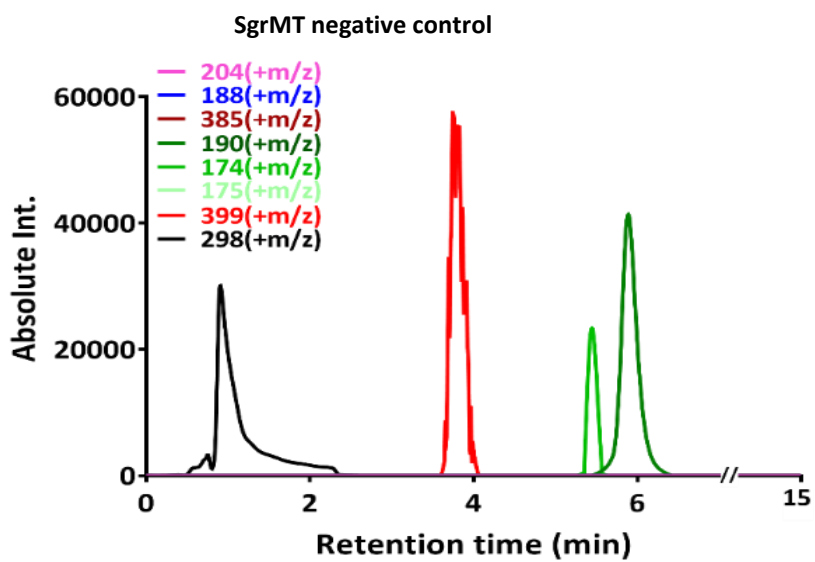
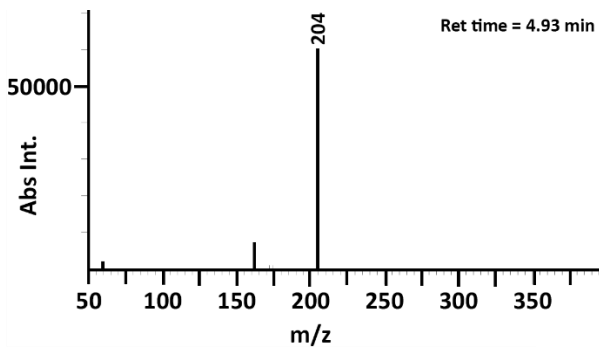
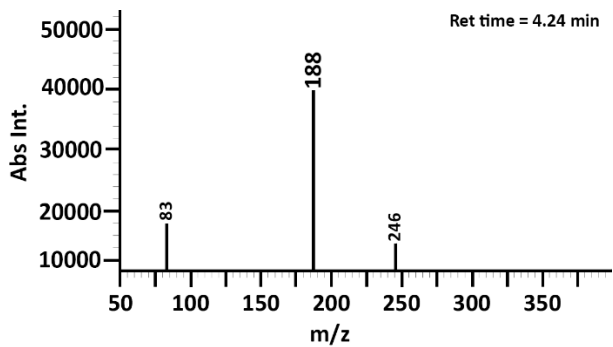


Figure 3.3.3-1: A) Extracted ion chromatograms of SgrMT reaction products B) Extracted ion chromatograms of SgrMT negative control.

A)



B)



C)

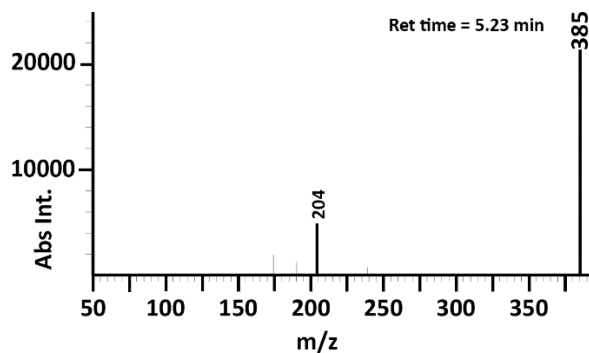
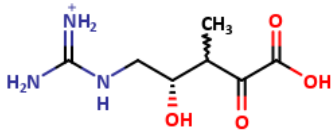
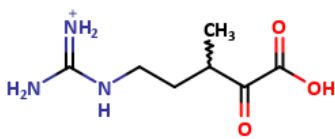
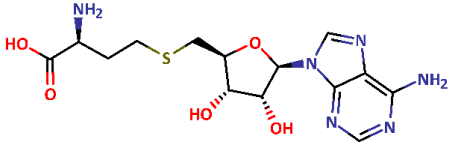
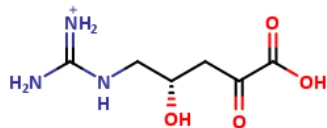
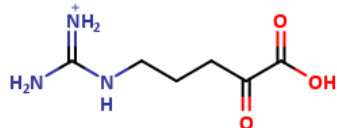
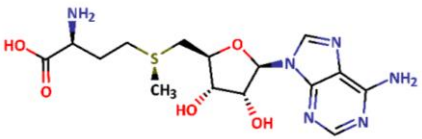
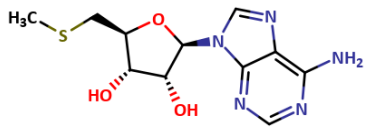
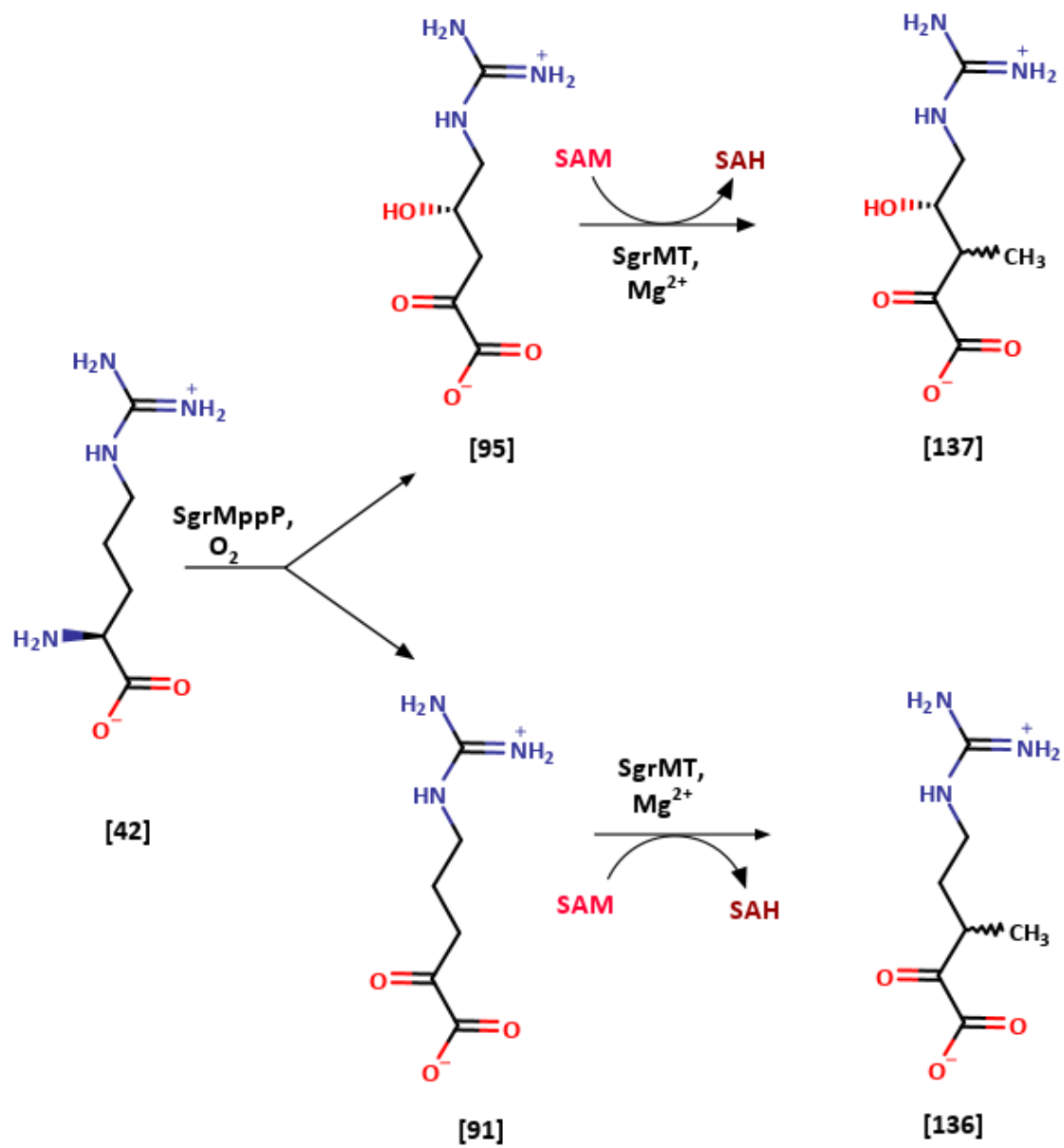


Figure 3.3.3.-2: Mass spectra of individual chromatogram at specified retention time. All the data shown are above 5% relative abundance threshold, considering the base peak 100%. A) mass spectrum of EIC 204 (m/z) in positive mode at 4.93 min retention time, B) mass spectrum of EIC 188 (m/z) in positive mode at 4.24 min retention time, C) mass spectrum of EIC 385 (m/z) in positive mode at 5.23 min retention time.

Table 3.3.3-1: The major traces of m/z along with their corresponding predicted structure, analyzed from LCMS data of SgrMT reaction products. The negative control refers to SgrMppP reaction products + SAM+ Mg²⁺ + heat denatured SgrMT.

m/z	Ret time (min)	Structure prediction	Chemical name	Presence in sample	Annotation in scheme 3.3.2-1
204	4.93		3-methyl-2-oxo-4(S)-hydroxy-5-guanidinovaleric acid	SgrMT reaction	[136]
188	4.24		3-methyl-2-oxo-5-guanidinovaleric acid	SgrMT reaction	[137]
385	5.23		S-adenosyl homocysteine	SgrMT reaction	[135]

190	5.92		2-oxo-4(S)- hydroxy-5- guanidinovale ric acid	negative control	[95]
174	5.28		2-oxo-5- guanidinovale ric acid	negative control	[91]
399	3.85		S-adenosyl methionine	negative	[134]
298	0.91		S-methyl adenosine	Both, SgrMT reaction and negative control	



Scheme 3.3.3-1: Predicted reaction catalyzed by SgrMT.

3.3.4 Determination of SgrMT products by NMR spectroscopy

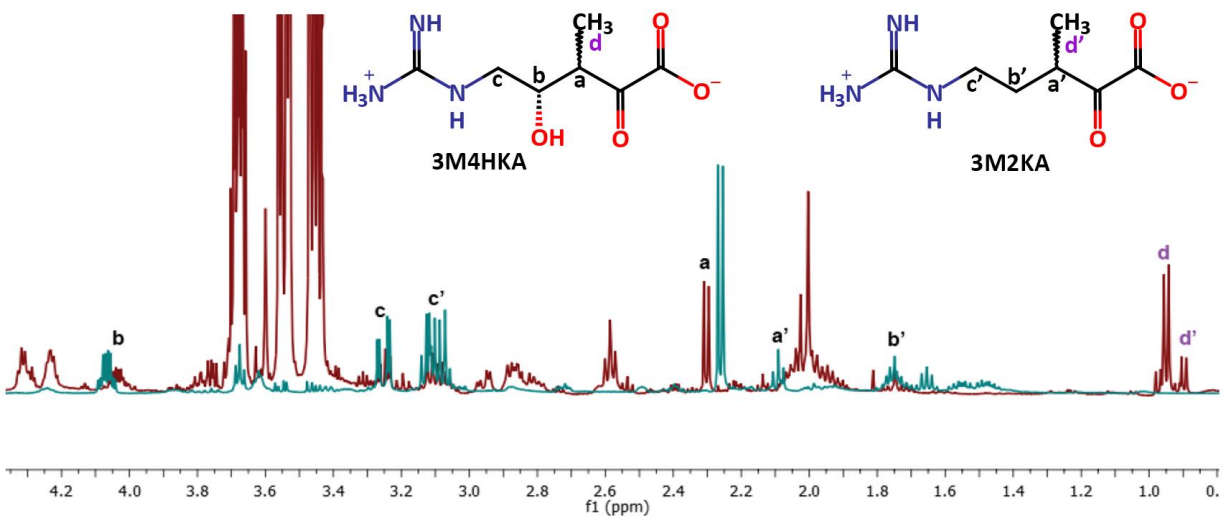
The SgrMT reaction products were concentrated to approximately 12 mM by evaporating the reaction mixtures to dryness and reconstituting the residue in D₂O. The SgrMppP reaction mixture (before addition of SgrMT) was used as a negative control (green spectrum in Figure 3.3.4-1A). There was subtle chemical shift observed in the SgrMT reaction. Such downfield shift is usual for SAM dependent reactions, possibly due to the ionic state of sulfur atom that is easily influenced by pH fluctuation in enzymatic reactions^{clxxxvii}. A pair of doublets are observed in the SgrMT reaction at 0.9 ppm (d and d' in Figure 3.3.4-1A). Such doublets are not observed at this chemical shift when SAM and SAH are analyzed by NMR at different pH values. Moreover, such up-field chemical shift is observed in case of aliphatic protons^{clxxxviii}. This observation is consistent with the fact that the possible methylation position is the β-carbon of **[95]** and **[91]**. The doublet with higher intensity corresponds to the major product of MppP **[95]** methylation (annotated as **d** in figure 3.3.3-1A). On the other hand, the doublet with lower intensity corresponds to the minor product of MppP **[91]** (annotated as d' in figure 3.3.4-1A).

The 2D-¹H correlation was acquired by 2D-¹H-COSY NMR of SgrMT Reaction with 2KA (figure 3.3.4-1B). As the MppP reaction product contains unreacted L-Arg, **[95]**, and **[91]**, it was difficult to assign all of these peaks in the 1D spectra. The presence of SAM and SAH further complicates these spectra. This problem was addressed by running reactions of SgrMT with pure **[91]** rather than the mixture from the SgrMppP reaction and the products were analyzed by COSY NMR. Figure 3.3.4-1B clearly shows the correlation of annotated β-hydrogen “a” with methyl hydrogen

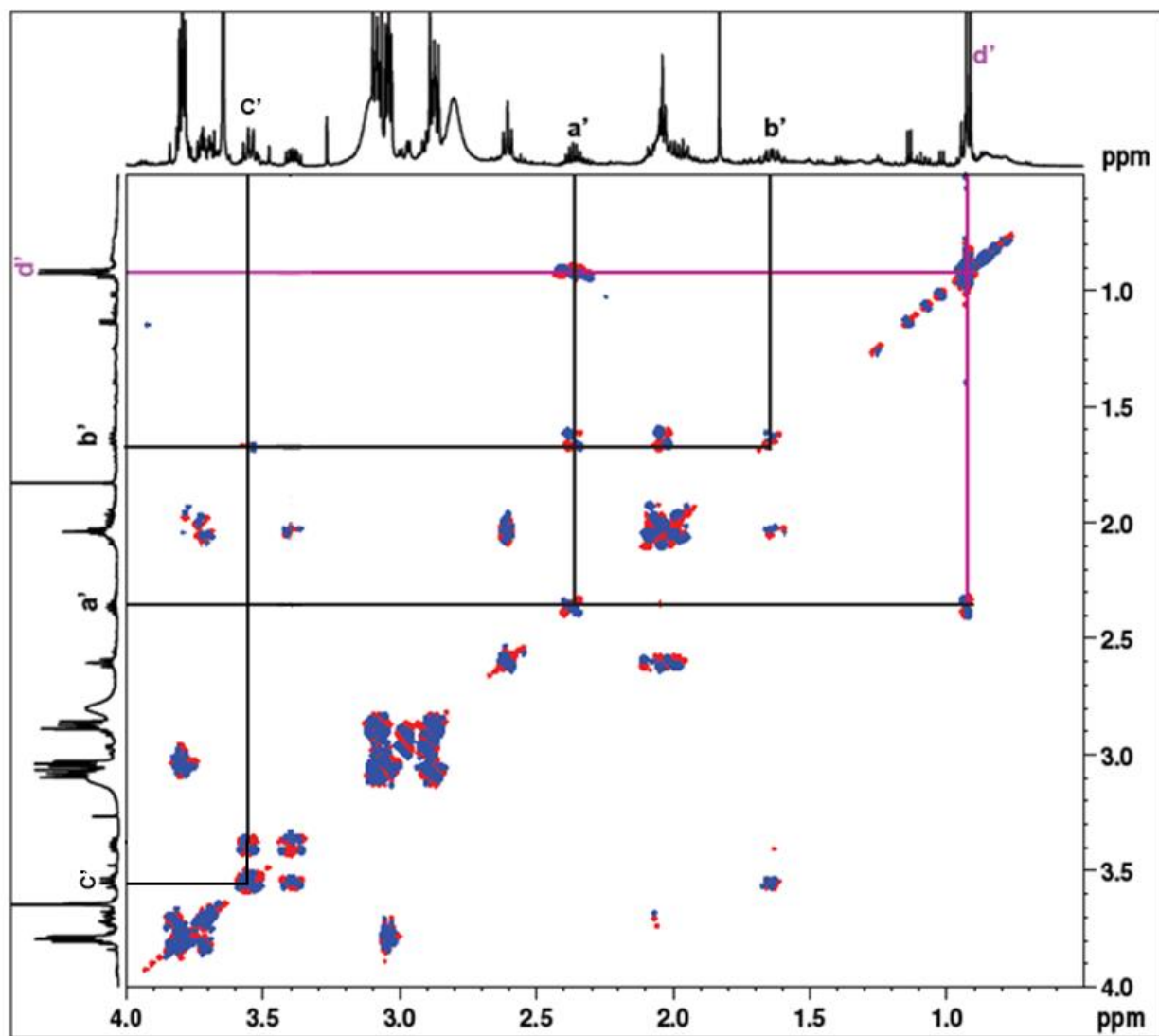
d', the β -H is also split by γ -H, annotated as b'. On the other hand, b' is coupling with the a' as well as the δ -H c'.

To get a cleared picture of the SgrMT product correlation, we also analyzed the 2D- ^1H - ^{13}C HSQC Spectrum of SgrMT Reaction with 2KA (Figure 3.3.4-1C). The X-axis refers to the ^1H -NMR and the Y-axis refers to the ^{13}C -NMR. The annotated methyl carbon β' shows up at 12 ppm, which is coupled with d' or, methyl hydrogens. In contrast, the β -C at 43 ppm is coupled with d' or the a' multiplet near 2.35 ppm along X-axis. The γ -C at 29.5 ppm is coupled with b' or the γ -Hs, which is the multiplet at 1.68 ppm along X-axis. Finally, the δ -C at 47ppm along Y-axis is coupled with c' or the γ -Hs, the multiplet at approximately 3.4 ppm.

A)



B)



c)

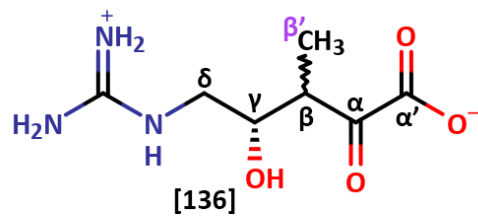
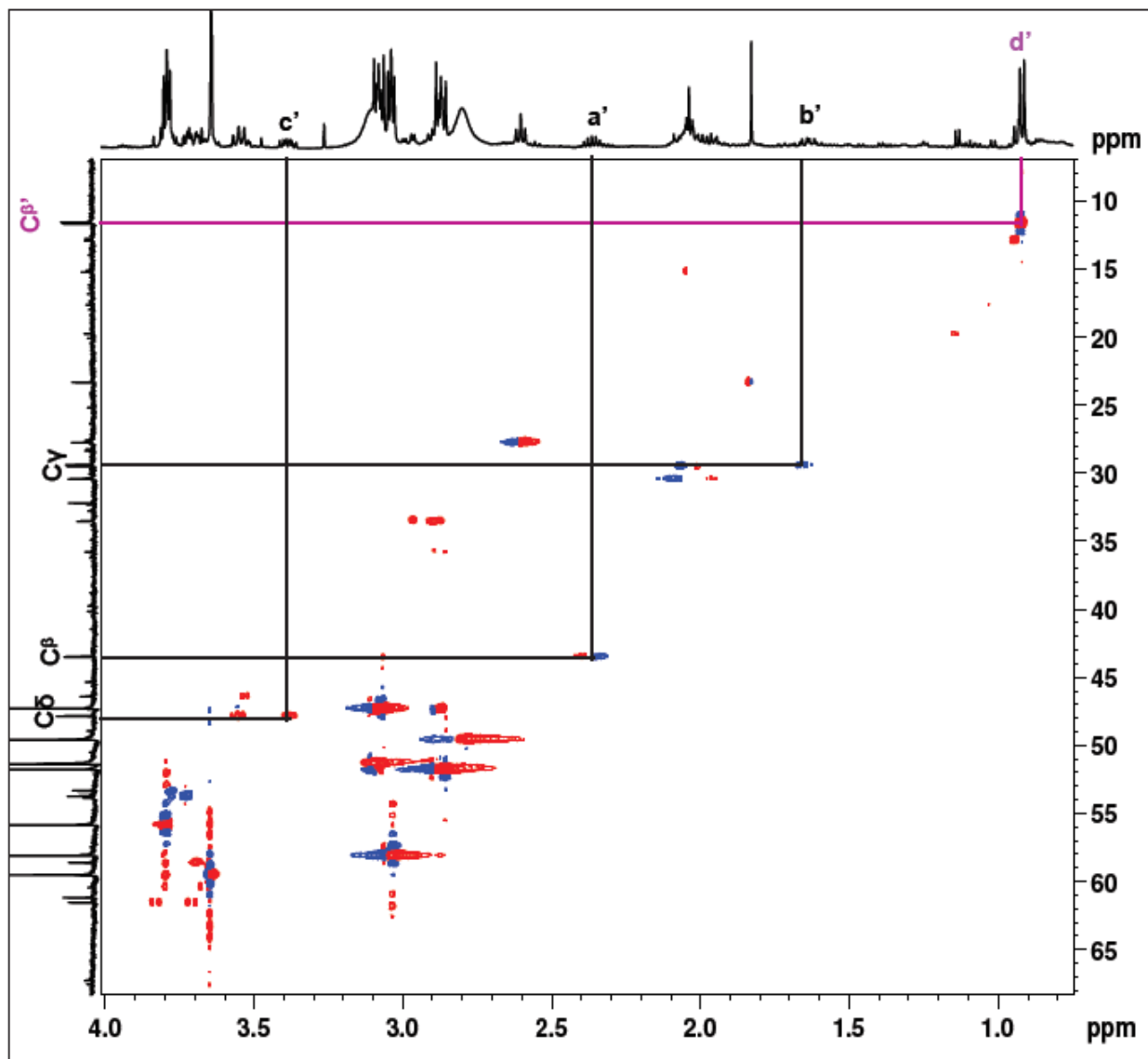
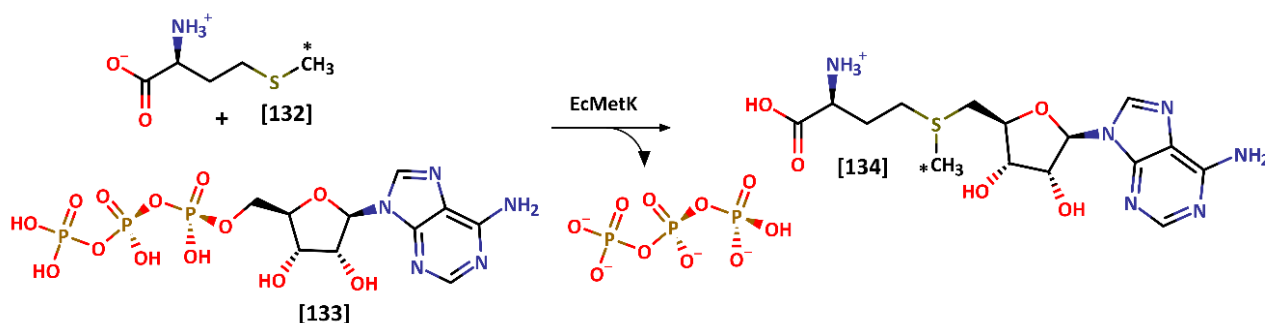


Figure 3.3.4-1: determination of SgrMT reaction products by NMR spectroscopy. A) ^1H -NMR of SgrMT reaction products. The green spectrum is the negative control or in other words SgrMppP reaction products only. The red spectrum is the SgrMT reaction products in presence of SAM and Mg^{2+} , where substrates are SgrMppP products. B) ^1H -COSY NMR of SgrMT Reaction with 2KA. C) ^1H - ^{13}C HSQC Spectrum of SgrMT Reaction with 2KA. The X-axis refers to the ^1H -NMR and the Y-axis refers to the ^{13}C -NMR.

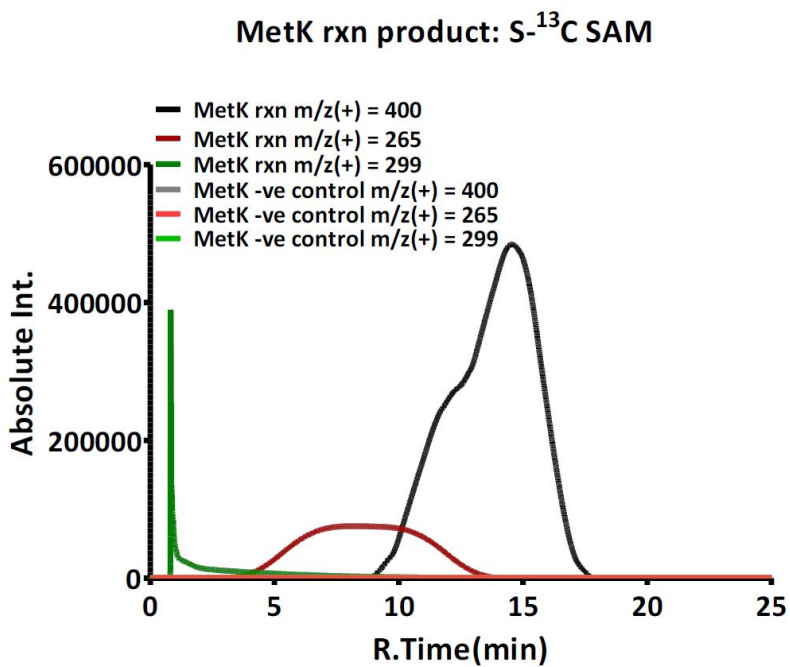
3.3.5 Confirmation of methyl-¹³C SAM synthesis by LC-MS

The purified EcMetK enzyme was used to generate methyl-¹³C SAM from methyl-¹³C L-Met and ATP (Scheme 3.3.5-1). The product of the reaction was analyzed using LC-MS. The results were compared with naturally abundant SAM [134]. The fragment fingerprint of naturally abundant SAM shows base peaks of $m/z = 399, 298,$ and 264 (Figure 3.3.5-1A) where all the three fragments retain the S-connected donor methyl group (METLIN ID 3289). The methyl-¹³C-labeled SAM shows base peaks of $m/z = 400, 299,$ and 265 instead, confirming the isotopic substitution. In case of methyl-¹³C labeled SAM, the specific ions are eluting across a larger retention timespan, possibly due to sample overload. Table 3.3.5-1 shows the structures of the major ions in the chromatograms.



Scheme 3.3.5-1: Reaction catalyzed by EcMetK, where the S-connected methyl group is ¹³C-isotope labelled

A)



B)

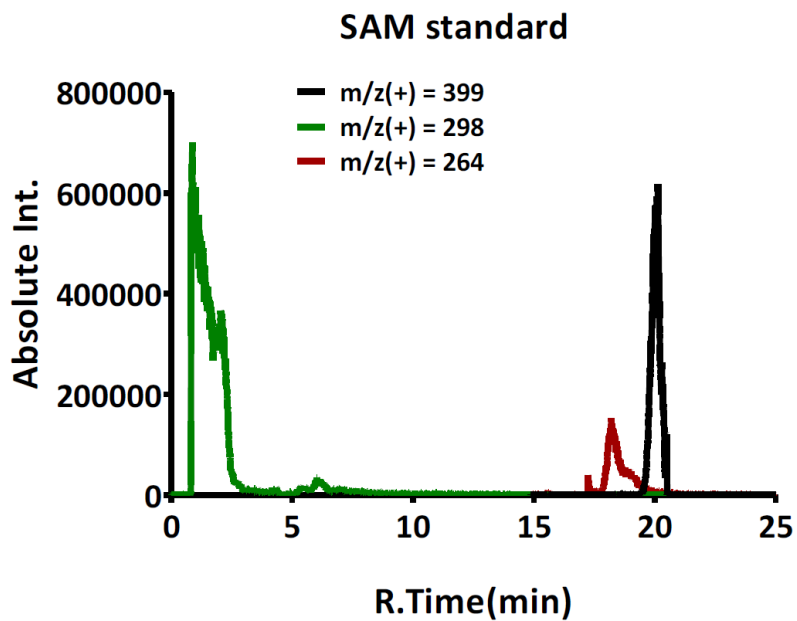
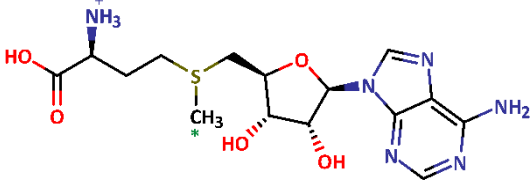
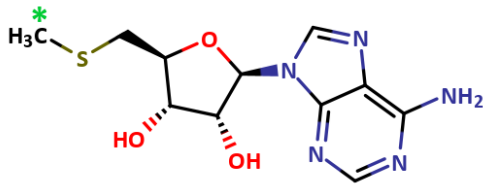
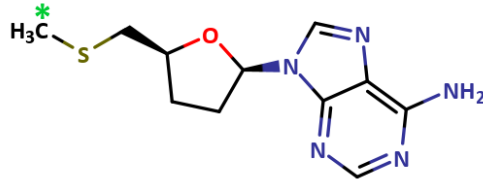


Figure 3.3.5-1 A) Chromatogram of EcMetK reaction product with S-¹³C labelled SAM, B) Chromatogram of naturally abundant SAM.

Table 3.3.5-1: Structure of major peaks determined from SAM mass spectroscopic fingerprint. The green asterisks indicates ¹³C-isotope label

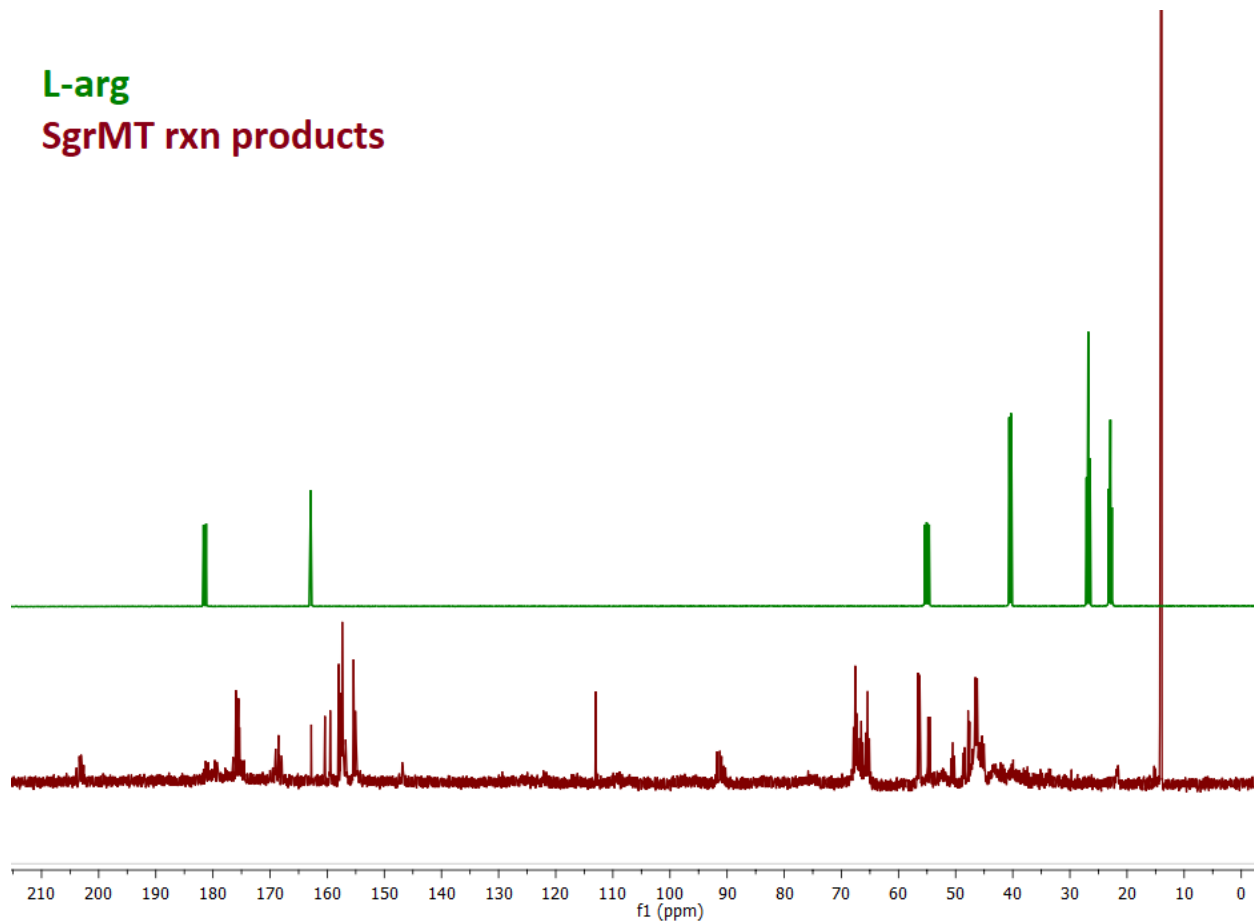
Structure of major peaks	m/z of natural compound	m/z of prepared S- ¹³ C SAM
	399	400*
	298	299*
	264	265*

3.3.6 Determination of methylation position by INADEQUATE NMR

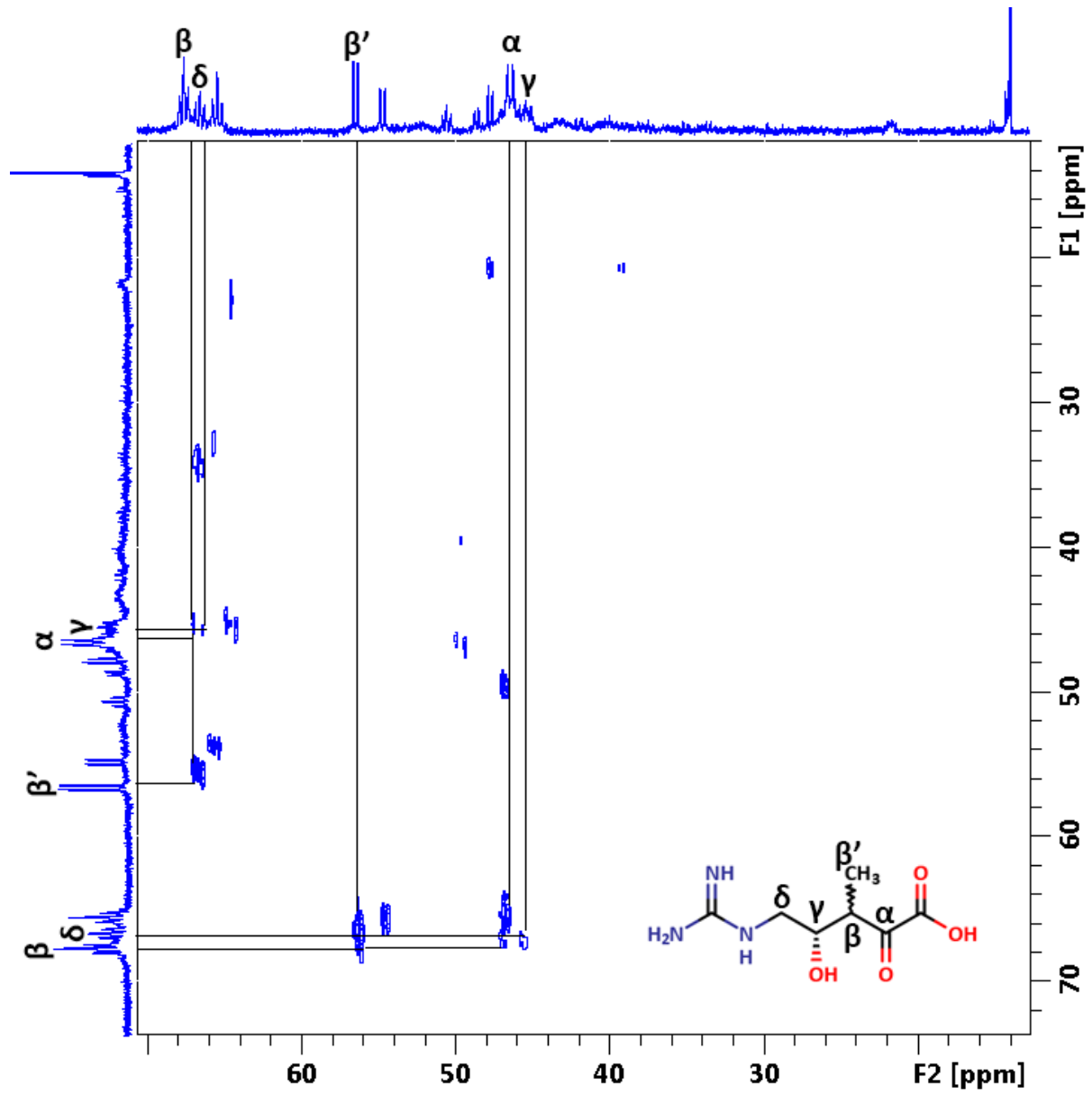
The preliminary NMR data of ^1H NMR, ^1H - ^1H correlation spectroscopy (COSY), and heteronuclear single quantum transfer correlation spectroscopy (HSQC) suggested that the SAM methyl group is installed at $\text{C}\beta$ of the substrate, but these experiments were not definitive. To unequivocally assign the methylation position on the substrate, we used 2D- ^{13}C - ^{13}C coupling NMR or Incredible Natural Abundance Double QUAntum Transfer Experiment (INADEQUATE) NMR^{clxxxix,cxc}. The distinguishing characteristic of the INADEQUATE experiment is that it is insensitive to the 0.01% natural abundance of ^{13}C isotope. Any species with at least 10% ^{13}C can be exclusively excited by INADEQUATE, thus the 2D coupling of adjacent carbons can be identified if only the carbons in the target compounds are all ^{13}C -labeled. Figure 3.3.6-1 shows the 2D ^{13}C - ^{13}C correlation of SgrMT reaction products. The doublet around 17ppm is assigned as the β' carbon. This shows that SgrMT methylates the β carbon of **[95]** or **[91]**. Since the αC in both case of **[136]** and **[137]** is a keto-group, it was supposed to be around 180 to 200 ppm region. We observe from the INADEQUATE correlation that the αC has a chemical shift to approximately 47 ppm. This huge chemical shift is probably due to the conversion of α -keto group into alcohol (C-OH)^{cxc}. Also we observed that there is only one set of doublet for methylation of **[91]** and **[95]**. This observation suggests that the SgrMT products (**[136]** and **[137]**) has only one stereoisomer of β -methyl group. Moreover, the absence of any additional doublet resembles that the **[136]** and **[137]** does not form any lactone over the course of reaction at pH 8.4. Additional experiment, i.e. structure of SgrMT with product **[136]** or **[137]** bound form is required to finally reveal the conformation of β -methyl group.

A)

L-arg
SgrMT rxn products



B)



c)

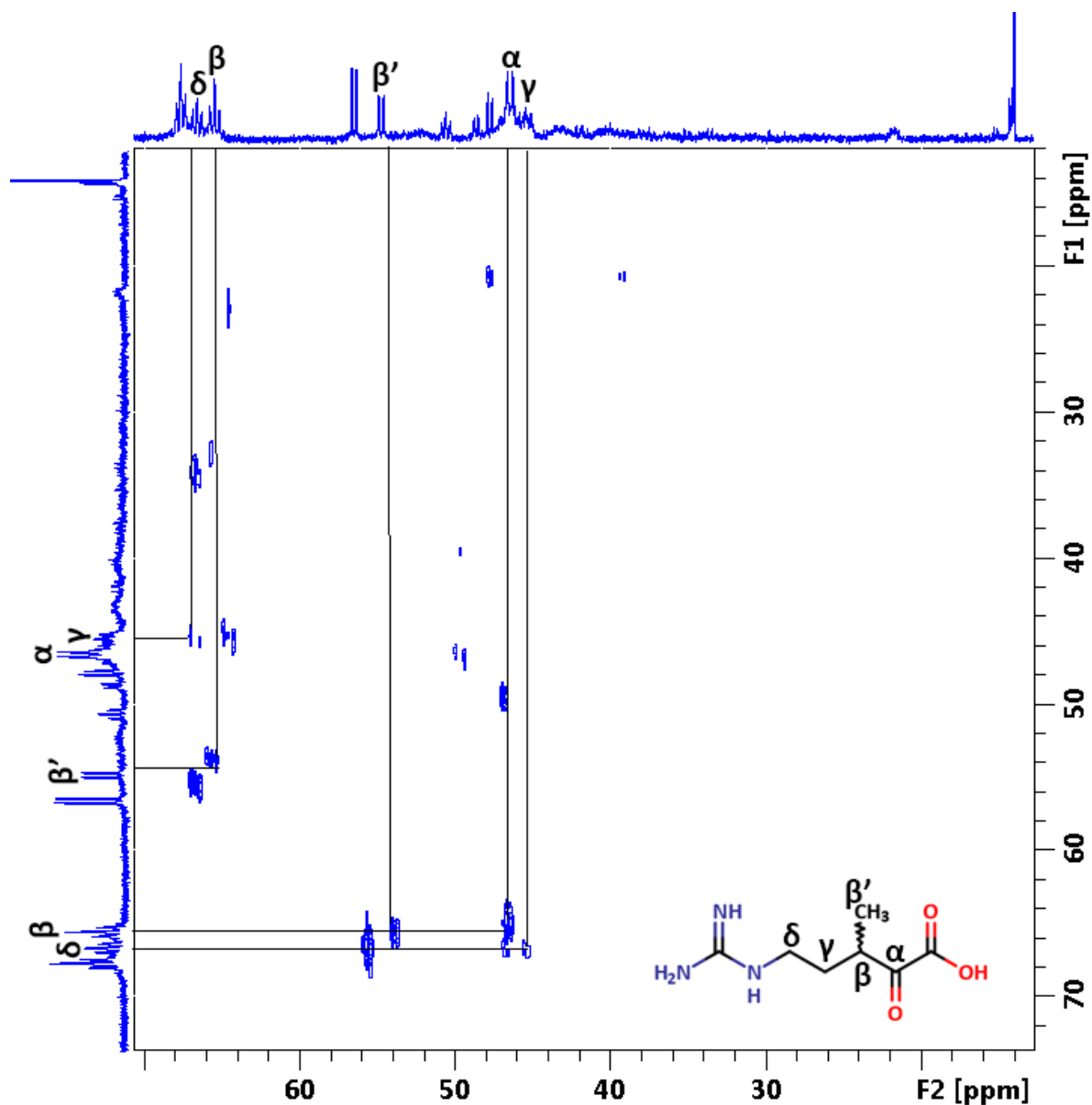


Figure 3.3.6-1: INADEQUATE (2D ^{13}C - ^{13}C coupling) spectra of SgrMT reaction products. A) ^{13}C -NMR of ^{13}C -isotope rich L-arg (green spectra) aligned with SgrMppP reaction product with the ^{13}C -isotope rich L-arg, followed by SgrMT reaction product (red spectra). B) 2D-INADEQUATE NMR correlation of SgrMT reaction product [136], C) 2D-INADEQUATE NMR correlation of SgrMT reaction products [137]. All the carbon of SgrMT products are labelled with ^{13}C -isotope.

3.3.7 SgrMT is not specific for any divalent metal

SgrMT requires both a divalent metal ion as well as SAM to catalyze its methylation reaction. The reaction was tested with Mg^{2+} , Fe^{2+} , Mn^{2+} , Ca^{2+} , and Zn^{2+} . Both LCMS and ^1H -NMR analyses show that the methylation occurs in presence of Mg^{2+} , Fe^{2+} and Mn^{2+} . In the reaction with Fe^{2+} , a small amount of ascorbate ($\sim 1 \mu\text{g}$) was used to keep the Fe^{2+} reduced^{cxci}. Figure 3.3.7-1 shows the ^1H -NMR of the SgrMT reaction in the presence of Fe^{2+} . The methyl hydrogen signal of SAM at 1.8 ppm goes down as it is consumed by SgrMT. Integration of the methyl proton singlet and the doublet of the SgrMT product, shows that both signals have approximately equal intensity.

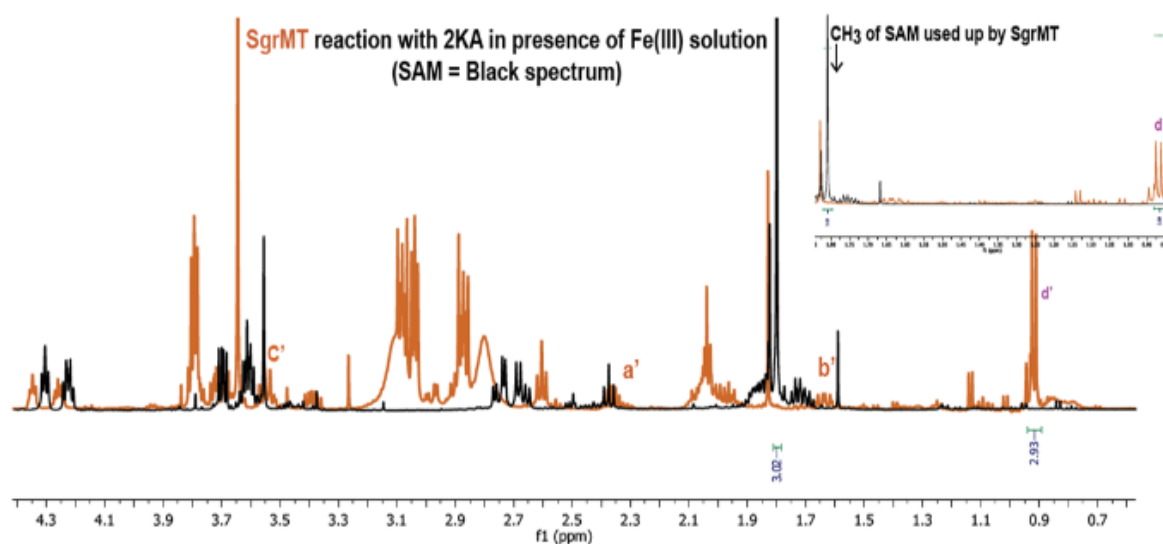
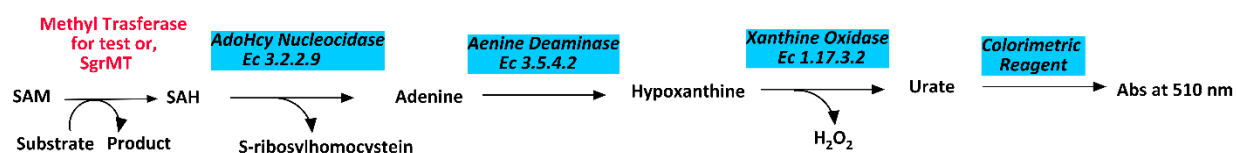


Figure 3.3.7-1: ^1H -NMR of SgrMT reaction product in presence of $\text{Fe}^{(III)}$ and Ascorbate. The reaction substrate was pure 2KA [91].

3.3.8 Steady State Kinetics of SgrMT with 2KA and 4HKA

The steady state kinetics of SgrMT were studied using a SAM-methyltransferase assay kit from G-Biosciences (SAM510: SA Methyltransferase Assay, Cat. # 786-430). The kit contains a proprietary enzyme mix that utilizes the SAM byproduct after methyl group donation, SAH, as a substrate for a series of coupled reactions. The reaction product along with the colorimetric reagent in the kit, gives a species with maximum absorbance at 510 nm (Scheme 3.3.8-1).



Scheme 3.3.8-1: reaction catalyzed by SAM-methyltransferase assay kit from G-Biosciences

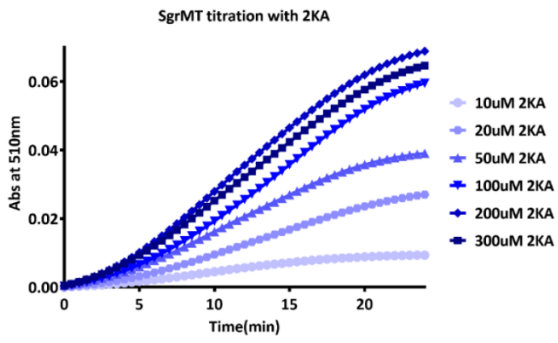
(SAM510: SA Methyltransferase Assay, Cat. # 786-430)

The rates of the SgrMT-catalyzed reaction were monitored separately for both **[91]** and **[95]** as substrate. Purified **[91]** was available for these experiments^{cxci}. We did not have access to pure **[95]**, so kinetics experiments used the reaction products of SgrMppP (approximately 120 μ M SgrMppP was reacted with 2mM L-arg, resulting in a mixture of **[95]**, **[91]**, and trace amounts of unreacted L-Arg). The ratio of **[95]** to **[91]** in the reaction product was estimated using ¹H-NMR spectroscopy. The reaction product contained 8% contamination of the minor product **[91]**; **[95]**

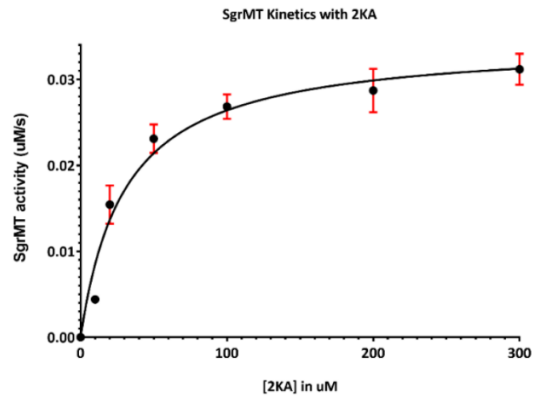
was the major product (~92%). The enzyme was removed and the reaction product was evaporated to dryness and reconstituted in 10 mM TRIS pH 8.0.

The reaction of SgrMT with **[91]** (Figure 3.3.8-1A) was slower, whereas reaction of SgrMT with **[95]** was approximately 4-fold faster (Figure 3.3.8-1A). The kinetics with **[91]** reached a plateau by 20 minutes; in contrast, the reaction with **[95]** reached a plateau by 5-6 minutes. Table 3.3.8-1 represents comparison of Michaelis–Menten kinetic fit of SgrMT with 2KA and 4HKA. The K_{cat} value of 2KA reaction is 30 μM , whereas the K_{cat} value for 4HKA is 58 μM . Moreover, the turnover number of 2KA reaction is 0.03 s^{-1} and for 4HKA reaction, it is 2.3 s^{-1} . From the value of turnover numbers, reaction with 4HKA seems approximately 77 folds faster than 2KA. In addition to that, the specificity constant for 4HKA ($4 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$) is also higher than that of 2KA ($1.1 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$). Thus SgrMT is more specific to 4HKA than 2KA as a substrate.

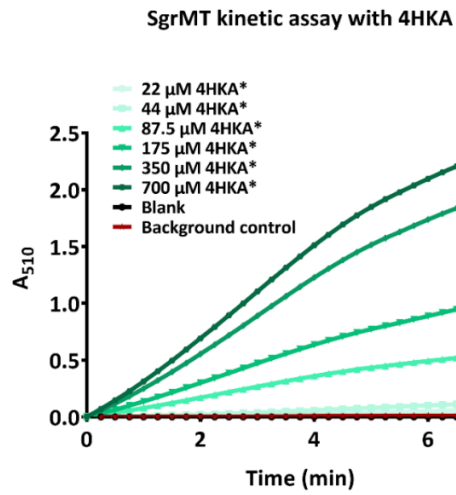
A)



C)



B)



D)

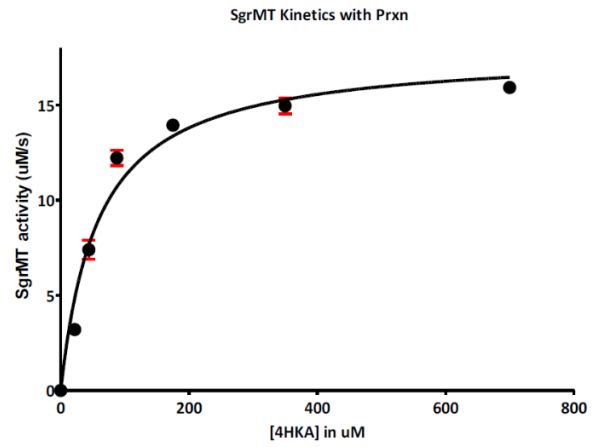


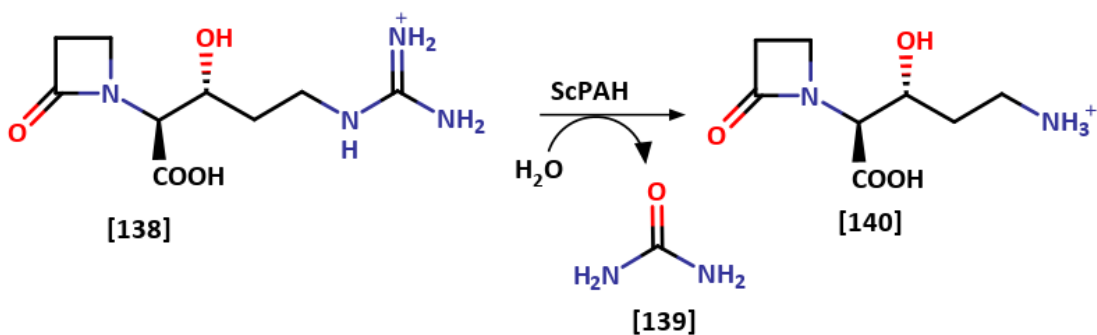
Figure 3.3.8-1: Steady state kinetic data of SgrMT. A) absorbance at 510nm vs time plotted for different concentration of 2KA substrate with SgrMT, B) absorbance at 510nm vs time plotted for different concentration of 4HKA substrate with SgrMT. The asterisk refers to the presence of 8% 2KA in the sample. C) Michaelis–Menten plot of SgrMT with 2KA, D) Michaelis–Menten plot of SgrMT with 4HKA*.

Table 3.3.8-1: comparison of Michaelis–Menten kinetic fit of SgrMT with 2KA and 4HKA

SgrMT Kinetics	2KA as Substrate	4HKA* as substrate
Michaelis-Menten constant, K_m (μM)	30.4 ± 3.2	$58.4^* \pm 7.3$
Turnover number, k_{cat} (S^{-1})	0.03 ± 0.002	$2.3^* \pm 0.3$
Specificity Constant, k_{cat}/K_m ($\text{M}^{-1}\text{s}^{-1}$)	$(1.1 \pm 0.01) \times 10^3$	$(4.0 \pm 0.3) \times 10^4$

3.3.9 SgrAH is a structural homolog of ScPAH

The next gene product adjacent to SgrMT is an amidinohydrolase homolog, SgrAH. It shares 51% sequence identity with another amidinohydrolase homolog from the biosynthesis of clavulanic acid in *Streptomyces clavuligerus*, procalvaminic acid hydrolase (ScPAH, PDB ID: 1GQ6). ScPAH is a Mg^{2+} -dependent enzyme. The function of ScPAH is to hydrolyze procalvaminic acid [138] to calvaminic acid [140], producing urea [139] in the process (scheme 3.3.9-1).



Scheme 3.3.9-1: reaction catalyzed by ScPAH

The superimposed structures of SgrAH and ScPAH (Figure 3.3.9-1) shows that the active sites are rich in acidic residues. Most of the other active site residues are retained in SgrAH homolog, except for Arg63 in SgrAH is substituted with His64 in ScPAH. Likewise, Glu249 of SgrAH is substituted with Ala151 in ScPAH. These differences likely reflect differences in the substrates of these enzymes. SgrAH retains the metal-binding triad of ScPAH, comprised of His159, His120 and Asp147 (SgrAH numbering). We could see Mn²⁺ bound in the metal-binding site. We made a search on checkmyblob^{cxciiv} using electron density, which shows that it could also be Mg or Co. We assume the possibility to be Mn is more because Mn was in the crystallization condition, as well as in the SgrAH storage buffer.

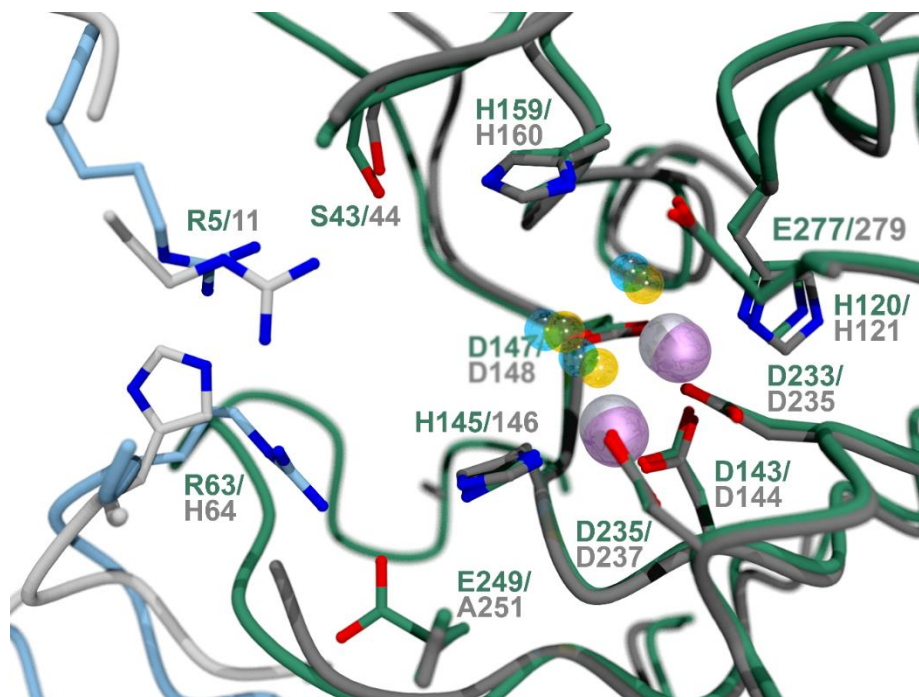
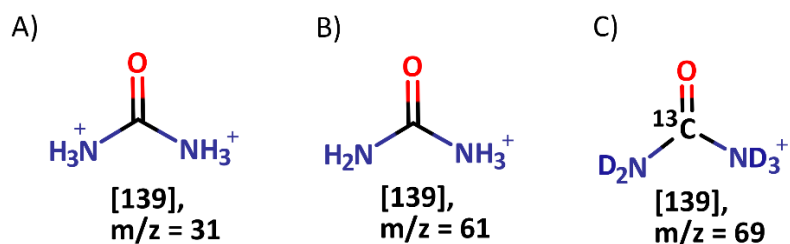


Figure 3.3.9-1: SgrAH (green chain) overlaid with ScPAH (gray chain, (PDB ID: 1GQ6) with RMSD value of 0.97 Å. The structure of SgrAH was determined at 2.1 Å resolution, R_{cryst} : 0.142, R_{free} : 0.184. The Light blue chain is the second chain of the SgrAH dimer. Similarly the light gray chain is the second chain of ScPAH dimer. The pink spheres are the Mn^{2+} ions associated with SgrAH, whereas the silver spheres are the Mn^{2+} ions of ScPAH.

3.3.10 SgrAH Reacts with the product(s) of SgrMT

Our hypothetical reaction sequence predicts that SgrAH might hydrolyze the product(s) of SgrMT. If this is the case, by analogy to ScPAH, the hydrolysis of the guanidinium group should produce urea as a product. We used LC-MS with a Shimadzu 8040 triple quadrupole instrument to look for the urea. We compared the mass spectra from reactions with SgrAH with published mass spectra for urea (METLIN ID: 6). One drawback of the Shimadzu 8040 instrument was that its lower resolution limit is $m/z = 50$. The di-cationic species of urea has $m/z = 31$. However, published data (METLIN ID = 6) also showed a significant signal for the mono-cationic form with $m/z = 61$. To increase the mass of urea as much as possible, we used the ^{13}C - and ^2D -enriched sample from the SgrMT reaction (see Section 3.3.6). The mono-cationic form of the labelled urea would have $m/z = 69$ (Scheme 3.3.10-1). We analyzed the species by extracted ion chromatogram of the SgrAH reaction sample.



Scheme 3.3.10-1: different form of urea in positive mode with their corresponding m/z (Ref: METLIN ID 6). A) di-cationic form of naturally abundant urea has m/z = 31, B) mono cationic form of urea has m/z = 61, C) mono cationic form of urea with ^{13}C and ^2D isotope labelled, has m/z = 69.

The SgrAH reaction products were analyzed using LC-MS, where the samples were ionized using ESI ionization method and separating the ions using ZIC[®] HILIC column. The Extracted ion chromatogram (in positive mode) from LCMS analysis, is presented in Figure 3.3.10-1. The expected mono-cationic form of isotope labelled urea is evident in the positive mode spectrum with m/z = 69. Apart from that, two more base peak were observed having m/z of 180 and 164. The ion at m/z = 180 is possibly the isotopically labeled product **[141]** of SgrAH hypothetical reaction (Scheme 3.1-1). The ion at m/z = 164 could be the isotope labeled product **[142]** (Scheme 3.3.10-2). For predicted compound **t** and **[142]**, isotopic effect was observed in mass spectrometry. Thus for **[141]**, m/z of 177,178,179 was present. Similarly for **[142]**, m/z of 161,162,163 was present. For clear presentation, we are showing the extracted ion

chromatogram (EIC) of m/z 69, 180 and 164. These ions ($m/z = 69, 180, \text{ and } 164$) were not observed in the negative control, which was an identical reaction except that the SgrAH had been heat-denatured by boiling for 10 min. Thus, it appears that SgrAH hydrolyzes methylated [95] (i.e. 2-oxo-3-methyl-4(S)-hydroxy-5-guanidinovaleric acid [136]) to give 2-oxo-3-methyl-4-hydroxy ornithine [141]. SgrAH can also hydrolyze the minor product of SgrMT derived from [91], 2-oxo-3-methyl-5-guanidinovaleric acid [137] to yield 2-oxo-3-methylornithine [142]. This finding suggests that SgrAH is acting on the SgrMT products. The predicted reaction of SgrAH presented in Scheme 3.3.10-2 appears to be confirmed. The major peaks observed in the SgrAH reaction LC-MS chromatogram are listed in Table 3.3.10-1.

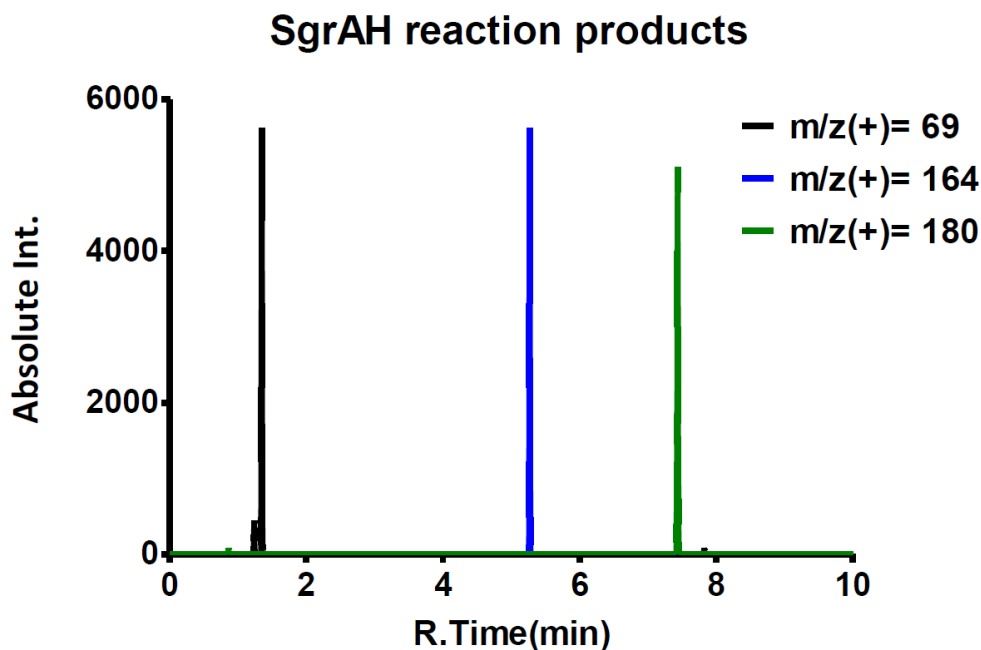
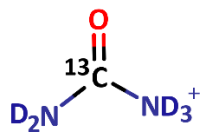
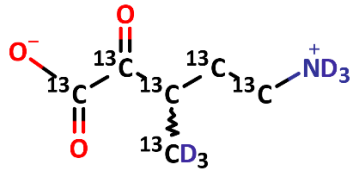
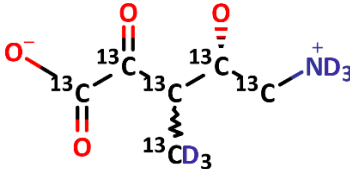
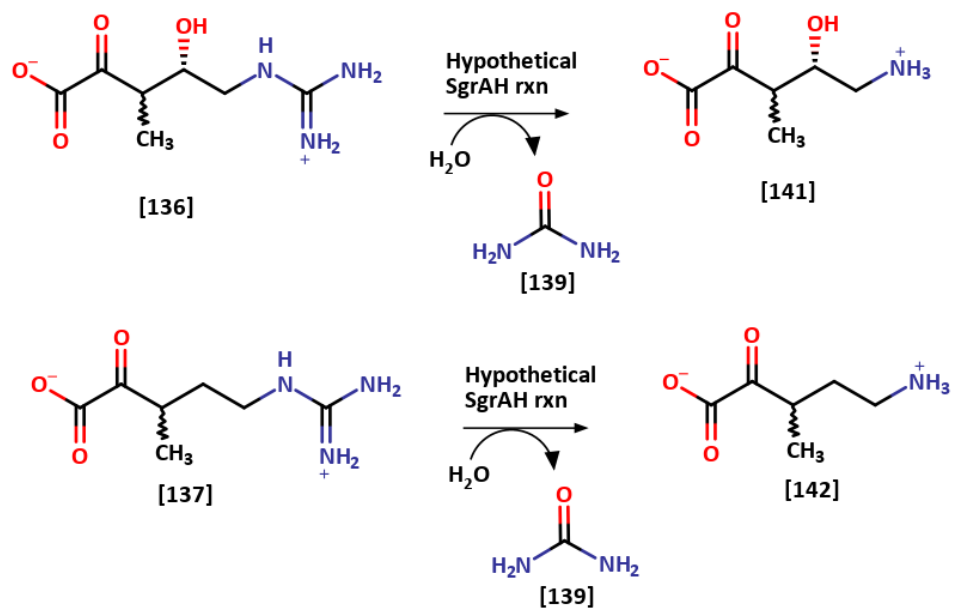


Figure 3.3.10-1: Extracted ion chromatogram of SgrAH reaction with ^{13}C - and ^2D -labelled SgrMT products.

Table 3.3.10-1: Prediction of structure from ^{13}C - and ^2D -labelled products from the SgrAH reaction.

m/z	Ret time (min)	Structure prediction	Chemical name	Annotation in scheme 3.3.9-1
69	1.24		Urea	[139]
164	5.31		2-oxo-3-methyl- 4(S)-hydroxy ornithine	[141]
180	7.42		2-oxo-3-methyl ornithine	[142]



Scheme 3.3.10-2: Predicted reaction catalyzed by SgrAH.

4 Conclusion

MppP is a PLP-dependent L-Arg oxidase. In *S. wadayamensis*, MppP is involved in L-End biosynthesis. The bioinformatics analysis of MppP homologs is compelling because it seems that this enzyme activity is present not only in the context of L-End biosynthesis, but also in gene clusters for unknown natural products that do not include L-End. The gene context of Pbr and Sgr MppP homolog contains totally different genes comparing to our prototypical SwMppP homolog. None of these gene clusters contain genes for MppQ or MppR homologs, so clearly, MppP is employed in biosynthetic contexts beyond L-enduracididine biosynthesis. We don't have much clue about rest of the enzymes in PbrMPCO. We could reveal some biochemical function of SgrMT and SgrAH from Sgr gene context. We have previously shown that SwMppP is involved in L-end biosynthesis, and another MppP homolog was also studied by Ryan lab, where they showed the involvement of MppP homolog in Indolmycin biosynthesis. They also recently showed that similar to PbrMPCO, gene cluster present on *Streptomyces eurocidicus* is involved in azomycin biosynthesis^{CXCV}. So, the presence of MppP homolog in these two gene contexts are significant. Depending on the bioinformatic analysis, we are rooting for antibiotic, we still need some more work to conclude the final product of these pathway.

The MppP-containing gene cluster in *P. brassicacaerum* includes four additional gene products. PbrMppP behaves exactly as SwMppP, both in terms of the mix of products produced as well as the kinetics. None of the remaining gene products are homologous to MppQ or MppR. The gene we call PbrHYP is a protein of unknown function with the PIN domain fold. In spite of a minuscule

sequence identity (<10 %) with the closest structural homolog VapC, the PIN domain fold of this bacterial toxin is conserved in PbrHYP. The substitution of two active site residues (D116, D96) results in disruption of metal-binding capacity, and thus the ribonuclease activity of the toxin VapC is not conserved in PbrHYP. On the contrary, PbrDHPS is a type-I aldolase homolog. The retention of the catalytic lysine as well as its orientation with **[91]** bound in the PbrDHPS active site, suggest that PbrDHPS might retain aldolase activity. The substitution of Asp187 with tryptophan might result in catalytic divergence. It is possible that the substitution of crucial Asp187 in PbrDHPS is adopted to accommodate a specific substrate.

The PbrOX is a putative mononuclear Fe^{II}-dependent oxygenase homolog. In the absence of a terminal electron acceptor or PbrOX substrate, it is hard to decipher an actual reaction cycle. Although the involvement of PbrFD as an intermediate electron transporter suggests that all three enzymes are presumably part of three-component oxygenase system in *P. brassicacearum*.

The work that remains to be done would be cloning the deletion plasmid into donor cell S17-1^{cxvii}, followed by mating of the donor cell with DF41 to produce a strain where the PbrMPCO is knocked out. Continuing towards the end of the project, comparative metabolomic analysis of these two strains, using LC-MS/MS would lead to identification of the final product of the PbrMPCO. The metabolomic analysis will also provide information about biochemical function of the rest of the PbrMPCO enzymes.

Like the PbrMPCO, the MppP-containing gene cluster from *S. griseofuscus* includes 5 additional open reading frames, but no *mppQ* or *mppR* homologs are present in the genome. Enzymatically, SgrMppP is similar to both SwMppP and PbrMppP. For the Sgr project, along with the structure

and function of SgrMppP, we also identified that SgrMT is a metal- and SAM-dependent methyltransferase that catalyzes the addition of a methyl group to the β -carbon of the SgrMppP products (i.e. 2-oxo-4(S)-hydroxy-5-guanidinovaleric acid and 2-oxo-5-guanidinovaleric acid). This is a development towards the final product identification of the SgrMPCO. We have also shown that SgrAH hydrolyzes the guanidinium group of the SgrMT reaction products to give 2-oxo-3-methyl-4(S)-hydroxyornithine and 2-oxo-3-methylornithine and urea. Our next goal is to identify the functions of the remaining enzymes in this pathway, one of which much use the 2-oxo-3-methyl-4(S)-hydroxyornithine as a substrate.

5 Reference

- ⁱ Daniel A. Jacobo-Velázquez^{1*}, M. G. I.-A. e. L. C.-Z. (2015). "Cross-talk between signaling pathways: The link between plant secondary metabolite production and wounding stress response." Science Reports **5**: 8608.
- ⁱⁱ Isah, T. (2016). "Anticancer Alkaloids from Trees: Development into Drugs." Pharmacognosy Reviews **10**(20): 90-99.
- ⁱⁱⁱ Nicholas Ekow Thomford ¹, † ID, Dimakatso Alice Senthebane ^{3,4}, Arielle Rowe ³, et al. (2018). "Natural Products for Drug Discovery in the 21st Century: Innovations for Novel Drug Discovery." Molecular Sciences **19**: 1578.
- ^{iv} David J. Newman^{*}, a. G. M. C. (2016). "Natural Products as Sources of New Drugs from 1981 to 2014." Journal of Natural Products **76**: 629-651.
- ^v Shi-Lin Chen^{1*}, H. Y., ^{3*}, Hong-Mei Luo², Qiong Wu^{2,4}, Chun-Fang Li² and André Steinmetz⁵ (2016). "Conservation and sustainable use of medicinal plants: problems, progress, and prospects." Chinese medicine **11**(37).
- ^{vi} Luch, A. (2009). "Molecular, clinical and environmental toxicology." Springer: 20.
- ^{vii} Kutchan, K. S. a. T. M. (2009). "Introduction to the Different Classes of Natural Products." Plant-derived Natural Products: Synthesis, Function, and Application.
- ^{viii} Liangcheng Du, C. S., Mei Chen, Daniel J Edwards, Ben Shen (2000). "The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase." Cell Chemical Biology **7**(8).
- ^{ix} Neal L Benowitz, P. J. (1990). "Nicotine metabolism in nonsmokers." Clinical Pharmacology & Therapeutics **48**(4): 473.
- ^x Langer, A. S. S. a. P. "Non-Proteinogenic Amino Acids, Natural Origin, Biological Functions." Asymmetric Synthesis of Non-Proteinogenic Amino Acids, First Edition.: 8.
- ^{xi} Margaret F. Roberts¹, D. S., and Michael Wink³ (2010). "BIOSYNTHESIS OF ALKALOIDS AND BETALAINS." Annual Plant Reviews **40**: 20-91.

^{xii} Sun-Young Kang¹, Oksik Choi^{1,3†}, Jae Kyung Lee^{1,2}, Bang Yeon Hwang², Tai-Boong Uhm³ and Young-Soo Hong^{1*} (2012). "Artificial biosynthesis of phenylpropanoic acids in a tyrosine overproducing *Escherichia coli* strain." *Microbial Cell Factories* **11**(53).

^{xiii} KEISHI OSAKABE^{*†‡}, C. C. T., LAIGENG LI^{*}, JACQUELINE L. POPKO^{*}, TOSHIAKI UMEZAWA^{*}, and R. H. S. DANIEL T. CARRAWAY[¶], CHANDRASHEKHAR P. JOSHI^{*}, AND VINCENT L. CHIANG^{*¶} (1999). "Coniferyl Aldehyde 5Hydroxylation and Methylation Direct Syringyl Lignin Biosynthesis in *Angiosperms*." *Proceedings of the National Academy of Sciences* **99**: 8955–8960.

^{xiv} Mitjans, M. P. V. a. M. (2017). "Lignins and Their Derivatives with Beneficial Effectson Human Health." *Interntional Journal of Molecular Sciences* **18**(1219).

^{xv} Carmen Rodríguez-García, 2 Cristina Sánchez-Quesada,^{1,2,3} Estefanía Toledo,^{4,5,6} Miguel Delgado-Rodríguez,^{1,2,7} and José J. Gaforio^{1,2,3,7,*} (2019). "Naturally Lignan-Rich Foods: A Dietary Tool for Health Promotion?" *Molecules* **24**(5).

^{xvi} Carla Ghelardini¹ , N. G., Gabriela Mazzanti^{2,*} (2001). "Local Anaesthetic Activity of Monoterpenes and Phenylpropanes of Essential Oils." *Planta Medica* **67**(6).

^{xvii} Jih-Jung Chen, W.-J. L., Chang-Hui Liao, and Po-Chuen Shieh (2007). "Anti-inflammatory Benzenoids from *Antrodia camphorata*." *Journal of Natural Products* **70**(6): 989-992.

^{xviii} L.K.A.MLealb, A.A.GFerreiraa, G.ABezerraa F.J.AMatosc, G.S.BVianaa (2000). "Antinociceptive, anti-inflammatory and bronchodilator activities of Brazilian medicinal plants containing coumarin: a comparative study." *Journal of Ethnopharmacology* **70**(2): 151-159.

^{xix} Blaine A. Pfeifer, C. C. C. W., Christopher T. Walsh, Chaitan Khosla (2003). "Biosynthesis of Yersiniabactin, a Complex Polyketide-Nonribosomal Peptide, Using *Escherichia coli* as a Heterologous Host." *Applied and Environmental Microbiology* **11**: 6698-6702.

^{xx} Chandra Risdian, 2,* Tjandrawati Mozef,³ and Joachim Wink^{1,*} (2019). "Biosynthesis of Polyketides in *Streptomyces*." *Microorganisms* **7**: 124.

^{xxi} Dewick, P. M. (2002). "Medicinal Natural Products." *John Wiley and sons* **2nd edition** 167.

^{xxii} Jullyana de S S Quintans 1 , B. M. S., Rosana P C Ferraz, Allan C A Oliveira, Thanany B da Silva, Leociley R A Menezes, Marília F C Sampaio, Ana Paula do N Prata, Manoel O Moraes, Claudia

Pessoa, Angelo R Antonioli, Emmanoel V Costa, Daniel P Bezerra (2013). "Chemical constituents and anticancer effects of the essential oil from leaves of *Xylopi* *laevigata* " Planta Med. **79**(2).

xxiii 1 Nor Hayati Abdullah, A. H. A. H., 1 and Yew Su Fong³ (2012). "Cardiovascular Activity of Labdane Diterpenes from *Andrographis paniculata* in Isolated Rat Hearts." *Natural Products for Medicine*.

xxiv Tholl, D. (2015). "Biosynthesis and Biological Functions of Terpenoids in Plants." Adv Biochem Eng Biotechnol **148**: 63-106.

xxv Fumitaka Kudo^a, A. M. a. T. E. ^b. (2014). "Biosynthesis of natural products containing β -amino acids " Natural Product Reports(8).

xxvi Zi-ming Feng, Z.-l. Z., Ya-nan Yang, Jian-shuang Jiang & Pei-cheng Zhang (2016). "Naturally occurring hybrids derived from γ -amino acids and sugars with potential tail to tail ether-bonds." Nature, Scientific Reports.

xxvii Hubert Lam, * Dong-Chan Oh,^{2,*†} Felipe Cava,^{1,*} Constantin N. Takacs,^{1,‡} Jon Clardy,² Miguel A. de Pedro,³ and Matthew K. Waldor^{1,§} (2010). "D-amino Acids Govern Stationary Phase Cell Wall Re-Modeling in Bacteria." HHS Public access **325**(5947).

xxviii Salzman, L. A., Weissbach, H., & Katz, E. (1964). "STUDIES ON THE MECHANISM OF SYNTHESIS OF D-VALINE BY *STREPTOMYCES ANTIBIOTICUS* " J. Biol. Chem **239**: 1864-1866.

xxix JONES, G. H. (1987). "Actinomycin Synthesis in *Streptomyces antibioticus*: Enzymatic Conversion of 3-Hydroxyanthranilic Acid to 4-Methyl-3-Hydroxyanthranilic Acid." Journal of bacteriology **169**(12): 5575-5578.

xxx Keller', A. S. a. U. (1994). "Epimerization of the D-Valine Portion in the Biosynthesis of Actinomycin D." Biochemistry **33**.

xxxi Theo H M Smits, F. R., Tim Kamber, Jochen Blom (2011). "Metabolic Versatility and Antibacterial Metabolite Biosynthesis Are Distinguishing Genomic Features of the Fire Blight Antagonist *Pantoea vagans* C9-1." PLOS ONE **6**(7).

xxxii VOLKER MAGNUS, S. S., SONJA ISKRIC, AND SERGIJE KVEDER (1982). "Metabolism of Tryptophan, Indole-3-acetic Acid, and Related Compounds in Parasitic Plants from the Genus *Orobanche*." Plant Physiol. **69**: 853-858.

xxxiii Dubinina, D. P. K. A. P. W. S. L. J. A. N. P. V. M. G. G. A. (1994). "Biological production and consumption of gaseous organic sulphur compounds " Biochemical Society transactions **22**(4): 1011-1

xxxiv J. Wróbel, J. (1976). Low molecular weight-Sulfur containing natural products: 143.

xxxv Jasmin Dischinger, Imke Wiedemann, Gabriele Bierbaum, Hans-Georg Sahl (2013). "Chapter 19- Lantibiotics." Handbook of Biologically Active Peptides (Second Edition): 119-128.

xxxvi Olivia McAuliffe a;b;1, R. P. R. c., Colin Hill a;b;* (2001). "Lantibiotics: structure, biosynthesis and mode of action." FEMS Microbiology Reviews **25**: 285-308.

xxxvii L J Ward 1 , J. C. B., G P Davey (1994). "Application of the Ligase Chain Reaction to the Detection of nisinA and nisinZ Genes in Lactococcus Lactis Ssp. Lactis " FEMS Microbiol Lett. **117**(1).

xxxviii Rick Rink, J. W., 1 Anneke Kuipers,1 Leon D. Kluskens,1 Arnold J. M. Driessen,2 Oscar P. Kuipers,3 and Gert N. Moll1,* (2007). "Production of Dehydroamino Acid-Containing Peptides by Lactococcus lactis ▽ ." Appl Environ Microbiol. 73(6).

xxxix W. C. CHAN, H. M. D., 2 N. HORN,2 K. MACLEAN,1 L.-Y. LIAN,3 B. W. BYCROFT,1 M. J. GASSON,2 AND G. C. K. ROBERTS3* (1996). "Structure-Activity Relationships in the Peptide Antibiotic Nisin: Role of Dehydroalanine 5." Applied and Environmental Microbiology **62**: 2966–2969.

xl Hillman§, L. S. a. J. D. (2008). "Therapeutic Potential of Type A (I) Lantibiotics, a Group of Cationic Peptide Antibiotics." HHS Public access **11**(5): 401-408.

xli Mohammad R. Islam*, J.-i. N., Takeshi Zendo* and Kenji Sonomoto*†1 (2012). "Antimicrobial mechanism of lantibiotics." Biochemical Society transactions **40**(6).

xlii Wolfgang M. M_ller, T. S., Paul Enslin, and Roderich D. S_ssmuth* (2010). "In Vitro Biosynthesis of the Prepeptide of Type-III Lantibiotic Labyrinthopeptin A2 Including Formation of a C_C Bond as a Post-Translational Modification**." Angew. Chem. Int. Ed. **49**: 2436 –2440.

xliii Sussmuth, G. M. S. a. R. D. (2011). "Synthetic studies toward labionin, a new α,α -disubstituted amino acid from type III lantibiotic labyrinthopeptin A2." J. Pept. Sci. **17**: 581–584.

xliv Geoffrey Férrir, M. I. P., Graciela Andrei, Dana Huskens, Bart Hoorelbeke, Robert Snoeck, Jos

Vanderleyden, Jan Balzarini, Stefan Bartoschek, Mark Brönstrup, Roderich D. Süßmuth, Dominique Schols (2013). "The Lantibiotic Peptide Labyrinthopeptin A1 Demonstrates Broad Anti-HIV and Anti-HSV Activity with Potential for Microbicidal Applications." *PLOS ONE* 8(5).

^{xlv} Routb, J. D. A. a. M. P. (2000). "The Road to Ribosomes, Filling Potholes in the Export Pathway." *J Cell Biol.* 151(5): 23–26.

^{xlvi} Yun Ding, J. P. T., Jinsha Liu, Shams Al-Azzam, Priyanka Pandya & Sepideh Afshar (2020). "Impact of non-proteinogenic amino acids in the discovery and development of peptide therapeutics." *Amino Acids* 52: 1207-1226.

^{xlvii} Wang H, F. D., Holm L, Rouhiainen L, Sivonen K (2014). "Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes." *Proc Natl Acad Sci* **111**(25): 9259–9264.

^{xlviii} Darcy J Atkinson, 2 Briar J Naysmith,^{1,2} Daniel P Furkert,^{1,2} and Margaret A Brimblecorresponding author^{1,2} (2016). "Enduracididine, a rare amino acid component of peptide antibiotics: Natural products and synthesis." *BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY* **12**: 2325-2342.

^{xlix} Haiyin He, † R. Thomas Williamson,† Bo Shen,† Edmund I. Graziani,† Hui Y. Yang,† Subas M. Sakya,‡ Pete J. Petersen,‡ and Guy T. Carter† (2002). "Mannopeptimycins, Novel Antibacterial Glycopeptides from *Streptomyces hygroscopicus*, LL-AC98." *JACS* **124**: 9729-9736.

^l Levine, D. P. (2006). "Vancomycin: a history." *Clinical Infectious Diseases, Volume 42* **42**: S5-S12.

^{li} Li-Yan Yin, J. H. C., 1, * Jacob K. Thomas,¹ Stuart Shapiro,² and Anne Schmitt-Hoffmann² (2008). "Efficacies of Ceftobiprole Medocaril and Comparators in a Rabbit Model of Osteomyelitis Due to Methicillin-Resistant *Staphylococcus aureus*." *Antimicrobial Agen and Chemithery* **52**(5): 1618-1622.

^{lii} M. P. Singh, P. J. P., ² W. J. Weiss,² J. E. Janso,¹ S. W. Luckman,¹ E. B. Lenoy,² P. A. Bradford,² R. T. Testa,^{1,2} and M. Greenstein¹ (2003). "Mannopeptimycins, New Cyclic Glycopeptide Antibiotics Produced by *Streptomyces hygroscopicus* LL-AC98: Antibacterial and Mechanistic Activities." *Antimicrobial Agen and Chemithery* **47**(1): 62-69.

^{liii} Nathan A. Magarvey, B. H., Min He, Michael Greenstein, and John A. Hucul* (2006). "Biosynthetic Pathway for Mannopeptimycins, Lipoglycopeptide Antibiotics Active against Drug-Resistant Gram-Positive Pathogens." *Antimicrobial Agen and Chemithery* **50**(6): 2167-2177.

^{liv} Alexey Ruzin, G. S., Anatoly Severin, Youjun Yang,[†] Russell G. Dushin, Alan G. Sutherland, Albert Minnick, Michael Greenstein, Michael K. May, David M. Shlaes,[‡] and Patricia A. Bradford (2004). "Mechanism of Action of the Mannopeptimycins, a Novel Class of Glycopeptide Antibiotics Active against Vancomycin-Resistant Gram-Positive Bacteria." Antimicrobial Agen and Chemotherapy **48**(3): 728-738.

^{lv} Watanakunakorn, C. (1984). "Mode of action and in-vitro activity of vancomycin " J Antimicrob Chemother(14): 7-18.

^{lvi} He, H. (2005). "Mannopeptimycins, a novel class of glycopeptide antibiotics active against gram-positive bacteria." Appl Microbiol Biotechnol **67**: 444–452.

^{lvii} Alessandro Paiardini, R. C., 1 Ashley M. Buckle,² and Barbara Cellini³ (2014). "PLP-Dependent Enzymes." BioMed Research International.

^{lviii} Kirsch, A. C. E. a. J. F. (2004). "Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations." Annual Review of Biochemistry **73**: 383-415.

^{lix} Silverman, R. B. (2002). "The organic chemistry of enzyme-catalyzed reactions (Rev. ed.)."

^{lx} Lanlan Han, A. W. S., Graham R. Moran, and Nicholas R. Silvaggi* (2015). "Streptomyces wadayamensis MppP Is a Pyridoxal 5'-Phosphate-Dependent L-arginine α -Deaminase, γ -Hydroxylase in the Enduracididine Biosynthetic Pathway." Biochemistry **54**(47): 7029–7040.

^{lxi} Vanderlinde, R. E. (1986). "Review of pyridoxal phosphate and the transaminases in liver disease " Ann Clin Lab Sci **16**(2): 79-93.

^{lxii} Bertoldi, M., Cellini, B., Montioli, R., and Borri Voltattorni, C. (2008). "Insights into the mechanism of oxidative deamination catalyzed by DOPA decarboxylase." Biochem **47**: 7187-7195.

^{lxiii} Riccardo Montioli, B. C., Mirco Dindo, Elisa Oppici, and Carla Borri Voltattorni* (2013). "Interaction of Human Dopa Decarboxylase with L-Dopa: Spectroscopic and Kinetic Studies as a Function of pH." Biomed Res Int.: 161456.

^{lxiv} Yamada T, K. J., Takata Y, Ogawa H, Pitot HC, Takusagawa F (2003). "Crystal structure of serine dehydratase from rat liver." Biochemistry **42**(44): 12854-12865.

lxv TohruYoshimura¹, N. (2003). "Amino Acid Racemases: Functions and Mechanisms." JOURNAL OF BIOSCIENCE AND BIOENGINEERING **96**(2): 103-109.

lxvi Lanlan Han, N. V., Sarah A. Oehm, Tyler G. Fenske, Alan W. Schwabacher, and a. N. R. Silvaggi* (2018). "Streptomyces wadayamensis MppP is a PLP-Dependent Oxidase, Not an Oxygenase." Biochemistry **57**(23): 3252-3264

lxvii Elesha R. Hoffarth, K. W. R. a. K. S. R. (2020). "Emergence of oxygen- and pyridoxal phosphate-dependent reactions." The FEBS Journal(287): 1403–1428.

lxviii Fassler J, C. P. (2011). "BLAST Glossary." Bethesda (MD): National Center for Biotechnology Information (US).

lix Sudhir Kumar, G. S., Michael Li, Christina Knyaz, and Koichiro Tamura (2018). "MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms." Molecular Biology and Evolution(25): 1547-1549.

lxx Ryan, J. B. H. P. D. K. S. (2019). "In vitro Reconstitution of the Biosynthetic Pathway to the Nitroimidazole Antibiotic Azomycin." Angew. Chem. Int. Ed.(58): 1-6.

lxxi Yi-Ling Du, R. S., Lona M Alkhalaf, Eugene Kuatsjah, Hai-Yan He, Lindsay D Eltis & Katherine S Ryan (2016). "A pyridoxal phosphate–dependent enzyme that oxidizes an unactivated carbon-carbon bond." Nature Chemical Biology **12**: 194-199.

lxxii Shanteri Singh, J. G. M., Changsheng Zhang,§ Craig A. Bingman, George N. Phillips, Jr., and Jon S. Thorson (2008). "Structure and Mechanism of the Rebecamycin Sugar 4'-O-Methyltransferase RebM." J Biol Chem. **283**(33): 22628–22636

lxxiii Meng-RuHo^a, K.-W., Wen-changLin^c (2012). "A unified framework of overlapping genes: Towards the origination and endogenic regulation." ELSEVIER **100**(4): 231-239.

lxxiv Kai Papenfort¹, a. B. B., 3,# (2016). "Quorum-Sensing Signal-Response Systems in Gram-Negative Bacteria." Nat Rev Microbiol. **14**(9): 576–588.

lxxv Jason B. Hedges, E. K., Yi-Ling Du, Lindsay D. Eltis, and Katherine S. Ryan* (2018). "Snapshots of the Catalytic Cycle of an O₂, Pyridoxal Phosphate-Dependent Hydroxylase." ACS Chem. Biol. **13**(4): 965–974.

lxxvi https://www.kegg.jp/kegg-bin/show_pathway?sen01057

^{lxxvii} Adrien Biessy¹, A. N., Jochen Blom², Geneviève Léger¹, Linda S. Thomashow³, Francisco M. Cazorla⁴, Dragana Josic⁵ and Martin Filion^{1*} (2019). "Diversity of phytobeneficial traits revealed by whole-genome analysis of worldwide-isolated phenazine-producing *Pseudomonas* spp." Environmental Microbiology **21**(1): 437-455.

^{lxxviii} Xie*, J. C. a. J. (2011). "Role and Regulation of Bacterial LuxR-Like Regulators." Journal of Cellular Biochemistry **112**: 2694–2702

^{lxxix} Stephen P. Diggle, K. W., 2 Andrée Lazdunski,³ Paul Williams,^{1,2} and Miguel Cámara^{1*} (2002). "Advancing the Quorum in *Pseudomonas aeruginosa*: MvaT and the Regulation of N-Acylhomoserine Lactone Production and Virulence Gene Expression†." JOURNAL OF BACTERIOLOGY **184**(10): 2576–2586.

^{lxxx} Lanlan Han ¹, N. V., Sarah A Oehm ¹, Tyler G Fenske ¹, Alan W Schwabacher ¹, Nicholas R Silvaggi ¹ (2018). "Streptomyces wadayamensis MppP is a PLP-Dependent Oxidase, Not an Oxygenase " Biochemistry **57**(23): 3252-3264.

^{lxxxi} Uddipan Das a, V. P. b., Udaya Kumar Tiruttani Subhramanyamc, Matthias Wilmannsb, Samudrala Gourinath d,†, Alagiri Srinivasan a,† (2014). "Crystal structure of the VapBC-15 complex from *Mycobacterium tuberculosis* reveals a two-metal ion dependent PIN-domain ribonuclease and a variable mode of toxin–antitoxin assemblyq " ELSEVIER **188**: 249-258.

^{lxxxii} Christian Mirwaldt*, I. K. r. a. R. H. (1995). "The Crystal Structure of Dihydrodipicolinate Synthase from *Escherichia coli* at 2.5 Å Resolution." Journal of Molecular Biology **246**: 227-239.

^{lxxxiii} Wupeng Yan^{#1}, H. S., Fuhang Song³, Yisong Guo⁶, Cheng-Hsuan Wu², Ampon Sae Her², Yi Pu^{2,4}, Shu Wang², Nathchar Naowarojna², Andrew Weitz⁶, Michael P. Hendrich⁶, Catherine E. Costello^{2,4}, Lixin Zhang³, Pinghua Liu², and Yan Jessie Zhang^{1,5} (2015). "Endoperoxide formation by an α -ketoglutarate-dependent mononuclear non-haem iron enzyme." PMC **527**(7579): 539–543.

^{lxxxiv} N. V. Zharikova*, T. R. I., E. I. Zhurenko, V. V. Korobov, and T. V. Markusheva (2018). "Bacterial Genes of Non-Heme Iron Oxygenases, Which Have a Rieske-Type Cluster, Catalyzing Initial Stages of Degradation of Chlorophenoxyacetic Acids." Russian Journal of Genetics **54**(3): 284-295.

^{lxxxv} Uddipan Das a, V. P. b., Udaya Kumar Tiruttani Subhramanyamc, Matthias Wilmannsb, Samudrala Gourinath d,†, Alagiri Srinivasan a,† (2014). "Crystal structure of the VapBC-15 complex from *Mycobacterium tuberculosis* reveals a two-metal ion dependent PIN-domain ribonuclease and a variable mode of toxin–antitoxin assemblyq " ELSEVIER **188**: 249-258.

^{lxxxvi} Dorota Matelska, K. S. a. K. G. (2017). "Comprehensive classification of the PIN domain-like superfamily." Nucleic Acids Research **45**(12).

^{lxxxvii} Christian Mirwaldt*, I. K. r. a. R. H. (1995). "The Crystal Structure of Dihydrodipicolinate Synthase from Escherichia coli at 2.5 Å Resolution." Journal of Molecular Biology **246**: 227-239.

^{lxxxviii} Wupeng Yan^{#1}, H. S., Fuhang Song³, Yisong Guo⁶, Cheng-Hsuan Wu², Ampon Sae Her², Yi Pu^{2,4}, Shu Wang², Nathchar Naowarojna², Andrew Weitz⁶, Michael P. Hendrich⁶, Catherine E. Costello^{2,4}, Lixin Zhang³, Pinghua Liu², and Yan Jessie Zhang^{1,5} (2015). "Endoperoxide formation by an α -ketoglutarate-dependent mononuclear non-haem iron enzyme." PMC **527**(7579): 539–543.

^{lxxxix} Daniel J. Ferraro, L. G., S. Ramaswamy * (2005). "Rieske business: Structure–function of Rieske non-heme oxygenases." ELSEVIER **338**(175–190).

^{xc} Tzong-Yuan Lin¹, T. W., Jae-Hun Jeoung, and Holger Dobbek² (2012). "Suppression of Electron Transfer to Dioxygen by ChargeTransfer and Electron Transfer Complexes in the FAD-dependent Reductase Component of Toluene Dioxygenase*." THE JOURNAL OF BIOLOGICAL CHEMISTRY **287**(45): 38338–38346

^{xcⁱ} Groves, D. o. O. d. O. L. Æ. M. R. (2009). "The three-component signalling system HbpS–SenS–SenR as an example of a redox sensing pathway in bacteria." Amino Acids **37**: 479–486.

^{xcⁱⁱ} Keri L. Colabroy *, I. R. S., Alexander H.S. Vlahos, Androo J. Markham, Matthew E. Jakubik (2014). "Defining a kinetic mechanism for L-DOPA 2,3 dioxygenase, a single-domain type I extradiol dioxygenase from Streptomyces lincolnensis." ELSEVIER(1844): 607–614.

^{xcⁱⁱⁱ} Minor, Z. O. a. W. (1997). ""Processing of X-ray Diffraction Data Collected in Oscillation Mode,"." Methods in Enzymology **276**(Macromolecular Crystallography, part A): 307-326.

^{xc^{iv}} T. Geoff G. Battye, a. L. K., a Owen Johnson, a Harold R. Powell, a and Andrew G. W. Leslie, * (2011). "iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM " Acta Crystallogr D Biol Crystallogr. **67**(Pt 4): 271–281

^{xc^v} P.R.Evans (2011). ""An introduction to data reduction: space-group determination, scaling and intensity statistics"." Acta Cryst. D **67**: 282-292.

^{xcvi} COLLABORATIVE COMPUTATIONAL PROJECT, N. (1994). "The CCP4 Suite: Programs for Protein Crystallography". Acta Cryst. D **50**: 760–763.

^{xcvii} Cowtana, P. E. a. K. (2004). "Coot: model-building tools for molecular graphics " Acta Crystallographica Section D, BIOLOGICAL CRYSTALLOGRAPHY **60**(12): 2126-2132.

^{xcviii} Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." Acta Crystallogr. Sect. D Biol. Crystallogr. **66**: 213-221.

^{xcix} P.V. Afonine, R. W. G.-K., N. Echols, J.J. Headd, N.W. Moriarty, M. Mustyakimov, T.C. Terwilliger, A. Urzhumtsev, P.H. Zwart, and P.D. Adams. (2012). "Towards automated crystallographic structure refinement with phenix.refine." Acta Crystallogr D Biol Crystallogr **68**: 352-367.

^c Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999). "Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation." J. Mol. Biol. **285**: 1735-1747.

^{ci} Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). "Features and development of Coot." Acta Crystallogr. Sect. D Biol. Crystallogr. **66**: 486-501.

^{cii} Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. (2008). "Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard." Acta Crystallogr D Biol Crystallogr **64**: 61-69.

^{ciii} SnapGene software (from Insightful Science; available at snapgene.com)

^{civ} Okamoto, A., Higuchi, T., Hirotsu, K., Kuramitsu, S., and Kagamiyama, H. (1994). "X-ray crystallographic study of pyridoxal 5'-phosphate-type aspartate aminotransferases from *Escherichia coli* in open and closed form." Journal of biochemistry **116**: 95-107.

^{cv} P Chelikani¹, I. F., P C Loewen (2004). "Diversity of structures and properties among catalases." Cellular and Molecular Life Sciences CMLS volume **61**(2): 192-208.

^{cv} Marina Evich, E. S., 1 Yujun George Zheng, 2 and Markus W. Germann^{corresponding author 1} (2016). "Effect of methylation on the side-chain pK a value of arginine." Protein Sci. **25**(2): 479-486.

^{cvi} Fast, T. W. L. a. W. (2010). "Guanidine-Modifying Enzymes in the Pentein Superfamily." Elsevier Ltd.

^{cvi} Munmun Nandi, a. C. B., b Ann Karen C. Brassinga, a Mark F. Belmonte, c W. G. Dilantha Fernando, d Peter C. Loewen, a and Teresa R. de Kievit (2016). "Pseudomonas brassicacearum Strain DF41 Kills Caenorhabditis elegans through Biofilm-Dependent and Biofilm-Independent Mechanisms." Appl Environ Microbiol. **82**(23): 6889–6898.

^{cix} M. Senissar, M. C. M., and D. E. Brodersen * (2017). "Structural conservation of the PIN domain active site across all domains of life." PROTEIN SCIENCE **26**: 1474–1492.

^{cx} Min AB, M. L., Sawaya MR, Habel J, Cascio D, Eisenberg D (2012). "The crystal structure of the Rv0301–Rv0300 VapBC-3 toxin–antitoxin complex from M. tuberculosis reveals a Mg(2)(1) ion in the active site and a putative RNA-binding site." Protein Sci **21**: 1754–1767.

^{cx} Dorota Matelska, K. S. a. K. G. (2017). "Comprehensive classification of the PIN domain-like superfamily." Nucleic Acids Research **45**(12).

^{cxii} Arcus VL, M. J., Robson J, Cook GM (2011). "The PIN-domain ribonucleases and the prokaryotic VapBC toxin-antitoxin array." Protein Eng. Des. Sel. **24**: 33-40.

^{cxiii} Xu K, D. E., Brodersen DE (2016). " Structural analysis of the active site architecture of the VapC toxin from Shigella flexneri." Proteins **84**: 892–899.

^{cxiv} Makarova KS, A. L., Galperin MY, Grishin NV, Tatusov RL, Wolf YI, Koonin EV (1999). "Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell." Genome Res **9**: 608–628.

^{cxv} Pandey DP, G. K. (2005). "Toxin–antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes." Nucleic Acids Res **33**: 966-976.

^{cxvi} Sung-Min Kang, D.-H. K., † Chenglong Jin, and Bong-Jin Lee* (2018). "A Systematic Overview of Type II and III Toxin-Antitoxin Systems with a Focus on Druggability." Toxins **10**(12): 515.

^{cxvii} Page R, P. W. (2016). "Toxin–antitoxin systems in bacterial growth arrest and persistence." Nat Chem Biol **12**: 208-214.

^{cxviii} Tollervey², C. S. a. D. (2013). "Threading the barrel of the RNA exosome." Trends in Biochemical Sciences **33**(10): 485-493.

^{cxix} Clissold PM, P. C. (2000). "PIN domains in nonsense-mediated mRNA decay and RNAi. ." Curr Biol **10**: R888-R890.

^{cxx} Mattison K, W. J., So M, Brennan RG (2006). "Structure of FitAB from *Neisseria gonorrhoeae* bound to DNA reveals a tetramer of toxin–antitoxin heterodimers containing pin domains and ribbon-helix–helix motifs." J Biol Chem **281** **281**: 37942–37951.

^{cxxi} Deo Prakash Pandey¹, K. G. (2005). "Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes." Nucleic Acids Res. **33**(3): 966-976.

^{cxixii} Bintou Ahmadou Ahidjo, D. K., Joanna L. McKenzie, Edith E. Machowski, Bhavna G. Gordhan, Vickery Arcus, Garth L. Abrahams, Valerie Mizrahi (2011). "VapC Toxins from *Mycobacterium tuberculosis* Are Ribonucleases that Differentially Inhibit Growth and Are Neutralized by Cognate VapB Antitoxins " PloS One **6**(6).

^{cxixiii} Oksana I. Demidenok, A. S. K. A. V. G. (2014). "Toxin–antitoxin vapBC locus participates in formation of the dormant state in *Mycobacterium smegmatis*." FEMS microbiol. lett. **352**: 69-77.

^{cxixiv} Lee IG, L. S., Chae S, Lee KY, Kim JH, Lee BJ (2015). "Structural and functional studies of the *Mycobacterium tuberculosis* VapBC30 toxin–antitoxin system: implications for the design of novel antimicrobial peptides." Nucleic Acids Res **43**: 7624–7637.

^{cxixv} Das U, P. V., Subhramanyam UK, Wilmanns M, Gourinath S, Srinivasan A (2014). "Crystal structure of the VapBC-15 complex from *Mycobacterium tuberculosis* reveals a two-metal ion dependent PIN-domain ribonuclease and a variable mode of toxin–antitoxin assembly." J Struct Biol **188**: 249–258.

^{cxixvi} Winsor GL, G. E., Lo R, Dhillon BK, Shay JA, Brinkman FS (2016). "Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database." Nucleic Acids Res.

^{cxixvii} Dobson, R. C., Griffin, M.D., Jameson, G.B., Gerrard, J.A. (2005). "The crystal structures of native and (S)-lysine-bound dihydrodipicolinate synthase from *Escherichia coli* with improved resolution show new features of biological significance." Acta Crystallogr D Biol Crystallogr **61**: 1116-1124.

^{cxxviii} Henrick, E. K. a. K. "Protein structure comparison service PDBeFold at European Bioinformatics Institute (<http://www.ebi.ac.uk/msd-srv/ssm>)."

^{cxxix} Christian Mirwaldt*, I. K. r. a. R. H. (1995). "The Crystal Structure of Dihydrodipicolinate Synthase from *Escherichia coli* at 2.5 Å Resolution." *Journal of Molecular Biology* **246**: 227-239.

^{cxxx} Rutter, W. J. (1964). *Fedn Proc.* **23**: 1248-1257.

^{cxxxi} Wierenga*, R. K. (2001). "The TIM-barrel fold: a versatile framework for efficient enzymes." *FEBS Letters* **492**: 193-198.

^{cxxxii} E.Grazi1T.ChengB.L.Horecker (1962). "The formation of a stable aldolase-dihydroxyacetone phosphate complex." *Biochemical and Biophysical Research Communications* **7**(3): 250-253.

^{cxxxiii} O'Connell, I. A. R. a. E. L. (1969). "Studies on the Interaction of Aldolase with Substrate Analogues " *Journal of Biological Chemistry* **244**: 126.

^{cxxxiv} Mirwaldt, C., Korndorfer, I., Huber, R. (1995). "The crystal structure of dihydrodipicolinate synthase from *Escherichia coli* at 2.5 Å resolution." *J Mol Biol* **246**: 227-239.

^{cxxxv} E Krissinel 1 , K. H. (2004). "Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions." *Acta Crystallogr D Biol Crystallogr* **12**: 2256-2268.

^{cxxxvi} Gary X. Shaw, * Yue Li,²* Genbin Shi,¹ Yan Wu,² Scott Cherry,¹ Danielle Needle,¹ Di Zhang,¹ Joseph E. Tropea,¹ David S. Waugh,¹ Honggao Yan,² and Xinhua Ji¹ (2014). "Structural enzymology and inhibition of the bifunctional folate pathway enzyme HPPK-DHPS from the biowarfare agent *Francisella tularensis*." *FEBS J.* **281**(18): 4123–4137.

^{cxxxvii} A. Maxwell Burroughsa, R. W. H., Neal C. Goebelc, Bilal H. Sayyedb, Tyler J. Voegtlineb, Alan W. Schwabacherb, T. Mark Zabriskiec, and Nicholas R. Silvaggi*,* (2013). "Structural and Functional Characterization of MppR, an Enduracididine Biosynthetic Enzyme from *Streptomyces hygroscopicus*: Functional Diversity in the Acetoacetate Decarboxylase-Like Superfamily." *Biochemistry* **52**(26): 4492–4506

^{cxxxviii} Lisa S. Mydy, R. W. H., Trevor M. Hagemann, Alan W. Schwabacher, and Nicholas R. Silvaggi* (2019). "Mechanistic Studies of the *Streptomyces bingchengensis* Aldolase-Dehydratase: Implications for Substrate and Reaction Specificity in the Acetoacetate Decarboxylase-like Superfamily." *Biochemistry* **58**(40): 4136–4147.

^{cxxxix} Lisa S. Mueller, R. W. H., Jenna M. Ochsenwald, Robert T. Berndt, Geoffrey B. Severin, Alan W. Schwabacher, and Nicholas R. Silvaggi* (2015). "Sbi00515, a Protein of Unknown Function from *Streptomyces bingchenggensis*, Highlights the Functional Versatility of the Acetoacetate Decarboxylase Scaffold." Biochemistry **54**: 3978–3988.

^{cxl} Lisa S. Mydy, R. W. H., Trevor M. Hagemann, Alan W. Schwabacher, and Nicholas R. Silvaggi* (2019). "Mechanistic Studies of the *Streptomyces bingchenggensis* Aldolase-Dehydratase: Implications for Substrate and Reaction Specificity in the Acetoacetate Decarboxylase-like Superfamily." Biochemistry **58**(40): 4136–4147.

^{cxli} Henrick, E. K. a. K. "Protein structure comparison service PDBeFold at European Bioinformatics Institute (<http://www.ebi.ac.uk/msd-srv/ssm>)."

^{cxlii} Rui, Z., Li, X., Zhu, X., Liu, J., Domigan, B., Barr, I., Cate, J.H., Zhang, W. (2014). "Microbial biosynthesis of medium-chain 1-alkenes by a nonheme iron oxidase." Proc Natl Acad Sci U S A **111**: 18237-18242.

^{cxliii} Zhe Ruia, b., Xin Lia,c, Xuejun Zhua,b, Joyce Liua,d, Bonnie Domigana,b, Ian Barre,f, Jamie H. D. Catee,f,g, and Wenjun Zhanga,b,g,1 (2014). "Microbial biosynthesis of medium-chain 1-alkenes by a nonheme iron oxidase." PNAS **111**(51): 18237–18242.

^{cxliv} Krebs C, F. D., Walsh CT, Bollinger JM Jr (2007). "Non-heme Fe(IV)-oxo intermediates." Acc Chem Res **40**: 484-492.

^{cxlv} Kovaleva EG, L. J. (2008). "Versatility of biological nonheme Fe(II) centers in oxygen activation reactions. ." Nat Chem Biol **4**: 186-193.

^{cxlvi} Vaillancourt FH, B. J., Eltis LD (2006). "The ins and outs of ring-cleaving dioxygenases. ." Crit Rev Biochem Mol Biol Direct. **41**: 241-267.

^{cxlvii} Bruijninx PCA, v. K. G., Klein Gebbink RJM (2008). "Mononuclear non-heme iron enzymes with the 2-His-1-carboxylate facial triad: recent developments in enzymology and modeling studies." Chem Soc Rev **37**: 2716–2744.

^{cxlviii} RP, H. (2004). "Fe(II)/alpha -ketoglutarate-dependent hydroxylases and related enzymes." Crit Rev Biochem Mol Biol Direct. **39**: 21-68.

^{cxlix} Bugg TDH, A. M., Hardiman EM, Singh R (2011). "The emerging role for bacteria in lignin degradation and bio-product formation." Curr Opin Biotechnol **22**: 394-400.

^{cl} Bassan A, B. M., Siegbahn PEM (2003). "Mechanism of aromatic hydroxylation by an activated FeIV=O core in tetrahydrobiopterin-dependent hydroxylases." Chem Eur J **9**: 4055–4067.

^{cli} Gibson DT, P. R. (2000). "Aromatic hydrocarbon dioxygenases in environmental biotechnology." Curr Opin Biotechnol **11**: 236-243.

^{clii} Ferraro DJ, G. L., Ramaswamy S (2005). "Rieske business: structure-function of Rieske non-heme oxygenases. ." Biochem Biophys Res Commun **338**: 175-190.

^{cliii} Juan Du, a. J. J. C., a,b,c,d and Garry R. Buettner^{a,c,*} (2012). "Ascorbic acid: Chemistry, biology and the treatment of cancer☆." Biochim Biophys Acta. **1826**(2): 443-457.

^{cliv} Laura Ledesma-García¹, A. S.-A., Milagros Medina², Francisca Reyes-Ramírez¹ & Eduardo Santero¹ (2016). "Redox proteins of hydroxylating bacterial dioxygenases establish a regulatory cascade that prevents gratuitous induction of tetralin biodegradation genes." Nature Scientific Reports **6**(23848).

^{clv} Cammack, J. R. M. a. R. (1992). "THE ELECTRON-TRANSPORT PROTEINS OF HYDROXYLATING BACTERIAL DIOXYGENASES." Annu. Rev. Microbiol. **46**: 277-305.

^{clvi} Fraaije¹, A. N. F. M. W. (2020). "An overview of microbial indigo-forming enzymes." Applied Microbiology and Biotechnology **104**: 925–933.

^{clvii} Murdock D, E. B., Serdar C, Thalen M (1993). "Construction of metabolic operons catalyzing the de novo biosynthesis of indigo in Escherichia coli." Biotechnology (NY) **11**: 381–386.

^{clviii} Toshiki Furuya, c. a. M. H., and Kuniki Kino (2013). "Reconstitution of Active Mycobacterial Binuclear Iron Monooxygenase Complex in Escherichia coli." Appl Environ Microbiol. **79**(19): 6033–6039.

^{clix} Fraaije¹, A. N. F. M. W. (2020). "An overview of microbial indigo-forming enzymes." Applied Microbiology and Biotechnology **104**: 925–933.

^{clx} N. V. Zharikova*, T. R. I., E. I. Zhurenko, V. V. Korobov, and T. V. Markusheva (2018). "Bacterial Genes of Non-Heme Iron Oxygenases, Which Have a Rieske-Type Cluster, Catalyzing Initial Stages of Degradation of Chlorophenoxyacetic Acids." Russian Journal of Genetics **54**(3): 284-295.

^{clxi} Sylvain Gervason, D. L., Amir Ben Mansour, Thomas Botzanowski, Christina S. Müller, Ludovic Pecqueur, Gwenaëlle Le Pavec, Agnès Delaunay-Moisan, Omar Brun, Jordi Agramunt, Anna

Grandas, Marc Fontecave, Volker Schünemann, Sarah Cianférani, Christina Sizun, Michel B. Tolédano & Benoit D'Autréaux (2019). "Physiologically relevant reconstitution of iron-sulfur cluster biosynthesis uncovers persulfide-processing functions of ferredoxin-2 and frataxin." Nature Communications(3566).

clxii Mair E A Churchill 1 , H. M. S., Charis L Uhlon (2011). "Defining the structure and function of acyl-homoserine lactone autoinducers " Methods Mol Biol. **692**: 159-171.

clxiii S B Nielsen 1, A. L., J U Andersen, U V Pedersen, S Tomita, L H Andersen (2001). "Absorption spectrum of the green fluorescent protein chromophore anion in vacuo." Phys Rev Lett **87**(22): 228102.

clxv Jesse M. Reinhardt, S. L. B. (2018). "Partitioning protein ParP directly links chemotaxis to biofilm dispersal in *Pseudomonas aeruginosa*." bioRxiv.

clxvi Laura M. Ketelboeter, S. L. B. (2017). "Characterization of 2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione resistance in pyomelanogenic *Pseudomonas aeruginosa* DKN343." PloS One.

clxvii Giavalisco P, K. K., Hummel J, Seiwert B, Willmitzer L (2009). "¹³C isotope-labeled metabolomes allowing for improved compound annotation and relative quantification in liquid chromatography-mass spectrometry-based metabolomic research. ." Anal Chem **81**(15): 6546–6551.

clxviii Muschet C, M. G., Prehn C, de Angelis MH, Adamski J, Tokarz J (2016). "Removing the bottlenecks of cell culture metabolomics: fast normalization procedure, correlation of metabolites to cell number, and impact of the cell harvesting method. ." Metabolomics **12**(10): 151.

clxix Patti G.J., T. R., Rinehart D., Cho K., Nikolskiy I., Johnson C., Siuzdak G. (2013). "A View from Above: The Cloud Plot for Visualizing Global Metabolomic Data." Analytical Chemistry. **85**(2): 798-804.

clxx TfSiteScan (RRID:SCR_010667)

clxxi Hongxia Wang 1, J. C. A., Jorge A Benitez, Anisia J Silva (2012). "Interaction of the histone-like nucleoid structuring protein and the general stress response regulator RpoS at *Vibrio cholerae* promoters that regulate motility and hemagglutinin/protease expression." J Bacteriol. **194**(5): 1205-1215.

clxxii Paolo Landini 1, T. E., Johannes Wolf, Stephan Lacour (2014). "sigmaS, a major player in the response to environmental stresses in Escherichia coli: role, regulation and mechanisms of promoter recognition." Environ Microbiol Re. **6**(1): 1-13.

clxxiii Suzuki, M., Takahashi, Y., Noguchi, A., Arai, T., Yagasaki, M., Kino, K., Saito, J. (2012). "The structure of L-amino-acid ligase from Bacillus licheniformis." Acta Crystallogr D Biol Crystallogr **68**: 1535-1540.

clxxiv Kim, S.-H., Hisano, T., Takeda, K., Iwasaki, W., Ebihara, A., Miki, K. (2007). "Crystal structure of the mutant of HpaB (T198I, A276G, and R466H)." J Biol Chem **282**: 33107-33117.

clxxv Zou XW, L. Y., Hsu NS, Huang CJ, Lyu SY, Chan HC, Chang CY, Yeh HW, Lin KH, Wu CJ, Tsai MD, Li TL (2014). "Crystal structure of methyltransferase from Streptomyces hygroscopicus complexed with S-adenosyl-L-methionine and phenylpyruvic acid." Acta Crystallogr. D Biol. Crystallogr. **70**: 1549-1560.

clxxvi Elkins J.M., C. I. J., Hernandez H., Doan L.X., Robinson C.V., Schofield C.J., Hewitson K.S. (2002). "Oligomeric structure of proclavaminc acid amidino hydrolase: evolution of a hydrolytic enzyme in clavulanic acid biosynthesis." Biochem. J. **366**: 423-434.

clxxvii Lanlan Han, N. V., Sarah A. Oehm, Tyler G. Fenske, Alan W. Schwabacher, and Nicholas R. Silvaggi* (2018). "Streptomyces wadayamensis MppP is a PLP-Dependent Oxidase, Not an Oxygenase." Biochemistry **57**(23): 3252–3264.

clxxviii Xiao-Wei Zou, a. Y.-C. L., a,b,c† Ning-Shian Hsu,a Chuen-Jiuan Huang,a Syue-Yi Lyu,a Hsiu-Chien Chan,a Chin-Yuan Chang,a Hsien-Wei Yeh,a Kuan-Hung Lin,a Chang-Jer and d. M.-D. T. Wu, c,e and Tsung-Lin Lia,b,f* (2014). "Structure and mechanism of a nonhaem-iron SAM-dependent C-methyltransferase and its engineering to a hydratase and an O-methyltransferase " Acta Crystallographica Section D D70: 1549–1560.

clxxix Sreejith Raran-Kurussi 1 , S. C., Di Zhang 1 , David S Waugh 2 (2017). "Removal of Affinity Tags with TEV Protease " Methods Mol Biol. **1586**: 221-230.

clxxx Cowtana, P. E. a. K. (2004). "Coot: model-building tools for molecular graphics " Acta Crystallographica Section D, BIOLOGICAL CRYSTALLOGRAPHY **60**(12): 2126-2132.

clxxxi Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010). "PHENIX:

a comprehensive Python-based system for macromolecular structure solution." *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**: 213-221.

clxxxii P.V. Afonine, R. W. G.-K., N. Echols, J.J. Headd, N.W. Moriarty, M. Mustyakimov, T.C. Terwilliger, A. Urzhumtsev, P.H. Zwart, and P.D. Adams. (2012). "Towards automated crystallographic structure refinement with phenix.refine." *Acta Crystallogr D Biol Crystallogr* **68**: 352-367.

clxxxiii Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999). "Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation." *J. Mol. Biol.* **285**: 1735-1747.

clxxxiv Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). "Features and development of Coot." *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**: 486-501.

clxxxv Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. (2008). "Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard." *Acta Crystallogr D Biol Crystallogr* **64**: 61-69.

clxxxvi Okamoto, A., Higuchi, T., Hirotsu, K., Kuramitsu, S., and Kagamiyama, H. (1994). "X-ray crystallographic study of pyridoxal 5'-phosphate-type aspartate aminotransferases from *Escherichia coli* in open and closed form." *Journal of biochemistry* **116**: 95-107.

clxxxvii Dunja Urbančiča, Anita Kotarč, Alenka Šmida, Marko Jukiča, Stanislav Gobeca, Lars-Göran Mårtensson-Janež, Plavec Irena Mlinarič-Raščana (2019). "Methylation of selenocysteine catalysed by thiopurine S-methyltransferase." *Biochimica et Biophysica Acta (BBA) - General Subjects* **1863**(1): 182-190.

clxxxviii A, G. H. K. V. N. (1997). "NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities." *J. Org. Chem* **62** (21): 7512–7515.

clxxxix Kunio HIKICHI, T. H., Mitsuhiro IKURA, Keiichiro HIGUCHI,* Keiji EGUCHI,* and Muneki OHUCHI* (1987). "Two-Dimensional INADEQUATE NMR Spectra of Ethylene-Propylene Copolymer." *Polymer Journal* **19**(11): 1317-1320.

cx Alex D Bain 1 , D. W. H., Christopher K Anand, Zhenghua Nie, Valerie J Robertson (2010). "Problems, artifacts and solutions in the INADEQUATE NMR experiment " *Magn Reson Chem.* **48**(8): 630-641.

^{cxci} K.R. Keshari, J. M. M. (2017). "Biological Applications of Hyperpolarized ¹³C NMR." Encyclopedia of Spectroscopy and Spectrometry (Third Edition).

^{cxcii} Keri L. Colabroy *, I. R. S., Alexander H.S. Vlahos, Androo J. Markham, Matthew E. Jakubik (2014). "Defining a kinetic mechanism for L-DOPA 2,3 dioxygenase, a single-domain type I extradiol dioxygenase from *Streptomyces lincolnensis* " Biochimica et Biophysica Acta(1844): 607–614.

^{cxciii} Vuksanovic, N. (2020). "Structural and Functional Characterization of L-Enduracididine Biosynthetic Enzymes." UWM Digital Commons.

^{cxciv} Kowiel M, B. D., Porebski PJ, Shabalin IG, Jaskolski M, Minor W (2019). "Automatic recognition of ligands in electron density by machine learning." Bioinformatics (Oxford, England) **35**(3): 452-461.

^{cxcv} Ryan*, J. B. H. K. (2019). "In vitro Reconstitution of the Biosynthetic Pathway to the Nitroimidazole Antibiotic Azomycin." Angew.Chem. **131**: 11773–11777.

^{cxcvi} Trine Aakvik Strand, R. L., 1 , * Kristin Fløgstad Degnes, 2 Malin Lando, 1 and Svein Valla 1 (2014). "A New and Improved Host-Independent Plasmid System for RK2-Based Conjugal Transfer." PloS One **9**(3): e90372.

Appendices

Appendix A: Crystallographic data collection and refinement statistics of Pbr proteins

	PbrMppP.PLP	PbrMppP.4H KA	PbrHyp (denzo for scalling)	PbrDHPS	PbrDHPS.2KA	PbrOx
Beamline	LS-CAT 21-ID- D	LS-CAT 21-ID- D	LS-CAT 21-ID- D	LS-CAT 21-ID- D	LS-CAT 21-ID- G	LS-CAT 21-ID- D
Space group	P 21 2 21	P 21 2 21	C 1 2 1	C 2 2 2	C 2 2 2	P 42
Cell dimensions						
a, b, c (Å)	86.68, 104.29, 199.76	86.68, 104.29, 199.76	149.8, 60.2, 96.7	110.16, 126.694, 72.227	110.16, 126.694, 72.227	94.2, 94.2, 43.5
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90, 125.02, 90	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0 90.0 90.0
Resolution (Å) (last shell) ^a	41.5-2.05 (2.14-1.8)	44.53-2.25 (2.29-2.0)	40.166-2.5 (2.31-1.83)	35.99 -1.2 (2.33-0.187)	41.7 – 2.8 (2.33-0.187)	47.08- 2.7
Wavelength (Å)	0.97921 (D)	0.97856 (D)	0.97849 (D)	0.97849 (D)	0.97921 (G)	0.97856 (D)
No. of reflections						
Observed	890997 (711649)	256371 (12494)	180163 (84329)	982564 (47738)	111722 (53434)	20510 (97125)
Unique	143020 (114176)	79816 (3984)	43157 (22868)	292765 (14556)	60681 (40713)	10735 (80257)
Completeness (%)	100.0 (100.0)	96.2 (98.0)	93.91 (96.47)	98.8 (99.5)	94.5.9 (97.1)	99.1 (100.0)
Rmerge (%) ^{a,b}	0.120 (0.624)	0.106 (0.726)	0.057 (0.193)	0.071 (0.183)	0.131 (0.488)	0.11 (0.28)
Multiplicity	6.2 (5.1)	3.2 (3.1)	4.8 (3.9)	3.3 (3.5)	4.8 (5.2)	4.1 (7.2)
$\langle I/\sigma(I) \rangle^a$	14.3 (6.1)	10.0 (2.3)	24.6 (3.1)	13.2 (5.3)	14.46 (6.3)	19.1 (3.7)
Model Refinement Statistics						
Reflections in work set	100876	76747	20203	297119	16525	19721
Reflections in test set	19402	3015	19402	96068	9054	13047
R _{cryst} (R _{free})	0.137 (0.166)	0.156 (0.197)	0.183 (0.231)	0.1856 (0.2335)	0.19 (0.243)	0.159 (0.202)
No. of residues	782	779	688	1434	1454	285

No. of solvent atoms	621	593	370	1032	930	102
Number of TLS groups	32	31	41	33	20	24
	PbrMppP.PLP	PbrMppP.4H KA	PbrHyp	PbrDHPS	PbrDHPS.2KA	PbrOx
Average <i>B</i> -factor (Å ²) ^c						
Protein atoms	32.0	34.1	28.8	31.2	32.4	25.3
Ligands	30.9	33.9	31.6	27.6	41.9	21.5
Solvent	41.7	35.2	33.7	37.3	36.9	33.4
RMS deviations						
Bond lengths (Å)	0.013	0.015	0.010	0.013	0.081	0.017
Bond angles (°)	1.395	1.565	1.141	0.73	0.9	1.462
Ramachandran statistics (favored/allowed/outliers)	98.3 / 1.7 / 0.0	98.4 / 1.6 / 0.0	97.2 / 2.5 / 0.7	98.3/ 2.0 / 0.4	98.2/ 1.8 / 0	97.2 / 2.5 / 0.3

^a Values in parentheses apply to the high-resolution shell indicated in the resolution row

^b $R = \frac{\sum (|F_{obs}| - scale * |F_{calc}|)}{\sum |F_{obs}|}$

^c Isotropic equivalent B factors, including contribution from TLS refinement

Appendix B: Crystallographic data collection and refinement statistics of Sgr proteins

	SgrMppP	SgrMppP.4HKA	SgrMT.2KA	SgrAH
Beamline	LS-CAT 21-ID-D	LS-CAT 21-ID-D	LS-CAT 21-ID-D	LS-CAT 21-ID-D
Space group	P 23 21 2	P 23 21 2	P 21 21 21	P 3 2 1
	Cell dimensions			
a, b, c (Å)	91.429, 91.429, 91.125	91.429, 91.429, 91.125	71.65, 75.76, 300.98	83.16, 83.16, 77.5
α , β , γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 120.0
Resolution (Å) (last shell) ^a	45.7 -2.3 (2.14-0.16)	47.7 -2.9 (2.2-0.17)	73.4-2.21 (2.29-0.22)	52.56 -1.59 (1.97- 0.92)
Wavelength (Å)	0.97921 (D)	0.97921 (D)	0.97856 (D)	0.97921 (D)
	No. of reflections			
Observed	1532933 (60186)	1325441 (78452)	34283 (10942)	30385 (13521)
Unique	2889 (2708)	3215 (2784)	4495 (3152)	2712 (19781)
Completeness (%)	100.0 (100.0)	99.2 (100.0)	90.2 (91.2)	95.64 (98.2)
Rmerge (%) ^{a,b}	0.106 (0.734)	0.015 (0.821)	0.106 (0.726)	0.057 (0.193)
Multiplicity	14.4 (11.5)	11.9 (10.5)	3.2 (3.1)	4.8 (3.9)
$\langle I/\sigma(I) \rangle^a$	30.7 (5.2)	17.5 (4.1)	10.0 (1.9)	24.7 (6.1)
	Model Refinement Statistics			
Reflections in work set	2807	31521	76747	20203
Reflections in test set	174	20140	3015	19402
R _{cryst} (R _{free})	0.16 (0.21)	0.142 (0.184)	0.22 (0.31)	0.142 (0.184)
No. of residues	772	778	673	1264
No. of solvent atoms	242	201	311	409

Number of TLS groups	23	21	37	29
Average <i>B</i> -factor (Å ²) ^c				
Protein atoms	32.0	29.2	34.1	28.8
Ligands	30.9	41.5	43.9	31.6
Solvent	35.6	29.7	36.6	33.0
RMS deviations				
Bond lengths (Å)	0.012	0.011	0.021	0.013
Bond angles (Å)	0.021	0.035	0.109	0.015

^a Values in parentheses apply to the high-resolution shell indicated in the resolution row

^b $R = \sum (| |F_{obs}| - scale * |F_{calc}| |) / \sum |F_{obs}|$

^c Isotropic equivalent *B* factors, including contribution from TLS refinement

CURRICULUM VITAE

Lamia Tabassum Badhon

Place of Birth: Dhaka, Bangladesh

Education:

- Ph.D., University of Wisconsin-Milwaukee, May-2021
Major: Biochemistry
Title of Thesis: Structure & function of enzymes in two uncharacterized gene clusters from *Pseudomonas brassicacearum* and *Streptomyces griseofuscus*
- M.Sc., University of Dhaka, June-2014
Major: Biochemistry & Molecular Biology
- B.Sc., University of Dhaka, June-2013
Major: Biochemistry & Molecular Biology

Awards:

- Outstanding poster award (1st position) in 36TH Midwest Enzyme Chemistry Conference (2016) of poster titled "Exploration of an MppP-Containing Operon in *Pseudomonas brassicacearum*"

Manuscripts in preparation:

- Structure of A newly identified PIN domain homolog, PbrHyp from *Pseudomonas brassicacaerum* (DF41)
- Structure of PbrDHPS, a type-I aldolase homolog in *Pseudomonas brassicacerum*
- Structure of PbrOx, a mononuclear non-heme Fe(II) binding protein in *Pseudomonas brassicacerum*
- Structural and functional characterization of SAM-dependent methyltransferase homolog from a putative biosynthetic cluster in *Streptomyces griseofuscus*
- Structural and functional characterization of an amidinohydrolase homolog from a putative biosynthetic cluster in *Streptomyces griseofuscus*

Poster presentations:

- "Structural and functional characterization of proteins of unknown function from a putative biosynthetic cluster in *Streptomyces griseofuscus* " 40th Midwest Enzyme Chemistry Conference'2020
- "Structure and functional prediction of a SAM and metal dependent methyltransferase in MppP gene cluster of *Streptomyces griseofuscus*" in 39th Midwest Enzyme Chemistry Conference'2019

-
- “Identification of A PLP-dependent L-arg oxidase in an uncharacterized gene cluster in *Pseudomonas brassicacearum*” in 38th Midwest Enzyme Chemistry Conference’2018
 - “Exploration of an MppP-Containing Operon in *Pseudomonas brassicacearum*” in 36th Midwest Enzyme Chemistry Conference’2016, at departmental research symposium’2017

Teaching & research experience:

- Spring-2016: Teaching assistant on CHEM 104
- Summer, Fall-2016-Spring, Fall-2017, Fall-2020: Teaching assistant on CHEM 102
- Spring, Fall-2018, Fall-2019, Spring-2020: Teaching assistant on CHEM 103
- Spring-2019: Teaching assistant on CHEM 603

Research experience as a graduate student in Dr. Silvaggi lab