**Distribution Agreement**

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Tamra C. Blue Date

The Fourth Wave of Biocatalysis: Biochemical and *in silico* Characterization of FeS Cluster Containing Metalloenzyme Superfamilies

By

Tamra C. Blue

Doctor of Philosophy

Chemistry

_____

Katherine M. Davis, Ph.D.
Advisor

_____

Vincent Conticello, Ph.D.
Committee Member

_____

William M. Wuest, Ph.D.
Committee Member

Accepted:

_____

Kimberly Jacob Arriola, Ph.D., MPH
Dean of the James T. Laney School of Graduate Studies

_____

Date

# The Fourth Wave of Biocatalysis: Biochemical and *in silico* Characterization of FeS Cluster Containing Metalloenzyme Superfamilies

By

Tamra C. Blue
B.S., Georgia State University, 2017
M.S., Emory University, 2021

Advisor: Katherine M. Davis, PhD

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Chemistry
2023

## Abstract

The Fourth Wave of Biocatalysis: Biochemical and *in silico* Characterization of FeS Cluster Containing Metalloenzyme Superfamilies

By Tamra C. Blue

The chemical manufacturing sector is one of the largest greenhouse gas and waste production zones in the United States. Enzyme biocatalysts have been at the forefront of industrial and pharmaceutical research as a green chemistry alternative to chemo-chemical methods. Enzymes are an attractive alternative as they have high stereospecificity and generate enantiomerically pure compounds without harsh solvents. However, a subset of enzymes has been traditionally underexplored, often due to their severe oxygen sensitivity. These metal cofactor-containing enzymes, or metalloenzymes, often have a propensity for radical initiation reactions catalyzing the stereoselective CH functionalization of unactivated carbons; however, they also display diverse reactivities.

This dissertation employs techniques honed in the most recent wave of biocatalysis innovation to unlock the untapped catalytic potential of iron-sulfur (FeS) metalloenzymes. These techniques highlight *in silico* approaches to enzymatic discovery and characterization. This includes genomics, bioinformatics, molecular docking, and molecular dynamic protocols coupled with traditional biochemical methods. The first chapter briefly introduces two metalloenzymes superfamilies of industrially interest and general *in silico* techniques.

The second chapter describes the implementation of a rationally designed protein library to probe the Old Yellow Enzyme (OYE) superfamily sequence space for novel enzyme candidates useful in the biosynthetic generation of monocarboxylic acid and decalone chiral building blocks.

Additionally, the study assessed the standard operating procedure utilized to systematically explore the superfamily for novel activity. While this study yielded no monocarboxylic acid-compatible enzymes, four enzymes displayed activity for decalone chiral building block biosynthesis. Of those four, three were classified as insoluble under our standard operating conditions, indicating that the conditions are not universal.

The third chapter details the characterization of a metalloenzyme OYE subfamily with a unique mutation in its metal binding motif. This mutant displayed novel *N*-methyl-proline oxidative demethylation activity while being able to retain monocarboxylic acid reduction activity. The fourth chapter details our exploration of the ribosomally synthesized and post-translationally modified peptide (RiPP) recognition mechanism of radical SAM SuiB from *Streptococcus suis*. SuiB is a tailoring enzyme within a RiPP biosynthetic gene cluster of SuiA. However, it lacks interaction with the RiPP recognition domain. This chapter provides evidence for the non-conical identification of the precursor peptides by the tailoring enzyme's bridging domain.

# The Fourth Wave of Biocatalysis: Biochemical and *in silico* Characterization of FeS Cluster Containing Metalloenzyme Superfamilies

By

Tamra C. Blue

B.S., Georgia State University, 2017

M.S., Emory University, 2020

Advisor: Katherine M. Davis, PhD

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Chemistry

2023

## Acknowledgments

I did not get here on my own, and the list is too long to name everyone who took the time to teach and mentor me. Whether your goal was to help or otherwise, thank you. You made me a better scientist and person.

To my current and former Ph.D. PIs (Dr. Katherine Davis, and Dr. Stefan Lutz) and committee members, thank you for introducing me to the world of enzymes and computational biochemistry. You both taught me the sacrifices it takes to be a researcher, a scholar, and a human.

To my mentors, co-workers, and students at Spelman College, thank you for reminding me why this degree has value.

To my father, mother, brother, sister, and husband: thank you for encouraging one step at a time every day.

# List of Frequently Used Abbreviations

| Abbreviation | Full Name |
|---|---|
| OYE | Old Yellow Enzyme |
| rSAM | Radical S-Adenosyl methionine |
| FMN | Flavin Mononucleotide |
| FAD | Flavin Adenine Dinucleotide |
| NADH | Nicotinamide Adenine Dinucleotide |
| SAM | S-Adenosyl methionine |
| SSN | Sequence Similarity Network |
| 2-ER | 2-enoate reductase |
| MD | Molecular dynamics |
| Docking | Molecular Docking |
| BGC | Biosynthetic gene cluster |
| *E. coli* | *Escherichia coli* |
| TIM | triosephosphate isomerase |

# Table of Contents

# Table of Figures

# Table of Tables

# **Chapter 1.** General Introduction

## 1.1. General Introduction

It has been 190 years since the discovery of the first enzyme.[1] Since then, enzyme biocatalysis development and utilization have progressed with human technological advancement.[2] Classified into three major waves of enzyme biocatalyst discovery, the first wave hallmarked the utilization of plant and animal matter to facilitate chemical transformations.[3] Since antiquity, *Saccharomyces cerevisiae (*Brewer's yeast) has been used to catalyze the anaerobic ethanol fermentation of grains to produce beer.[4] For almost as long, goat stomach extracts and *Rhizomucor miehei* mold have been used for cheese-making.[5] However, the instability and difficulty storing plant/animal matter long-term made consistent chemical transformations difficult, thus motivating the start of the second wave of biocatalysis.

The second wave encompassed the exploitation of DNA sequencing to heterologously express, isolate, and subsequently purify enzyme biocatalysts from living matter (Figure 1.1).[3] The first purified flavoenzyme, old yellow enzyme 1 (OYE1), was isolated from Brewer's yeast in 1932.[6, 7] OYE1 and its homologs are ene-reductases that catalyze the reduction of α/β unsaturated alkenes adjacent to electron-withdrawing groups (ketones, aldehydes, etc.).[8, 9] The isolation of OYE1 led to further breakthroughs in the field of biocatalysis. In 1935, the first catalytically active cofactor, flavin mononucleotide (FMN), was discovered. Without FMN, the OYE1 apoprotein remained structurally stable however was catalytically inactive until FMN was reconstituted.[10, 11] The second wave also highlighted the beginnings of protein engineering via site-directed mutagenesis, allowing for modifications to the protein sequence to improve substrate scope,[12] alter enantioselectivity of the product,[13, 14] optimize thermostability,[12, 15] and elucidate reaction mechanisms.[16]

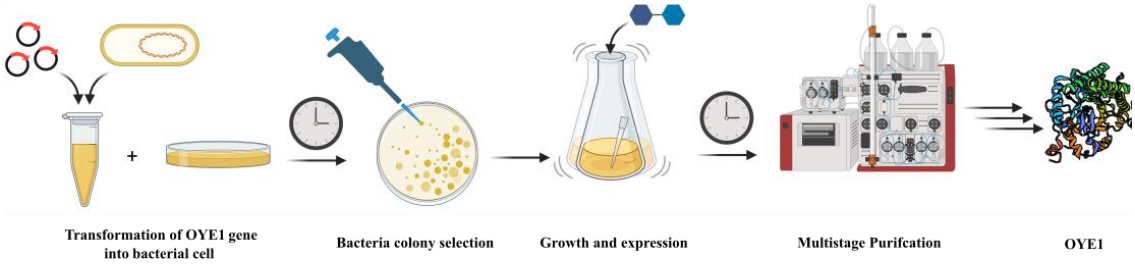



*Figure 1.1. Modern acquisition of OYE1 by heterologous expression*

The third wave came crashing in with expanded approaches to protein engineering primarily driven by error-prone polymerase chain reaction (ER-PCR) and directed evolution techniques. While ER-PCR increases genetic diversity by allowing for the random introduction of mutations to genetic information, directed evolution couples sequential rounds of genetic diversification to selection for target activity. YqjM from *Bacillus subtilis*, an OYE1 homolog, underwent iterative saturation mutagenesis (ISM) directed evolution to expand its substrate scope to include 3-methylcyclohexenone.[17] Opposed to random genetic diversification with ER-PCR, ISM has higher effectiveness as amino acid (AA) multi-sites are rationally selected.[17, 18]

We are currently living in the fourth wave of enzyme biocatalyst discovery. Integrating bioinformatics, artificial intelligence, and other computational tools to validate techniques honed in the previous eras highlights this wave. Fourth-wave techniques have been further utilized to discover novel biocatalysis, identify enzymes with target activity, and elucidate de novo 3D structure/function from a primary sequence. OYE1 and its homologs, Old Yellow Enzymes (OYEs), have been ever-present throughout the previous waves (Figure 1.2). As the field dove into the fourth wave, OYEs continued to ride the wave, and a newer enzyme superfamily, the Radical

*S*-Adenosyl-L-Methionine (rSAM) superfamily, emerged with a splash. In this chapter, these two metalloenzyme superfamilies will be introduced, and fourth-wave techniques will be discussed in the context of tools that attempt to bridge the gap between metalloenzyme sequence, structure, and function.

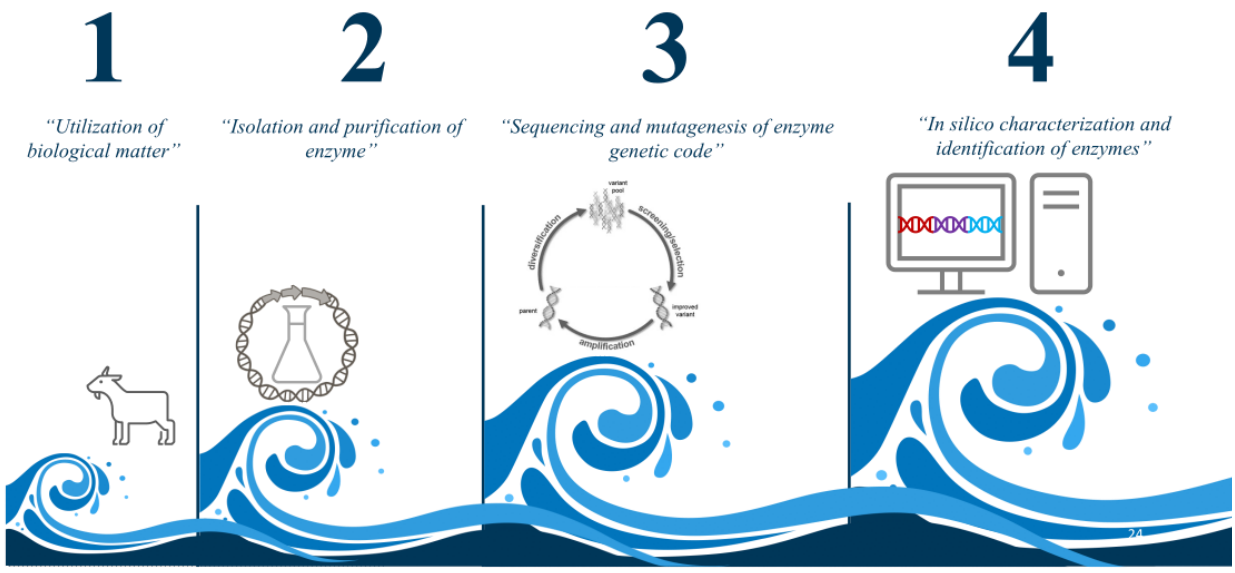



*Figure 1.2. Waves of Biocatalysis discovery and innovation*

### 1.2. Metalloenzyme Biocatalysts

Biocatalysts, as the name implies, are macromolecules of biological origin that increase the rate of a chemical reaction without undergoing permanent alteration. Enzymes are protein biocatalysts that catalyze reactions by decreasing the activation energy of the system. This decrease is often accomplished by reducing proximity between reaction actors, encouraging optimal substrate confirmation, and/or directly participating during reactivity.[19] To date, ~35% of enzyme structures cataloged in the Protein Data Bank (PDB) have a metal ion bound to one of its chains.[20-22] Metal ions play a variety of native functions in enzymes, including aiding in protein folding, mediating

charge transfer, and participating directly in catalysis.[23] Enzymes that contain metal cofactors are referred to as metalloenzymes. One of the most abundant metal cofactors is the iron-sulfur (FeS) cluster, existing as [2Fe-2S], [3Fe-4S], [4Fe-4S], [4Fe-3S], or [8Fe-7S] stoichiometries.[24-27] The clusters commonly coordinate to the protein by cysteinyl sidechain binding motifs. However, some systems tolerate non-cysteine FeS cluster coordination sites, such as histidine,[28, 29] arginine, lysine, serine,[30] and water.[31] The most common FeS cluster is the [4Fe-4S] cluster which readily shuttles between the $[4Fe\text{-}4S]^{+} \rightleftharpoons [4Fe\text{-}4S]^{+2} \rightleftharpoons [4Fe\text{-}4S]^{+3}$ charge states (Figure 1.3).

FeS clusters are well-known for their role in the electron transport chain of the mitochondria, where the metal clusters transport electrons from Complex I-III to the heme protein cytochrome c before subsequent transfer to molecular oxygen.[32] Lesser known is their role as gene expression regulators. SoxR from *Escherichia coli (E. coli)* is a redox-sensitive transcriptional activator. In times of oxidative stress, the FeS cluster of SoxR is oxidized. It induces the expression of SoxS, a transcription initiator. However, under reductive conditions, SoxR loses the ability to activate s*oxS* transcription and instead aids in DNA repair.[33-35]

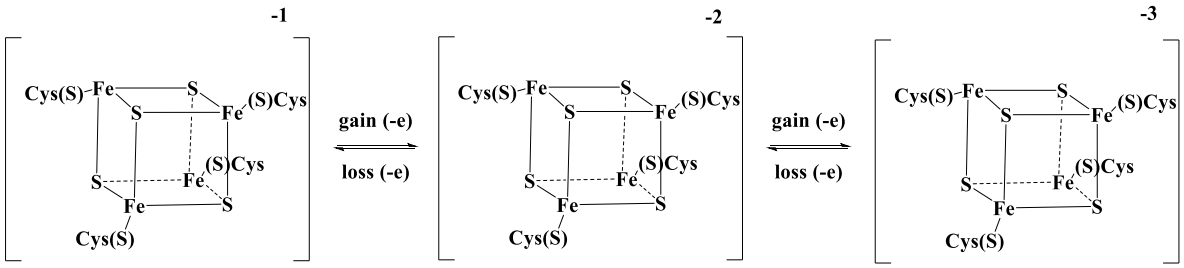



*Figure 1.3. Oxidation states of FeS clusters*

### 1.3. Old Yellow Enzymes

An industrially relevant metalloenzyme-containing superfamily is the OYE superfamily. Most OYEs are single-domain nicotinamide adenine dinucleotide (NAD(P)H) dependent

oxidoreductases that catalyze the asymmetric reduction of activated alkenes, providing chiral products.[36, 37] Substrate reduction occurs via a bi-bi-ping-pong mechanism in which a nicotinamide cofactor initially reduces the FMN.[9] The substrate is subsequently protonated by hydride transfer from the reduced FMN. An additional hydride transfer from the N5-atom of the flavin to the β-carbon of the activated alkene follows, allowing for protonation of the α-carbon by nearby water or tyrosine 197 in OYE1 (Figure 1.4).[38] This specific mechanism leads to anti-addition hydrogenation, meaning the hydrogen atoms approach the alkene from opposite sides.[39]

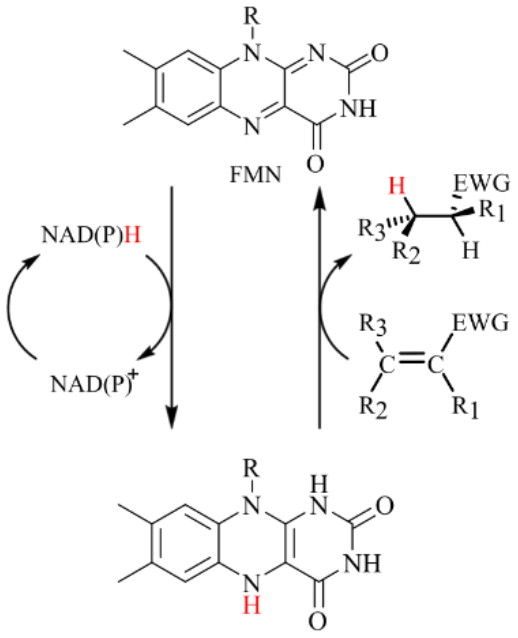



*Figure 1.4. OYE ene-reductase scheme*

This mechanism readily occurs with substrates possessing aldehydes, ketones, anhydrides, nitros, cyclic imides, β-cyanoacrylates, and β-nitro acrylates.[38] As a result, OYEs have been employed to produce diverse products, including anti-inflammatory drugs, macrocyclic antibiotics, and anticonvulsants.[40-42] However, a significant limitation of classical and engineered OYEs is their inability to readily reduce unsaturated free carboxylic groups as substrates. This ability to reduce

alkenes adjacent to carboxylic acids is henceforth referred to as non-traditional OYE chemistry, as most known OYEs are not traditionally known to perform such activity.[9]

The over >115,000 homologous members of the OYE superfamily are identified from highly conserved secondary structure motifs: an $(\alpha,\beta)_8$-triosephosphate isomerase barrel (TIM barrel) fold domain, with an FMN binding motif within the barrel near the carboxy-terminal ends of the β-strands in OYE1.[43-46] Additional secondary structural elements exist depending on the enzyme. In OYE1, four β-strands and five α-helices form a "lid" that covers the N-terminal end of the barrel, while additional structural elements form segments that aid in binding the flavin and forming a ligand-binding pocket.[7, 8]

Historically, the OYE superfamily was divided into three main classes based on sequence homology and structural features: "classical OYE" as represented by OYE1, "thermophilic-like OYE" as represented by YqiM from *B. subtilis*, and "2- enoate reductase" as represented by *E. coli* 2,4-Dienoyl CoA Reductase (DCR).[8] The 2-Enoate Reductase (2-ER) family inhabits almost a quarter of the OYE superfamily. Like all OYEs, its members contain an FMN binding motif and a highly conserved TIM. However, 2-ERs are multidomain metalloenzymes composed of a traditional OYE-like FMN substrate binding domain, a FeS cluster binding domain (Cys334, Cys337, Cys341, and Cys353 in 2,4-dienoyl-CoA reductase (DCR) from *E. coli*, and often a flavin adenine dinucleotide (FAD) binding domain (Figure 1.5).[47-49] A highly conserved [4Fe-4S] cluster binding domain links the FAD and FMN domains.

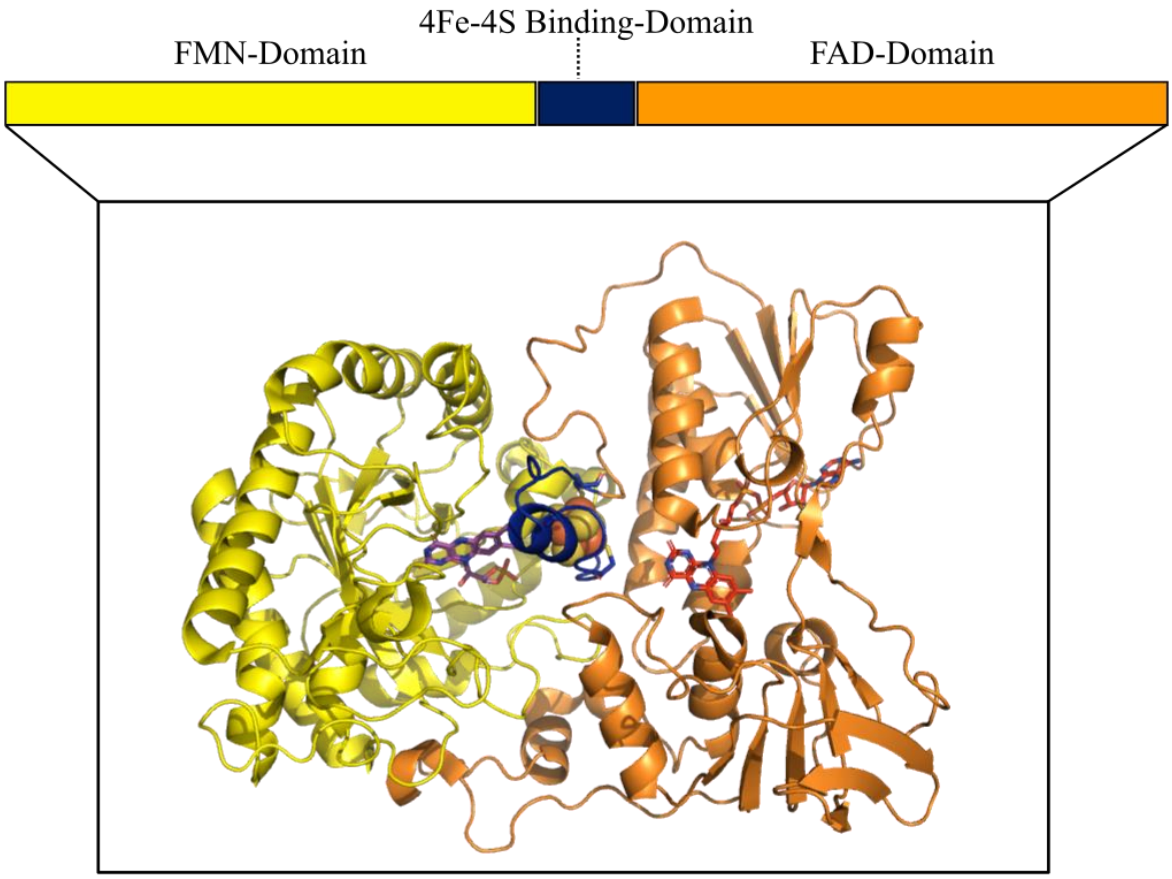



*Figure 1.5. Structure of 2-ER representative DCR from E. coli (PDB:1ps9). FAD domain (orange), FMN domain (yellow), and [4Fe-4S] binding domain (navy)*

The [4Fe-4S] cluster is responsible for the family's oxygen-sensitive nature. This oxygen sensitivity and the considerable molecular weight of the homologs have made the family less desirable for characterization. However, 2-ERs house activity is coveted for the enantiomerically pure biosynthesis of carboxylic acid chiral synthons.[9, 49] The [4Fe-4S] cluster is predicted to play a role in 2-ER's unique substrate specificity, as traditional OYE cannot readily perform this activity with monocarboxylic acids. While the reduction of unsaturated carboxylic acids is the staple of 2-ER chemistry, the literature reveals that 2-ERs facilitate other diverse chemistries such as oxidative deamination[50] and oxidative demethylation,[51, 52] both of which catalytically depend on the [4Fe-4S]

cluster. The techniques of the fourth wave of biocatalysis exploration are uniquely compatible with the identification and characterization of novel 2-ERs.

### 1.4. rSAM Enzymes

Another industrially relevant metalloenzyme superfamily is the rSAM superfamily. The bioinformatic discovery of the initial >600 superfamily members is a clear example of how fourth-wave techniques have propelled the field of biocatalysis.[53] Members of the rSAM superfamily catalyze diverse radical reactions initiated by a 5′-deoxyadenosyl radical (dAdo•) intermediate. SAM cleavage is mediated by coordination to the canonical [4Fe-4S] housed within the conserved partial $(\alpha/\beta)_6$-TIM barrel representing the rSAM domain.[54] While rSAMs can accommodate additional auxiliary FeS clusters, the rSAM [4Fe-4S] cluster uniquely contains a SAM binding open coordination site. This site is a consequence of the highly conserved rSAM tri-cysteinyl FeS binding motif: $CX_3CX_2C$. In the rSAM mechanism, an electron transfer reaction occurs between the $[4Fe\text{-}4S]^+$ and SAM to catalyze the reductive cleavage of SAM to form the highly reactive dAdo• intermediate and methionine.[53, 55]

Some rSAM enzymes are of particular industrial interest as they play a significant role in the biosynthesis of ribosomally synthesized and post-translationally modified peptides (RiPP). RiPPs comprise a superfamily of natural products that display diverse bioactivity as antibiotics, antifungals, antivirals, and anticancer agents.[56-59] Another structural domain commonly observed in rSAM enzymes is involved in the production of these molecules: namely, the RiPP Recognition Element (RRE). While, RiPP precursor peptides and RRE domains lack a conserved sequence identity, the RRE presents a conserved winged helix-turn-helix (wHTH) secondary structure composed of three α-helixes and a three to four-strand β-sheet.[60-62] This conserved secondary

structure has been a target for RiPP genome mining as RiPP precursor peptides are often encoded directly upstream from the RRE, whether discrete or embedded in another enzyme.[60] However, it is not understood how enzymes that lack an RRE, or how enzymes that show no evidence of utilizing their RRE, recognize their precursor peptide. Fourth-wave techniques are uniquely compatible with visualizing and identifying novel peptide binding modes that may be too fleeting for protein crystallography.

### 1.5. Computational Sequence Analysis

A protein's primary sequence encodes its structure, and its structure influences its function. This function could refer to a protein's native function or binding affinity towards a particular ligand class.[63] The relationship between the primary sequence of a protein and its function is known as "the protein sequence-function relationship." Bioinformatic tools for sequence analysis have exploited this relationship to identify members as novel members of the OYE superfamily, discover the rSAM superfamily, and classify both.[38, 53, 64] Bioinformatics is the application of computer science, physics, biology, chemistry, and statistics to understand, group, and draw conclusions about biological data.[65, 66] The combination of DNA sequencing, structural biology, and bioinformatics, in recent years, has completely revolutionized biomolecular science.[67] The discovery of the rSAM superfamily and the RiPP superfamily, as previously mentioned, are key examples.[53, 68, 69]

#### 1.5.1. Sequence Similarity Networks

UniProt is a protein database containing >240 million nonredundant AA sequences in 2023.[70] Large databases like Uniprot and InterPro utilized Basic Local Alignment Search Tool (BLAST) algorithms to find similarity between sequences and annotate functions from conserved domains

and features. The quality of the sequence similarity is classified by the expected value (E-value), which is the chance that the search would result in random hits given the size of the dataset, which means that a higher number, i.e., 10, correlates to up to 10 "random hits" that can appear during the search and is less accurate than an E-value of 1. The lower the E-value, the more significant the similarity match is.[71] This E-value can be utilized to visualize extensive datasets as an alternative to phylogenetic trees bioinformatically. A protein sequence similarity network (SSN) organizes proteins by sequence identity into clusters composed of nodes and edges. Nodes represent homologous sequences within a predefined sequence identity threshold. In contrast, edges represent the similarity between two nodes, quantified as the E-values between two node representatives. Nodes appear as geometric shapes (i.e., eclipse, square, or triangles), while edges appear as lines that connect the nodes. Edges only exist between two nodes if the E-value is within the assigned threshold. The SSN allows for quick visualization and comparisons of proteins in large datasets.[72]

### 1.5.2. Function Prediction

In the context of the sequence-function relationship, the lower the E-value threshold in an SSN, the higher the sequence similarity within a cluster. Atkinson and coworkers identified that at a high level of stringency (depending on the system and size of the data set), it can be assumed that the clusters contain isofuntional enzymes.[73] Sequence analysis is commonly used to identify domains and motifs involved in function based on conserved AA residues and homology-based annotation. For example, rSAMs have been identified by the highly conserved partial TIM barrel domain with a $CX_3CX_2C$ FeS binding domain in the Interpol database (IPR007197). However, reasonable caution should be utilized when using the primary sequence homology alone to classify functions. Protein sequences with low sequence identity could have similar secondary structures, such as the

RRE domain of rSAMs, and protein sequences with high similarity can have opposing substrate specificity.[74] Recent studies identified extensive enzyme misannotation in some protein databases averaging 5%-63% functional misannotations across multiple superfamilies.[75] These studies displayed that UniProtKB/Swiss-Prot had the lowest levels of misannotation, at times being zero depending on the superfamily.

### 1.5.3. Structure Prediction

The ability to predict secondary structure from primary sequence has recently advanced with the implementation of artificial intelligence software, particularly from AlphaFold. Prior to AlphaFold, the most common form of structure prediction was homology modeling (comparative modeling) which involved a BLAST search (or similar) to identify similar protein sequences. The sequences would then be aligned to identify highly conserved domains with known structural information. These parameters produce reasonable secondary structures but bias novel sequences to resemble known protein structures.[76] Additionally, homology modeling is impossible if no similar structural data is available. In juxtaposition, AlphaFold is a non-homology modeling algorithm that relies on predictions of a neural network trained from over 170,000 sequences.[77]

### 1.5.4. Molecular Docking

Molecular docking is a computational tool that relies primarily on the “lock and key” mechanism to generate enzyme: ligand binding complexes. Through iterative sampling and ligand conformation variations, lock and key algorithms place the mobile ligand within the binding site of the rigid body model to produce plausible binding models. The docking complexes are scored and clustered to generate the most optimum structures. Docking can account for far more than steric effects, including electrostatics, van der Waals, hydrogen bonding interaction, polarity,

hydrophobicity, and system energy, to name a few.[78] Docking has been a powerful tool in identifying potential substrates for catalysis and elucidating mechanistic inquiries in both superfamilies. In particular, OYE3 was mutated to gain enantioselectivity in the production of (R)-citronellal from a racemic mix of (E/Z)-citral Molecule docking was used to visualize the point mutation, resulting in enatiopreference for the E-citral. Docking revealed that the mutant OYE3 forced Z-citral in a flipped binding orientation against the *si*-face of FMN, resulting in activity loss. Alternatively, the E-citrals docking mimicked wtOYE3 binding conformation against the *re*-face, retaining activity.[79]

### 1.5.5. Molecular Dynamics Simulations

While molecular docking provides snapshots of potential binding models, these models are limited as protein systems are dynamic molecules. Often upon peptide binding, the enzyme undergoes conformational changes to accommodate. Dynamic change is not readily observed in docking software. Instead, molecular dynamics (MD) is employed to visualize system xyz coordinates change over time, often with the docking models as $t_0$. For example, when the triosephosphate isomerase (TIM) from *Trypanosoma cruzi* is docked in an open conformation, the system readily receives substrates; what is difficult to infer from docking is that upon substrate binding, TIM enters a closed conformation via movements of its loops and cannot accept additional substrate.[80, 81] Molecular docking is limited to providing static ligand-bound complexes. These snapshots cannot capture the dynamic nature of enzyme systems. MD is employed for the computational prediction and analysis of protein conformation changes over time. MD utilizes classical mechanics approaches based on Newtonian physics to determine how all atoms in a biological system move due to preset forces acting on each atom over time. MD simulations rely on molecular mechanics (MM) force fields which apply experimental and quantum mechanical data to increase

physiological relevance and assume that each atom is a ball on a spring to reduce computational load.[82, 83]

### 1.6. **Heterologous Expression of Metalloenzymes**

Fourth-wave techniques are invaluable tools to validate and support conclusions observed in experimentation. Expressing [4Fe-4S] cluster containing metalloenzymes in non-native cell lines has been a challenge due, in part, to the lack of chaperones necessary to assemble and insert the metal cluster.[84] To adjust for this limitation, techniques have been developed to aid in the heterologous expression of metalloenzymes. The most common method of metal cluster incorporation is the reconstitution of the FeS cluster post-expression and/or co-expression of the protein of interest with iron-sulfur-cluster-assembly (isc) machinery.[85] Genomic analysis reveals that homologous *isc* operons exist across species but vary depending on the microorganism, with some organisms containing all systems, two, one, or a fragment of one system.[86-88]. There are three known systems involved in the biogenesis of prokaryotic FeS metalloenzymes.[84] (1)The NIF system encoded in *Azotobacter vinelandii* by operon *nifUSVWZM* is responsible for the maturation of the FeS cluster containing nitrogenase.[89] While the (2) ISC system from *A. vinelandii* and the (3) SUF systems in *E. coli* encodes the *iscSUA-hscBA-fdx* operon and *sufABCDSE* operon, respectively, and generate housekeeping proteins that aid in the production of a variety of FeS-containing proteins.[86] Known organisms with FeS assembly machinery share a similar generalized mechanism. First, a cysteine desulfurase (NifS, IscS, or SufSE) extracts sulfur from L-cysteine, which is then combined with iron to assemble clusters within a scaffold protein (NifU, IscU, or SufB). Finally, a transfer step occurs where the assembled FeS cluster is transferred to the recipient protein via energy-dependent proteins (HscAB or SufCD) and FeS carrier proteins (IscA or SufA).[86-88]

To aid in the heterologous expression of metalloenzymes, *isc* assembly machinery is often incorporated into non-native cell lines via the co-transformation of an *isc* plasmid and the protein of interest. The pDB1282 plasmid employed in Chapter 3 encodes six genes from the *iscSUA* operon of *A. vinelandii* hosted on a pARA13 backbone. The genes include *iscS, iscU, iscA*, *hscA, hscB, and fdx (*Figure 1.6*)*.[90] This plasmid has been shown to increase soluble expression of 2Fe-2S and [4Fe-4S] containing metalloenzymes when co-expressed with the protein of interest.[91-93]

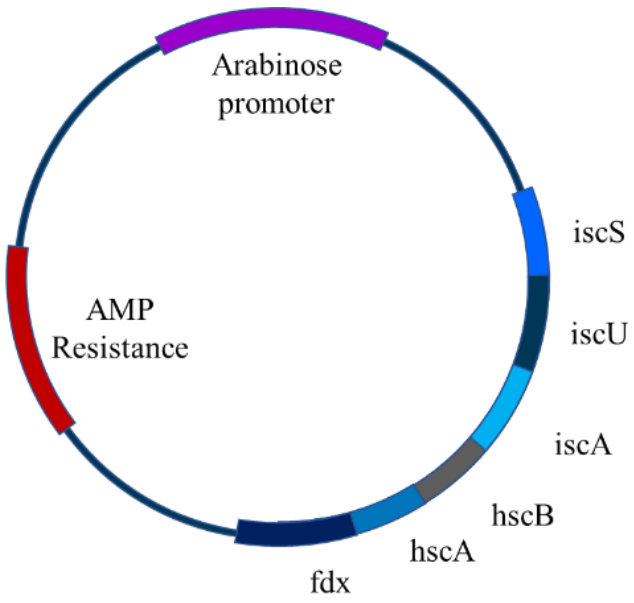



*Figure 1.6. pDB1282 plasmid components (Cartoon diagram)*

### 1.7. Aims and Scope of the Dissertation

The overall theme of this dissertation is the characterization of two metalloenzyme superfamilies with fourth-wave techniques in order to elucidate the role of dynamic movement and/or the FeS cluster and its effect on enzyme reactivity.

**In chapter two**, I detail the use of bioinformatics to systematically explore the untapped sequence space of the OYE superfamily.

**In chapter three,** I characterized a member of the OYE 2ER family containing a mutation within the FeS-cluster binding motif in order to understand the role of the FeS cluster in 2ER OYEs.

**Finally, in chapter four,** I use molecular docking and molecular dynamics to explore the role of conformational changes associated with the binding of the RiPP precursor peptide SuiA to the RiPP modifying protein SuiB from *Streptococcus suis*.

**1.8. Bibliography**

(1) Armstrong, E. Enzymes: A Discovery And Its Consequences. *Nature* **1933**, *131* (3311), 535-537.

(2) Heckmann, C. M.; Paradisi, F. Looking Back: A Short History Of The Discovery Of Enzymes And How They Became Powerful Chemical Tools. *Chemcatchem* **2020**, *12* (24), 6082-6102.

(3) Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. Engineering The Third Wave Of Biocatalysis. *Nature* **2012**, *485* (7397), 185-194.

(4) Raihofer, L.; Zarnow, M.; Gastl, M.; Hutzler, M. A Short History Of Beer Brewing: Alcoholic Fermentation And Yeast Technology Over Time: Alcoholic Fermentation And Yeast Technology Over Time. *EMBO Reports* **2022**, *23* (12), E56355.

(5) Ilany, J.; Netzer, A. Milk-Clotting Activity Of Proteolytic Enzymes. *Journal Of Dairy Science* **1969**, *52* (1), 43-46.

(6) Warburg, O. Ein Zweites Sauer-Stoffubertragendes Ferment Und Sein Absorptionspektrum. *Naturwissenschafte* **1932**, *20*.

(7) Fox, K. M.; Karplus, P. A. Old Yellow Enzyme At 2 A Resolution: Overall Structure, Ligand Binding, And Comparison With Related Flavoproteins. *Structure* **1994**, *2* (11), 1089-1105.

(8) Toogood, H. S.; Gardiner, J. M.; Scrutton, N. S. Biocatalytic Reductions And Chemical Versatility Of The Old Yellow Enzyme Family Of Flavoprotein Oxidoreductases. *Chemcatchem* **2010**, *2* (8), 892-914.

(9) Toogood, H. S.; Scrutton, N. S. Discovery, Characterization, Engineering, And Applications Of Ene-Reductases For Industrial Biocatalysis. *ACS Catalysis* **2018**, *8* (4), 3532-3549.

(10) Theorell, H. Das Gelbe Oxydationsferment. Biochemische Zeitschrift **1935**, *278*, 263-290.

(11) Massey, V. The Chemical And Biological Versatility Of Riboflavin. *Biochemical Society Transactions* **2000**, *28* (4), 283-296.

(12) Balke, K.; Beier, A.; Bornscheuer, U. T. Hot Spots For The Protein Engineering Of Baeyer-Villiger Monooxygenases. *Biotechnology Advances* **2018**, *36* (1), 247-263.

(13) Ying, X.; Yu, S.; Huang, M.; Wei, R.; Meng, S.; Cheng, F.; Yu, M.; Ying, M.; Zhao, M.; Wang, Z. Engineering The Enantioselectivity Of Yeast Old Yellow Enzyme OYE2y In Asymmetric Reduction Of (E/Z)-Citral To (R)-Citronellal. *Molecules* **2019**, *24* (6).

(14) Van Den Heuvel, R. H.; Fraaije, M. W.; Ferrer, M.; Mattevi, A.; Van Berkel, W. J. Inversion Of Stereospecificity Of Vanillyl-Alcohol Oxidase. *Proceedings Of The National Academy Of Sciences* **2000**, *97* (17), 9455-9460.

(15) Riedel, A.; Mehnert, M.; Paul, C. E.; Westphal, A. H.; Van Berkel, W. J.; Tischler, D. Functional Characterization And Stability Improvement Of A 'Thermophilic-Like'ene-Reductase From Rhodococcus Opacus 1CP. *Frontiers In Microbiology* **2015**, *6*, 1073.

(16) Brown, B. J.; Deng, Z.; Karplus, P. A.; Massey, V. On The Active Site Of Old Yellow Enzyme. Role Of Histidine 191 And Asparagine 194. *Journal of Biological Chemistry* **1998**, *273* (49), 32753-32762.

(17) Bougioukou, D. J.; Kille, S.; Taglieber, A.; Reetz, M. T. Directed Evolution Of An Enantioselective Enoate-Reductase: Testing The Utility Of Iterative Saturation Mutagenesis. *Advanced Synthesis & Catalysis* **2009**, *351* (18), 3287-3305.

(18) Acevedo-Rocha, C. G.; Hoebenreich, S.; Reetz, M. T. Iterative Saturation Mutagenesis: A Powerful Approach To Engineer Proteins By Systematically Simulating Darwinian Evolution. *Methods in Molecular Biology* **2014**, *1179*, 103-128.

(19) Cooper, G.; Adams, K. *The Cell: A Molecular Approach*; Oxford University Press, 2022.

(20) Putignano, V.; Rosato, A.; Banci, L.; Andreini, C. Metalpdb In 2018: A Database Of Metal Sites In Biological Macromolecular Structures. *Nucleic Acids Research* **2017**, *46* (D1), D459-D464.

(21) Andreini, C.; Cavallaro, G.; Lorenzini, S.; Rosato, A. Metalpdb: A Database Of Metal Sites In Biological Macromolecular Structures. *Nucleic Acids Research* **2013**, *41* (Database Issue), D312-319.

(22) Zheng, H.; Chruszcz, M.; Lasota, P.; Lebioda, L.; Minor, W. Data Mining Of Metal Ion Environments Present In Protein Structures. *Journal Of Inorganic Biochemistry* **2008**, *102* (9), 1765-1776.

(23) Tainer, J. A.; Roberts, V. A.; Getzoff, E. D. Protein Metal-Binding Sites. *Current Opinion In Biotechnology* **1992**, *3* (4), 378-387.

(24) Barras, F.; Loiseau, L.; Py, B. How Escherichia Coli And Saccharomyces Cerevisiae Build Fe/S Proteins. *Advances In Microbial Physiology* **2005**, *50*, 41-101.

(25) Shi, R.; Hou, W.; Wang, Z.-Q.; Xu, X. Biogenesis Of Iron–Sulfur Clusters And Their Role In DNA Metabolism. *Frontiers In Cell And Developmental Biology* **2021**, *9*, Review.

(26) Lanzilotta, W. N.; Christiansen, J.; Dean, D. R.; Seefeldt, L. C. Evidence For Coupled Electron And Proton Transfer In The [8Fe-7S] Cluster Of Nitrogenase. *Biochemistry* **1998**, *37* (32), 11376-11384.

(27) Beinert, H.; Holm, R. H.; Munck, E. Iron-Sulfur Clusters: Nature's Modular, Multipurpose Structures. *Science* **1997**, *277* (5326), 653-659.

(28) Volbeda, A.; Charon, M.-H.; Piras, C.; Hatchikian, E. C.; Frey, M.; Fontecilla-Camps, J. C. Crystal Structure Of The Nickel–Iron Hydrogenase From Desulfovibrio Gigas. *Nature* **1995**, *373* (6515), 580-587.

(29) Iwata, S.; Saynovits, M.; Link, T. A.; Michel, H. Structure Of A Water Soluble Fragment Of The 'Rieske' Iron–Sulfur Protein Of The Bovine Heart Mitochondrial Cytochrome Bc1 Complex Determined By MAD Phasing At 1.5 Å Resolution. *Structure* **1996**, *4* (5), 567-579.

(30) Brereton, P. S.; Duderstadt, R. E.; Staples, C. R.; Johnson, M. K.; Adams, M. W. W. Effect Of Serinate Ligation At Each Of The Iron Sites Of The [Fe4S4] Cluster Of Pyrococcus Furiosus Ferredoxin On The Redox, Spectroscopic, And Biological Properties. *Biochemistry* **1999**, *38* (32), 10594-10605.

(31) Robbins, A. H.; Stout, C. D. Structure Of Activated Aconitase: Formation Of The [4Fe-4S] Cluster In The Crystal. *Proceedings Of The National Academy Of Sciences* **1989**, *86* (10), 3639-3643.

(32) Read, A. D.; Bentley, R. E.; Archer, S. L.; Dunham-Snary, K. J. Mitochondrial Iron-Sulfur Clusters: Structure, Function, And An Emerging Role In Vascular Biology. *Redox Biol* **2021**, *47*, 102164.

(33) Lee, P. E.; Demple, B.; Barton, J. K. DNA-Mediated Redox Signaling For Transcriptional Activation Of Soxr. *Proceedings Of The National Academy Of Sciences* **2009**, *106* (32), 13164-13168.

(34) Ding, H.; Hidalgo, E.; Demple, B. The Redox State Of The [2Fe-2S] Clusters In Soxr Protein Regulates Its Activity As A Transcription Factor. *Journal Of Biological Chemistry* **1996**, *271* (52), 33173-33175.

(35) Gaudu, P.; Weiss, B. Soxr, A [2Fe-2S] Transcription Factor, Is Active Only In Its Oxidized Form. *Proceedings Of The National Academy Of Sciences* **1996**, *93* (19), 10094-10098.

(36) Hall, M.; Stueckler, C.; Hauer, B.; Stuermer, R.; Friedrich, T.; Breuer, M.; Kroutil, W.; Faber, K. Asymmetric Bioreduction Of Activated C=C Bonds Using Zymomonas Mobilis NCR

Enoate Reductase And Old Yellow Enzymes OYE 1–3 From Yeasts. *European Journal Of Organic Chemistry* **2008**, *2008* (9), 1511-1516.

(37) White, D. W., Iamuri, S., Joud, P., Blue, T.C., Copp, J., Lutz, S. The Hidden Biocatalytic Potential Of The Old Yellow Enzyme Family. *Biorxiv* **2023**.

(38) Scholtissek, A.; Tischler, D.; Westphal, A. H.; Van Berkel, W. J. H.; Paul, C. E. Old Yellow Enzyme-Catalysed Asymmetric Hydrogenation: Linking Family Roots With Improved Catalysis. *Catalysts* **2017**, *7* (5), 130.

(39) Lonsdale, R.; Reetz, M. T. Reduction Of A,B-Unsaturated Ketones By Old Yellow Enzymes: Mechanistic Insights From Quantum Mechanics/Molecular Mechanics Calculations. *Journal Of The American Chemical Society* **2015**, *137* (46), 14733-14742.

(40) Pietruszka, J.; Schölzel, M. Ene Reductase-Catalysed Synthesis Of (R)-Profen Derivatives. *Advanced Synthesis & Catalysis* **2012**, *354* (4), 751-756.

(41) Korpak, M.; Pietruszka, J. Chemoenzymatic One-Pot Synthesis Of Γ-Butyrolactones. *Advanced Synthesis & Catalysis* **2011**, *353* (9), 1420-1424.

(42) Winkler, C. K.; Clay, D.; Davies, S.; O'Neill, P.; Mcdaid, P.; Debarge, S.; Steflik, J.; Karmilowicz, M.; Wong, J. W.; Faber, K. Chemoenzymatic Asymmetric Synthesis Of Pregabalin Precursors Via Asymmetric Bioreduction Of B-Cyanoacrylate Esters Using Ene-Reductases. *The Journal Of Organic Chemistry* **2013**, *78* (4), 1525-1533.

(43) Saito, K.; Thiele, D. J.; Davio, M.; Lockridge, O.; Massey, V. The Cloning And Expression Of A Gene Encoding Old Yellow Enzyme From Saccharomyces Carlsbergensis. *Journal of Biological Chemistry* **1991**, *266* (31), 20720-20724.

(44) Akeson, A.; Theorell, H. Molecular Weight And FMN Content Of Crystallin Old Yellow Enzyme. *Archives of Biochemistry and Biophysics* **1956**, *65* (1), 439-448.

(45) Fraaije, M. W.; Mattevi, A. Flavoenzymes: Diverse Catalysts With Recurrent Features. *Trends in Biochemical Sciences* **2000**, *25* (3), 126-132.

(46) Karplus, P. A.; Fox, K. M.; Massey, V. Flavoprotein Structure And Mechanism. 8. Structure-Function Relations For Old Yellow Enzyme. *The FASEB Journal* **1995**, *9* (15), 1518-1526.

(47) Hubbard, P. A.; Liang, X.; Schulz, H.; Kim, J. J. The Crystal Structure And Reaction Mechanism Of Escherichia Coli 2,4-Dienoyl-Coa Reductase. *Journal Of Biological Chemistry* **2003**, *278* (39), 37553-37560.

(48) Mordaka, P. M.; Hall, S. J.; Minton, N.; Stephens, G. Recombinant Expression And Characterisation Of The Oxygen-Sensitive 2-Enoate Reductase From Clostridium Sporogenes. *Microbiology (Reading)* **2018**, *164* (2), 122-132.

(49) Tu, X.; Hubbard, P. A.; Kim, J.-J. P.; Schulz, H. Two Distinct Proton Donors At The Active Site Of Escherichia Coli 2,4-Dienoyl-Coa Reductase Are Responsible For The Formation Of Different Products. *Biochemistry* **2008**, *47* (4), 1167-1175.

(50) Reed, T.; Lushington, G. H.; Xia, Y.; Hirakawa, H.; Travis, D. M.; Mure, M.; Scott, E. E.; Limburg, J. Crystal Structure Of Histamine Dehydrogenase From Nocardioides Simplex. *Journal Of Biological Chemistry* **2010**, *285* (33), 25782-25791.

(51) Boyd, G.; Mathews, F. S.; Packman, L. C.; Scrutton, N. S. Trimethylamine Dehydrogenase Of Bacterium W3A1 Molecular Cloning, Sequence Determination And Over-Expression Of The Gene. *FEBS Letters* **1992**, *308* (3), 271-276.

(52) Loechel, C.; Basran, A.; Basran, J.; Scrutton, N. S.; Hall, E. A. H. Using Trimethylamine Dehydrogenase In An Enzyme Linked Amperometric Electrode Part 1. Wild-Type Enzyme Redox Mediation. *Analyst* **2003**, *128* (2), 166-172, 10.1039/B211895E.

(53) Sofia, H. J.; Chen, G.; Hetzler, B. G.; Reyes-Spindola, J. F.; Miller, N. E. Radical SAM, A Novel Protein Superfamily Linking Unresolved Steps In Familiar Biosynthetic Pathways With Radical Mechanisms: Functional Characterization Using New Analysis And Information Visualization Methods. *Nucleic Acids Research* **2001**, *29* (5), 1097-1106.

(54) Broderick, J. B.; Duffus, B. R.; Duschene, K. S.; Shepard, E. M. Radical S-Adenosylmethionine Enzymes. *Chemical Reviews* **2014**, *114* (8), 4229-4317.

(55) Oberg, N.; Precord, T. W.; Mitchell, D. A.; Gerlt, J. A. Radicalsam.Org: A Resource To Interpret Sequence-Function Space And Discover New Radical SAM Enzyme Chemistry. *ACS Bio & Med Chem Au* **2022**, *2* (1), 22-35.

(56) Do, T.; Link, A. J. Protein Engineering In Ribosomally Synthesized And Post-Translationally Modified Peptides (Ripps). *Biochemistry* **2023**, *62* (2), 201-209.

(57) Zhong, G.; Wang, Z. J.; Yan, F.; Zhang, Y.; Huo, L. Recent Advances In Discovery, Bioengineering, And Bioactivity-Evaluation Of Ribosomally Synthesized And Post-Translationally Modified Peptides. *ACS Bio Med Chem Au* **2023**, *3* (1), 1-31.

(58) Li, Y.; Rebuffat, S. The Manifold Roles Of Microbial Ribosomal Peptide–Based Natural Products In Physiology And Ecology. *Journal Of Biological Chemistry* **2020**, *295* (1), 34-54.

(59) Walker, J. A.; Hamlish, N.; Tytla, A.; Brauer, D. D.; Francis, M. B.; Schepartz, A. Redirecting Ripp Biosynthetic Enzymes To Proteins And Backbone-Modified Substrates. *ACS Central Science* **2022**, *8* (4), 473-482.

(60) Burkhart, B. J.; Hudson, G. A.; Dunbar, K. L.; Mitchell, D. A. A Prevalent Peptide-Binding Domain Guides Ribosomal Natural Product Biosynthesis. *Nature Chemical Biology* **2015**, *11* (8), 564-570.

(61) Davis, K. M.; Schramma, K. R.; Hansen, W. A.; Bacik, J. P.; Khare, S. D.; Seyedsayamdost, M. R.; Ando, N. Structures Of The Peptide-Modifying Radical SAM Enzyme Suib Elucidate The Basis Of Substrate Recognition. *Proceedings Of The National Academy Of Sciences* **2017**, *114* (39), 10420-10425.

(62) Brennan, R. G. The Winged-Helix DNA-Binding Motif: Another Helix-Turn-Helix Takeoff. *Cell* **1993**, *74* (5), 773-776.

(63) Song, H.; Bremer, B. J.; Hinds, E. C.; Raskutti, G.; Romero, P. A. Inferring Protein Sequence-Function Relationships With Large-Scale Positive-Unlabeled Learning. *Cell Systems* **2021**, *12* (1), 92-101.E108.

(64) Peters, C.; Frasson, D.; Sievers, M.; Buller, R. Novel Old Yellow Enzyme Subclasses. *Chembiochem* **2019**, *20* (12), 1569-1577.

(65) Bayat, A. Science, Medicine, And The Future: Bioinformatics. *British Medical Journal* **2002**, *324* (7344), 1018-1022.

(66) Steffen-Munsberg, F.; Vickers, C.; Kohls, H.; Land, H.; Mallin, H.; Nobili, A.; Skalden, L.; Van Den Bergh, T.; Joosten, H.-J.; Berglund, P.; Et Al. Bioinformatic Analysis Of A PLP-Dependent Enzyme Superfamily Suitable For Biocatalytic Applications. *Biotechnology Advances* **2015**, *33* (5), 566-604.

(67) Alderson, R. G.; De Ferrari, L.; Mavridis, L.; Mcdonagh, J. L.; Mitchell, J. B.; Nath, N. Enzyme Informatics. *Current Topics in Medicinal Chemistry* **2012**, *12* (17), 1911-1923.

(68) Hudson, G. A.; Burkhart, B. J.; Dicaprio, A. J.; Schwalen, C. J.; Kille, B.; Pogorelov, T. V.; Mitchell, D. A. Bioinformatic Mapping Of Radical S-Adenosylmethionine-Dependent Ribosomally Synthesized And Post-Translationally Modified Peptides Identifies New Cα, Cβ,

And Cγ-Linked Thioether-Containing Peptides. *Journal of the American Chemical Society* **2019**, *141* (20), 8228-8238.

(69) Kloosterman, A. M.; Medema, M. H.; Van Wezel, G. P. Omics-Based Strategies To Discover Novel Classes Of Ripp Natural Products. *Current Opinion In Biotechnology* **2021**, *69*, 60-67.

(70) Uniprot Consortium, T. Uniprot: The Universal Protein Knowledgebase. *Nucleic Acids Research* **2018**, *46* (5), 2699-2699.

(71) Johnson, M.; Zaretskaya, I.; Raytselis, Y.; Merezhuk, Y.; Mcginnis, S.; Madden, T. L. NCBI BLAST: A Better Web Interface. *Nucleic Acids Research* **2008**, *36* (Suppl_2), W5-W9.

(72) Copp, J. N.; Akiva, E.; Babbitt, P. C.; Tokuriki, N. Revealing Unexplored Sequence-Function Space Using Sequence Similarity Networks. *Biochemistry* **2018**, *57* (31), 4651-4662.

(73) Atkinson, H. J.; Morris, J. H.; Ferrin, T. E.; Babbitt, P. C. Using Sequence Similarity Networks For Visualization Of Relationships Across Diverse Protein Superfamilies. *PLOS One* **2009**, *4* (2), E4345.

(74) Jacobson, M. P.; Kalyanaraman, C.; Zhao, S.; Tian, B. Leveraging Structure For Enzyme Function Prediction: Methods, Opportunities, And Challenges. *Trends Biochemistry Science* **2014**, *39* (8), 363-371.

(75) Schnoes, A. M.; Brown, S. D.; Dodevski, I.; Babbitt, P. C. Annotation Error In Public Databases: Misannotation Of Molecular Function In Enzyme Superfamilies. *PLOS Computational Biology* **2009**, *5* (12), E1000605.

(76) Muhammed, M. T.; Aki-Yalcin, E. Homology Modeling In Drug Discovery: Overview, Current Applications, And Future Perspectives. *Chemical Biology & Drug Design* **2019**, *93* (1), 12-20.

(77) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A. Highly Accurate Protein Structure Prediction With Alphafold. *Nature* **2021**, *596* (7873), 583-589.

(78) Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. HADDOCK: A Protein−Protein Docking Approach Based On Biochemical Or Biophysical Information. *Journal Of The American Chemical Society* **2003**, *125* (7), 1731-1737.

(79) Wang, T.; Wei, R.; Feng, Y.; Jin, L.; Jia, Y.; Yang, D.; Liang, Z.; Han, M.; Li, X.; Lu, C.; Et Al. Engineering Of Yeast Old Yellow Enzyme OYE3 Enables Its Capability Discriminating Of (E)-Citral And (Z)-Citral. *Molecules* **2021**, *26* (16).

(80) Vázquez-Jiménez, L. K.; Juárez-Saldivar, A.; Gómez-Escobedo, R.; Delgado-Maldonado, T.; Méndez-Álvarez, D.; Palos, I.; Bandyopadhyay, D.; Gaona-Lopez, C.; Ortiz-Pérez, E.; Nogueda-Torres, B.; Et Al. Ligand-Based Virtual Screening And Molecular Docking Of Benzimidazoles As Potential Inhibitors Of Triosephosphate Isomerase Identified New Trypanocidal Agents. *International Journal Of Molecular Sciences* **2022**, *23* (17), 10047.

(81) Derreumaux, P.; Schlick, T. The Loop Opening/Closing Motion Of The Enzyme Triosephosphate Isomerase. *Biophysical Journal* **1998**, *74* (1), 72-81.

(82) Ponder, J. W.; Case, D. A. Force Fields For Protein Simulations. *Advances In Protein Chemistry* **2003**, *66*, 27-85.

(83) Karplus, M.; Petsko, G. A. Molecular Dynamics Simulations In Biology. *Nature* **1990**, *347* (6294), 631-639.

(84) Blanc, B.; Gerez, C.; Ollagnier De Choudens, S. Assembly Of Fe/S Proteins In Bacterial Systems: Biochemistry Of The Bacterial ISC System. *Biochimica Et Biophysica Acta (BBA) - Molecular Cell Research* **2015**, *1853* (6), 1436-1447.

(85) Mccarthy, E. L.; Booker, S. J. Biochemical Approaches For Understanding Iron-Sulfur Cluster Regeneration In Escherichia Coli Lipoyl Synthase During Catalysis. *Methods Enzymology* **2018**, *606*, 217-239.

(86) Takahashi, Y.; Tokumoto, U. A Third Bacterial System For The Assembly Of Iron-Sulfur Clusters With Homologs In Archaea And Plastids * 210. *Journal Of Biological Chemistry* **2002**, *277* (32), 28380-28383.

(87) Nachin, L.; Loiseau, L.; Expert, D.; Barras, F. Sufc: An Unorthodox Cytoplasmic ABC/Atpase Required For [Fe—S] Biogenesis Under Oxidative Stress. *The EMBO Journal* **2003**, *22* (3), 427-437.

(88) Olson, J. W.; Agar, J. N.; Johnson, M. K.; Maier, R. J. Characterization Of The Nifu And Nifs Fe− S Cluster Formation Proteins Essential For Viability In Helicobacter Pylori. *Biochemistry* **2000**, *39* (51), 16213-16219.

(89) Zheng, L.; Cash, V. L.; Flint, D. H.; Dean, D. R. Assembly Of Iron-Sulfur Clusters: IDENTIFICATION OF AN <Em>Iscsua-Hscba-Fdx</Em> GENE CLUSTER FROM <Em>AZOTOBACTER VINELANDII</Em>*. *Journal Of Biological Chemistry* **1998**, *273* (21), 13264-13272.

(90) Mccarty, R. M.; Krebs, C.; Bandarian, V. Spectroscopic, Steady-State Kinetic, And Mechanistic Characterization Of The Radical SAM Enzyme Quee, Which Catalyzes A Complex Cyclization Reaction In The Biosynthesis Of 7-Deazapurines. *Biochemistry* **2013**, *52* (1), 188-198.

(91) Knox, H. L.; Chen, P. Y.-T.; Blaszczyk, A. J.; Mukherjee, A.; Grove, T. L.; Schwalm, E. L.; Wang, B.; Drennan, C. L.; Booker, S. J. Structural Basis For Non-Radical Catalysis By Tsrm, A Radical SAM Methylase. *Nature Chemical Biology* **2021**, *17* (4), 485-491.

(92) Lee, K. H.; Saleh, L.; Anton, B. P.; Madinger, C. L.; Benner, J. S.; Iwig, D. F.; Roberts, R. J.; Krebs, C.; Booker, S. J. Characterization Of Rimo, A New Member Of The Methylthiotransferase Subclass Of The Radical SAM Superfamily. *Biochemistry* **2009**, *48* (42), 10162-10174.

(93) Mccarthy, E. L.; Booker, S. J. Destruction And Reformation Of An Iron-Sulfur Cluster During Catalysis By Lipoyl Synthase. *Science* **2017**, *358* (6361), 373-377.

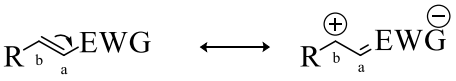
# **Chapter 2.** Expanding the Industry Applications of the Old Yellow Enzyme Superfamily

*This chapter is unpublished work in collaboration with Dr. Janine Copp from the University of British Columbia and members of the Former Lutz Lab at Emory University, Dr. David White, Dr. Samantha Iamurri, and Parisa Keshavarz-Joud.*

## 2.1.Introduction

Chiral building blocks (CBBs) are valuable intermediates in the synthesis of biologically relevant natural products.[1, 2] As fragments of larger synthons, CBBs offer a starting point in synthesizing complex synthons with challenging synthetic routes. The ability to synthesize enantiopure CBBs is of great interest, as a compound's chirality directly influences the substance's industrial use. For example, Citalopram, an FDA-approved selective serotonin reuptake inhibitor (SSRI), has two enantiomeric forms. However, only the S-(+) enantiomer of Citalopram displays therapeutic activity as an SSRI class antidepressant.

**A.**




**B.**

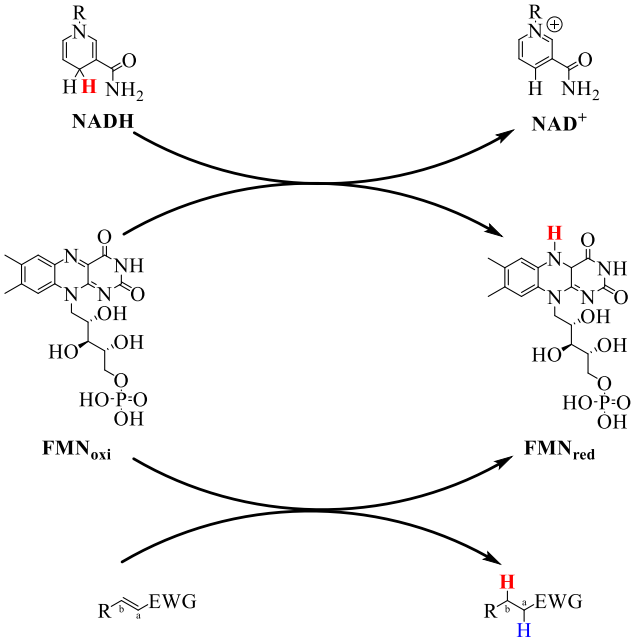



*Figure 2.1.OYE Mechanism A. Electron withdrawing group (EWG) promotes activation of the α/β unsaturated alkene B. Scheme of substrate reduction.*

The asymmetric reduction of unsaturated compounds is one of the most efficient methods of forming enantiopure CBBs. Asymmetric reductions involve the hydrogenation of an activated alkene to generate a stereospecific product.[3] To facilitate reactivity, organocatalysts or biocatalysts are often employed. Biocatalysts are more appealing than organocatalysts as they perform catalysis at similar efficiency while requiring fewer solvents and heavy metals.[4-6]

As mentioned in Chapter 1, OYEs are nicotinamide-dependent oxidoreductases that catalyze the asymmetric reduction of diverse α/β unsaturated alkenes adjacent to EWG via a flavin cofactor (Figure 2.1).[7-9] Due to OYE biocatalyst's stereospecificity and sizable substrate scope resulting from the commutable EWG, OYEs were employed in the biosynthesis of industrially relevant chiral building blocks.[1, 10] In 2013, Winkler and coworkers substantiated this by developing a novel OYE-conducted biosynthetic route to (S)-3-aminomethyl-5-methyl hexanoic acid (common name: Pregabalin), an FDA-approved anticonvulsant.[10] OYE ene-reductases have been further attested for industrial use by the use in the biosynthesis of (R)-profens, a class of non-steroidal anti-inflammatory drugs or NSAIDs, and natural product building blocks.[11, 12]

Historically, OYEs were collated into three distinct groups: classical OYEs, thermophilic-like OYEs, and 2-ER OYEs. In this classification system, classical OYEs are represented by OYE1 and OYE2 and can present in solution as monomers and dimers.[13] Thermophilic-like OYEs, as represented by YqjM, exist in solution in a wide range of oligomeric states.[9] 2-ER OYEs, as represented by DCR from *E. coli*, are identified by their multiple domains and characteristic [4Fe-4S] coordination site. However, in 2019 they were reorganized into five classes (Class I-V), where Class I represents classical OYEs, and Class II represents thermophilic-like OYEs. Classes are organized by mutually conserved sequences, distinct covalent interactions with FMN, and species of origin.[9] Affording the newly designated class III, to be represented by YqiG from *B. subtilis*,

class IV, to be represented by Ppo-ER3 from *Paenibacillus polymyxa*, and class V to be represented by Lla-ER from *Lactococcus lactis*.[9]

In contrast to the 126 OYEs characterized (via biochemical, computational or structural methods) in literature as of 2018, 70,366 protein sequences (PF00724) were annotated as homologous members of a larger OYE superfamily in the Pfam database.[14, 15] While it is recognized that a startling number of incorrectly annotated proteins abide within automated annotation systems, we hypothesized that the high sequence similarity among OYEs and the well-documented conserved residues reduce associated annotation error.[8, 9, 16] It was, therefore, impossible to render the full catalytic ability of the OYE superfamily by considering less than 0.2% of the superfamily.

Known OYE homologs share similar reaction schemes; however, they exhibit diversity in substrate scope, enatiopreference, or propensity to dimer/trimerize. To quickly and efficiently explore this unchartered sequence space, the Lutz lab developed a systematic method of classifying and characterizing members of the OYE superfamily utilizing classical biochemical techniques and a bioinformatic tool, sequence similarity networks (SSN).[17] Sourcing sequence data cataloged in the PFam database under "PF00724", the OYE superfamily SSN was generated to conserve previous OYE classifications: classical OYEs (Cluster 1), 2-enoate reductases OYEs (Cluster 2), and thermophilic-like OYEs (Cluster 3) (Figure 2.2). SSNs are optimal for this study as they allow for the quick visualization of extensive sequence data and promote homolog function prediction by grouping proteins of similar sequence identities into intraconnected clusters.[18] As mentioned in Chapter 1, the sequence function relationship affords the assumption of cluster isofuctionality when a SSN is encoded with high stringency.

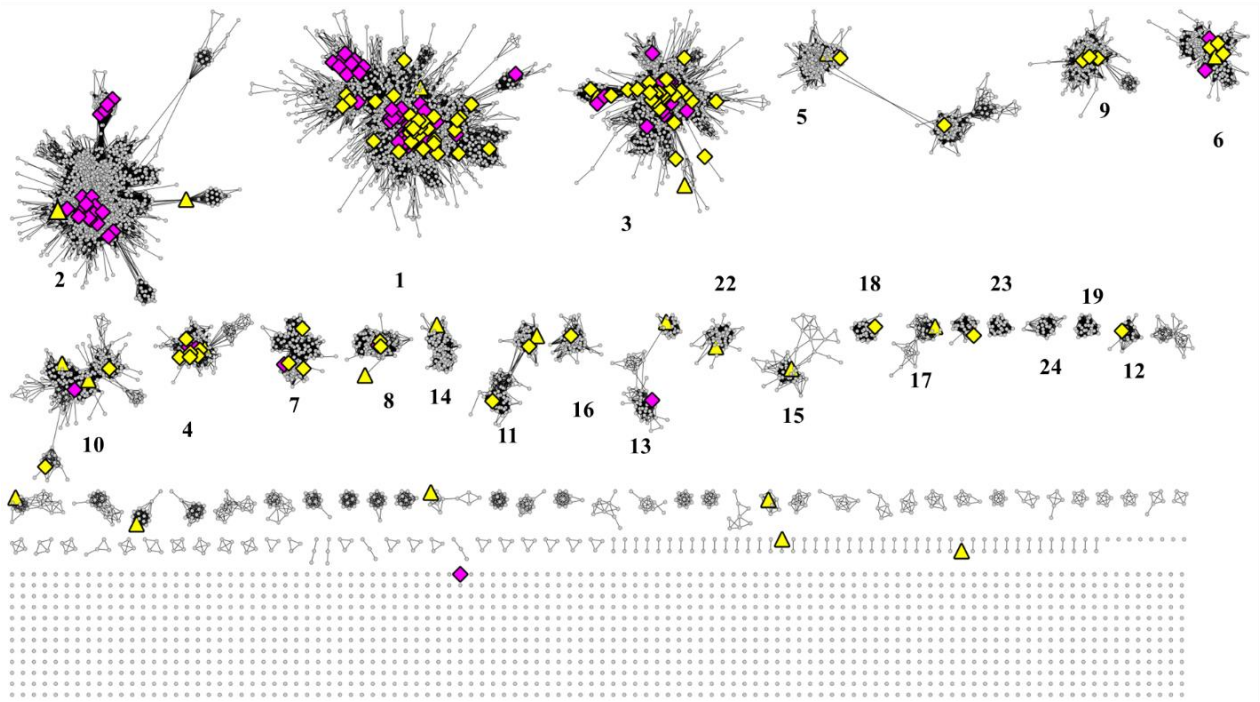



*Figure 2.2. 2018 Sequence Similarity Networks. OYE superfamily SSN 50% ID $e^{-85}$ curated in 2018. Gray circles represent nodes, black lines connecting nodes represent edges, and groups of interconnected nodes and edges represent clusters. OYE homologs characterized in the literature (pink diamonds) and OYE protein library (yellow diamonds and triangles indicating two separate batches).*

From this SSN, 125 OYEs were selected from under-sampled diverse clusters to create a protein library that could represent the biochemical breath of the OYE superfamily. This protein library was screened against substrate panels to investigate the propensity to (1) perform previously described OYE activity (Figure 2.3, Mix I), (2) generate chiral centers on pharmaceutically relevant furanones (Figure 2.3. Substrate Mixes. Mix I (including (S)-carvone, ketoisophorone, cinnamaldehyde, and 4-phenylbut-3-yn-2-one), Mix II (substituted furanones), Mix III(precursor to the Roche ester, nitriles, and nitro alkene), Mix IV (alkane products of MixI), and Mix V (, Mix II), (3) substrates with difficult chemistry for OYEs (Figure 2.3, Mix III), and (4) exhibit desaturase activity (i.e., oxidation) (Figure 2.3, Mix IV) and on sterically diverse industrially relevant substrates (Figure 2.3, gray circle)

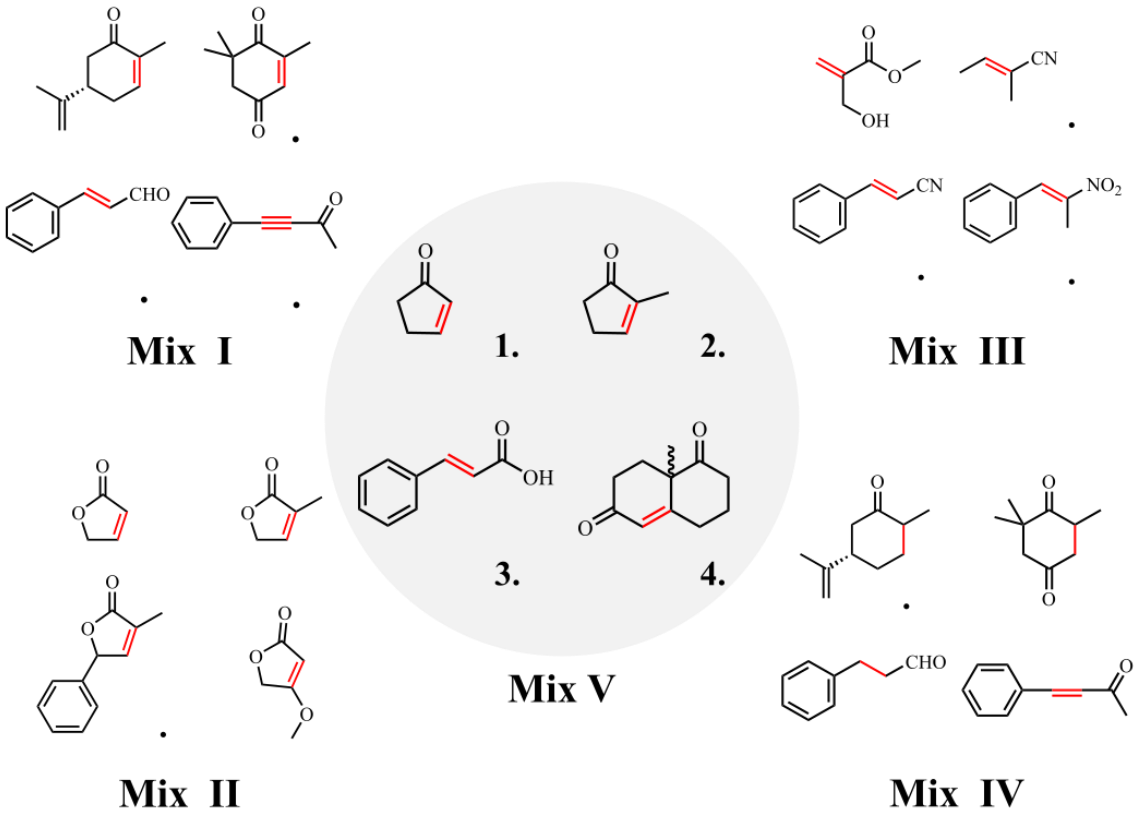



*Figure 2.3. Substrate Mixes. Mix I (including (*S*)-carvone, ketoisophorone, cinnamaldehyde, and 4-phenylbut-3-yn-2-one), Mix II (substituted furanones), Mix III(precursor to the Roche ester, nitriles, and nitro alkene), Mix IV (alkane products of MixI), and Mix V (cyclopent-2-en-one (1) and 2-methylcyclopent-2-en-1one (2), a model MCA cinnamic acid (3), and The Wieland–Miescher ketone (4))*

However, at the time, the OYE superfamily had yet to be interrogated for activity leading to intermediates of sterically hindered decalone-derived natural products such as 8a-methyl-3,4,8,8a-tetrahydronaphthalene-1,6(2H,7H)-dione, commonly known as the Wieland–Miescher ketone **(4).** This bioactive compound bicyclic diketone decalone building block applied in the total synthesis of steroids and higher terpenoids (i.e., sesquiterpenoids, di-terpenoids, and tri-terpenoids) with antibacterial, antiviral, antifungal, and antineurodegenerative properties.[19-23] The generation of chiral centers on unsaturated aliphatic mono-carboxylic acid (MCA) compounds is of similar

interest, as MCAs are versatile chiral building blocks for the production of industrial and cosmetic compounds. However, while OYEs readily facilitate the reduction of alkenes adjacent to dicarboxylic acids, limited reactivity has been observed for MCA systems.[7, 24] In contrast, organocatalysts of MCAs often require strong acids due to difficulties during separation; thus, a biosynthetic pathway to produce enantiopure MCAs is needed.[25]

To ascertain biosynthetic pathways to MCA and decalone chiral building blocks, we utilized the OYE protein library described above to screen the superfamily for the target activity. Cyclopent-2-en-one **(1)** 2-methylcyclopent-2-en-1one **(2)**, a model MCA cinnamic acid **(3)**, and the Wieland–Miescher ketone **(4)** were included in the panel. **(1)** and **(2)** act as internal controls for traditional OYE chemistry as they have been shown to be readily converted in literature (Figure 2.4).

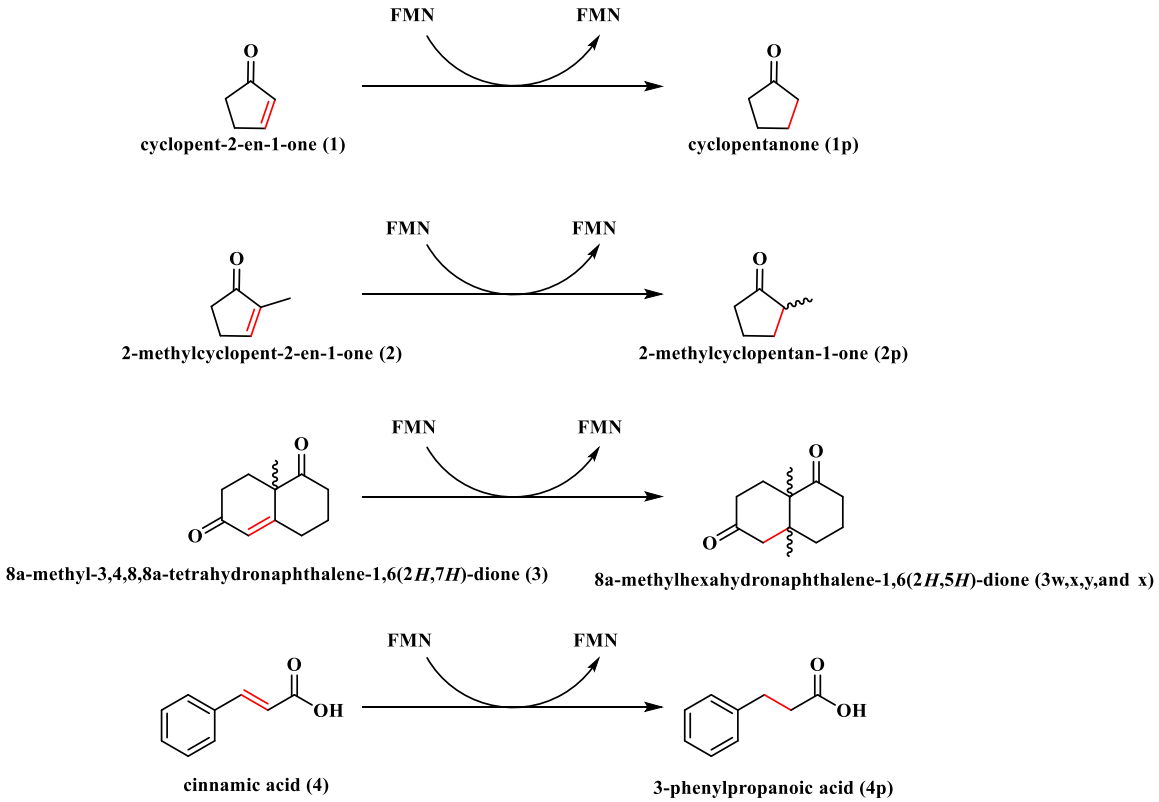



*Figure 2.4. Current substrate panel and expected product after incubation with ene-reductase. The reactive site is indicated by a red bond.*

## 2.2.Results and Discussion

### 2.2.1. OYE Protein Library Composition

The protein library comprises representatives that span 32 clusters across the OYE SSN, including two singletons and 28 “major” clusters, defined as any cluster containing over 100 non-repetitive sequences (Figure 2.2. 2018 Sequence Similarity Networks. OYE superfamily SSN 50% ID e-85 curated in 2018. Gray circles represent nodes, black lines connecting nodes represent edges, and groups of interconnected nodes and edges represent clusters. OYE homologs characterized in the literature (pink diamonds) and OYE protein library (yellow diamonds and triangles indicating two separate batches).. However, 54.03% of the superfamily’s sequences reside in clusters 1 and 3, so novel sequence space within those clusters was also of interest when designing the library. The library has sequences that span from 100-1500 AA. The library contains previously characterized OYEs, which act as benchmarks, and 119 novel protein sequences. The protein members were codified to reduce bias until all biochemical assays were assessed (Table S2.1). The 125-member protein library comprises OYE homologs originating from Bacteria, Plantae, Fungi, Archaea, Protista, Metazoic, Chromista organisms, and Metagenomic DNA. The bioinformatic evaluation shows that the OYE superfamily is primarily (~70%) composed of enzymes of bacterial origin to date.

### 2.2.2. IVTT to efficiently produce protein.

The PURExpress® In-Vitro Transcription/Translation (IVTT) system from New England BioLabs (NEB) was utilized efficiently to obtain protein without large-scale purification. The IVTT system contains purified intercellular components (i.e., DNA polymerases, tRNAs, ribosomes, AAs, etc.) necessary for transcription and subsequent translation of a plasmid in under 3 hrs. On average, the

system produces 10–200 µg/mL of protein of interest, which is more than adequate for detection by analytical chromatography.[26] This methodology has several advantages to traditional heterogeneous transformation and expression. The most obvious is a time cost reduction, where traditional methods consume a minimum of 72 hrs to produce crude protein. An additional advantage is crude protein quality, where the IVTT system has been shown to reduce detectable background activity.

The IVTT reaction vials were spiked with an RNAase inhibitor, FMN, and FAD to enhance enzyme solubility. FMN and FAD are essential OYE cofactors that increase yield when introduced during protein expression.[27] Upon completion of in-vitro OYEs synthesis, protein members were anaerobically incubated overnight in a reaction mixture composed of the substrate, a nicotinamide source, glucose dehydrogenase, glucose, and buffer. Reactions were then quenched and analyzed via gas chromatography-mass spectrometry (GC-MS) for detectable product formation.

### 2.2.3. Biochemical Evaluation of Protein Library

Enzymatic reduction of **(2)** and **(4)** generates more than one enantiomeric product (

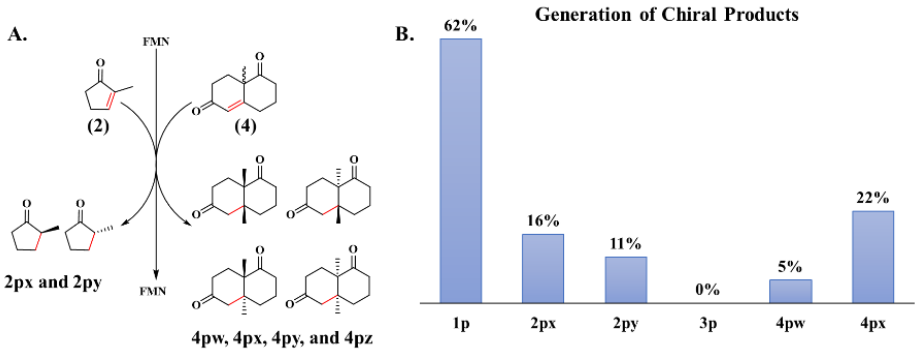



Figure 2.5A). With a chiral column, we were able to detect the formation of each product. However, further differentiation is not possible at this time, so product enantiomers will be referred

to as w, x, y, or z when appropriate. These letters are in order of retention time instead of exact isomer. 62% of the OYE library generated detectable products or was “active” with at least one substrate in the panel (Figure 2.5B).

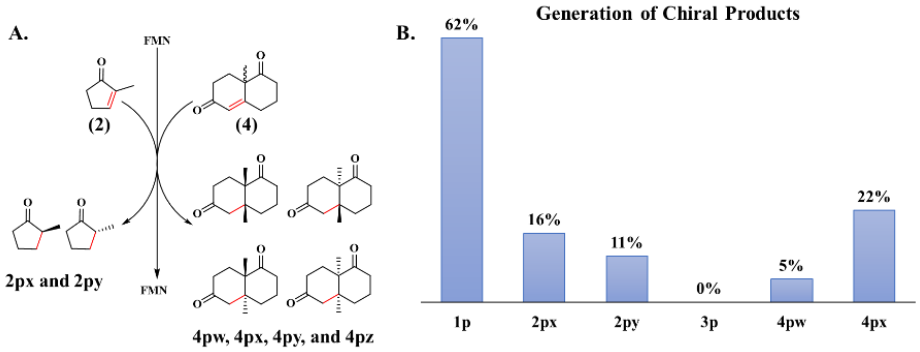



*Figure 2.5. A. Proposed chiral products of (2) and (4). B. Percentage of OYE library members with the ability to generate each product where x and y represent different enantiomers of each compound.*

Although OYEs share high sequence similarity, the predominance of enantiomer formation varies with the enzyme system. OYE1, for example, was found to consume 100% of **(2)**; however, it generates a racemic mix of **(2px):(2py),** while 3D1, a thermophilic-like OYE of bacterial origin, also consumed 100% **(2)** however produced 93% (**2px):(2py)** (Figure 2.6).

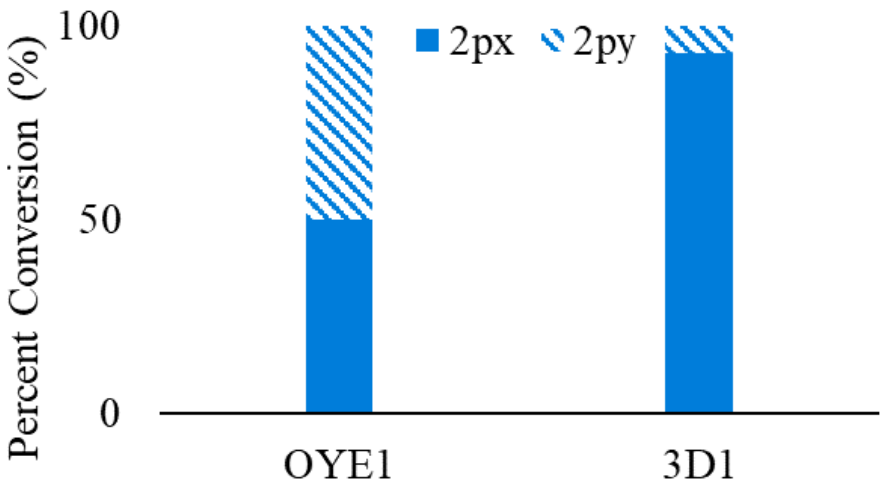



*Figure 2.6. OYE homologs have unique product enantiomeric preference*

No enzyme was found to produce **(3p)**, indicating the inability of our panel to form chiral centers on MCA substrates. However, 35% of our library was able to generate a chiral center at >1% conversion with (**4).** Although **(4)** can produce four enantiomeric products indicated as **(4pw)**, **(4px)**, **(4py)**, and **(4pz)** (Figure 2.5A), we only detected two. The enzymes in the library prefer generating the **(4px)** enantiomer with enzyme 3F3, providing the best conversion of 28.8 ± 0.7%. 3F3 is a codification of a previously characterized Cluster 1 enzyme, N-ethylmaleimide reductase (NemA) from *E. coli* (Accession code: P77258). While NemA has been shown to be involved in nitroglycerin metabolism, it had not previously been demonstrated to reduce sterically hindered enones or generate decalone chiral building blocks.

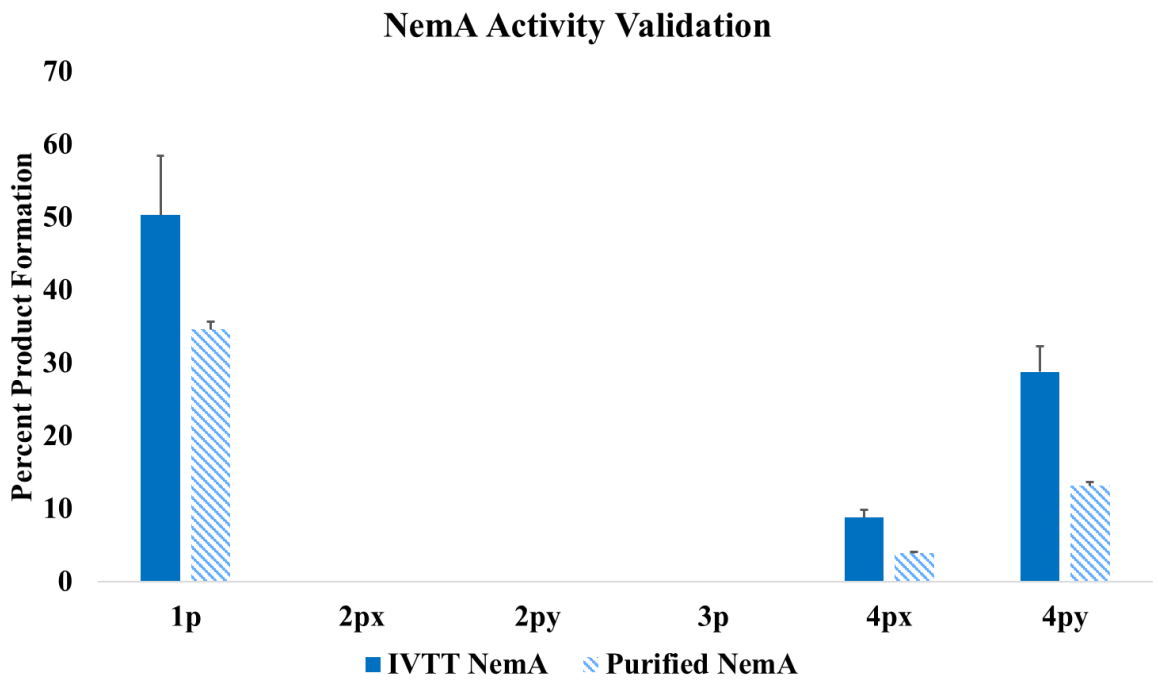



*Figure 2.7. Validation of activity observed in enzyme expressed with IVTT system compared to the purified enzyme.*

To confirm activity from the IVTT system, NemA was heterologously expressed in *E. coli* BL21 cells and purified by anion exchange column. While activity in the IVTT system appeared ~31.1% higher, the substrate trends remained the same (Figure 2.7.). Differences in activity, in part, result from higher enzyme concentration in the IVTT system (10–200 μg/mL) versus the purified system (9.879 μg/mL). Additionally, some enzyme systems tend to lose activity as they undergo purification stages.[28]

### 2.2.4. Solubility characterization of the OYE Protein Library

To validate the overall ability of the OYE protein library in conjunction with our standard operating procedure to represent the breadth of the OYE superfamily's activity, the solubility under the study's standard expression procedure was assessed. This protocol included aerobic expression

and purification followed by an enzymatic reaction with the substrate, NAD(P)H, and a nicotinamide regeneration system, all under anaerobic conditions. Genes of all members were codon optimized for *E. coli* and inserted into a pet28a plasmid. Each plasmid was heterologously expressed, analyzed on sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) gel, and then visually quantified on a five-point scale. Akin to a five-point Likert scale, each library member was ranked qualitatively from most soluble to least soluble. The scale was then converted to a numerical system spanning from 0-5 based on the ratio of the enzyme, at the corresponding molecular weight, within the soluble (S) or insoluble (I) fraction post-cell lysis (Figure 2.8A). On this scale, (0)-zero represented a protein with "no expression," as indicated by no protein in either the soluble or insoluble fractions. (1)-one represented "insoluble" expression where the entirety of expressed protein appeared in the insoluble fraction, (2)-two represented "mostly insoluble expression, where the enzyme predominantly expressed insolubly, and there was detectable soluble protein. (3)-three represented a "50:50 soluble" expression where, as the name implies, approximately half of the protein appeared in the soluble fraction while the remainder expressed insolubly. (4)-three represented a protein with "soluble" expression where either the entirety of the enzyme appeared in the soluble fraction or predominantly expressed soluble, and there was detectable insoluble protein (Figure 2.8A).

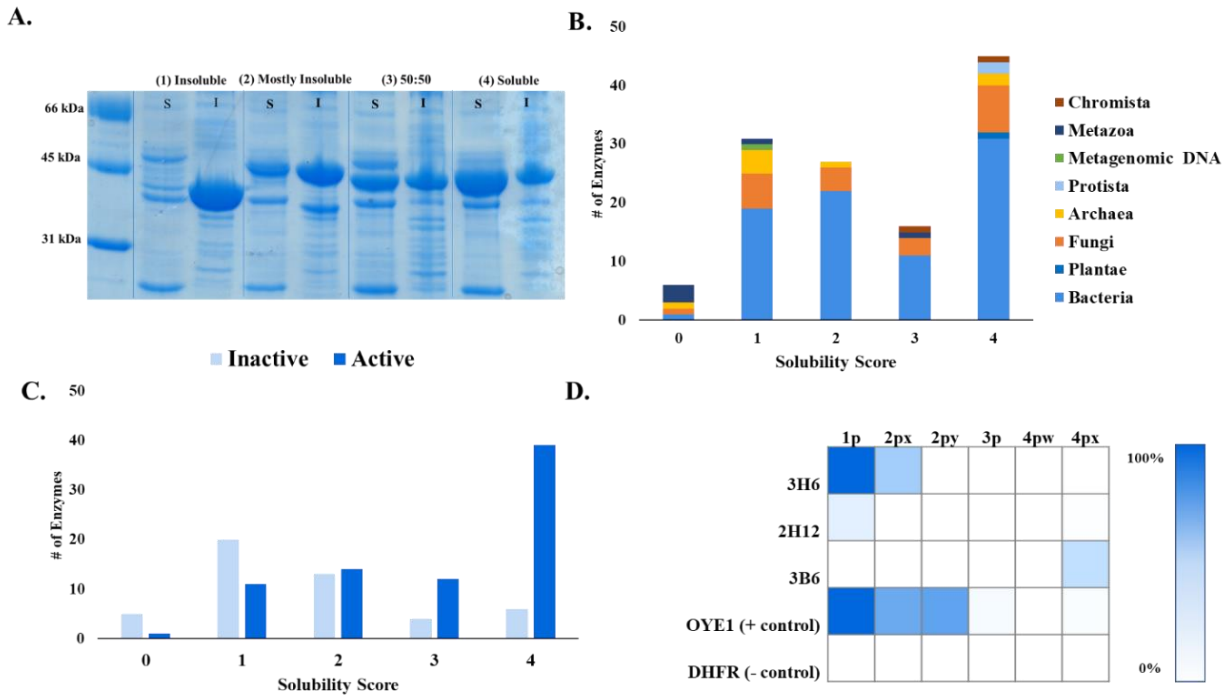



*Figure 2.8. Characterization of Reactivity. A) (1) Insoluble represented by protein A0A0M4QXT8 (C3), (2) Mostly insoluble represented by A0A2G2QFR0 (D3), (3) 50:50 represented by protein W9ATX1 (E3), (4) soluble represented by protein A0A2D5KJ34 (F3). B) solubility score by species of origin. C) Active enzymes by solubility score. D) OYE library members identified as active with new substrate panel with substrate conversion.*

### 2.2.5. Solubility Relative to Activity and Species of Origin

As a result of this study, it was determined that the library is reasonably soluble, with 34.1% expressing soluble, 15.1% 50:50 soluble, 21.4% mostly insoluble, 24.6% insoluble, and 4.8% of the library not expressing under the experimental conditions. Those ranked 2, 3, and 4 (~70% of the library) are determined to be purifiable upon overexpression. ~8% of the library members that were classified as not expressing or expressed insolubly correlated to species of origin, where, unsurprisingly, those of eukaryotic origin were predominantly non-expressing under our conditions (Figure 2.7.B). This may indicate that these proteins are aggregating to the level of phase change, making the SDS page gel appear as if the enzyme was not expressing.[29] However, 27.1% of enzymes ranked 0 or 1 displayed activity as described in David et al. The active no-

expressing and insoluble enzymes (Figure 2.8C) indicate that the use of purified intercellular components may overcome insoluble overexpression by traditional biochemical means. After probing the OYE library for activity with the substrate panel, three OYEs (J3P9F5 (3H6), V7I5A8 (2H12), and C0ZVN3 (3B6)) previously designated as inactive in White et al. were found to be active (Figure 2.7.D). However, it is essential to note that >80% of the no expression and >70% of the insoluble designated enzymes displayed no activity under our conditions. In addition, as the solubility of the enzymes increased, so did the ratio of active to inactive enzymes, indicating that initial solubility screening and subsequent exchanging of library members would improve panel quality (Figure 2.8C).

A significant flaw in implementing this panel to discover novel OYE activity is the meticulous conservation expression conditions. These expression conditions, aerobic with excess flavin, may be adequate for the soluble expression of traditional OYEs; however, they may dramatically inhibit the expression of non-traditional OYEs. 2% of our enzyme library originates from Cluster 2, which is comprised of multidomain 2-enoate reductases that have been bioinformatically confirmed to share an additional conserved FeS cluster binding motif: $CX_{2-3}CX_{2-3}CX_{12}C$. Under the standard operating procedure, these bacteria lack FeS assembly and incorporation chaperons. This lack of a properly incorporated FeS cluster could result in insoluble expressing protein, especially if the cluster acts as a structural support. It is difficult to determine whether soluble inactive enzymes are inherent to the system or a result of incompatible expression conditions.

**2.3.Conclusion**

To date, the OYE superfamily has been shown to contain >120,000 oxidoreductases, many of which perform industrially relevant chemistries. In 2018, we generated an SSN curated from the

PFam entry PF00724 (now hosted on UniProt and InterPro under the same accession code). From this OYE SSN, an OYE homolog protein library spanning 30 clusters, 20 of which were unexplored, was generated to characterize unexplored OYE sequence space. However, without initial screening, proteins may be present outside of the solution where interaction with the substrate is undetectable. It is evident from the results of the described solubility study that the expression and reaction conditions employed did not allow for the full catalytic ability of the OYE superfamily representatives to be observed. The standard expression procedure that afforded quick and efficient screening did not account for the unique expression conditions required for soluble expression of non-bacterial enzymes or enzymes with iron-sulfur cluster metal cofactors. This conclusion is validated by inactive non-expressing enzymes primarily originating from archaea and eukaryotic (fungi and metazoa) species. Additionally, all bioinformatically predicted FeS-containing enzymes presented as insoluble, mostly insoluble, or inactive. Rescreening non-expressing and insoluble OYE library members with tailored expression conditions would provide additional insights into the actual catalytic ability of the OYE superfamily. This study also sustained that while a solubility test is not a definitive predictor of activity, it does provide a quick initial validation of a protein library member's ability to represent a broader superfamily.

Interestingly, upon expanding the substrate panel to include MCA and decalone chiral building blocks, three novel OYEs, previously designated as inactive in our original study, and a characterized OYE, NemA, displayed the generation of a chiral center on **(4).** These results indicate that OYE library members characterized as inactive with reasonable solubilities (2-4) may have a substrate scope beyond our selection criteria and could perform novel activity. These results demonstrated that the OYE superfamily has the ability to convert sterically hindered alkenes and

enones. These results are further supported by reports of a biosynthetic route to **(4)** derivatives via dehydrogenases and metagenomic ene-reductases.[30, 31]

However, there was no observable activity for MCA compounds. Members of the OYE superfamily have observed the generation of chiral centers on MCA compounds. However, the activity is thought to be isolated to members of the 2-ER metalloenzyme family.[32] Due to the oxygen sensitivity of a majority of metalloenzymes, these enzymes are not appealing for industrial use. It is of future interest to explore how the presence of a metal cofactor or lack thereof aids in the generation of MCA chiral building blocks.

## 2.4.Experimental

### 2.4.1. Materials

Unless otherwise noted, all substrates were purchased from Sigma Aldrich. The *E. coli* optimized genes of the 125-member OYE protein library incorporated into pET28a vectors were received from the DOE Full Joint Genome Institute.

### 2.4.2. Transformation and Protein overexpression

Acquired OYE library members in pET28a plasmids were transformed into chemically competent E.coli DH5α ($OYE_x$:DH5α) and BL21 ($OYE_x$:BL21) cells. $OYE_x$:BL21 cells and $OYE_x$:DH5α cells were used to inoculate in Luria-Bertani (LB) Broth supplemented with kanamycin (50 μg•$mL^{-1}$) to grow overnight while shaking at 250 r.p.m, at 37 °C. $OYE_x$:DH5α cells were grown overnight add harvested for additional plasmids following the QIAquick Kit procedure (Qiagen). Aliquots of the $OYE_x$:BL21 cells overnight pre-cultures were used to inoculate 5 mL LB broth. Cultures shook 250 r.p.m at 37 °C until $OD_{600}$ = 0.6-0.7, then temperatures were reduced to

22 °C, and expression was initiated by IPTG (0.3 mM). Expression was allowed to continue overnight (approximately 16-18 hrs), and then samples were chilled at 4 °C.

### 2.4.3. Solubility Test

Expressed cells were harvested by centrifugation at 4000 g for 15 min at 4 °C. After discarding the supernatant, the pellet was resuspended in 250 µL chilled BugBuster® Protein Extraction Reagent (Millipore) and incubated for 10–15 min on ice while rocking. Soluble and insoluble fractions were collected by centrifugation at 4000 g for 15 min at 4 °C. The supernatants (soluble protein fractions) were transferred into fresh tubes, while the remaining pellet was resuspended in a volume equal to the soluble fraction and represented the insoluble protein fractions. Both protein fractions were analyzed on a 5 %:12 % stacking SDS-PAGE gel.

### 2.4.4. IVTT of OYE Library Members

IVTT reactions were assembled according to the PURExpress in vitro protein synthesis kit on a reduced scale and supplemented with additional components to enhance soluble expression. IVTT reaction vials contained library member plasmids as the template DNA (200 ng), PURExpress Solution A, PURExpress Solution B, 20 U of Murine RNAse Inhibitor (NEB), 100 µM FMN, 100 µM FAD, 5 µM GroEL, 10 µM GroES, and OmniPur RNAse Free Water (EMD). All reactions were run simultaneously with OYE1 in pET14b as a positive control, dihydrofolate reductase (NEB), and no enzyme as a negative control. Reaction mixtures were incubated at 37 °C for 2.5 hours. Following synthesis, reactions were quenched by cooling on ice for 5 minutes before being added to the activity assay.

### 2.4.5. Expression and purification of 3F3 (NemA)

3F3:BL21 cells were used to inoculate 10 mL LB Broth supplemented with kanamycin (50 $\mu g \bullet ml^{-1}$) to grow overnight while shaking at 250 r.p.m, at 37 °C. An aliquot of the overnight culture was used to inoculate 500 mL LB broth supplemented with kanamycin (50 $\mu g \bullet mL^{-1}$) and excess FMN. Cultures shook 250 r.p.m until $OD_{600}$ = 0.6-0.7 at 37 °C, then temperatures were reduced to 22 °C, and expression was induced by IPTG (200 $\mu g \bullet mL^{-1}$). Expression was allowed to complete overnight (approximately 16-18 hrs). Expressed cells were harvested by centrifugation at 4000 g for 15 min at 4 °C.

After discarding the supernatant, the pellet was resuspended in chilled buffer A (40 mM Tris-HCl pH 8 and 10 mM NaCl, 2.5 mL/g pellet) supplemented with DNase I (Millipore) and protease inhibitor cocktail. The mixture was sonicated on ice for 4 mins (on-10sec/off-20 sec) and centrifuged at 4000 g for 30 mins. The clarified lysate was loaded on a 5 mL HiTrap Q FF anion exchange column (Cytiva), washed with two column volumes (CVs) of buffer A, followed by a linear gradient 15-18 CVs to 100% Buffer B (40 mM Tris-HCl pH 8.0, 1 M NaCl). Dual UV absorption at 280 nm (protein) and 450 nm (flavin) monitored protein with intact flavin elution. Product fractions were collected on ice, concentrated, and buffer-exchanged using a Millipore filter unit (MWCO: 10 kDa), then further purified by size exclusion chromatography (Superdex 200, 10/300 GL column) with Buffer C (40 mM Tris HCl pH 8.0, 300 mM NaCl). Protein elution was monitored as described above. Fractions containing the OYE library members were collected and assessed for purity by SDS-PAGE. Protein concentration was quantified by beers law at UV absorbance at 280 nm.

### 2.4.6. Biotransformation

Each substrate was dissolved in dimethyl sulfoxide (DMSO) so that the final concentration of DMSO was <0.5 % to aid in substrate solubilization while not affecting enzyme activity. Glucose dehydrogenase and glucose were supplemented into the reaction mix to allow for nicotinamide regeneration. The substrate-specific activity of each OYE was determined anaerobically using solutions degassed in an anaerobic chamber for a minimum of 24 hrs. Reaction vials contained protein extract (250 nM from purified enzyme or 4 µL from IVTT reaction vial), 200 mM substrate in DMSO, NADH (10 µM), NADPH (10 µM), glucose dehydrogenase (2 U), glucose (100 mM) in 50 mM Tris-HCl buffer pH 7.5. Reactions underwent at ambient temperature for 24 hrs, after which the reactions were quenched with 0.5 mM cyclohexanone in ethyl acetate.

### 2.4.7. Analytical methods

Substrates percent conversions were analyzed by gas chromatography-mass spectrometry (GC-MS). GC-MS spectra were obtained on a Shimadzu QP2010 SE instrument equipped with a chiral CycloSil-B column (30 m x 0.32 mm/0.25 µm, Agilent, Santa Clara, CA), an after-column splitter, a flame ionization detector (FID) detector (detector temperature 200 °C, split ration 1:1) and GC-MS detector, using helium as a carrier gas (column flow 3.69 mL/min). The GC-MS had an interface temperature of 200 °C, MS mode, EI; detector voltage, 0.2 kV; mass range, 12-250; scan speed, 833 u/s. The initial oven temperature was held at 55 °C, ramped up at 2.50 °C $min^{-1}$ to 100 °C, again ramped up at 20 °C $min^{-1}$ to 140 °C, then finally increased at 5.00 °C $min^{-1}$ to 210 °C and held for 5 min. Retention times of the internal standard cyclohexanone, substrates, and products (if available) were compared to the retention times of T0 substrates and confirmed by MS Trace analysis. As seen in the equation below, percentage conversions were calculated from the

ratios of product to reactant areas. All GC-MS data were acquired by GC-MS Lab Solutions accompanying software packages (Shimadzu).

**Eq 1:** $\%Conversion = \frac{Area\ of\ Product}{Area\ of\ Unreacted\ Substarte + Area\ of\ Product} x100$

## 2.5 Supplemental Information

*Table 2.1 Summary of OYE Library.*

| UniProt | Generation | Kingdom | Cluster | Solubility | Active |
|---|---|---|---|---|---|
| **B9T8J4** | ▲81 3D7 | Plantae | 3 | 4 | Y |
| **M5B1C9** | 2E12 | Bacteria | 61 | 2 | |
| **V7I5A8** | 2H12 | Bacteria | 15 | 3 | Y |
| **W7ZMG5** | 2H7 | Bacteria | 8 | 3 | |
| **A0YFJ6** | 3A3 | Bacteria | 11 | 2 | Y |
| **B9Y7J4** | 3A4 | Bacteria | 20 | 2 | |
| **Q9JN79** | 3A5 | Bacteria | 4 | 0 | |
| **Q9WYQ9** | 3A7 | Bacteria | 14 | 2 | |
| **R7SZ38** | 3A8 | Fungi | 22 | 1 | |
| **S6WB48** | 3B1 | Bacteria | 2 | 2 | |
| **Q24RN8** | 3B2 | Bacteria | 63 | 2 | |
| **P54524** | 3B4 | Bacteria | 6 | 4 | Y |
| **C0ZVN3** | 3B6 | Bacteria | 7 | 1 | Y |
| **A0A017SC09** | 3B7 | Fungi | 9 | 1 | Y |
| **V4RWU7** | 3C4 | Bacteria | 8 | 4 | Y |
| **Q4AA30** | 3C5 | Bacteria | 13 | 1 | |
| **A8FEW4** | 3D1 | Bacteria | 3 | 3 | Y |

| | | | | | |
|---|---|---|---|---|---|
| **Q4WZ70** | 3D4 | Fungi | 1 | 2 | Y |
| **W0DA85** | 3E1 | Bacteria | 3 | 1 | Y |
| **K6TQP5** | 3E2 | Bacteria | 25 | 3 | Y |
| **P77258** | 3F3 | Bacteria | 1 | 4 | Y |
| **M2Z4K1** | 3F7 | Bacteria | 26 | 2 | |
| **C3MNJ2** | 3F8 | Archaea | 2 | 1 | |
| **K0B444** | 3G1 | Bacteria | 34 | 2 | |
| **Q8PUE0** | 3G3 | Archaea | 10 | 2 | Y |
| **Q2TJB8** | 3G4 | Protista | 1 | 4 | Y |
| **S5YUA6** | 3G5 | Bacteria | 29 | 3 | |
| **J7L5I6** | 3G6 | Bacteria | 5 | 2 | |
| **C0QFC4** | 3H3 | Bacteria | 10 | 2 | |
| **Q02899** | 3H5 | Fungi | 1 | 4 | Y |
| **J3P9F5** | 3H6 | Fungi | 3 | 3 | Y |
| **E3HDQ4** | A1 | Bacteria | 21 | 4 | |
| | | | | | |
| **E9E863** | A10 | Fungi | 1 | 1 | |
| **A0A1Q8DL33** | A11 | Bacteria | 17 | 4 | |
| **A0A1W0XDZ4** | A12 | Metazoa | 1 | 0 | |
| **A0A1H1DXC0** | A2 | Bacteria | 3 | 4 | Y |
| **A0A1H3H521** | A3 | Archaea | 3 | 0 | |
| **A0A1Q3H1G5** | A4 | Bacteria | 1 | 4 | Y |
| **A0A104JBL3** | A5 | Bacteria | 1 | 3 | Y |
| **U9YMT6** | A6 | Bacteria | 6 | 1 | |

| | | | | | |
|---|---|---|---|---|---|
| **M2VAD5** | A7 | Bacteria | 3 | 2 | |
| **A0A1C4MT72** | A8 | Bacteria | 1 | 2 | Y |
| **A0A2D1JR17** | A9 | Bacteria | 6 | 2 | Y |
| **A0A062X679** | B1 | Bacteria | 3 | 1 | |
| **A0A1G6QTQ7** | B10 | Bacteria | 3 | 4 | Y |
| **A0A1Y1W9M1** | B11 | Fungi | 3 | 0 | |
| **B8LTL5** | B12 | Fungi | 3 | 4 | Y |
| **A5VBP0** | B2 | Bacteria | 1 | 4 | Y |
| **A0A0R0D6E4** | B3 | Bacteria | 1 | 4 | Y |
| **A0A1X6MY35** | B4 | Fungi | 1 | 4 | |
| **C7ZM49** | B5 | Fungi | 1 | 3 | Y |
| **A0A2G0VWT1** | B6 | Bacteria | 8 | 2 | Y |
| **A0A2C2U3H6** | B7 | Bacteria | 6 | 4 | Y |
| **A0A1Z9EHY4** | B8 | Archaea | 3 | 4 | Y |
| **A0A0U1LWY3** | B9 | Fungi | 1 | 2 | Y |
| **R4K9R9** | C1 | Bacteria | 3 | 2 | Y |
| **G4HUZ7** | C10 | Bacteria | 7 | 1 | |
| **Q6CUW9** | C11 | Fungi | 3 | 4 | |
| **A0A1S9RTM4** | C12 | Fungi | 3 | 4 | |
| **A0A0P7GE52** | C2 | Bacteria | 1 | 4 | Y |
| **A0A0M4QXT8** | C3 | Bacteria | 3 | 1 | Y |
| **A0A1Y1YHM2** | C4 | Fungi | 1 | 4 | Y |
| **A0A0Q7WCQ9** | C5 | Bacteria | 1 | 2 | |
| **A0A0B7GCC5** | C6 | Bacteria | 1 | 4 | Y |
| **A0A0F9NP68** | C7 | Metagenomic DNA | 10 | 1 | |

| | | | | | |
|---|---|---|---|---|---|
| **D8M5A9** | C8 | Protista | 3 | 4 | Y |
| **W8NW54** | C9 | Fungi | 1 | 4 | |
| **A0A0H4KFX4** | D1 | Bacteria | 3 | 2 | Y |
| **A0A226DQD1** | D10 | Metazoa | 3 | 0 | |
| **A0A0B7N6T2** | D11 | Fungi | 3 | 3 | Y |
| **V5TJE2** | D12 | Archaea | 18 | 1 | |
| **A0A2E0KBB9** | D2 | Bacteria | 1 | 4 | Y |
| **A0A2G2QFR0** | D3 | Bacteria | 1 | 2 | Y |
| **A0A245ZID8** | D4 | Bacteria | 11 | 4 | Y |
| **A0A087B547** | D5 | Bacteria | 3 | 1 | Y |
| **A0A0D6I6F4** | D6 | Bacteria | 4 | 2 | Y |
| **A0A137SCF2** | D7 | Bacteria | 1 | 2 | Y |
| **A0A1D1W904** | D8 | Metazoa | 3 | 1 | |
| **J3GED7** | D9 | Bacteria | 3 | 3 | Y |
| **A0A1Q4W807** | E1 | Bacteria | 3 | 1 | |
| **K5Y813** | E10 | Fungi | 3 | 2 | Y |
| **A0A1H1PA63** | E11 | Bacteria | 7 | 1 | |
| **Q3JA76** | E12 | Bacteria | 23 | 1 | |
| **A0A1Q5C6Q4** | E2 | Bacteria | 3 | 4 | Y |
| **W9ATX1** | E3 | Bacteria | 3 | 3 | Y |
| **A0A239MCT0** | E4 | Bacteria | 1 | 1 | |
| **A0A098SST3** | E5 | Bacteria | 4 | 3 | Y |
| **A0A0C2I951** | E6 | Bacteria | 1 | 4 | Y |
| **U6SIA5** | E7 | Bacteria | 6 | 4 | Y |
| **J3EXQ1** | E8 | Archaea | 8 | 1 | |

| | | | | | |
|---|---|---|---|---|---|
| **A0A1A0W0R3** | E9 | Bacteria | 4 | 1 | Y |
| **A0A166JDD2** | F1 | Bacteria | 1 | 4 | Y |
| **A0A2A3FEX1** | F10 | Bacteria | 7 | 1 | |
| **L7ITH5** | F11 | Fungi | 9 | 2 | Y |
| **Q23090** | F12 | Metazoa | 9 | 3 | |
| **A0A0D5AF70** | F2 | Bacteria | 3 | 1 | |
| **A0A2D5KJ34** | F3 | Archaea | 11 | 4 | Y |
| **A0A1I6Q9M7** | F4 | Bacteria | 1 | 4 | Y |
| **A0A1X2BF87** | F5 | Bacteria | 1 | 4 | Y |
| **A0A2A9JE68** | F6 | Bacteria | 3 | 4 | Y |
| **K3WFC3** | F7 | Chromista | 1 | 4 | Y |
| **A0A100YAA1** | F8 | Bacteria | 1 | 1 | |
| **D0NQX8** | F9 | Chromista | 1 | 3 | Y |
| **A0A014MP39** | G1 | Bacteria | 1 | 2 | |
| **E3LUM4** | G10 | Metazoa | 9 | 0 | Y |
| **G3JMK4** | G11 | Fungi | 5 | 1 | |
| **A0A136H5V4** | G2 | Bacteria | 1 | 4 | Y |
| **A0A0E3SPA7** | G3 | Archaea | 1 | 1 | Y |
| **B5XWM6** | G4 | Bacteria | 1 | 4 | Y |
| **Q87XC7** | G5 | Bacteria | 1 | 4 | Y |
| **A0A161P8U3** | G6 | Bacteria | 4 | 4 | Y |
| **A0A0S3C0K6** | G7 | Bacteria | 12 | 4 | Y |
| **A0A151VNU3** | G8 | Fungi | 16 | 1 | Y |
| **A0A0A2JDP3** | G9 | Fungi | 24 | 1 | |
| **A0A1I5G981** | H1 | Bacteria | 3 | 3 | Y |

| | | | | | |
|---|---|---|---|---|---|
| **A0A1H1V287** | H10 | Bacteria | 5 | 1 | Y |
| **A0A261Y278** | H11 | Fungi | 9 | 4 | Y |
| **A0A1R0ZM98** | H2 | Bacteria | 1 | 2 | |
| **A0A1M7EKV3** | H3 | Bacteria | 1 | 4 | Y |
| **P71278** | H4 | Bacteria | 1 | 4 | Y |
| **A0A0J0YHR8** | H5 | Bacteria | 4 | 4 | Y |
| **A0A1M5T1E8** | H6 | Bacteria | 10 | 1 | |
| **A0A061QIQ1** | H7 | Bacteria | 1 | 4 | Y |
| **A0A2G4DT26** | H8 | Bacteria | 3 | 3 | Y |
| **A0A0R3L3J0** | H9 | Bacteria | 3 | 1 | Y |

## 2.6. Bibliography

(1) Winkler, C. K.; Tasnádi, G.; Clay, D.; Hall, M.; Faber, K. Asymmetric Bioreduction Of Activated Alkenes To Industrially Relevant Optically Active Compounds. *Journal Of Biotechnology* **2012**, *162* (4), 381-389.

(2) Marino, S. T.; Stachurska-Buczek, D.; Huggins, D. A.; Krywult, B. M.; Sheehan, C. S.; Nguyen, T.; Choi, N.; Parsons, J. G.; Griffiths, P. G.; James, I. W.; Et Al. Synthesis Of Chiral Building Blocks For Use In Drug Discovery. *Molecules* **2004**, *9* (6), 405-426.

(3) Trost, B. M. Asymmetric Catalysis: An Enabling Science. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101* (15), 5348-5355.

(4) Carlone, A.; Bernardi, L.; Mccormack, P.; Warr, T.; Oruganti, S.; Cobley, C. J. Asymmetric Organocatalysis And Continuous Chemistry For An Efficient And Cost-Competitive Process To Pregabalin. *Organic Process Research & Development* **2021**, *25* (12), 2795-2805.

(5) Abdussalam-Mohammed, W.; Qasem Ali, A.; O. Errayes, A. Green Chemistry: Principles, Applications, And Disadvantages. *Chemical Methodologies* **2020**, *4* (4), 408-423.

(6) Li, C.-J.; Trost, B. M. Green Chemistry For Chemical Synthesis. *Proceedings Of The National Academy Of Sciences* **2008**, *105* (36), 13197-13202.

(7) Toogood, H. S.; Gardiner, J. M.; Scrutton, N. S. Biocatalytic Reductions And Chemical Versatility Of The Old Yellow Enzyme Family Of Flavoprotein Oxidoreductases. *Chemcatchem* **2010**, *2* (8), 892-914.

(8) Scholtissek, A.; Tischler, D.; Westphal, A. H.; Van Berkel, W. J. H.; Paul, C. E. Old Yellow Enzyme-Catalysed Asymmetric Hydrogenation: Linking Family Roots With Improved Catalysis. *Catalysts* **2017**, *7* (5), 130.

(9) Peters, C.; Frasson, D.; Sievers, M.; Buller, R. Novel Old Yellow Enzyme Subclasses. *Chembiochem* **2019**, *20* (12), 1569-1577.

(10) Winkler, C. K.; Clay, D.; Davies, S.; O'Neill, P.; Mcdaid, P.; Debarge, S.; Steflik, J.; Karmilowicz, M.; Wong, J. W.; Faber, K. Chemoenzymatic Asymmetric Synthesis Of Pregabalin Precursors Via Asymmetric Bioreduction Of B-Cyanoacrylate Esters Using Ene-Reductases. *The Journal Of Organic Chemistry* **2013**, *78* (4), 1525-1533.

(11) Korpak, M.; Pietruszka, J. Chemoenzymatic One-Pot Synthesis Of Γ-Butyrolactones. *Advanced Synthesis & Catalysis* **2011**, *353* (9), 1420-1424.

(12) Pietruszka, J.; Schölzel, M. Ene Reductase-Catalysed Synthesis Of (R)-Profen Derivatives. *Advanced Synthesis & Catalysis* **2012**, *354* (4), 751-756.

(13) Shi, Q.; Wang, H.; Liu, J.; Li, S.; Guo, J.; Li, H.; Jia, X.; Huo, H.; Zheng, Z.; You, S.; Et Al. Old Yellow Enzymes: Structures And Structure-Guided Engineering For Stereocomplementary Bioreduction. *Applied Microbiology And Biotechnology* **2020**, *104* (19), 8155-8170.

(14) Finn, R. D.; Tate, J.; Mistry, J.; Coggill, P. C.; Sammut, S. J.; Hotz, H.-R.; Ceric, G.; Forslund, K.; Eddy, S. R.; Sonnhammer, E. L. L.; Et Al. The Pfam Protein Families Database. *Nucleic Acids Research* **2007**, *36* (Suppl_1), D281-D288.

(15) Uniprot Consortium, T. Uniprot: The Universal Protein Knowledgebase. *Nucleic Acids Research* **2018**, *46* (5), 2699-2699.

(16) Schnoes, A. M.; Brown, S. D.; Dodevski, I.; Babbitt, P. C. Annotation Error In Public Databases: Misannotation Of Molecular Function In Enzyme Superfamilies. *PLOS Computational Biology* **2009**, *5* (12), E1000605.

(17) White, D. W., Iamuri, S., Joud, P., Blue, T.C., Copp, J., Lutz, S. The Hidden Biocatalytic Potential Of The Old Yellow Enzyme Family. *Biorxiv* **2023**.

(18) Atkinson, H. J.; Morris, J. H.; Ferrin, T. E.; Babbitt, P. C. Using Sequence Similarity Networks For Visualization Of Relationships Across Diverse Protein Superfamilies. *PLOS One* **2009**, *4* (2), E4345.

(19) Wieland, P.; Miescher, K. Über Die Herstellung Mehrkerniger Ketone. *Helvetica Chimica Acta* **1950**, *33* (7), 2215-2228.

(20) Eder, U.; Sauer, G.; Wiechert, R. New Type Of Asymmetric Cyclization To Optically Active Steroid CD Partial Structures. *Angewandte Chemie International Edition In English* **1971**, *10* (7), 496-497.

(21) Bradshaw, B.; Bonjoch, J. The Wieland-Miescher Ketone: A Journey From Organocatalysis To Natural Product Synthesis. *Synlett* **2012**, *2012* (03), 337-356.

(22) Cox-Georgian, D.; Ramadoss, N.; Dona, C.; Basu, C. *Therapeutic And Medicinal Uses Of Terpenes*; Medicinal Plants. 2019 Nov 12:333-59. Doi: 10.1007/978-3-030-31269-5_15.,

(23) Hagiwara, H. Aspects In The Total Syntheses Of Higher Terpenoids Starting From Wieland–Miescher Ketone And Its Derivative: A Review. *Natural Product Communications* **2020**, *15* (5), 1934578X20925340.

(24) Stueckler, C.; Hall, M.; Ehammer, H.; Pointner, E.; Kroutil, W.; Macheroux, P.; Faber, K. Stereocomplementary Bioreduction Of A,B-Unsaturated Dicarboxylic Acids And Dimethyl Esters Using Enoate Reductases: Enzyme- And Substrate-Based Stereocontrol. *Organic Letters* **2007**, *9* (26), 5409-5411.

(25) Shi, Q.; He, N.; Gao, F.; Song, Y.; Yu, Y.; Wan, H. Adsorption Of CO On Zn-Cu(I)/HMCM-41. In *Studies In Surface Science And Catalysis*, Sayari, A., Jaroniec, M. Eds.; Vol. 141; Elsevier, 2002; Pp 229-234.

(26) Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T. Cell-Free Translation Reconstituted With Purified Components. *Nature Biotechnology* **2001**, *19* (8), 751-755.

(27) Toogood, H. S.; Scrutton, N. S. Discovery, Characterization, Engineering, And Applications Of Ene-Reductases For Industrial Biocatalysis. *ACS Catalysis* **2018**, *8* (4), 3532-3549.

(28) Mohan Rao, T. J.; Goyal, A. Purification, Optimization Of Assay, And Stability Studies Of Dextransucrase Isolated From Weissella Cibaria JAG8. *Preparative Biochemistry & Biotechnology* **2013**, *43* (4), 329-341.

(29) Nikfarjam, S.; Jouravleva, E. V.; Anisimov, M. A.; Woehl, T. J. Effects Of Protein Unfolding On Aggregation And Gelation In Lysozyme Solutions. *Biomolecules* **2020**, *10* (9).

(30) Bertuletti, S.; Bayout, I.; Bassanini, I.; Ferrandi, E. E.; Bouzemi, N.; Monti, D.; Riva, S. Biocatalytic Approaches To The Enantiomers Of Wieland–Miescher Ketone And Its Derivatives. *European Journal Of Organic Chemistry* **2021**, *2021* (29), 3992-3998.

(31) Dobrijevic, D.; Benhamou, L.; Aliev, A.; Mendez Sanchez, D.; Dawson, N.; Baud, D.; Tappertzhofen, N.; Moody, T.; Orengo, C.; Hailes, H.; Et Al. Metagenomic Ene-Reductases For The Bioreduction Of Sterically Challenging Enones. *RSC Advances* **2019**, *9*, 36608-36614.

(32) Mordaka, P. M.; Hall, S. J.; Minton, N.; Stephens, G. Recombinant Expression And Characterisation Of The Oxygen-Sensitive 2-Enoate Reductase From Clostridium Sporogenes. *Microbiology (Reading)* **2018**, *164* (2), 122-132.

# **Chapter 3.** The OYE Superfamily Rides the 4th Wave of Biocatalysis

## 3.1.Introduction

The asymmetric reduction of alkenes to produce enantioselective chiral products is of great interest to pharmaceutical and agricultural industries.[1] However, traditional means of meeting this demand often rely on chemical syntheses that utilize precious metal catalysts and/or harsh conditions with high environmental and technical costs. Emerging as a cheaper green chemistry alternate, the utilization of "ene" reductases has increased in interest.[2-4] One of the most widely used "ene"-reductases are OYEs.[5]

A significant limitation of traditional and engineered OYEs is their inability to accept unsaturated free carboxylic groups as substrates, as seen in Chapter 2.[4] Unsaturated carboxylic acid substrates are of industrial interest due to their involvement in the production of polycaprolactam (nylon 6,6).[6, 7] The production of Nylon, 6,6 is a multibillion-dollar industry that contributes to the approximate 186 million metric tons of greenhouse gas emissions linked to U.S. chemical manufacturing in 2019.[7] The major contributor to this number is the harsh nature of the synthesis of adipic acid, a Nylon, 6,6 co-monomer. Adipic acid is industrially synthesized by the oxidization of a cyclohexanol–cyclohexanone mixture by concentrated nitric acid producing the equivalent of 3.1 million metric tons of carbon dioxide equivalent (CO2e) per year of the greenhouse gas, nitrous oxide.[7] This results in a conversion rate of 5-10 %.[8] An alternative and more efficient route to the synthesis of the monomer is through the biosynthetic reduction of alkenes adjacent to two carboxylic acids in cis,cis-mucconic acid (Figure 3.1).[9] Such a pathway has been reported using an ene-reductase *in vivo*; however, it was unreproducible *in vitro*.[10]

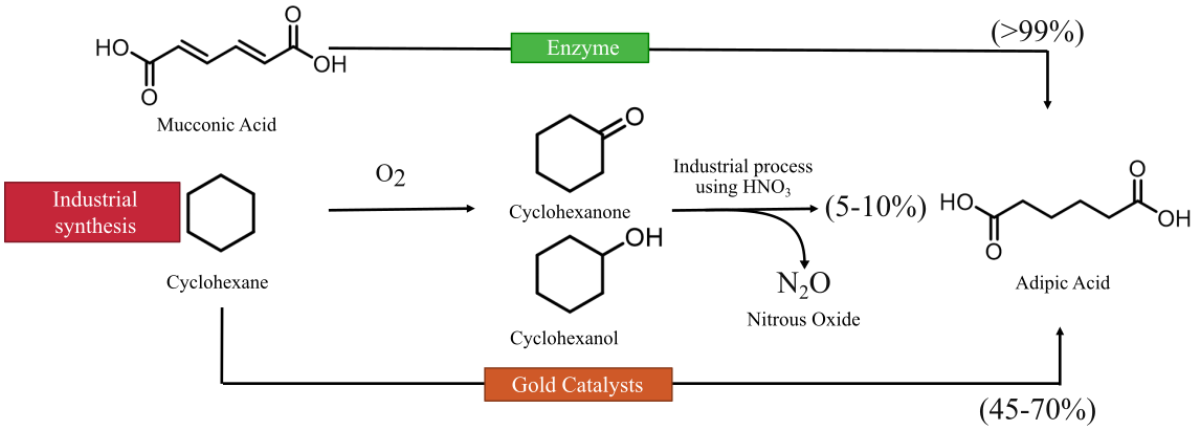



*Figure 3.1. Alternative routes scheme. Enzyme biocatalysts in the production of nylon 6,6 compared to an inorganic catalysis and organic synthesis.*

The 2-ER family is a severely underexplored portion of the OYE superfamily that has shown the ability to catalyze the reduction of α/β alkenes adjacent to carboxylic acid EWGs. While the reduction of unsaturated carboxylic acids (2-ER activity) is the staple of the family, the literature reveals that 2-ERs facilitate other diverse chemistries, such as oxidative deamination of histamine[11] and the oxidative demethylation of trimethylamine.[12] The chemistries performed by 2-ERs are diverse and significant; however, the characterization and industrial utilization of them has remained underexplored due to the enzymes' large size, difficulties with heterologous expression, and oxygen sensitivity. Understanding and developing novel methods of harnessing the chemistries of 2-ERs will have broad impacts as green chemistry alternatives for the asymmetric reduction of industrially and pharmaceutically relevant substrates.

SSNs have been used to gain a broad overview of the OYE superfamily, such as when Shi and coworkers reviewed 59,660 protein sequences annotated under UniProt accession codes PF00724 via an SSN.[2] Later, we (the Lutz Lab) expanded that to 70,366 sequences. As referred to in Chapter 2, we utilized the SSN to refine the classification of OYEs and provided a means to systematically

characterize the unexplored sequence-function space of protein superfamilies for novel activity.[13, 14] However, in both of the beforementioned studies, the 2-ER family within the OYE superfamily was intentionally excluded or limitedly explored. This severely under-characterized family of multidomain flavoproteins has historically been avoided due to the oxygen sensitivity of the embedded [4Fe-4S] cluster and difficulty assembling said cluster during heterologous expression. Such challenges are uniquely compatible with the methods of the fourth wave of biocatalysis.

Herein, we use traditional biochemical approaches, bioinformatic analysis, and computational tools to explore and characterize the 2-ER family for novel features. We provide validation for the use of SSNs as a longitudinal classification system for protein superfamilies and identify a previously uncharacterized deviation in the highly conserved [4Fe-4S] cluster binding motif hallmarked by the OYE 2-ER family. Finally, to our knowledge, we provided the first characterization of an *N*-methly-proline demethylase within the 2-ER family.

## 3.2.Results and Discussion

### 3.2.1. Validation of Longitudinal Consistency Classification of OYE Superfamily by SSN

Previous to grouping by SSN clusters, the OYEs were organized on phylogenetic trees by "classes" dictated by species of origin and sequence homology. In Scholtissek 2017, the 63 known OYEs were separated into Class I (OYEs originating from plants and bacteria), Class II (Classical OYEs), and Class III (thermophilic-like and mesophilic OYEs originating from various bacteria).[15, 16] Peter and coworkers expanded this classification further in 2019 by discovering six additional OYE ene-reductases and using sequence alignments to identify differences in highly conserved motifs.[17] This allowed for redefining Class I as Classical OYEs and Class II as thermophilic-like OYEs while also introducing Classes Ia, Ib, V, VI. While the phylogenetic trees were sufficient for these

smaller data set studies as they provide compact ordering, visualization becomes challenging with the massive sequence data attributed to protein superfamilies. SSNs overcome this challenge by providing a method for organizing this sequence information by representing protein sequences as nodes connected by edges within a set sequence identity threshold.[18] Isolated collections of nodes and edges make up clusters, which, when appropriately designed, represent families/subfamilies with assumed iso-functionality.[19]

The OYE SSN curated from PF00724 in 2018 by White et al. (2018-SSN) was generated via the methodology described in Copp 2018, with sequences clustering to 50 %ID and subsequent all-by-all pairwise alignment at an e-value cutoff of $1\times10^{-85}$.[14] This SSN displayed major clusters (<85% of the superfamily), where the first three clusters represented the three most prominent and well-known OYE subclasses. Due to SSN's inherent reliance on sequence identity, the 2018-SSN readily mimicked Peter et al.'s classification based on conserved motifs while increasing the accuracy of the grouping by including extensive homologous data set.

SSN clusters are numbered according to cluster size. Their relative placement in the network corresponds to the number of edges. It is reasonable to assume that the cluster number within a network will fluctuate along with the discovery of additional novel OYE sequences.[13] However, whether the overall clustering interconnectedness between member OYEs will remain consistent is integral to assessing whether the SSN classification method can act as a classification method that can grow alongside OYE superfamily. The OYE SSN Cluster 1, representing Classical OYEs, contained 54 characterized OYEs with PDB structures and/or had been "reviewed" as defined by SwissProt.[20] Cluster 3, Thermophilic-like OYE subclass, contained 20 sequences of similar classification, while Cluster 2, OYE 2-Enoate Reductase (2-ER) subclass, contained eight characterized sequences. If the SSN method is a consistent classification system, we expect these

intra-cluster connections to persist as additional sequences are identified even as cluster numbers deviate.

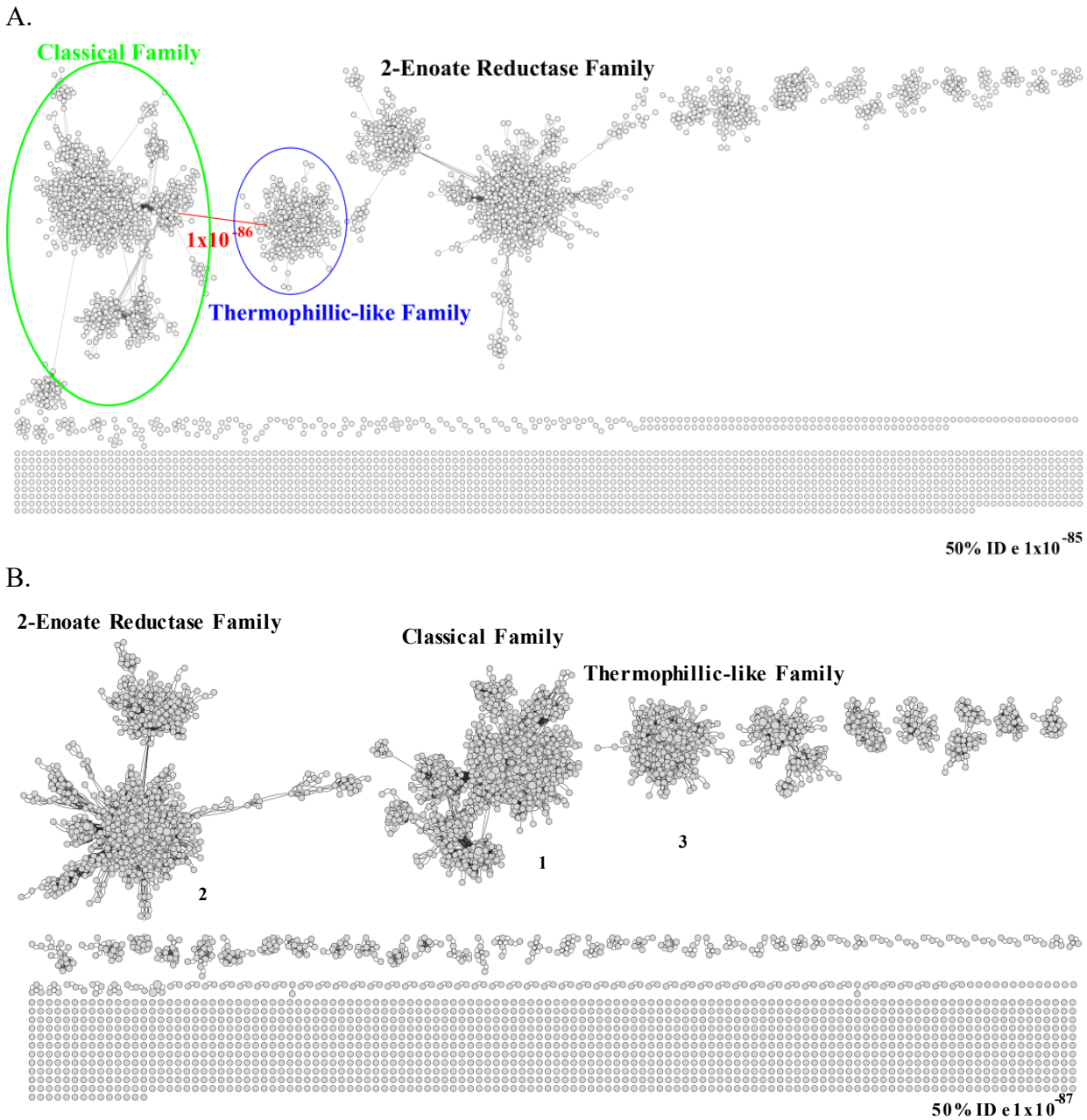



*Figure 3.2. 2023 Sequence Similarly Networks. A. SSN of OYE superfamily with 115k sequences (50 % ID e 1x10$^{-85}$) displaying Cluster 1 (green circle) and 3 (blue circle)linked by a single edge at x10$^{-86}$ B.SSN of the OYE superfamily with 115k sequences (50 % ID e 1x10$^{-87}$)*

From 2018 to 2023, the sequences annotated under PF00724 have increased by ~40 % (115,314). The number of characterized OYEs has increased proportionately. An updated SSN network of 115,314 homologous OYE sequences with a 50 %ID and an edge threshold of $10^{-85}$ was constructed, mimicking the 2018 OYE SSN (Figure 3.2A). While the bioinformatic grouping of known OYEs via SSNs remained reasonable over five years, clusters 1 and 3 formed a single edge representing e $10^{-86}$. This edge formation indicates the close evolutionary relationship between the two clusters. For studies that require implied cluster isofunctionality,[19] this longitudinal fusion can be overcome by a modest decrease in the e-value to increase the stringency of edges. Increasing to an edge threshold of $10^{-87}$ provided such (Figure 3.2B). The 2023SSN at $10^{-87}$ Clusters that contained characterized classical OYEs, (2018SSN:Cluster 1, 2023SSN:Cluster 1), thermophilic-like-OYEs, (2018SSN:Cluster 3, 2023SSN:Cluster 3), and 2-ERs (2018SSN:Cluster 2, 2023SSN:Cluster 2) remained clustered at 100% retention rate. More deviance was observed for clusters with fewer nodes.

### 3.2.2. Bioinformatics reveals non-canonical Fe/S-cluster binding motif

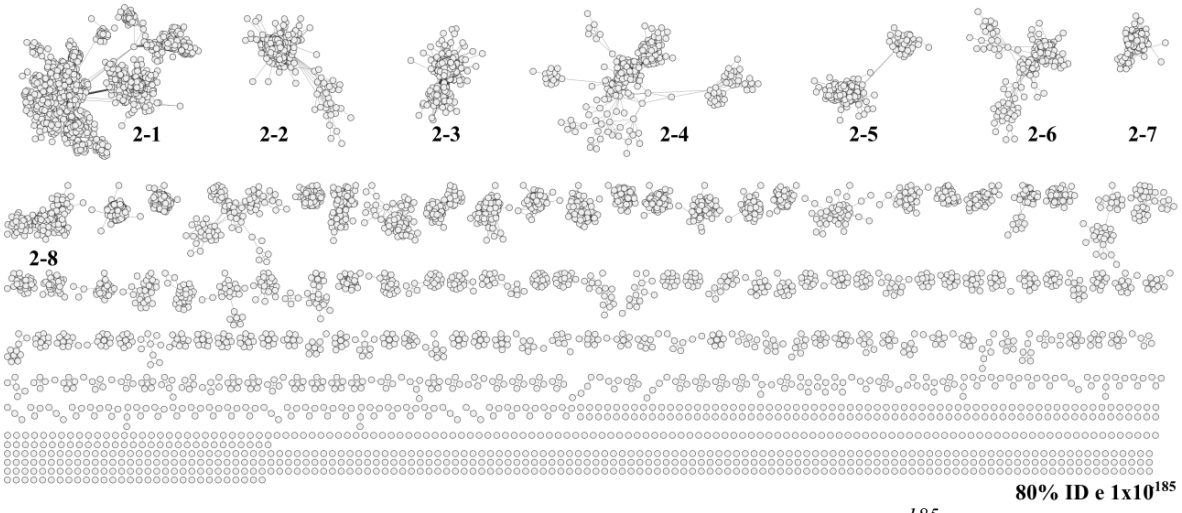



*Figure 3.3. SSN of the 2-ER Family (Cluster 2) (80%ID e threshold 1x10$^{-185}$)*

To further investigate the architectural and functional diversity of OYE 2-ERs, the associated raw sequence data associated with Cluster 2 in the broader 2023 OYE SSN was extracted and reclustered to ID80. This cluster, representing 27,868 nonredundant sequences through 8519 nodes and 816312 edges, was subjected to increased stringency resulting in a subcluster network for the 2-ER Family (Figure 3.3). Subclusters are visualized as eight major clusters comprising a total of >100 nodes. The remaining nodes are classified as statistically irrelevant for the purposes of this analysis. Although Cluster 2 makes up almost a quarter of the entire superfamily, it is one of the least explored, with <0.02% classified as "reviewed" in UniProtKB/Swiss-Prot to date.[20] Meaning annotation of activity and structure has been certified by the database. The predominant means of characterization of the family have been activity assays and limited structural information with only six crystal structures (PDB: 1PS9[21], 1DJN[22], and 3K30[11], 6DE6, 6L6J[23], and 6QKG[24]).

This SSN displayed that the homologous members of the 2-ER family predominantly originate from bacteria (95%). The remaining 5% comprises archaea and eukaryotes (fungi, protozoa, and metazoan, Figure 3.4A). Cluster 2 is composed of sequences ranging in length from 42-1519 AA. On average, the cluster's sequence length is between 600-800 AA (Figure 3.4B), making the average 2-ER twice as large as traditional OYEs (AA 300-400 AA), which are OYEs outside of Cluster 2. This size difference is of no surprise as 2-ERs are well known for being bi-domain, commonly possessing both a FAD and FMN binding domain.[11, 21, 25]

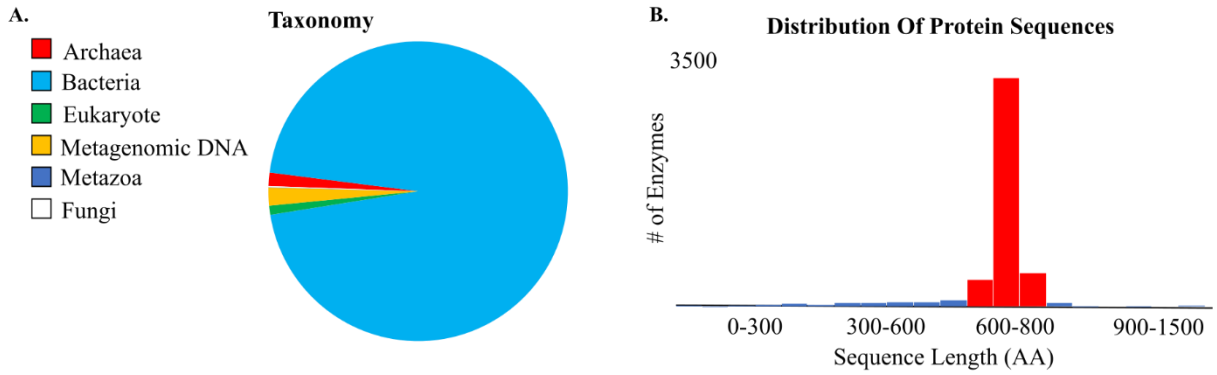



*Figure 3.4. A) taxonomy of 2-ER Family and B) Average length of sequences*

Our study displays that members of the 2-ER family often contain domains that extend beyond the addition of a single FAD domain, as predominantly displayed in literature, with the most common containing additional: an alternative adenosine diphosphate (ADP) domain, a secondary NADH binding domain, and a distinct pyridine nucleotide-disulfide oxidoreductase domain. Further exploration of AAs directly involved in catalytic turnover within the FMN domain displays notable variations in active site Tyr 197 (OYE1).[26] Tyr197 donates a proton to the α-carbon of the enolate intermediate to form the saturated product. Traditional OYEs conserve this residue at 99%. Cluster 2 predominantly follows that trend where in subclusters 2-1 and 2-3, this residue is, as expected, 99% Tyr. However, there are variations in subclusters 2-2 and 2-3. Interestingly, in subcluster 2-3, the highly conserved Tyr residue is 99% His, while subcluster 2-2 displays higher variability in this position than any subcluster displaying 26% Ser, 23% Thr, 23% Gly, and 15% Tyr. While Ser and Thr residues can still donate a proton to facilitate this reaction, characterized enzymes from subcluster 2-2 enzymes perform oxidative deamination chemistry. Whether alkene reduction activity is retained has yet to be explored in literature.

There is also variation in the canonical [4Fe-4S] cluster binding motif $C(X_2)C(X_{2-3})C(X_{11-12})C$ attributed to 2-ERs.[25] The four cysteines involved in [4Fe-4S] incorporation are predicted to be a highly conserved motif in all 2-ERs with no known deviations to date. However, after a multiple sequence alignment analysis of each subcluster of cluster 2, we discovered that subcluster 2 (2-2) has a pronounced variation at the second cysteine (C338 in DCR (PDB:1PS9)) in the motif. Unique to that subcluster, that position is held by alanine (92%) (Figure 3.5). Various other nonpolar AAs hold the remaining 8%. There are no additional cysteines nearby that would indicate a shift in the motif. In addition, no reported 2-ERs have originated from subcluster 2-2 to date. Due to the unique motif and active site residue variations seen in subcluster 2-2 and having no characterized representatives to date, we have selected subcluster 2-2 for further characterization.

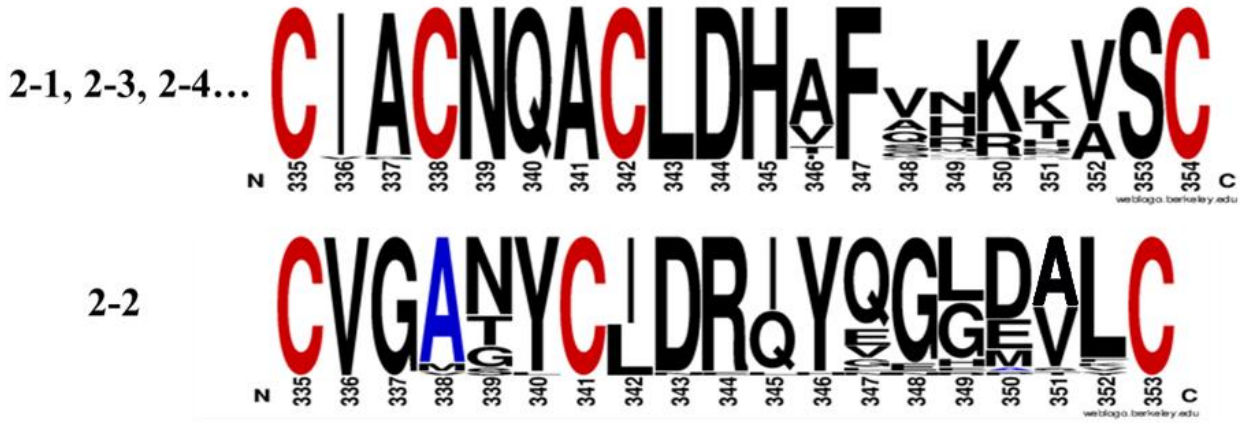



*Figure 3.5. Bioinformatic cluster Analysis of the FeS binding domain of 2-ERs. The top sequence represents the highly conserved region in all other major subclusters. The bottom sequence is the FeS binding region in cluster 2-2.*

### 3.2.3. Expression of Alanine mutation Representative

BurkOYE (Transcript ID: BAO89787) encoded from the BRPE67_CCDS01850 (*abc3*) gene in Burkholderia sp. RPE67, recently reclassified as *Burkholderia insecticola* or *Caballeronia insecticola*), was selected as a representative of subcluster 2-2. BurkOYE contains the 2-2 alanine

(A338) and histidine (H177) mutation while presenting a sequence length on the lower end of the average 2-ER sequence length to provide a higher probability of soluble expression. In 2-ER literature, the oxygen sensitivity of 2-ERs has been satisfied by performing protein expression fully anaerobically. [21, 25] However, these methods are time-consuming and do not address the poor incorporation of the [4Fe-4S] cluster, providing limitations on an industrial scale. Inspired by the heterologous expression methods of another pharmaceutically relevant metalloenzyme superfamily, the radical *S*-adenosyl-L-methionine (SAM) superfamily, an alternative method of producing soluble enzyme was explored.

Members of the rSAM superfamily are often co-expressed with a pDB1282 plasmid. The pDB1282 plasmid encodes six genes (*iscS, iscU, iscA, hscA, hscB*, and *fdx*) from the *isc* operon of *A. vinelandii* which contains all the components necessary for FeS cluster production and incorporation (Figure 3.6A). Upon arabinose-mediated co-expression, IscS, a cysteine desulfurase, extracts sulfur from L-cysteine to be incorporated as the sulfur ion FeS cluster. Excess $FeCl_3$ is supplemented directly into the growth media as the Fe source. IscU and IscA are believed to natively perform "housekeeping" functions related to assembling/repairing FeS metalloproteins and/or acting as scaffolding proteins to the maturing protein. HscB and HscA are homologous to members of a heat-shock-cognate molecular chaperone system and are hypothesized to facilitate the proper incorporation of FeS clusters into the target protein. Fdx is a ferredoxin believed to be involved in maintaining appropriate redox states during FeS cluster construction and incorporation.[27]

The pDB1282 plasmid was co-transformed with pet28a plasmid encoding BurkOYE with N-terminal SUMO tag and Strep-tag to aid solubility and purification (Figure 3.6B). To identify the optimum expression conditions for BurkOYE, three methods were explored during biochemical

analysis: (1) pseudo-anaerobic expression by utilization of excess L-cysteine and co-expression of the pDB1282 plasmid with BurkOYE under aerobic conditions, (2) fully anaerobic expression with co-expression of the pDB1282 plasmid, and (3) fully aerobic with no pDB1282 plasmid induction. The third condition was to test if the alanine variance would allow for oxygen tolerance displayed by traditional OYEs.

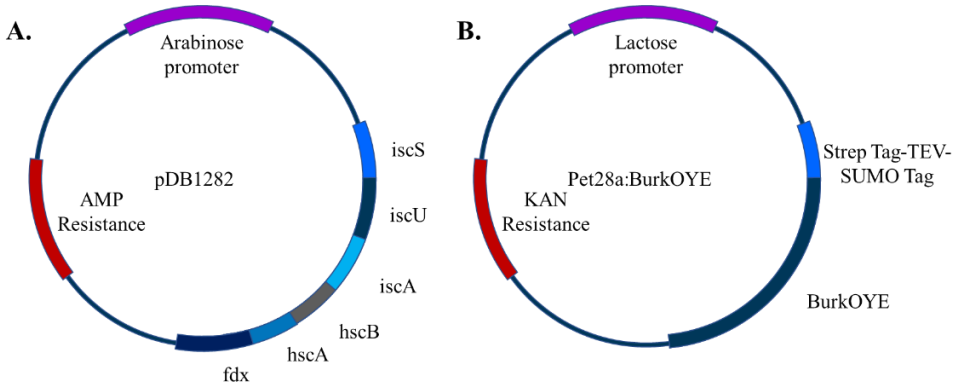



*Figure 3.6. Plasmid Cartoons. A. pDB1282 vector Cartoon Diagram Encoding for Fe-S cluster Assembly. B. pET28 vector Cartoon Diagram Encoding for Metalloenzyme of interest, BurkOYE*

### 3.2.4. Optimization of the BurkOYE activity for library screening

Before investigating depth of biotransformation, the conditions for optimal activity of BurkOYE were determined using BurkOYE co-expressed with pDB1282 plasmid under fully anaerobic conditions with *R*-carvone **(3)** as the exploratory substrate. Carvone is a well-studied substrate for OYEs with the ability to be converted by most known OYEs.[4, 5, 13] Optimum buffer pH, oxygen sensitivity, and nicotinamide dependency of BurkOYE were determined. The optimum buffer pH for activity was between 8.0–8.5 (Figure 3.7).

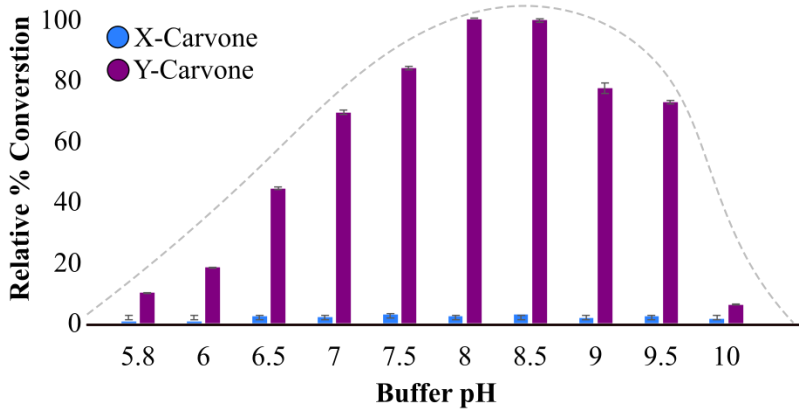



*Figure 3.7 pH profile of BurkOYE*

To determine nicotinamide dependency, the reduction of **(2)** was measured using a NAD(P)H-dependent spectrophotometric assay, using 50mM-350mM nicotinamide adenine dinucleotide (NADH) or nicotinamide adenine dinucleotide phosphate (NADPH) as the hydrogen donors under fully anaerobic conditions. **(2)** was reduced to dihydrocarvone **(2p)** with an initial rate of 0.37520 $\mu mol \bullet min^{-1}$ and 0.59908 $\mu mol \bullet min^{-1}$ for NADPH and NADH respectively (Figure 3.8). The $K_D$ of NADH was 92.84 ± 0.08, while $K_D$ NADPH was 167.25 ± 0.18. No product formation was observed in the absence of enzyme or nicotinamide. Therefore, future experiments were undertaken using 50 mM phosphate buffer pH 8.5 with NADH as the electron source.

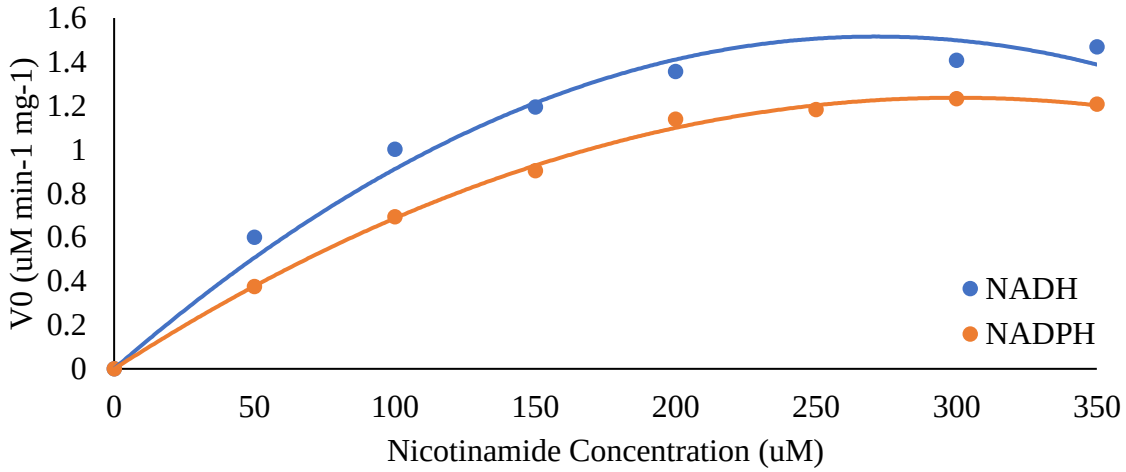



*Figure 3.8. NADH:NADPH preference kinetics study*

The effect of oxygen on BurkOYE activity was investigated by exposing samples of protein extracts to air for time intervals of 15 min for 1 hr., keeping the samples on ice throughout the exposure time. The protein extracts were then tested for carvone reduction using NADH as the electron donor under anaerobic conditions at 25 °C, measuring substrate and product concentrations by gas chromatography mass spectroscopy (GC-MS) (Figure 3.9). BurkOYE lost 25% activity after 1 hr of exposure to air, and activity was not restored by reintroduction of anaerobic conditions. This is evidence that the metal cofactor is required for reduction activity.

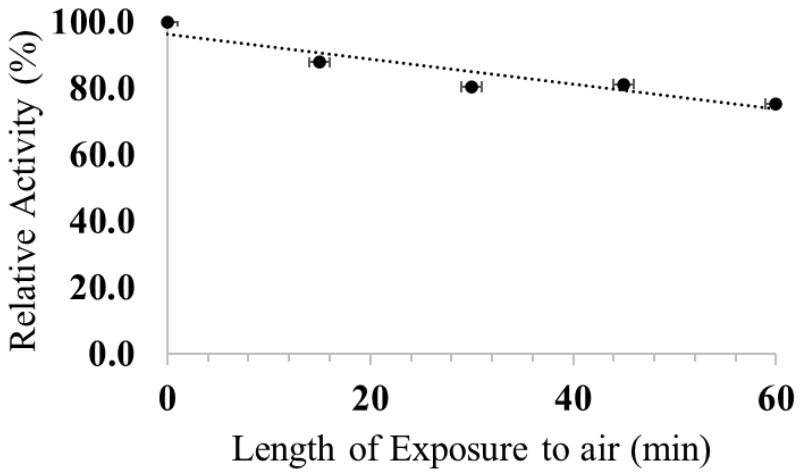



*Figure 3.9.Oxygen sensitivity of BurkOYE with R-Carvone (reduction)*

### 3.2.5. Biochemical Evaluation of BurkOYE

To probe the substrate scope of BurkOYE, the activity of the enzyme with a substrate library was explored. This library was composed of structurally diverse aliphatic and cyclic alkenes bearing ketones, aldehydes, and carboxylic acid EWGs (Figure 3.10). The substrate panel was composed of cinnamic acid **(1)**, 3-phenyl propionic acid **(2)**, *R*-carvone **(3)**, *S*-carvone **(4),** 2-methyl-2-cyclopentenone **(5)**, cyclopent-2-enone **(6)**, the Wieland Wiescher ketone **(7)**, and levidione **(8)**

**(**Table 3.1**)**. Each substrate was selected to probe for a specific attribute of known OYEs. Substrate **(1)** probed whether the alanine mutation affected the retention 2-ER activity (i.e. the conversion of alkenes adjacent to carboxylic acid EWGs). Substrates **(3)** and **(4)**, probed whether the enzyme retained activity well displayed in aerobic OYEs, referred to as traditional OYE activity. Substrates **(5), (6),** and **(7)** probed for traditional OYE activity with tolerance of small substrates vs larger more sterically hindered substrates. Substrates **(2)** and **(8)** probed for OYE desaturase activity.

Due to poor water solubility of some substrates, stock solutions of all the substrates were prepared in dimethyl sulfoxide (DMSO). When testing the stability of an OYE homologs with DMSO, ethanol, dimethylformamide (DMF), and ethyl acetate cosolvents, DMSO was shown to have the least impact on activity.[28]

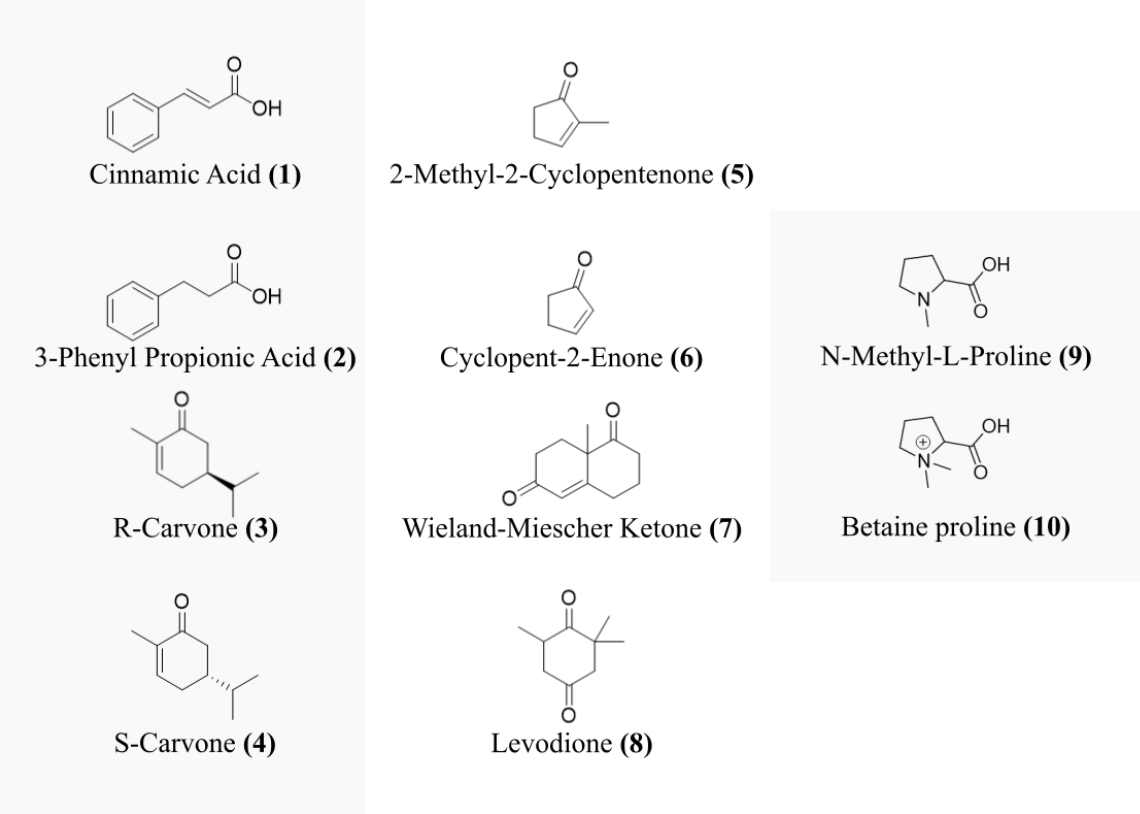



*Figure 3.10. Substrate panel for BurkOYE*

Biotransformations of BurkOYE (under the three expression conditions) with the substrate library were conducted in a pH 8.0 100mM phosphate buffer supplemented with NADH and a GDH nicotinamide regeneration system. The regeneration system consists of GDH and excess D-glucose which is used as a hydride source for nicotinamide reduction.[29] Assays were performed anaerobically to prevent side reactions between flavin and oxygen. Due to low protein expression yield, fully anaerobic BurkOYE assays were run with clarified lysate. Pseudo-anaerobic and aerobic expressions, which resulted in higher protein yield, underwent multistage purification.

**Table 3.1.** Substrate Range of BurkOYE 2-Enoate Reductase. All values are from reconstituted protein excluding those marked with "*"

| | Expression Condition | | |
|---|---|---|---|
| | **Anaerobic** | **Pseudo-Anaerobic** | **Aerobic** |
| **(1)** | 100±0.00% | 73.41±12.25% | 7.93±3.26% |
| **(2)** | N.D. | N.D. | N.D. |
| **(3)** | 10.80±0.50% | 0.54±0.01% | < 0.5% |
| **(4)** | < 0.5% | < 0.5% | < 0.5% |
| **(5)** | 2.12±0.26% | 1.82±0.22% | < 0.5% |
| **(6)** | 1.50±0.42% | N.D. | N.D. |
| **(7)** | N.D. | N.D. | N.D. |
| **(8)** | N.D. | N.D. | N.D. |
| **(9)** | 22.16±2.24% | 13.66±0.32% | 1.94±0.12% |
| **(10)** | NMP: 1.50±0.12%<br>Proline: < 0.5% | NMP: 0.63±0.01%<br>Proline: < 0.5% | *NMP: 7.06±0.38%<br>*Proline: < 0.5% |

BurkOYE expressed under pseudo anaerobic conditions displayed 73.41±12.25% conversion of **(1)** to **(2)** confirming the retention of 2-ER activity.[25] Iron-sulfur cluster reconstitution resulted in up to 18% higher activity than as-purified enzyme in this reaction. Desaturase activity was not observed for **(2)** or **(8)**. Neither did BurkOYE display a tolerance for sterically hindered substrates, **(7)**. The enzyme did, however, display tolerance for small substrates **(5),** 2.11±0.04%. Reactivity with **(3)**, 0.54±0.01%, and **(5)**, indicate a retention, abet limited, of traditional OYE activity. Fully anaerobic expression of BurkOYE displayed initially undetected or oftentimes 10-fold increase in activity when compared to pseudo anaerobic conditions displaying 10.8±0.50%, 10.3±0.30% and 2.40± 0.40% percent conversion for **(3)**, **(5)** and **(6)** respectively. The difference in reactivity for **(3)** and **(4)** indicates an enantiopreference *R*-conformation substrates as supported by Km, where the Km of **(3)** was 167.25 ± 37.1 while **(4)** was 270.553 ±59.91. Desaturase activity and conversion of sterically hindered substrates remained unobserved. Interestingly, BurkOYE expressed and purified under aerobic conditions displayed conversion of **(1)** to **(2**) at a precent conversion of 6.52±4.17%. While this is an 89% decrease in activity, to our knowledge this represents the second 2-ER to display activity after prolonged oxygen contamination after *C. acetobutylicum* 2-enoate reductase described in Sun et al. Such activity is highly sought after to produce Nylon-6 precursors.[30] Iron-sulfur reconstitution yielded a 21.6% higher activity; however, the enzyme never recovered the maximum observed activity for BurkOYE with this substrate.

After limited activity with the initial eight substrates, excluding **(1)**, additional substrates were considered. To guide substrate selection a Basic Local Alignment Search Tool (BLAST) search was performed.[31] Identifying the closest relative of BurkOYE (~98% sequence similarity) was computationally annotated to be a *N*-methyl-proline demethylase. To probe this activity in BurkOYE, *N*-methyl-L-proline **(9)** and proline betaine **(10)** were introduced to the library (Table

3.1). **(9)** demethylation activity was observed at 8.59±0.20% conversion with the pseudo-anaerobic expression and 31.6±1.24% with the anaerobic expression. While BurkOYE is not the only 2-ER to perform an oxidative demethylation mechanism it is the first to accept **(9)** as a substrate. This is significant as **(9)** and **(10)** are implicated in a number of homeostasis functions including acting as an osmoprotectant that aids gram negative bacteria response to osmotic stress.[32, 33]

Unexpectantly aerobic expression had the highest **(10)** demethylase activity at 6.99±0.38% conversion to the first demethylation product. While the two products of **(10)** were detectable, the conversion bottlenecked at the first demethylation resulting in <0.5% conversion for every expression condition. This indicates that the first methylation step of **(10)** may undergo an oxygen-dependent mechanism, however further studies are needed to understand the role of $O_2$ in BurkOYE activity. Based on the results of our substrate panel, we propose that BurkOYE displays two primary pathways: **(1)** a reductive pathway resulting in the conversion of α/β alkenes adjacent to EWGs and **(2)** an oxidative pathway resulting in the demethylation of **(9)** to proline.

In White et al., we identified 33 OYEs, including NemA from *E. coli* and pentaerythritol tetranitrate reductase (PETNR) from *Enterobacter cloacae*, that exhibited desaturase activity and reductase activity.[13] We propose FMN reduction by NADH or NADPH initiates the reductive pathway. The reduced FMN then performs a hydride transfer to the β-C of the α/β alkene ( This supports our proposed molecular oxygen dependent demethylation scheme (Figure 3.11C).

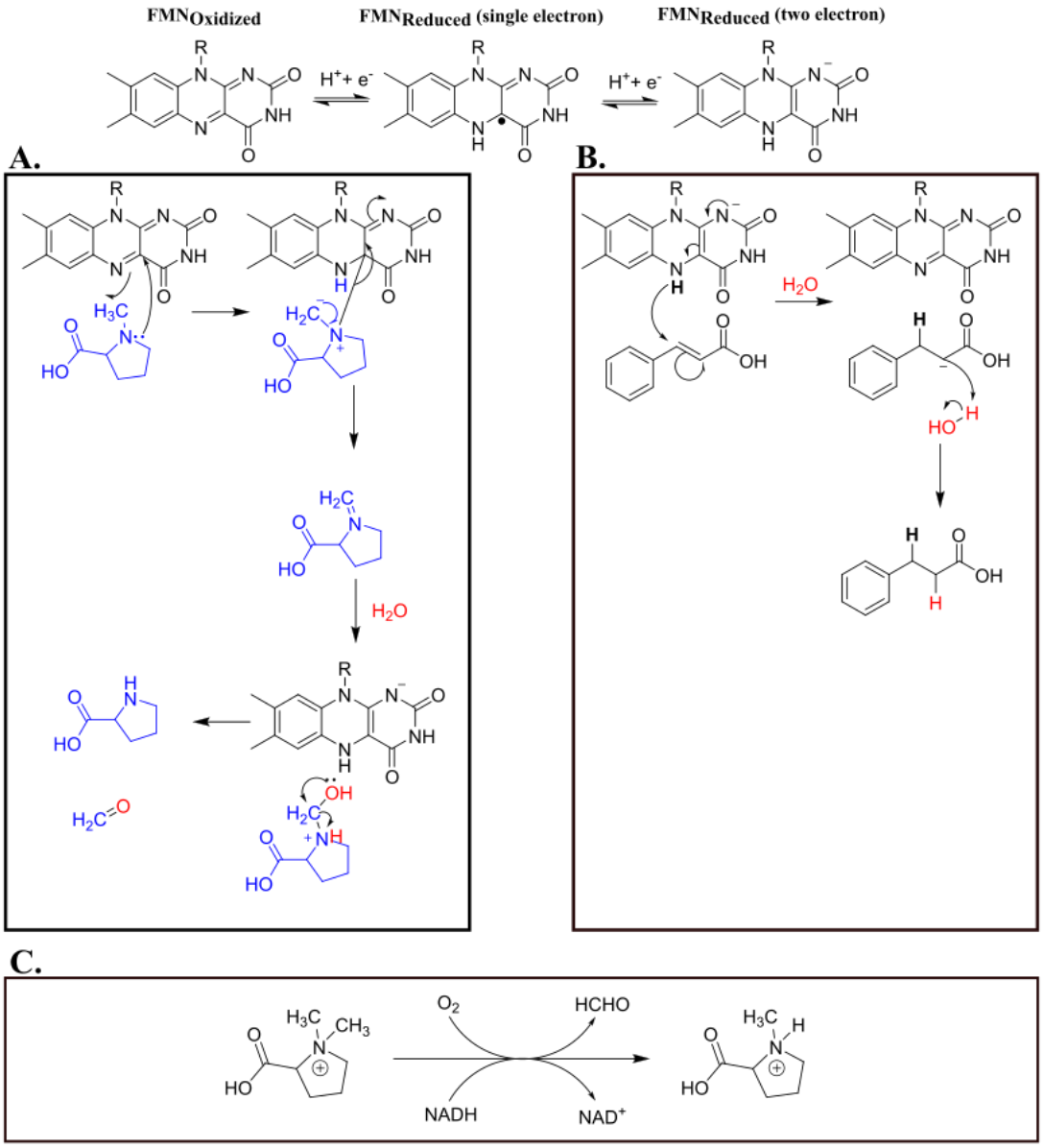



*Figure 3.11*B). Negative charges are satisfied by hydrogen abstraction by a nearby water or protonated residue. The [4Fe-4S] cluster appears to be involved in this mechanism, as greater oxygen exposure during protein expression displayed reduced activity. This is typically a result of oxidative degradation of an oxygen-sensitive metal center.[34]

We propose that the oxidative pathway, on the other hand, relies on an oxidized FMN similar to BurkOYE homolog trimethylamine dehydrogenase (TMADH).[22, 35, 36] We propose two possible

mechanisms based on known homolog kinetic and UV studies. The first could follow a similar pathway as TMADH where the substrate reduces the FMN to promote the production of an imine intermediate that spontaneously hydrolyzes to form the demethylated product This supports our proposed molecular oxygen dependent demethylation scheme (Figure 3.11C).

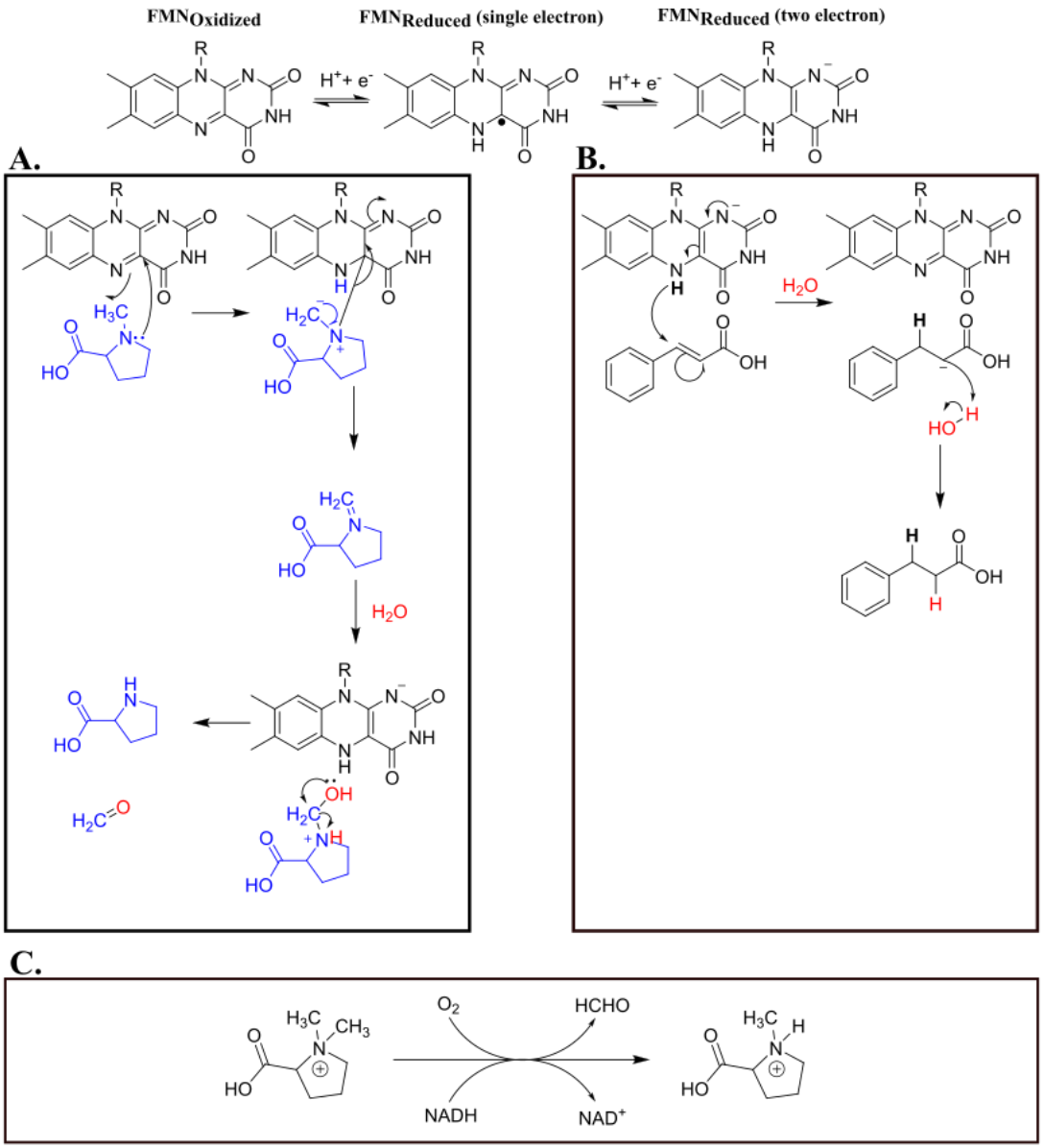



*Figure 3.11*A). Alternatively, BurkOYE could rely on molecular oxygen which would allow for quaternary ammonium oxidative demethylation of **(10)**, such as in the mechanism of nicotinamide-dependent stachydrine demethylase from *Sinorhizobium meliloti* Figure 3.11*C*).[37] In TMADH, the

FMN is regenerated upon electron transfer to the [4Fe-4S] cluster.[35] Because of this, we hypothesize that BurkOYE may be able to mitigate $O_2$ damage to its [4Fe-4S] cluster as a result of its native function. However, additional studies are needed to prove such. Interestingly, we observe that the oxidation pathway is still dependent on nicotinamide as assays without the GDH nicotinamide regeneration system, resulting in 92% reduction of product formation. This supports our proposed molecular oxygen dependent demethylation scheme (Figure 3.11*C*).

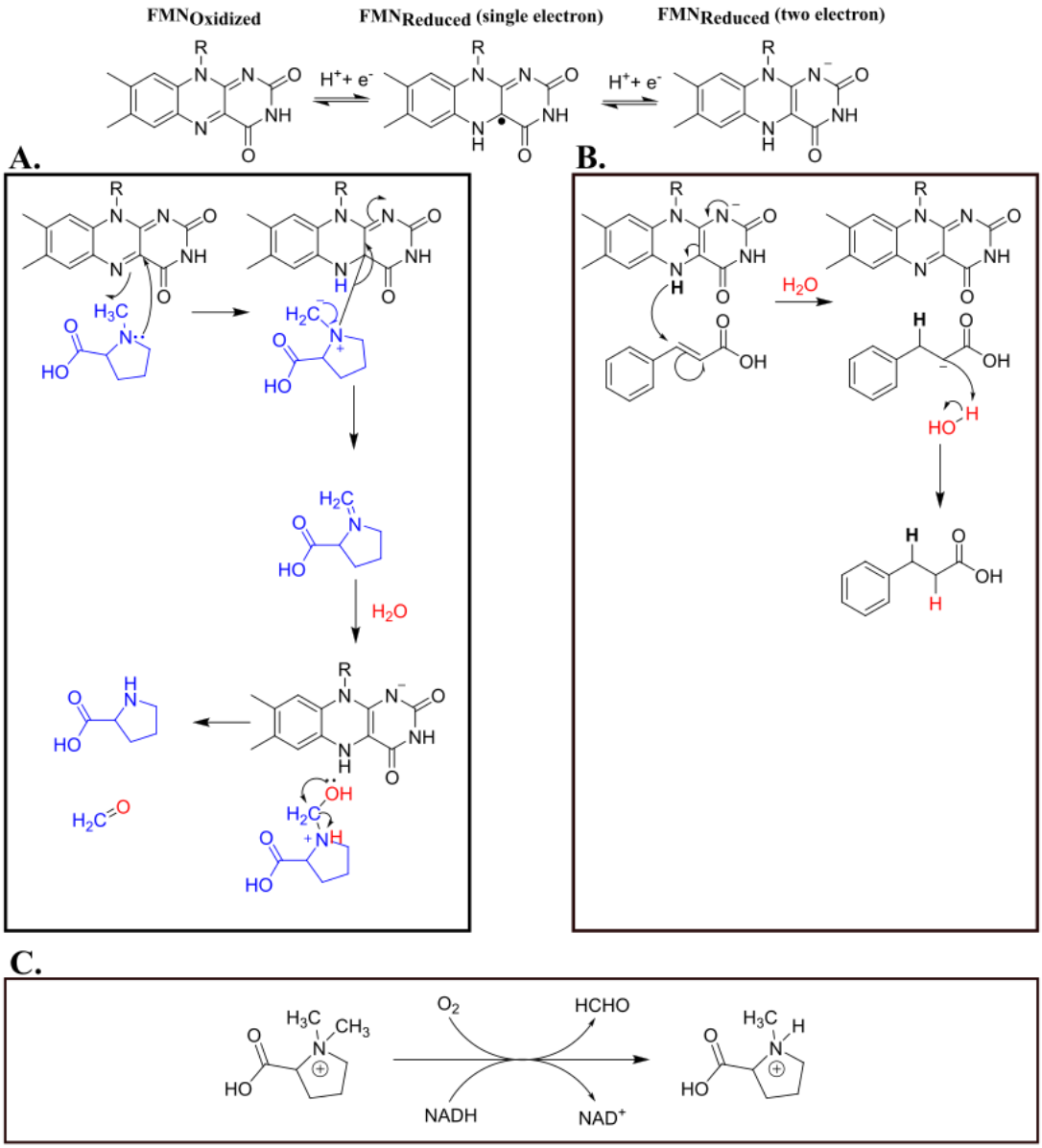


*Figure 3.11 Proposed mechanism for the oxidation of N-methlyproline and reduction of cinnamic acid by BurkOYE.* ***(A)*** *Proposed oxidative demethylation mechanism if homologous 2-ER trimethylamine dehydrogenase*[38] ***(B)*** *Proposed reduction mechanism is analogous to those displayed by traditional OYEs*[5] *.* ***(C)*** *Proposed molecular oxygen dependent demethylation scheme*

### 3.2.6. Validation of [4Fe-4S] Cluster Incorporation

To verify the incorporation of an iron-sulfur cluster in the BurkOYE expression conditions, the iron atoms per enzyme molecule was detected via inductively coupled plasma mass spectrometry (ICP-MS). Enzyme samples were thermally denatured to release Fe ions into the solution. ICPMS analysis displays that fully anaerobic and pseudo-anaerobic expressions incorporated 4.18±0.09 and 4.82±0.03 mols of Fe per mol of enzyme, respectively, corresponding to one fully incorporated [4Fe-4S] cluster.

*Table 3.2. ICPMS Results displaying mol of Fe/mol of enzyme per expression condition*

| | *Expression Condition* | | |
|---|---|---|---|
| *Reconstitution* | Anaerobic | Pseudo-Anaerobic | Aerobic |
| (+) | -- | 4.82±0.03 | 0.74±0.00 |
| (-) | 4.18±0.09 | 3.02±0.20 | 0.41±0.00 |

Prior to reconstitution, BurkOYE under pseudo-anaerobic expression conditions incorporated 3.02±0.20 mols of Fe, supporting a predominate composition of [3Fe-4S] clusters which can emerge from oxidative damage of [4Fe−4S] clusters (Figure 3.12. [4Fe-4S] cluster degradation by molecular oxygen).[39] As expected, the aerobic expression did not incorporate any Fe, even post-reconstitution.

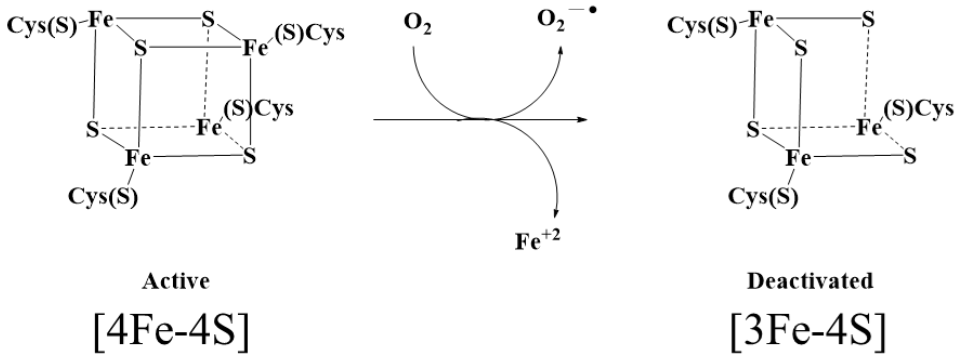



*Figure 3.12. [4Fe-4S] cluster degradation by molecular oxygen*

Interestingly, this [3Fe-4S] cluster sample retained >90% of the reactivity with **(1)** when compared to the **[**4Fe-4S] cluster. In dichotomy, the with **(9)** ~30% activity was retained when compared to the **[**4Fe-4S] cluster. This supports or previous hypothesis that the iron sulfur cluster is not as directly involved in the reductive reaction pathway as the oxidative demethylation pathway.

As expected, fully aerobic expression conditions resulted in no detectable Fe in the sample. As the sample were stable on ice for over 5 hrs, displaying no precipitation, we conclude that the [4Fe-4S] cluster is not involved in protein structural support. Two mutant constructs were generated to further explore BurkOYE's dependence on the [4Fe-4S] cluster for reactivity, AllAla-BurkOYE and AC-BurkOYE. The AllAla-BurkOYE construct mutated all predicted ligating cysteines to an alanine. The AC-BurkOYE construct mutated A348 to C348. Both constructs were unstable (i.e crashing out of solution) and biocatalytically inactive under pseudo-anaerobic expression conditions.

### 3.2.7. Structural Prediction of Burk OYE with AlphaFold

After many attempts, we were unable to produce BurkOYE crystals that successfully diffracted during X-ray crystallography. Inspired, instead, by the high accuracy of artificial intelligence (AI)-generated protein structural models, we employ AlphaFold for the structural characterization of BurkOYE.[40, 41] AlphaFold does not inherently include ligands and cofactors like FMN, FAD, or the iron-sulfur cluster. The predicted "active" flavo binding residues for ligand interaction are classified as inaccessible by High Ambiguity Driven protein-protein DOCKing (HADDOCK) making *ab initio* docking with the software unreliable. Aligning the BurkOYE AlphaFold model with OYE1, DCR, and TMADH results in a RMSD of 1.714 Å, 1.980 Å and 1.584 Å, respectively (Figure 3.13). This indicates that TMADH displays a secondary structure most similar to the BurkOYE AlphaFold model.

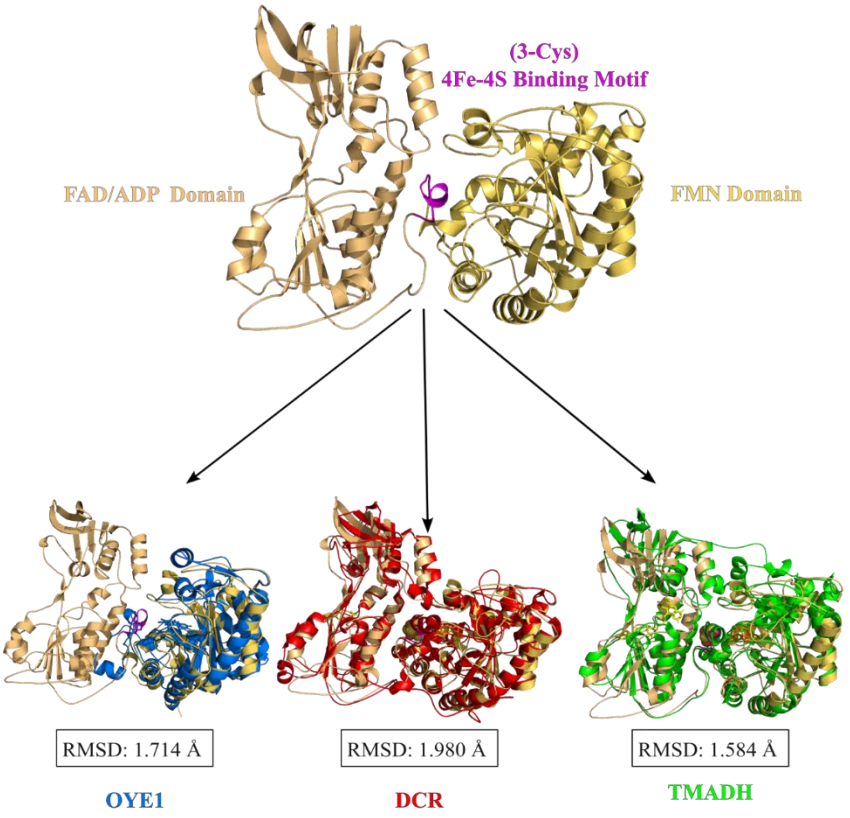



*Figure 3.13. Alignments of BurkOYE with OYE1 (blue), DCR (red), and TMADH (green)*

### 3.2.8. Genomic Neighborhood Analysis of BurkOYE

While BurkOYE displays traditional OYE chemistry, i.e., the reduction of α/β alkenes adjacent to electron-withdrawing groups (EWG), biochemical analysis revealed reactivity as an *N*-methlyproline demethylase. To hypothesize a native function of such activity in *Burkholderia insecticola,* a genomic neighborhood analysis of the genes around gene *abc3*, which encodes for BurkOYE, was performed using the Ensembl Bacteria database for *Burkholderia* sp. RPE67 (GCA_000828875) (Figure 3.14. Genomic neighborhood of BurkOYE).[42] The Ensembl Bacteria is a hub for bacterial and archaeal genomes linked to the International Nucleotide Sequence Database Collaboration. The genomic analysis exploits the evolutionary tendency of bacteria to cluster genes related to a particular function into operons. The technique allows for the estimation of enzymatic function via analysis of nearby protein-encoding genes involved in the same response pathway.[43] This genomic context is made more reliable when coupled to homolog analysis as characterized homologous proteins apart of closely related operons are more likely to display identical activity.[44]

*abc3* is located on the forward strand of Chromosome 3, position 198,377-199,339 of *B. insecticola*. Upstream to the *abc3* gene is BRPE67_CCDS01820 (abc1), which encodes for a LysR-type transcriptional regulator, followed by a BRPE67_CCDS01830 (*abc2*), which encodes for a small-conductance mechanosensitive channel (MscS) mechanosensitive ion channel (UniProt accession code: A0A158J5F7). Members of the MscS mechanosensitive ion channel family are transmembrane proteins that aid in preserving prokaryotes during hypo-osmotic stress. MscS mechanosensitive ion channel's primary role in this mechanism is sensing mechanical changes in the lipid bilayer of the bacteria and, in some cases regulating the export of ions and

osmoprotectants.[33] However, the transportation of substrates across the cellular membrane is more often performed by ATP-binding cassette (ABC) transporters. ABC transporters are ATP-binding proteins that hydrolyze ATP to power the transportation of highly water-soluble substrates into and out of the cytoplasm and are invaluable in proline betaine-dependent osmotic stress response.[45, 46] 326 bp downstream from *abc3* lies BRPE67_CCDS01860 (*abc4*), a dipeptide ATP-binding cassette (ABC) transporter.

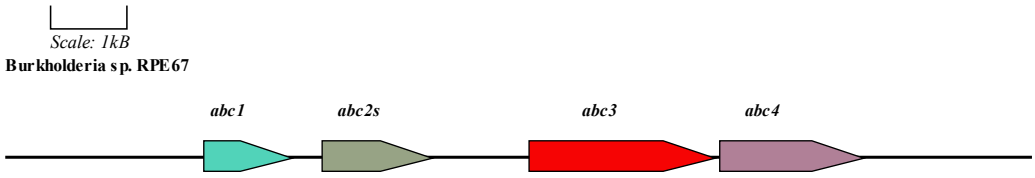



*Figure 3.14. Genomic neighborhood of BurkOYE*

In *Sinorhizobium meliloti,* a gram-negative bacteria root symbiont of the alfalfa plant, proline betaine-dependent osmotic stress response is well detailed. Proline betaine is abundant in soil and transported into the cytoplasm via its ABC transporter, Hut, and other secondary transporters during times of high external osmolarity/low internal osmolarity. At times of low external osmolarity, proline betaine influx reduces, and a catabolism pathway is activated where proline betaine is demethylated to *N*-methlyproline then proline.[32, 47] These complementing pathways allow the bacteria to protect itself against high external osmolarity by accumulating the nonionic osmoprotective compound into the cytoplasm, reducing the influx of water.[48] The genes in the genomic neighborhood of our system are consistent with the framework for an ABC operon responsible for this stress response as indicated by containing three molecular components: 1) a membrane protein, of which we propose is the MscS mechanosensitive ion channel protein, 2) an ATP-binding protein, of which we propose is the dipeptide ABC transporter, and 3) the substrate-

binding protein, of which we hypothesize is the role of BurkOYE.[32, 49] The genomic neighborhood analysis and biochemical assay support BurkOYE's involvement in the osmotic stress response.

### 3.3. Conclusion

Biocatalysis have experienced many waves of innovation congruent to humanities technological advancements. Through each wave OYE homologs have been utilized to advance the general knowledge of protein biocatalysts and generate industrially relevant oxidoreductases. While OYEs are an extensively studied superfamily of enzymes, significant portions of the sequence space remain functionally and structurally unexplored. This is unsurprising, as advancements in sequencing technology have led to an exponential growth in available protein sequence data , while the magnitude of the scientific community nor time to perform experimentation has not followed suit.[50] Particularly in the OYE superfamily where 2-ER family, the second largest protein family (by number of non-redundant protein sequence), remained severely unexplored due to inherent difficulties with aerobic heterologous expression. The tools and techniques that highlight the fourth wave of biocatalysis innovation (proteomics, metagenomics, genomics, AI structure modeling and function prediction, bioinformatic, and various biocomputational tools) were uniquely capable of supplementing experimental bottlenecks when exploring the 2-ER family.[51, 52]

Thus far our five-year longitudinal analysis of the OYE SSN supports that SSNs are reasonable bioinformatic tools as evolving classification system for protein superfamilies. With this SSN we were able to perform a bioinformatics analysis of the 2-ER family and identified a novel alanine mutation in the highly conserved iron sulfur binding motif. Upon biochemical analysis of a 2-ER representative with the beforementioned mutation, BurkOYE, we demonstrate that the A mutation in BurkOYE does not impede [4Fe-4S] cluster incorporation or stability of the enzyme. This leads

to further questions of the purpose of the native open coordination site. While it is difficult to ascertain the native function of most OYEs, genomic analysis and biocatalytic exploration supports BurkOYE being a part of the stress response system of Burkholdia, thus understating how this enzyme may lead to further methods of combating antibiotic resistance in gram-negative bacteria.

This paper details a protocol for the soluble and active expression of an oxygen sensitive 2-ER OYE in an aerobic environment. This may be useful as an initial screening of 2-ERs for industrial applications. Further analysis with a probing substrate library detected two reactive pathways in BurkOYE: **1)** a reductive pathway, more akin to traditional OYEs reduction of activated alkenes, and **2)** an oxidative demethylation pathway. The ability to perform oxidative and reductive mechanisms in OYEs is not unobserved in traditional OYEs. [13] Even in the 2-ER family there are three other known 2-ERs with oxidative pathways Histamine dehydrogenase (HADH) from *Nocardiodes simplex,* which catalyzes the oxidative deamination of histamine to give imidazole acetaldehyde, TMADH and DMADH, from *M. methylotrophus*.[11, 22] TMADH catalyzes the oxidative demethylation of trimethylamine to dimethylamine and formaldehyde, while DMADH catalyzes a similar reaction with dimethylamine as the substrate. To our knowledge, BurkOYE is the only 2-ER to date to perform oxidative demethylation with *N*-methyl-proline as its substrate, the only 2-ER to be confirmed to retain reduction activity with traditional OYE substrates (>10% conversion), and carboxylic acid EWGs (>40% conversion), while also being able to perform an oxidative pathway (>30% conversion). BurkOYE is currently the only 2-ER proposed to utilize molecular oxygen in one of its mechanisms.

## 3.4.Experimental

### 3.4.1. Materials

Unless otherwise noted, all substrates and cells were purchased from Sigma Aldrich. The *E. coli* optimized gene of BurkOYE incorporated into a pET28a vector was received from the Genescript. The pDB1282 plasmid bearing the *isc* operon was kindly donated by Dr. Squire Booker.

### 3.4.2. Local SSN generation

SSN was generated on a 4 core i5 Dell laptop with an Ubuntu 2020 operating system. The 2023 OYE dataset of all enzymes associated with Accession Code: PF00724 was obtained April of 2023 (115k sequences) from UniProt Protein Database.[53] The dataset was reduced by clustering to 50% identity using Cluster Database at High Identity with Tolerance (CD-HIT) software[54] and further refined to remove sequences below 100 AAs and over 1500 AAs. BLAST+ e value matrix was generated with an edge cutoff of $1xe^{-87}$ and visualized in Cystoscope.[55] The 2-ER SSN was generated by taking sequences associated with cluster two of the 2023 OYE SSN and creating an 80% ID database from those sequences with an edge cutoff of $1xe^{-200}$.

### 3.4.3. Transformation, Expression, and Purification of BurkOYE

A pET-28a vector bearing the BurkOYE gene (European Nucleotide Archive Accession Code: BAO89787) from *Burkholderia sp. RPE67* with a N-terminal SUMO solubility tag linked by a TEV cleavage site to a strep II purification tag was designed and obtained from Genescript. The two plasmids were co-transformed into BL21(DE3) Electrocompetent Cells and allowed to grow in SOC Outgrowth Medium (New England BioLabs) for 2 hrs at 37 °C before transfer onto an LB agar plate containing ampicillin ($100\ \mu g \bullet mL^{-1}$) and kanamycin ($50\ \mu g \bullet mL^{-1}$) and incubation overnight at 37 °C.

All expression conditions were performed using the BurkOYE: pDB1282 BL21(DE3) cells.

### 3.4.4. Anaerobic Expression

Cells from cryostock were used to inoculate 10 mL LB Broth supplemented with ampicillin (100 μg•mL$^{-1}$) and kanamycin (50 μg•mL$^{-1}$) to grow overnight while shaking aerobically at 250 r.p.m, at 37 °C. Aliquots of the overnight culture were used to inoculate (4) 100 mL pre-culture flasks. Pre-culture shook aerobically 250 r.p.m until $OD_{600}$ = 0.9-0.1 at 37 °C and then rested at 4°C overnight prior to spun down at 4000xg for 30 mins. Pellets were transferred into a vinyl anaerobic chamber and used to inoculate (4) flasks containing 1.0 L degassed LB broth supplemented with ampicillin (100 μg•mL$^{-1}$), kanamycin (50 μg•mL$^{-1}$) and excess FMN/FAD. LB cultures spun on a stir plate at 200 r.p.m until $OD_{600}$ = 0.5-0.6 at RT, then pDB1282 expression was induced by 0.2 % arabinose and media was supplemented with 40 mg $FeCl_3$ and 30 mg L-cysteine. Cultures shook 250 r.p.m until $OD_{600}$ = 0.8-0.9 then pet28a-BurkOYE expression was induced by 200 μM IPTG. All chemical supplements were degassed by water bath sonication for 15mins and remained in the anaerobic chamber for 4-5 hrs prior to use. Expression was allowed to complete overnight (16-18 hrs). Cells were harvested by centrifugation at 7500 xg in 15 min intervals at 4 °C then flash frozen in liquid nitrogen and stored at -80 °C until anaerobic purification.

### 3.4.5. Pseudo-Anaerobic Expression

Cells from cryostock were used to inoculate 250 mL LB Broth supplemented with ampicillin (100 μg•mL$^{-1}$) and kanamycin (50 μg•mL$^{-1}$) to grow overnight while shaking at 250 r.p.m, at 37 °C. Aliquots of the overnight culture were used to inoculate (6) flasks containing 2 L LB broth supplemented with ampicillin (100 μg•mL$^{-1}$), kanamycin (50 μg•mL$^{-1}$) and excess FMN/FAD. Cultures shook 250 r.p.m until $OD_{600}$ = 0.5-0.6 at 37 °C, then temperatures were reduced to 18 °C,

and pDB1282 expression was induced by 0.2 % arabinose. Media was supplemented with 80mg $FeCl_3$ and 60 mg L-cysteine. Cultures shook 250 r.p.m until $OD_{600}$ = 0.8-0.9 then pet28a-BurkOYE expression was induced by 200 µM IPTG. Expression was allowed to complete overnight (16-18 hrs). Cells were harvested by centrifugation at 7500 xg in 15 min intervals at 4 °C then flash frozen in liquid nitrogen and stored at -80 °C until anaerobic purification.

### 3.4.6. Aerobic Expression

Cells from cryostock were used to inoculate 250 mL LB Broth supplemented with ampicillin (100 µg•$mL^{-1}$) and kanamycin (50 µg•$mL^{-1}$) to grow overnight while shaking at 250 r.p.m, at 37 °C. Aliquots of the overnight culture were used to inoculate (6) flasks containing 2 L LB broth supplemented with ampicillin (100 µg•$mL^{-1}$), kanamycin (50 µg•$mL^{-1}$) and excess FMN/FAD. Cultures shook 250 r.p.m until $OD_{600}$ = 0.5-0.6 at 37 °C, then temperatures were reduced to 18 °C, and the pet28a-BurkOYE expression was induced by 200 µM IPTG. Expression was allowed to complete overnight (16-18 hrs). Cells were harvested by centrifugation at 7500 xg in 15 min intervals at 4 °C the frozen at -20°C until aerobic purification.

### 3.4.7. Purification

After thawing pellet on ice in anaerobic chamber, the pellet was resuspended in chilled strep binding buffer (40 mM Tris-HCl pH 8 and 10mM NaCl, 2.5 mL/g pellet) supplemented with DNase I (Millipore) and 1:1000 1x protease inhibitor cocktail. The mixture was sonicated on ice five cycles of 1min on/ 8mins off, replenishing ice as needed. Clarified lysate was obtained by

centrifugation in tightly sealed and parafilmed tubes for 1hr at 20,000 xg. Due to low expression yield, the clarified lysate for the anaerobic expression was further clarified by 500 µL Millipore filter units: MWCO: 30 kDa followed by MWCO: 10 kDa. This product was used for biotransformation.

Clarified lysate from pseudo-anaerobic expression was incubated on ice for 1 hr with excess FMN and FAD then loaded on a 5 mL StrepTrap HP streptactin sepharose column (Cytiva) attached to an Akta Go fast protein liquid chromatography (FPLC) instrument housed within the anaerobic chamber. The column was washed with two CVs of strep binding buffer, followed by elution with 3 CVs of 100 % strep binding buffer (40 mM Tris-HCl pH 8.0, 0.5 mM desthiobiotin). Protein elution was monitored by UV absorption at 280 nm. Product fractions were concentrated and buffer-exchanged using a Millipore filter unit (MWCO: 10 kDa), then further purified by anion exchange on a 5 mL HiTrap Q FF column. The column was washed with two CVs of strep binding buffer, followed by a linear gradient 15-18 CVs to 50% anion elution buffer (40 mM Tris-HCl pH 8.0, 2 M NaCl). Protein elution was monitored as described above. Fractions containing the BurkOYE were collected and assessed for purity by SDS-PAGE. Protein concentration was quantified by Bradford assay.

Aerobic Purification was performed as detailed for the pseudo-anaerobic expression purification above however outside of the anaerobic chamber in a 4 °C cold room. Purification was performed on an BioRad FPLC with dual UV absorption allowing for monitoring of protein with intact flavin: 280 nm (protein) and 450 nm (flavin). Product fractions were collected and quantified as described above.

### 3.4.8. pH Optimization

**(3)** was dissolved in DMSO. The pH-specific activity of BurkOYE was determined anaerobically using degassed solutions. Triplicate reaction vials contained protein extract (500 nM from clarified anaerobically expressed BurkOYE), 250 mM **(3)**, NADH (10 μM), glucose dehydrogenase (2 U), glucose (100 mM) in 100 mM phosphate buffer pH 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, and 10. Reactions underwent at RT for 24 hrs, after which the reactions were quenched with 0.5 mM cyclohexanone in ethyl acetate.

### 3.4.9. Biotransformation

Each substrate was dissolved in DMSO so that the final concentration of DMSO was <0.5 %. Glucose dehydrogenase and glucose were supplemented into the reaction mix to allow for nicotinamide regeneration. The substrate-specific activity of each BurkOYE expression condition was determined anaerobically using solutions degassed in an anaerobic chamber for a minimum of 24 hrs. Reaction vials contained protein extract (500 nM from purified enzyme or clarified whole cell lysate), 250 mM substrate in DMSO, NADH (10 μM), glucose dehydrogenase (2 U), glucose (100 mM) in 100 mM phosphate buffer pH 8.0. Reactions underwent at RT for 24 hrs, after which the reactions containing substrates **(3)**, **(4)**, **(5)**, **(6)**, **(7)**, and **(8)** were quenched with 0.5mM cyclohexanone in ethyl acetate. The remaining substrates reactions were freeze quenched at -20 °C overnight and immediately analyzed.

### 3.4.10. Analytical methods

Substrates percent conversions were analyzed by gas chromatography-mass spectrometry (GC-MS) and liquid chromatography-mass spectrometry (LC–MS). GC-MS spectra for **(3)**, **(4)**, **(5)**, **(6)**, **(7)**, and **(8)** were obtained on a Shimadzu QP2010 SE instrument under the same conditions as the previous chapter.

LC-MS spectra for **(1)**, **(2)**, **(9)**, and **(10)** were obtained on a Thermo scientific LTQ Orbitrap Velos equipped with a C-18 column**. (1)** and **(2)** were collected in positive mode, while **(9)** and **(10)** were collected in negative mode.

The determination of iron in BurkOYE solutions was performed with a quadrupole mass spectrometer inductively coupled plasma mass spectrometer (ICPMS, Perkin Elmer NexION 2000) at the Mass Spectrometry Center, Department of Chemistry, Emory University. Protein solutions were thermally denatured at 95 °C and diluted 1:100. External standards were 0, 1.000, 10.000, 100.000, and 1000.000 ppb Fe/mL respectively.

### 3.4.11. BurkOYE AlphaFold Prediction

BurkOYE AlphaFold model was generated in the AlphaFold structure prediction tool within USCF ChimeraX 1.5.[41]

### 3.4.12. Genomic Neighborhood Analysis of BurkOYE

Genomic neighborhood of BurkOYE was visualized via Gene Graphics: a genomic neighborhood data visualization web application inputting BurkOYE Protein ID: BAO89787 at a bp region of 10,000.[56] Punitive enzyme function annotation was provided through Uniprot.

## 3.5.Supplemental Information

*Table 3.3. FASTA Burk OYE Sequence*

>Burkholderia sp. RPE67. Proteobacteria Bacteria
MTTASSDPLLQPFTLKHLVLKNRVMSTSHASRLTRDEFPQEVYQRYHEEKAKGGL AL
TMFGGSSNVSLDSPNTFQQINLSADAVVPHLRRFSDRIHTHGAALMCQITHLGRRGDA
YTEPWLPMIAPSPVRETLHRAMPQAIHDADIKRVIRDFGLAAKRCRDGGLDGIETHAG
GHLIGQFMDPTVNLRTDKYGGSTANRVRFAIEVHEEIRKQVGDDFVVGFRFALEDGC
SFEEGLEMSRILQGTGLFDFFNVTFGRMDTKMALAVNSMPGMFVPSAPWLPKAAAFK
RAVDLPVFHAAKIADLATARYAIREGLLDMVGMTRAHIAEPHLVKLVEAGKEDEARP

CVGASFCRNFRATCIHNPATSRETYLTHDIPKAAQTKRVLIVGAGPAGLEAARICATR
GHDVTVLEANSTAGGQLLLAATGSWRRDLIGIVDWRVSALERLSVDVRYNHYAELS
DVLDHGADVVIIATGGLPNLDALPGAEHCKSVFDALTETPPREGSVIVYDGTGRHNAY
LCAERYVDAGLDVSLALIDSMPAQETGGRGDDQVWMRNIARWDVPVRTNIELIEVTA
SSNGKRRAVFQHHLTNERVELEADHVVVERGMLAVEDLFEAARMYSANDGYTDLE
AFATGKPQPGHEAEEGQFHLYRIGDASASRDIHTAIYDAYRLCLAL

**3.6. Bibliography**

(1) Winkler, C. K.; Tasnádi, G.; Clay, D.; Hall, M.; Faber, K. Asymmetric Bioreduction Of Activated Alkenes To Industrially Relevant Optically Active Compounds. *Journal Of Biotechnology* **2012**, *162* (4), 381-389.

(2) Shi, Q.; Wang, H.; Liu, J.; Li, S.; Guo, J.; Li, H.; Jia, X.; Huo, H.; Zheng, Z.; You, S.; Et Al. Old Yellow Enzymes: Structures And Structure-Guided Engineering For Stereocomplementary Bioreduction. *Applied Microbiology And Biotechnology* **2020**, *104* (19), 8155-8170.

(3) Li, C.-J.; Trost, B. M. Green Chemistry For Chemical Synthesis. *Proceedings Of The National Academy Of Sciences* **2008**, *105* (36), 13197-13202.

(4) Toogood, H. S.; Scrutton, N. S. Discovery, Characterization, Engineering, And Applications Of Ene-Reductases For Industrial Biocatalysis. *ACS Catalysis* **2018**, *8* (4), 3532-3549.

(5) Toogood, H. S.; Gardiner, J. M.; Scrutton, N. S. Biocatalytic Reductions And Chemical Versatility Of The Old Yellow Enzyme Family Of Flavoprotein Oxidoreductases. *Chemcatchem* **2010**, *2* (8), 892-914.

(6) Castellan, A.; Bart, J. C. J.; Cavallaro, S. Industrial Production And Use Of Adipic Acid. *Catalysis Today* **1991**, *9* (3), 237-254.

(7) Nicholson, S. R.; Rorrer, N. A.; Uekert, T.; Avery, G.; Carpenter, A. C.; Beckham, G. T. Manufacturing Energy And Greenhouse Gas Emissions Associated With United States Consumption Of Organic Petrochemicals. *ACS Sustainable Chemistry & Engineering* **2023**, *11* (6), 2198-2208.

(8) Alshammari, A. S. Heterogeneous Gold Catalysis: From Discovery To Applications. *Catalysts* **2019**, *9* (5), 402.

(9) Curran, K. A.; Leavitt, J. M.; Karim, A. S.; Alper, H. S. Metabolic Engineering Of Muconic Acid Production In Saccharomyces Cerevisiae. *Metabolic Engineering* **2013**, *15*, 55-66.

(10) Joo, J. C.; Khusnutdinova, A. N.; Flick, R.; Kim, T.; Bornscheuer, U. T.; Yakunin, A. F.; Mahadevan, R. Alkene Hydrogenation Activity Of Enoate Reductases For An Environmentally Benign Biosynthesis Of Adipic Acid. *Chemical Science* **2017**, *8* (2), 1406-1413, 10.1039/C6SC02842J.

(11) Reed, T.; Lushington, G. H.; Xia, Y.; Hirakawa, H.; Travis, D. M.; Mure, M.; Scott, E. E.; Limburg, J. Crystal Structure Of Histamine Dehydrogenase From Nocardioides Simplex. *Journal Of Biological Chemistry* **2010**, *285* (33), 25782-25791.

(12) Loechel, C.; Basran, A.; Basran, J.; Scrutton, N. S.; Hall, E. A. H. Using Trimethylamine Dehydrogenase In An Enzyme Linked Amperometric Electrode Part 1. Wild-Type Enzyme Redox Mediation. *Analyst* **2003**, *128* (2), 166-172, 10.1039/B211895E.

(13) White, D. W., Iamuri, S., Joud, P., Blue, T.C., Copp, J., Lutz, S. The Hidden Biocatalytic Potential Of The Old Yellow Enzyme Family. *Biorxiv* **2023**.

(14) Copp, J. N.; Akiva, E.; Babbitt, P. C.; Tokuriki, N. Revealing Unexplored Sequence-Function Space Using Sequence Similarity Networks. *Biochemistry* **2018**, *57* (31), 4651-4662.

(15) Scholtissek, A.; Tischler, D.; Westphal, A. H.; Van Berkel, W. J. H.; Paul, C. E. Old Yellow Enzyme-Catalysed Asymmetric Hydrogenation: Linking Family Roots With Improved Catalysis. *Catalysts* **2017**, *7* (5), 130.

(16) Peters, C.; Frasson, D.; Sievers, M.; Buller, R. Novel Old Yellow Enzyme Subclasses. *Chembiochem* **2019**, *20* (12), 1569-1577.

(17) Kumar Roy, T.; Sreedharan, R.; Ghosh, P.; Gandhi, T.; Maiti, D. Frontispiece: Ene-Reductase: A Multifaceted Biocatalyst In Organic Synthesis. *Chemistry – A European Journal* **2022**, *28* (21), E202282162.

(18) Winkler, C. K.; Clay, D.; Davies, S.; O'Neill, P.; Mcdaid, P.; Debarge, S.; Steflik, J.; Karmilowicz, M.; Wong, J. W.; Faber, K. Chemoenzymatic Asymmetric Synthesis Of Pregabalin Precursors Via Asymmetric Bioreduction Of B-Cyanoacrylate Esters Using Ene-Reductases. *The Journal Of Organic Chemistry* **2013**, *78* (4), 1525-1533.

(19) Atkinson, H. J.; Morris, J. H.; Ferrin, T. E.; Babbitt, P. C. Using Sequence Similarity Networks For Visualization Of Relationships Across Diverse Protein Superfamilies. *PLOS One* **2009**, *4* (2), E4345.

(20) Bairoch, A.; Apweiler, R. The SWISS-PROT Protein Sequence Database And Its Supplement Trembl In 2000. *Nucleic Acids Research* **2000**, *28* (1), 45-48.

(21) Tu, X.; Hubbard, P. A.; Kim, J.-J. P.; Schulz, H. Two Distinct Proton Donors At The Active Site Of Escherichia Coli 2,4-Dienoyl-Coa Reductase Are Responsible For The Formation Of Different Products. *Biochemistry* **2008**, *47* (4), 1167-1175.

(22) Trickey, P.; Basran, J.; Lian, L. Y.; Chen, Z.; Barton, J. D.; Sutcliffe, M. J.; Scrutton, N. S.; Mathews, F. S. Structural And Biochemical Characterization Of Recombinant Wild Type And A

C30A Mutant Of Trimethylamine Dehydrogenase From Methylophilus Methylotrophus (Sp. W(3)A(1)). *Biochemistry* **2000**, *39* (26), 7678-7688.

(23) Liu, G.; Wang, W.; He, F.; Zhang, P.; Xu, P.; Tang, H. Structural Insights Into 6-Hydroxypseudooxynicotine Amine Oxidase From Pseudomonas Geniculata N1, The Key Enzyme Involved In Nicotine Degradation. *Applied and Environmental Microbiology* **2020**, *86* (19).

(24) Willistein, M.; Bechtel, D. F.; Müller, C. S.; Demmer, U.; Heimann, L.; Kayastha, K.; Schünemann, V.; Pierik, A. J.; Ullmann, G. M.; Ermler, U.; Et Al. Low Potential Enzymatic Hydride Transfer Via Highly Cooperative And Inversely Functionalized Flavin Cofactors. *Nature Communications* **2019**, *10* (1), 2074.

(25) Mordaka, P. M.; Hall, S. J.; Minton, N.; Stephens, G. Recombinant Expression And Characterisation Of The Oxygen-Sensitive 2-Enoate Reductase From Clostridium Sporogenes. *Microbiology (Reading)* **2018**, *164* (2), 122-132.

(26) Kohli, R. M.; Massey, V. The Oxidative Half-Reaction Of Old Yellow Enzyme. The Role Of Tyrosine 196. *Journal Of Biological Chemistry* **1998**, *273* (49), 32763-32770.

(27) Zheng, L.; Cash, V. L.; Flint, D. H.; Dean, D. R. Assembly Of Iron-Sulfur Clusters. Identification Of An Iscsua-Hscba-Fdx Gene Cluster From Azotobacter Vinelandii. *Journal Of Biological Chemistry* **1998**, *273* (21), 13264-13272.

(28) Pesic, M.; Fernández-Fueyo, E.; Hollmann, F. Characterization Of The Old Yellow Enzyme Homolog From Bacillus Subtilis (Yqjm). *Chemistryselect* **2017**, *2* (13), 3866-3871.

(29) Weckbecker, A.; Hummel, W. Glucose Dehydrogenase For The Regeneration Of NADPH And NADH. In *Microbial Enzymes And Biotransformations*, Barredo, J. L. Ed.; Humana Press, 2005; Pp 225-238.

(30) Sun, J.; Lin, Y.; Shen, X.; Jain, R.; Sun, X.; Yuan, Q.; Yan, Y. Aerobic Biosynthesis Of Hydrocinnamic Acids In Escherichia Coli With A Strictly Oxygen-Sensitive Enoate Reductase. *Metabolic Engineering* **2016**, *35*, 75-82.

(31) Johnson, M.; Zaretskaya, I.; Raytselis, Y.; Merezhuk, Y.; Mcginnis, S.; Madden, T. L. NCBI BLAST: A Better Web Interface. *Nucleic Acids Research* **2008**, *36* (Suppl_2), W5-W9.

(32) Alloing, G.; Travers, I.; Sagot, B.; Rudulier, D. L.; Dupont, L. Proline Betaine Uptake In <I>Sinorhizobium Meliloti</I>: Characterization Of Prb, An Opp-Like ABC Transporter Regulated By Both Proline Betaine And Salinity Stress. *Journal Of Bacteriology* **2006**, *188* (17), 6308-6317.

(33) Bashir, A. M.; Hoffmann, T.; Kempf, B.; Xie, X.; Smits, S. H. J.; Bremer, E. Plant-Derived Compatible Solutes Proline Betaine And Betonicine Confer Enhanced Osmotic And Temperature Stress Tolerance To Bacillus Subtilis. *Microbiology* **2014**, *160 Pt 10*, 2283-2294.

(34) Brown, A. C.; Suess, D. L. M. An Open-Cuboidal [Fe3S4] Cluster Characterized In Both Biologically Relevant Redox States. *Journal Of The American Chemical Society* **2023**, *145* (4), 2075-2080.

(35) Jang, M.-H.; Basran, J.; Scrutton, N. S.; Hille, R. The Reaction Of Trimethylamine Dehydrogenase With Trimethylamine *. *Journal Of Biological Chemistry* **1999**, *274* (19), 13147-13154.

(36) Basran, J.; Jang, M.-H.; Sutcliffe, M. J.; Hille, R.; Scrutton, N. S. The Role Of Tyr-169 Of Trimethylamine Dehydrogenase In Substrate Oxidation And Magnetic Interaction Between FMN Cofactor And The 4Fe/4S Center *. *Journal Of Biological Chemistry* **1999**, *274* (19), 13155-13161.

(37) Daughtry, K. D.; Xiao, Y.; Stoner-Ma, D.; Cho, E.; Orville, A. M.; Liu, P.; Allen, K. N. Quaternary Ammonium Oxidative Demethylation: X-Ray Crystallographic, Resonance Raman, And UV–Visible Spectroscopic Analysis Of A Rieske-Type Demethylase. *Journal Of The American Chemical Society* **2012**, *134* (5), 2823-2834.

(38) Basran, J.; Sutcliffe, M. J.; Scrutton, N. S. Optimizing The Michaelis Complex Of Trimethylamine Dehydrogenase: Identification Of Interactions That Perturb The Ionization Of Substrate And Facilitate Catalysis With Trimethylamine Base *. *Journal Of Biological Chemistry* **2001**, *276* (46), 42887-42892.

(39) Outten, F. W. Iron-Sulfur Clusters As Oxygen-Responsive Molecular Switches. *Nature Chemical Biology* **2007**, *3* (4), 206-207.

(40) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A. Highly Accurate Protein Structure Prediction With Alphafold. *Nature* **2021**, *596* (7873), 583-589.

(41) Goddard, T. D.; Huang, C. C.; Meng, E. C.; Pettersen, E. F.; Couch, G. S.; Morris, J. H.; Ferrin, T. E. UCSF Chimerax: Meeting Modern Challenges In Visualization And Analysis. *Protein Science* **2018**, *27* (1), 14-25.

(42) Cunningham, F.; Allen, J. E.; Allen, J.; Alvarez-Jarreta, J.; Amode, M R.; Armean, Irina M.; Austine-Orimoloye, O.; Azov, Andrey G.; Barnes, I.; Bennett, R.; Et Al. Ensembl 2022. *Nucleic Acids Research* **2021**, *50* (D1), D988-D995.

(43) Lawrence, J. G. Shared Strategies In Gene Organization Among Prokaryotes And Eukaryotes. *Cell* **2002**, *110* (4), 407-413.

(44) Calhoun, S.; Korczynska, M.; Wichelecki, D. J.; San Francisco, B.; Zhao, S.; Rodionov, D. A.; Vetting, M. W.; Al-Obaidi, N. F.; Lin, H.; O'Meara, M. J.; Et Al. Prediction Of Enzymatic Pathways By Integrative Pathway Mapping. *Elife* **2018**, *7*.

(45) Ly, A.; Henderson, J.; Lu, A.; Culham, D. E.; Wood, J. M. Osmoregulatory Systems Of Escherichia Coli: Identification Of Betaine-Carnitine-Choline Transporter Family Member Betu And Distributions Of Betu And Trkg Among Pathogenic And Nonpathogenic Isolates. *Journal of Bacteriology* **2004**, *186* (2), 296-306.

(46) Detmers, F. J.; Lanfermeijer, F. C.; Poolman, B. Peptides And ATP Binding Cassette Peptide Transporters. *Research in Microbiology* **2001**, *152* (3-4), 245-258.

(47) Burnet, M. W.; Goldmann, A.; Message, B.; Drong, R.; El Amrani, A.; Loreau, O.; Slightom, J.; Tepfer, D. The Stachydrine Catabolism Region In Sinorhizobium Meliloti Encodes A Multi-Enzyme Complex Similar To The Xenobiotic Degrading Systems In Other Bacteria. *Gene* **2000**, *244* (1-2), 151-161.

(48) Bremer, E. Coping With Osmotic Challenges : Osmoregulation Through Accumulation And Release Of Compatible Solutes In Bacteria. **2000**;

(49) Tomii, K.; Kanehisa, M. A Comparative Analysis Of ABC Transporters In Complete Microbial Genomes. *Genome Research* **1998**, *8* (10), 1048-1059.

(50) Alley, E. C.; Khimulya, G.; Biswas, S.; Alquraishi, M.; Church, G. M. Unified Rational Protein Engineering With Sequence-Based Deep Representation Learning. *Nature Methods* **2019**, *16* (12), 1315-1322.

(51) Poppe, L.; Vértessy, B. G. The Fourth Wave Of Biocatalysis Emerges— The 13 Th International Symposium On Biocatalysis And Biotransformations. *Chembiochem* **2018**, *19* (4), 284-287.

(52) Bornscheuer, U. T. The Fourth Wave Of Biocatalysis Is Approaching. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2018**, *376* (2110).

(53) Uniprot Consortium, T. Uniprot: The Universal Protein Knowledgebase. *Nucleic Acids Research* **2018**, *46* (5), 2699-2699.

(54) Li, W.; Godzik, A. Cd-Hit: A Fast Program For Clustering And Comparing Large Sets Of Protein Or Nucleotide Sequences. *Bioinformatics* **2006**, *22* (13), 1658-1659.

(55) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment For Integrated Models Of Biomolecular Interaction Networks. *Genome Research* **2003**, *13* (11), 2498-2504.

(56) Harrison, K. J.; Crécy-Lagard, V.; Zallot, R. Gene Graphics: A Genomic Neighborhood Data Visualization Web Application. *Bioinformatics* **2018**, *34* (8), 1406-1408.

# **Chapter 4.** *In Silico* Elucidation of RiPP Recognition in Radical S-adenosyl-L-methionine (SAM) SuiB

*This work was performed in collaboration with Stacey Jones (CD validation of SuiA) and the Seyedsayamdost Lab at Princeton University (fluorescence assay).*

## 4.1.Introduction

Since its discovery in 1928, penicillin has saved an estimated 80-200 million lives.[1] Natural products are bioactive secondary metabolites isolated from biological organisms. Secondary metabolites are compounds nonessential to organismal growth and/or reproduction but provide an evolutionary advantage aiding survival.[2-4] Isolated from *Penicillium rubens*, penicillin is one of the most well-known secondary metabolites.[5] The broad-spectrum antibiotic destabilizes the DD-transpeptidase mediated cross-linking necessary to maintain the peptidoglycan cell walls of prokaryotes.[6, 7] The discovery of penicillin marked the beginning of the modern "era of antibiotics" and soon after spawned a parallel era of antibiotic resistance.[8]

In his 1945 Nobel lecture, Sir Alexander Flemming warned that microbes have an innate genetic plasticity allowing them to gain resistance to antimicrobials when exposed to insufficient concentrations.[9] His warning was clear, stating: *"[A] time may come when penicillin can be bought by anyone in the shops. Then there is the danger that the ignorant man may easily underdose himself and by exposing his microbes to non-lethal quantities of the drug make them resistant.*"

*-Sir Alexander Flemming, 1945 Nobel lecture, Penicillin*

Sir Flemming's fears were realized when the first case of antibiotic resistance was observed, and additional resistant species continued to emerge through the 21$^{st}$ centruy.[6, 10] In addition to underuse, the overuse of antibiotics during times of nonbacterial infections notably exacerbates antibiotic resistance.[6, 11] The emergence of pathogenic bacterial resistance to modern antibiotics is recognized as a significant public health threat affecting humans worldwide.[11-13] Described as the "silent pandemic", antibiotic resistance is a leading cause of death worldwide.[14] According to the Centers for Disease Control (CDC), more than five million deaths worldwide were attributed to

antibiotic resistance in 2019.[15, 16] Developing novel antimicrobial agents coupled with an appropriate application strategy is crucial to combating antibiotic resistance.[17]

Ribosomally synthesized and posttranslationally modified peptides (RiPPs) comprise a superfamily of natural products that display diverse bioactivity as antibiotics (anti-Gram-positive and anti-Gram-negative), antifungals, antivirals, and anticancer agents.[18-21] In juxtaposition to many other major classes of natural products (e.g. terpenoids, alkaloids, and polyketides), RiPPs are structurally diverse peptides (<10kDa) biosynthesized from a RiPP biosynthetic gene cluster (BGC) via the ribosome.[22, 23] RiPP precursor peptides are generally comprised of an N-terminal leader sequence and a C-terminal core sequence (Figure 4.1A); however, some precursor peptides contain an additional N-terminal signaling and/or C-terminal follower sequence depending on RiPP species of origin and the RiPP class.[19] While RiPP-modifying enzymes posttranslationally modify the core sequence, the leader/follower sequence is maintained, functioning as a recognition region. The precursor peptide gains bioactivity and is "matured" to a RiPP post-functionalization and cleavage of the leader/follower sequence.[24] The ribosomal origin of RiPPs and leader-dependent recognition during RiPP biosynthesis provide versatility important for bioengineering efforts, e.g. the RiPP core can be adjusted while leaving tailoring enzyme recognition unaffected.

The leader sequence is commonly identified for modification by a RiPP recognition element (RRE) domain. As discussed in Chapter 1, RREs are characterized by a conserved winged helix-turn-helix (wHTH) secondary structure motif composed of three α-helixes and a three to four-stranded β-sheet (Figure 4.1B).[25-27] Binding interactions are typically observed between the third α-helix and β-strand of the RRE, with the leader sequence extending the β-sheet.[26] It is believed that RREs recognize the precursor peptide by binding to highly conserved recognition sequences in the leader region. These recognition sequences are specific to each class of RiPP. Lasso peptides, for

example, contain a highly conserved YxxP motif.[28, 29] As many RiPPs rely on RREs for maturation, the structural motif has become a target for the bioinformatic identification of novel RiPPs classes.[30, 31]

Various RiPP tailoring enzymes exist in RiPP BGCs, depending on the RiPP product. These affiliated biosynthetic enzymes include peptidases, cyclases, protases, and C-H functionalizing enzymes. Many C-H functionalizing RiPP modifying enzymes belong to the radical *S*-adenosyl-L-methionine (rSAM) superfamily (PFam: PF04055).[32-34] For example, genes encoding for rSAM enzymes are found in the biosynthetic gene clusters of thiopeptides,[35] proteusins,[36] and bottromycins.[37] Specializing in radical-initiated C-H functionalization, the >750,000 members of the rSAM superfamily rely on a $[4Fe\text{-}4S]^{+}$ cluster to reductively cleave SAM to form a highly reactive 5′-deoxyadenosyl radical (dAdo•) intermediate and methionine (Figure 4.1C).[38-40] The dAdo• intermediate is then used to catalyze diverse reactivity, including carbon methylations, methylthiolations, sulfur insertions, carbon insertions, decarboxylations, epimerizations, rearrangements, and protein/peptide modifications.[41] A large subset of rSAMs are RiPPs modifying enzymes.[26, 42, 43]

One such RiPP modifying rSAM is SuiB from *Streptococcus suis* (Uniprot accession code: A0A0Z8EWX1). SuiB catalyzes the instillation of a lysine-tryptophan crosslink between two unactivated carbons (βC of the Lys2 residue and the C7-indole of Tyr6) on its precursor peptide, SuiA (Figure 4.1A,C).[26] The same lysine-tryptophan crosslink is observed in the StrB from *Streptococcus thermophilus* biosynthesis of streptide, a signaling peptide implicated in quorum sensing.[26, 42, 44] SuiB displays an N-terminal RRE region (residues 1-106), a conserved SAM domain (residues 107-310), a bridging domain (residues 311-346), and a SPASM domain (residues 347-437) (Figure 4.1D).[26, 39] The SAM domain displays a six (β/α) partial triose-phosphate

isomerase (TIM) barrel, which houses the highly conserved $CX_3CX\Phi C$ motif, where Φ represents Phe123 in SuiB. The three cysteines (Cys117, Cys121, and Cys124) coordinate a [4Fe-4S] cluster leaving one open coordination site for SAM cleavage.[39] SuiB also possesses a C-terminal SPASM domain with eight cysteines that allow for the incorporation of two auxiliary (Aux) [4Fe-4S] clusters: Aux I (Cys321, Cys 365, Cys347, and Cys419) and Aux II (Cys406, Cys409, Cys415, and Cys437). These clusters are positioned 16.0 Å and 27.4 Å from the SAM domain cluster, respectively, and 9.9 Å from each other. The auxiliary clusters are proposed to aid electron transfer, where Aux I acts as an oxidant to the Lys-Trp radical (Figure 4.1C). AuxI then shuttles the electron to Aux II which releases it to a protein electron acceptor.[26, 44, 45]

The RRE region of SuiB displays the expected wHTH motif containing a three-stranded antiparallel β-sheet connected to four α-helices (Figure 4.1B). Interestingly, while SuiB contains an RRE domain, no evidence for the SuiA leader sequence interacting with the RRE has been published to date. This is surprising, as previously reported structures of RRE-containing tailoring enzyme:peptide complexes depict formation of the characteristic β-sheet with the associated RRE.[25, 26] The crystal structure of SuiB (PDB: 5V1T), by contrast, shows primary interactions in the catalytic barrel, where the leader sequence interfaces with the so-called ‘bridging region’ of SuiB, suggesting either a vestigial or more nuanced role for the RRE.[26] Understanding the basis for peptide recognition and recruitment in SuiB is therefore important information for downstream bioengineering efforts.

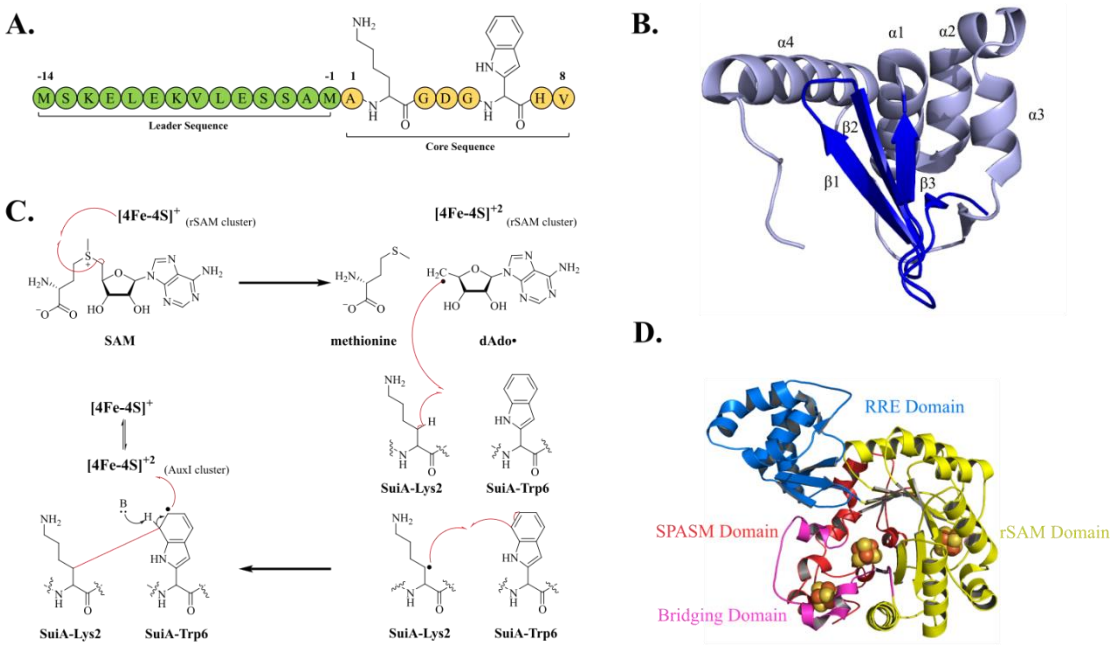



*Figure 4.1 Structure and Mechanism of SuiB and SuiA. , (A) Scheme precursor peptide SuiA depicting leader and core sequence, (B) wHTH motif in SuiB displaying three-stranded antiparallel β-sheet (β1, β2, and β3) connected to three α-helixes (α1, α2, α3 and α4), (C) Proposed mechanism of Lys-Trp crosslink, and (D) SuiB contains a RRE region, rSAM domain, SPASM Domain, Bridging domain, and three [4Fe-4S] cluster metallo- cofactors*

## 4.2.Results and Discussion

### 4.2.1. *suiA* Genomic Neighborhood Analysis

RREs have been identified in >50% of all prokaryotic RiPP class BGCs, either as a protein domain within the RiPP modifying enzyme (i.e., the N-terminus of TfxB in *Rhizobium* )[46, 47] or as a discrete protein (i.e., PqqD from *Klebsiella pneumoniae*).[25, 48] The crystal structure of SuiB depicts an N-terminal RRE, however, lacks crystallographic evidence of RiPP interaction. We, therefore, first conducted a genomic neighborhood analysis to determine if an additional discrete RRE that could in RRE binding is located in the *suiA* BGC.

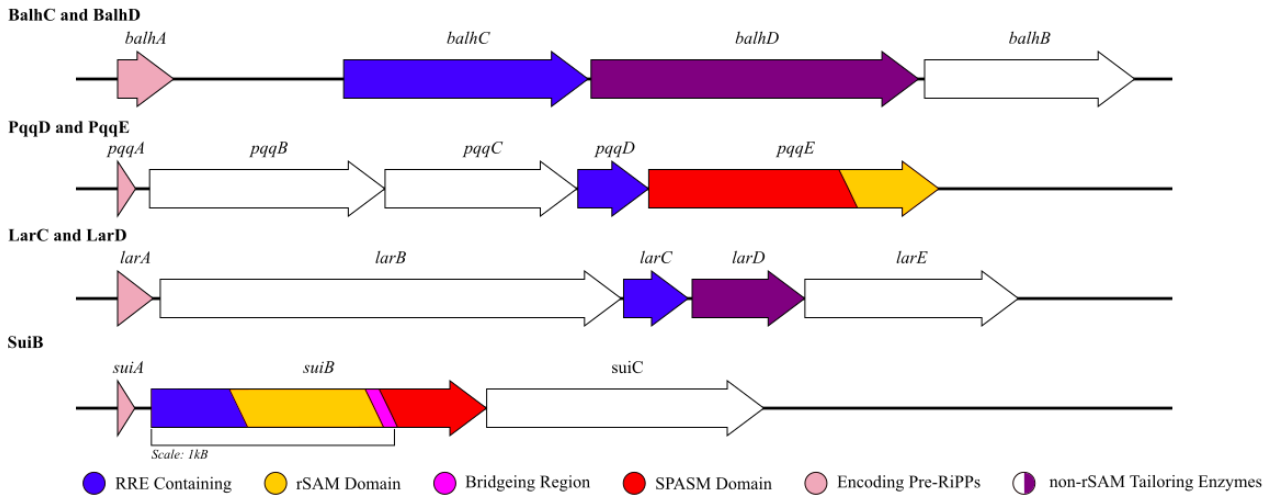



*Figure 4.2. The BGC clusters of precursor peptides balhA, pqqA, larA, and suiA. Genes that encode for: enzymes that contain RRE regions (Blue), precursor peptides (pink), and non-rSAM tailoring enzymes (white or purple). rSAM tailoring enzymes, PqqE and SuiB, present: the rSAM domain (yellow), the bridging region (magenta), and the SPASM domain (red). Genomic neighborhood generated in Gene Graphics web application.*[49]

After genomic analysis of three RiPP modifying enzymes and their proximity to the discrete RRE (LarC and LarD from *Rhodococcus jostii*,[50] BalhC and BalhD from *Bacillus thuringiensis str. Al-Hakam*,[51] and PqqD and PqqE from *K. pneumoniae*),[52] it was observed that discrete RREs are expressed directly upstream from the tailoring enzyme (Figure 4.2). The RRE then acts as a peptide chaperone by complexing with the latter, which performs the respective catalysis.[51, 52] This is not observed in *suiB*, where the proceeding gene is *suiA,* indicating that it is unlikely that a discrete RRE assists in binding (Figure 4.2).

### 4.2.2. Truncated SuiB RRE:SuiA-Fl fluorescence assay

To determine if the embedded SuiB RRE interacts with SuiA, fluorescence emission studies were performed in collaboration with Alessio Caruso from Seyedsayamdost Lab at Princeton University. A SuiB mutant composed solely of the RRE region and a TEV cleavable NusA expression tag with

an N-terminal poly-histidine tag was generated. A fluorescently labeled SuiA derivative (SuiA-Fl) was generated with 6-carboxyfluorescein (6-FAM) ligated to the N-terminus of SuiA. Potential binding interactions between SuiA-Fl and the purified and cleaved RRE product were monitored from 510-600 nm through a fractional saturation assay with constant SuiA-Fl (2 μM) and increasing concentration of RRE (0μM to 5μM). Upon binding of SuiA-Fl to RRE, there is expected to be a relative fluorescence signal change proportional to the concentration of RRE. The assay resulted in an estimated $K_D$ of a similar magnitude to other known RRE-precursor peptide interactions (0.82 ± 0.3 μM). However, it is approximately double the $K_D$ value from PqqD and PqqA from *Methylobacterium extorquens* ($K_D$ = 0.39 ± 0.08 μM), perhaps indicating a weaker interaction in SuiB.[52] The assay was repeated with wildtype SuiB (SuiBwt) and SuiA-Fl. However, oxidation of SuiBwt during fluorescence analysis yielded irreproducible data.

### 4.2.3. Characterization of RRE:SuiA interactions by *ab initio* docking

To account for the inherent oxygen sensitivity attributed to metalloenzymes, computational docking (HADDOCK 2.2 webserver) was performed to visualize binding interactions of SuiA to the RRE on SuiBwt. HADDOCK uses stochastic global-energy optimization to perform *ab initio* docking of unbound components.[53] In order to avoid biasing our results by considering interaction regions common to other RRE-precursor peptide complexes, we utilized the Random Ambiguous Interaction Restraints (AIRs), which randomly samples solvent-accessible residues and iteratively defines them as “active” binding targets to allow for a broad sampling of binding interactions. This procedure produces thousands of statistically possible, but not necessarily physiologically relevant, conformations. Thus, HADDOCK employs a clustering algorithm for ease of analysis that defines the top or best complexes by the “HADDOCK score.” The HADDOCK score is the summation of multiple energy terms, including van der Waals (Evdw), electrostatic (Eelec),

desolvation (Edesol), and restraint violation energies (Eair). This energy-based clustering is beneficial in analyzing protein systems as protein function depends on the system's energy landscape.[54] As the HADDOCK score is a weighted value that will vary as emphasis is placed on user-defined interactions, it is not designed to be a proxy for binding energy. However, the HADDOCK score does allow for guided analysis of structures that display interactions of interest.

Previously reported RRE and peptide complexes depicted the peptide forming a β-sheet with β3 on the RRE. Additionally, protein-protein interactions are generally governed by hydrophobic effects. Due to this, we biased the HADDOCK score towards increased Evdw, while reducing weight for Eelec, similar to the protein:ligand systems protocol. Under these conditions, no interactions were detected with the RRE region in either the SuiBwt:SuiA or RRE:SuiA complexes. However, the SuiBwt:SuiA and catalytic-barrel:SuiA complexes replicated the leader strand's crystallographically observed binding interactions with the bridging region between: SuiA(-5):SuiBTyr345, SuiA(-4):SuiBGly302, and SuiA(-3):SuiBPhe345 (Figure 4.3).

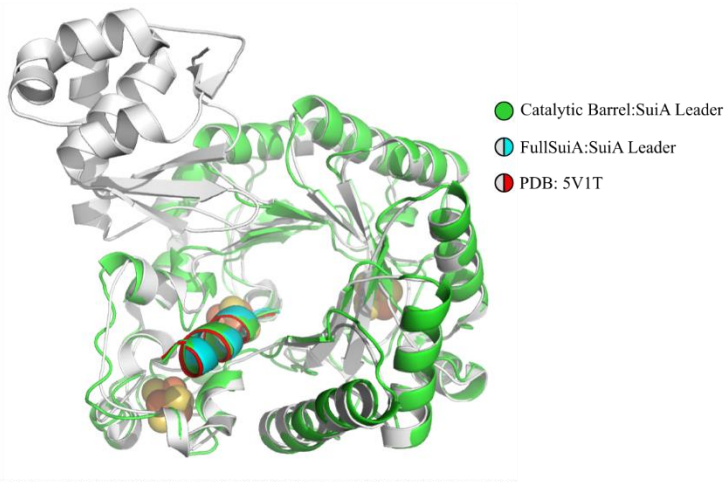



*Figure 4.3. Docking of SuiA with SuiB fragments. The docking complex of SuiA with the catalytic barrel of SuiB (green) and docking complex of SuiA (blue) with full SuiB model (white) alignment with SuiB:SuiA crystal structure (PDB: 5V1T) (white SuiB, red SuiA)*

### 4.2.4. Validation of docking approach for RiPP precursor and RRE

To validate our docking method's ability to reasonably model precursor peptide:RRE binding, the same docking parameters were utilized to reproduce the precursor peptide:RRE interactions of CteB, from *Clostridium thermocellum*. CteB is a rSAM RiPP modifying enzyme that catalyzes the formation of a sactionine thioether crosslink between Cys32 and Thr37 of its precursor peptide, CteA.[55] Composed of an N-terminal RRE, rSAM domain and C-terminal SPASM domain, CteB has structural similarities to SuiB. However, crystallographic characterization of the former captured binding of CteA to the associated RRE.[55] To perform this control, the structures of CteB and the leader strand of CteA were extracted from PDB: 5WGG. Docking runs were performed in the same method as described above and successfully generated a complex of CteA interacting with CTeB:RRE (Figure 4.4). This indicates that our method can capture RRE:precursor peptide interactions.

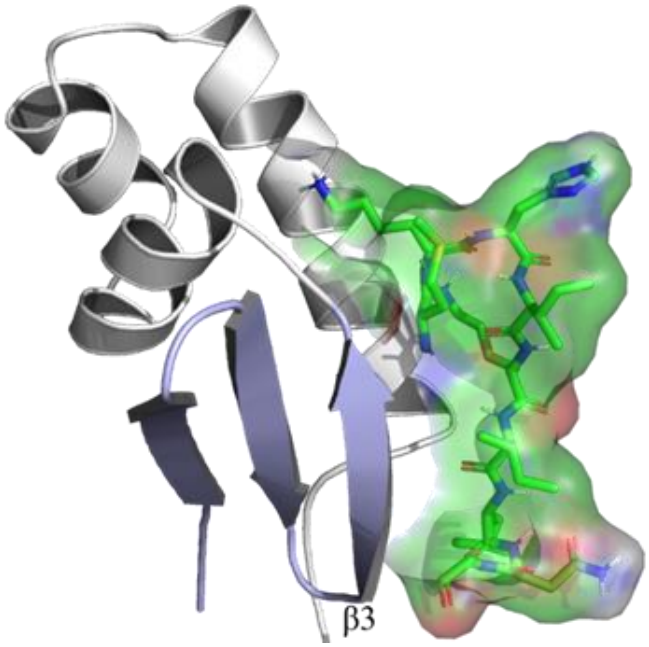



*Figure 4.4. Docking model of CteB:CteA showing interaction of CteA by β3*

### 4.2.5. Capturing SuiB Protein dynamics via MD

With validation of our docking protocol, we considered how protein dynamics might influence the SuiA interaction with SuiB. While molecular docking is a powerful tool to visualize ligand interactions, a limitation of the tool is that protein systems are not static.[56] It is even rarer for binding interactions not to couple with or precede some dynamic change.[54] SuiB is no exception. Davis and coworkers crystallographic exploration of the SuiB:SuiA complex revealed high mobility of the RRE region and dynamic changes of two loops (L1 and L2) in the rSAM domain upon SuiA binding.[26] To explore how SuiB protein dynamics affect binding of the peptide substrate, molecular dynamics simulations of SuiB in the presence of the SuiA leader strand (With SuiA) and absence of SuiA leader strand (Without SuiA) were conducted (Figure 4.4).

To confirm the validity of our prepared SuiB model (See section 4.4.2), MD simulations were run for 5 ns, after which they were analyzed by monitoring the radius of gyration ($R_g$), root-mean-square-deviation (RMSD), and root-mean-square-fluctuation (RMSF) across the time course. RMSD is a measure of displacement if the protein in comparison to the $t_0$ energy minimized (em) structure. RMSF is a more in-depth RMSD calculation performed for each atom, which allows us to identify specific domain conformation changes. All analyses were performed in comparison to the energy minimized structure which was prior to temperature and pressure coupling. Due to the short length of the simulation, a hydrogen bond (H-bond) analysis was not performed.

RMSD analysis showed that, on average, SuiB absent SuiA ($RMSD_{AVE}$= 0.22±0.04 Å) displays greater mobility than in the presence of the peptide ($RMSD_{AVE}$= 0.17±0.03 Å) (Figure 4.5C). Upon RMSF analysis, we determined that the simulated conformational changes were consistent with those reported in Davis et al.[26] Upon introduction of the peptide, large scale dynamic movements around L1 (residues 125-134) and L2 (residues 279-285) were detected (Figure 4.5B). Our study

captured an ~40 and ~200-fold increase in dynamic motion at L1 and L2, respectively. Peptide binding additionally appears to stabilize movements in the bridging region by a factor of ~250, supporting the bridging region playing a significant role in leader strand binding. Whether the bridging region has the ability to recognize and recruit the leader sequence has yet to be explored. The simulations also display significant movement of the RRE regardless of peptide binding in the barrel. Supporting that when the peptide is in the catalytic barrel primary interactions are with the bridging domain as opposed to the RRE region. Movement predominantly occurs in the three α-helixes of the RRE motif, however, it is noticeably absent in the region that links the RRE domain to the catalytic barrel (residues 101-107) (Figure 4.5A, RRE region). Movement in this region is observed in the crystal structure which may indicate that a longer simulation may provide a more accurate depiction of the system.

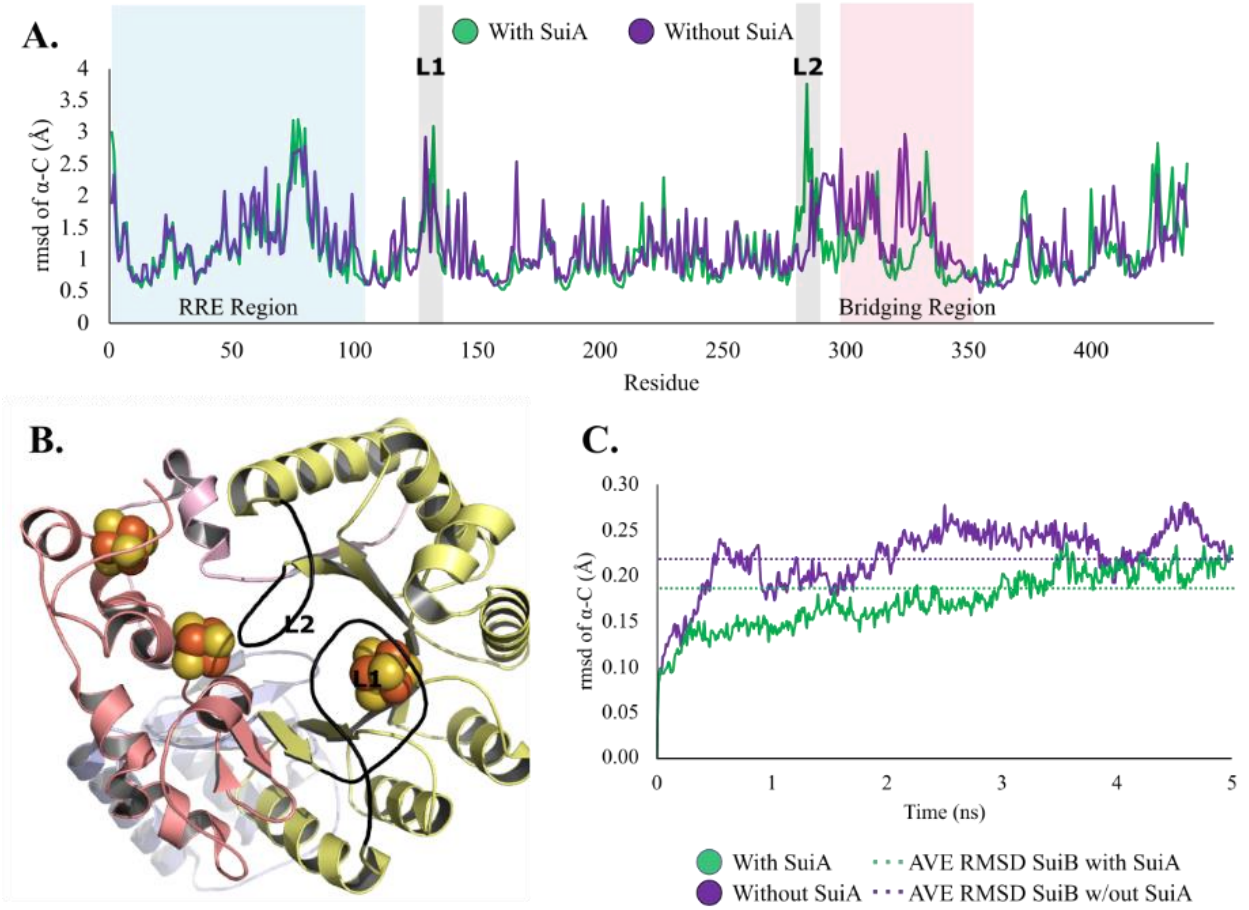



*Figure 4.5. MD simulations of SuiB with and without SuiA. (A) RMSF analysis of SuiB individual residues compared. (B) L1 and L2 are a part of the rSAM domain and fluctuate upon binding of SuiA. (C) RMSD analysis of SuiB with and without peptide.*

### 4.2.6. Characterizing SuiA Secondary Structure

Similar to proteins, peptides are also dynamic in solution. To understand how the dynamic nature of SuiA influences binding to the RRE, it is imperative to obtain a structure of SuiA prior to interaction with SuiB. While peptide crystallography is a viable option to obtain such, bioinformatic and computational tools allow for generation of similar information with less time cost while, more importantly, allowing for visualization of the dynamic peptide conformational landscape. These conformational changes may depict more physiologically relevant conformations which when coupled with MD and docking will allow us to capture SuiA:RRE interactions.

We first utilized the secondary structure prediction software JPred4 to predict SuiA's conformation absent interactions with SuiB. JPred4 is a secondary structure prediction webserver that relies on a neural network, Jnet, trained by iterative PSI-BLAST results to predict the secondary structure of uncharacterized peptides from homologous sequences of known structure.[57, 58] In addition to secondary structure prediction, JPred4 also comments on likely solvent accessibility and coiled-coil regions.[59] The Jnet output consists of (1) *Jnet_0, 5, and 25*, which are predictions of AA solvent accessibility; (2) *Jnet Rel*, which is an internal prediction accuracy metric from 0-9 where 9 has the highest predicted accuracy; (3) *Lupas* 14, 21, and 28 which represent coil-coil predictions; (4) *jhmm* , which uses hidden Markov models to output predicted secondary structure; and (5) *Jnet*, the final secondary structure prediction which takes into account the other components for a consensus output.

JPred4, was used to predict the secondary structure of the leader and core region of SuiA resulting in a *Jnet* of --HHHHHHHH---------- (Figure 4.6). SuiA produced no BLAST Uniref90 hits, so the JPred prediction was made solely using Markov models derived from the primary sequence. The resultant "H" indicates helical conformation for a portion of the leader sequence (residues (-12)-(-5)), while the "–" indicates disordered regions for the remaining segments of the leader sequence and the entire core region. The *Jnet Rel* estimates that the H designation of the leader sequence is reliable. This supports the 5V1T crystal structure, where all portions of SuiA leader sequence were helical. However, up until SuiB, no leader sequence had been crystallographically captured in a helical conformation. This cast doubt on how SuiA conformation varied in solution, thus motivating MD simulations.

OrigSeq : MSKELEKVLESSSAMAKGDGWHV : OrigSeq

Jnet : --HHHHHHHH---------- : Jnet
jhmm : --HHHHHHHH---------- : jhmm

Lupas 14 : ---------- : Lupas 14
Lupas 21 : ---------- : Lupas 21
Lupas 28 : ---------- : Lupas 28

Jnet_25 : ----B--BB--B-BB-------- : Jnet_25
Jnet_5 : ---------- : Jnet_5
Jnet_0 : ---------- : Jnet_0
Jnet Rel : 985899987515534678876 58 : Jnet Rel

*Figure 4.6. SuiA Jpred Prediction Results*

An 80 ns MD simulation of the full SuiA sequence with initial coordinates from 5V1T (Figure 4.7) depicts the SuiA leader sequence readily fluctuating between helical and disordered states. However, the modeled in SuiA core does not adopt a coiled conformation after its initial loss at 5 ns (Supplemental Information: Table 4.2.). SuiA adopted an α-helical state in both the leader and core during only 8% of the simulation; a partially helical, partially disordered state for 22% of the simulation; and a completely disordered loop for 70% (Figure 4.7).

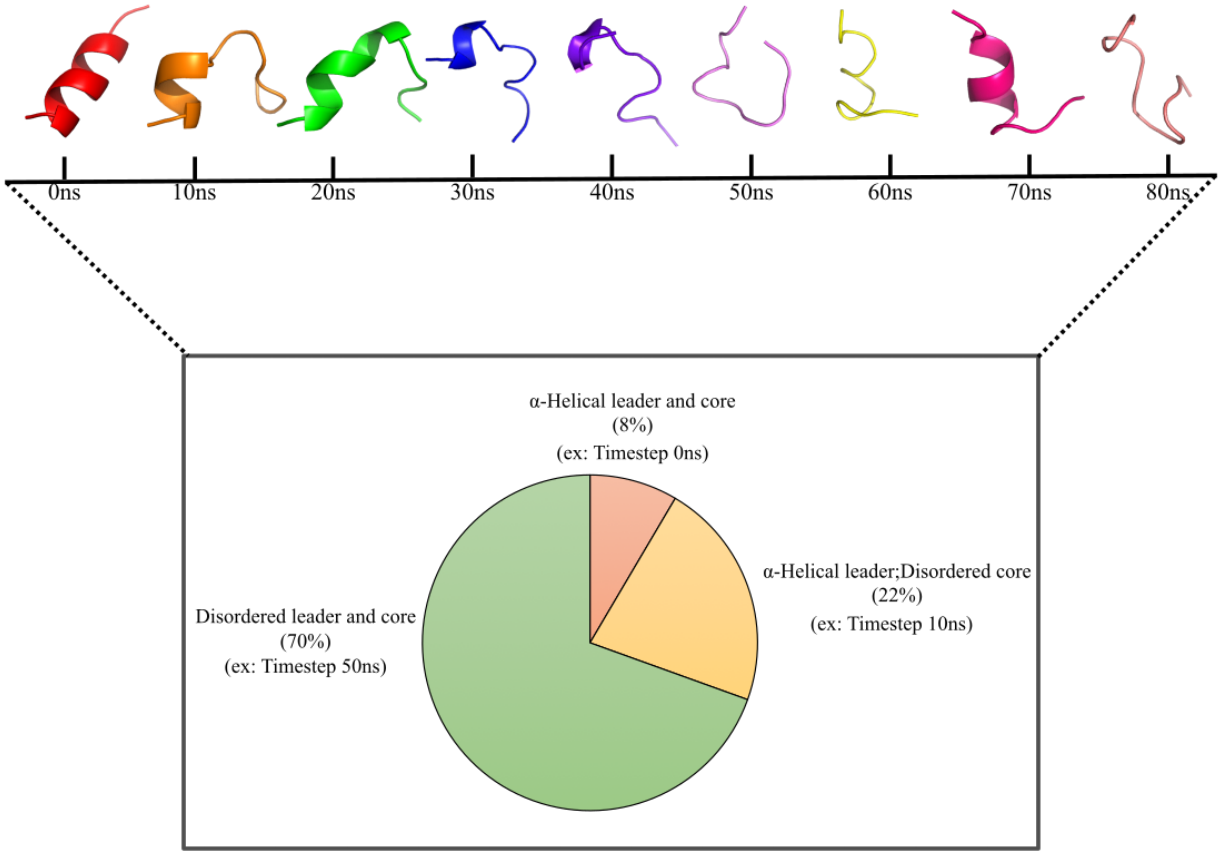



*Figure 4.7. The conformation changes of SuiA over time*

Computational results are further supported by far-ultra-violet-circular dichroism (Far-UV-CD) spectroscopy studies. CD is a form of absorption spectroscopy in which the light source is modulated to rotate left and right circularly. Chiral molecules are biased toward the absorption of one direction of the polarized light.[60] Far-UV-CD has readily been used to elucidate the secondary structure of proteins and peptides. At the far UV range (190-250 nm), α-helices, β-sheets, and disordered/random coil secondary conformations all have characteristic spectra. α-helical structures display a CD spectrum with a positive band at ~190 nm and two negative bands at ~208 nm and ~222 nm. β-sheets exhibit a positive band at ~198 nm and a single negative band ranging

from ~214-218 nm, depending on the type of structure. Random coils structure are characterized by a negative band below 200 nm.[60]

Three samples were prepared for CD in water: **(1)** SuiA, **(2)** RRE, and **(3)** a 1:1 ratio of SuiA:RRE. The resultant spectrum for SuiA displayed a peak characteristic of a disordered loop with a negative absolute minimum at 199.9 nm. The RRE spectrum exhibited peaks attributed to α-helices (minima at 207.6 nm and 222.4 nm) and β-sheets (maximum at 198.8 nm), consistent with the wHTH motif. The spectrum for the RRE in the presence of SuiA, upon subtraction of the RRE signal, indicated that SuiA remained in the disordered conformation (Figure 4.8). Based on these results and the results of the beforementioned computational studies, we conclude that SuiA is an intrinsically disordered peptide prior to interaction with SuiB, and that interaction with the RRE does not induce a significant conformational change. As such, we propose that the change in secondary structure occurs upon interaction with the catalytic barrel.

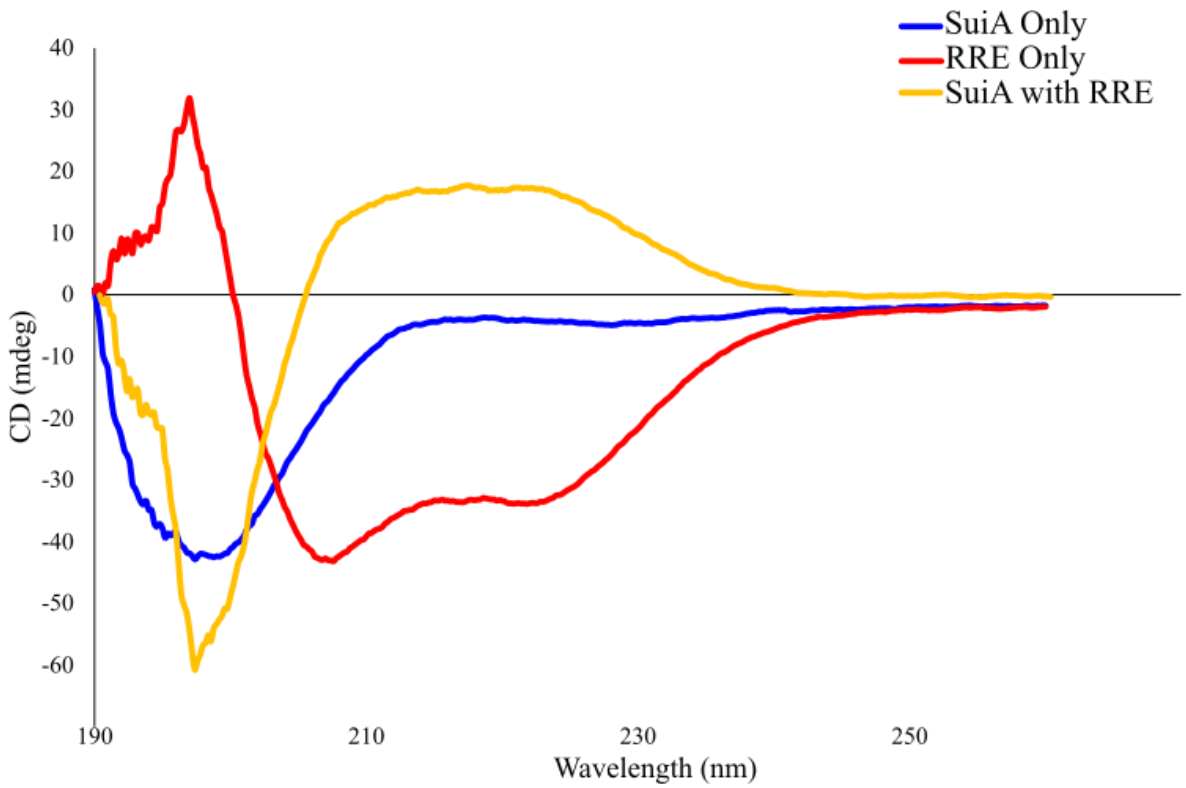



*Figure 4.8. SuiA CD studies. SuiA (blue) depicts a disordered conformation. The RRE region (red) primarily displays β-sheet and α-helical characteristics (198.8nm, 207.6nm, and 222.4). When SuiA is combined with the RRE (gray) disordered characteristics (202.2nm) and slight α-helical characteristics (222.8nm) remain.*

### 4.2.7. Exploration of Disordered SuiA with SuiB

With the knowledge of SuiA's intrinsically disordered nature, docking runs were repeated with SuiA conformations extracted from our 80 ns MD simulation at intervals of 5 ns (Figure 4.7). Under *ab intio* docking conditions, the partially disordered conformation of SuiA from 25 ns and 30ns displayed binding interactions with the RRE. Complex A, which was the docking of SuiA-25 ns and SuiB, displayed interactions between the following residues: SuiB-Lys22:SuiA-Glu(-9), SuiB-Glu29:SuiA-Lys2, SuiB-Glu29:SuiA-Ala1, SuiB-Arg64:SuiA-Glu(-5), SuiB-Arg64:Ser(-3), and SuiB-Arg64:Ala(-2). Glu29, a residue on β-3, did not display interaction with the leader sequence of SuiA. Binding interactions with the bridging domain were also observed: SuiB-

336:SuiA-Lys-(-12), SuiB-Phe337:SuiA-Lys(-12), and SuiB-Tyr345:SuiA-Lys(2) (Figure 4.9A). Complex B, which was the docking of SuiA-30ns and SuiB, displayed greater interactions below 4 Å to the bridging region: SuiB-Asn-332:SuiA-Ser(-4), SuiB-Tyr345:SuiA-Asp4, SuiB-Arg348:SuiA-Val8, SuiB-Tyr345:SuiA-Trp6, and SuiB-Glu329:SuiA-Lys(-12), SuiB-Ser334:SuiA-Glu(-11), SuiB-Ser338:SuiA-Ala1. There was one interaction of the RRE β-3 SuiB-Arg27:SuiA-Glu(3). However, again, there were no interactions with the leader sequence.

Due to the proximity of peptides to the RRE β-sheets, we ran further MD simulations to assess if additional binding interactions form. Simulations were run with a single monomer of SuiA in a cubane water-filled boundary with $Na^+$ to neutralize excess negative charges. However, in both models, the peptide dissociates from the RRE. Complex A t0 depicted a more helical SuiA starting conformation than Complex B, and by 100 ns, the peptide lost order and began losing hydrogen bonding contacts (3-4 Å) with SuiB-RRE β3. While the peptide did gain interactions with the bridging domain, simulation elongation to 200 ns depicted SuiA dissociating from the protein (Figure 4.9).

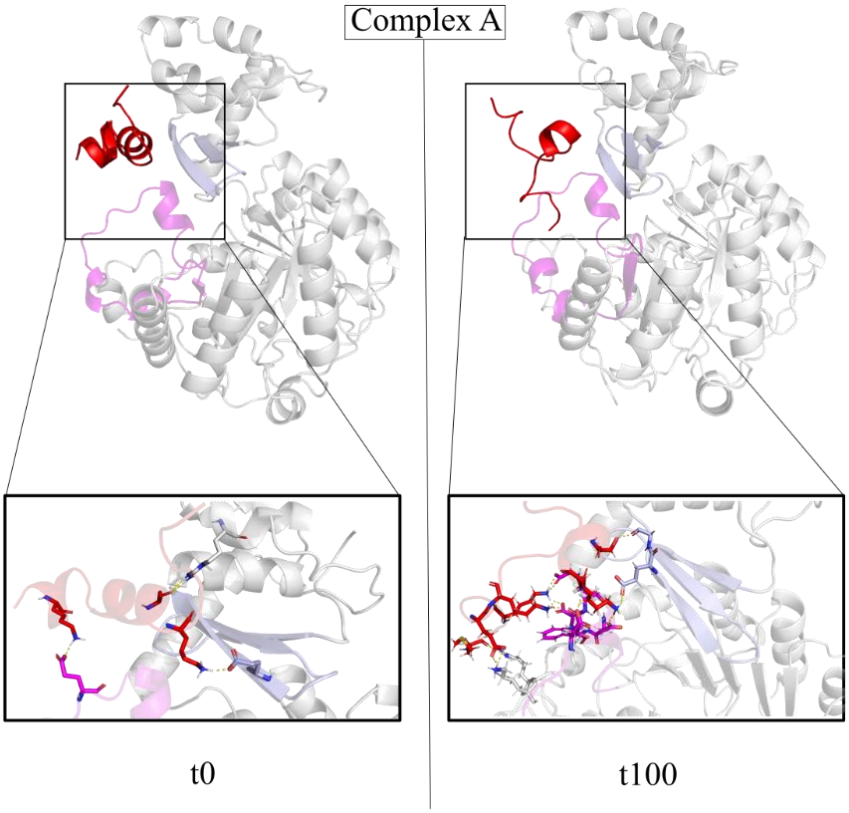



*Figure 4.9. Molecular dynamics simulation of Complex A 100ns*

.

Interestingly, after 300 ns, Complex B begins to move towards the catalytic barrel, gaining additional interactions with the bridging domain, again implicating the bridging domain in peptide binding and recognition (Figure 4.10).

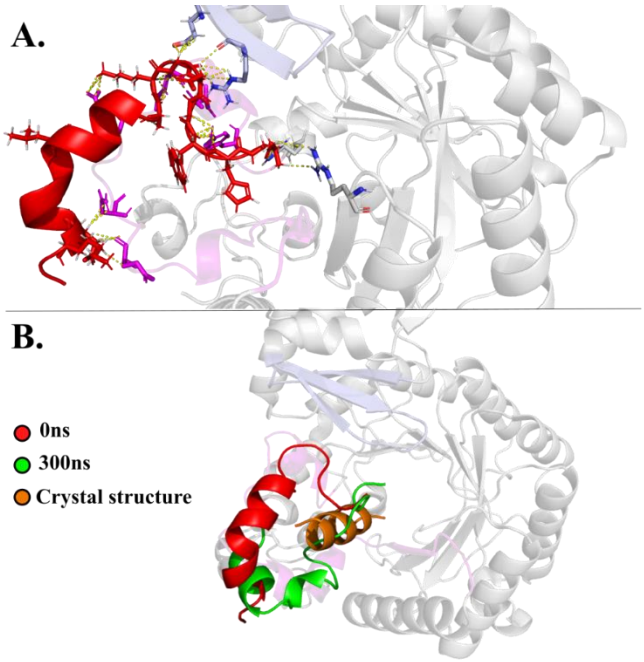



*Figure 4.10. Complex B docking and MD simulations. (A) SuiA binding interactions with RRE and bridging domain. (B) SuiA movement with Complex B as t0 and the complex after 300 ns MD simulation (md300ns) in reference to the crystal structure coordinates.*

### 4.2.8. Hydrophobicity Analysis of SuiB and SuiA

The hydrophobic effect is considered one of the driving factors that encourage protein folding.[61, 62] Because of the major role of hydrophobicity in protein intramolecular interactions, it is also unsurprising of that hydrophobicity plays a major role in intermolecular protein-protein and protein-peptide interactions. Hydrophobicity compatibility between protein and binding partner (i.e. hydrophobic region on protein to hydrophobic region on binding partner) is often an indication of protein-peptide binding sites.[63] This hydrophobicity compatibility has been noted as significant in RRE and lasso peptide leader recognition as "knobs-into-holes" where the hydrophobicity residues (knobs) on the RiPP sterically fit in the hydrophobic pockets on the RRE (holes).[64]

In order to understand how hydrophobic effects, or lack thereof, may contribute to SuiA:SuiB-RRE binding we perform a hydrophobicity analysis of the system. (Figure 4.11). Chimera X was

used to color AAs according to Kyte-Doolittle hydrophobicity from blue (representing most hydrophilic), white, then orange (representing most hydrophobic) (Figure 4.11A).[65]

SuiB displays a hydrophobic pocket in the bridging region (bridging pocket) composed of Phe324, Tyr327, Leu330, Ile335, Phe337, Tyr345, Leu411, and Ile414 (Figure 4.11B). This pocket is compatible with two hydrophobic knobs on the helical SuiA leader strand composed of residues Leu(-10), Val(-7), and Leu(-6) (Figure 4.11B,C). SuiB also displays a smaller hydrophobic pocket in the catalytic barrel with residues in the rSAM domain (rSAM pocket) which may be compatibility with the Met(-1)(Figure 4.12A). This hydrophobic bridging pocket could be involved in peptide recognition, as a similar but less pronounced pocket exists in CteB, our comparative rSAM, around residue Phe325 (Figure 4.12B).

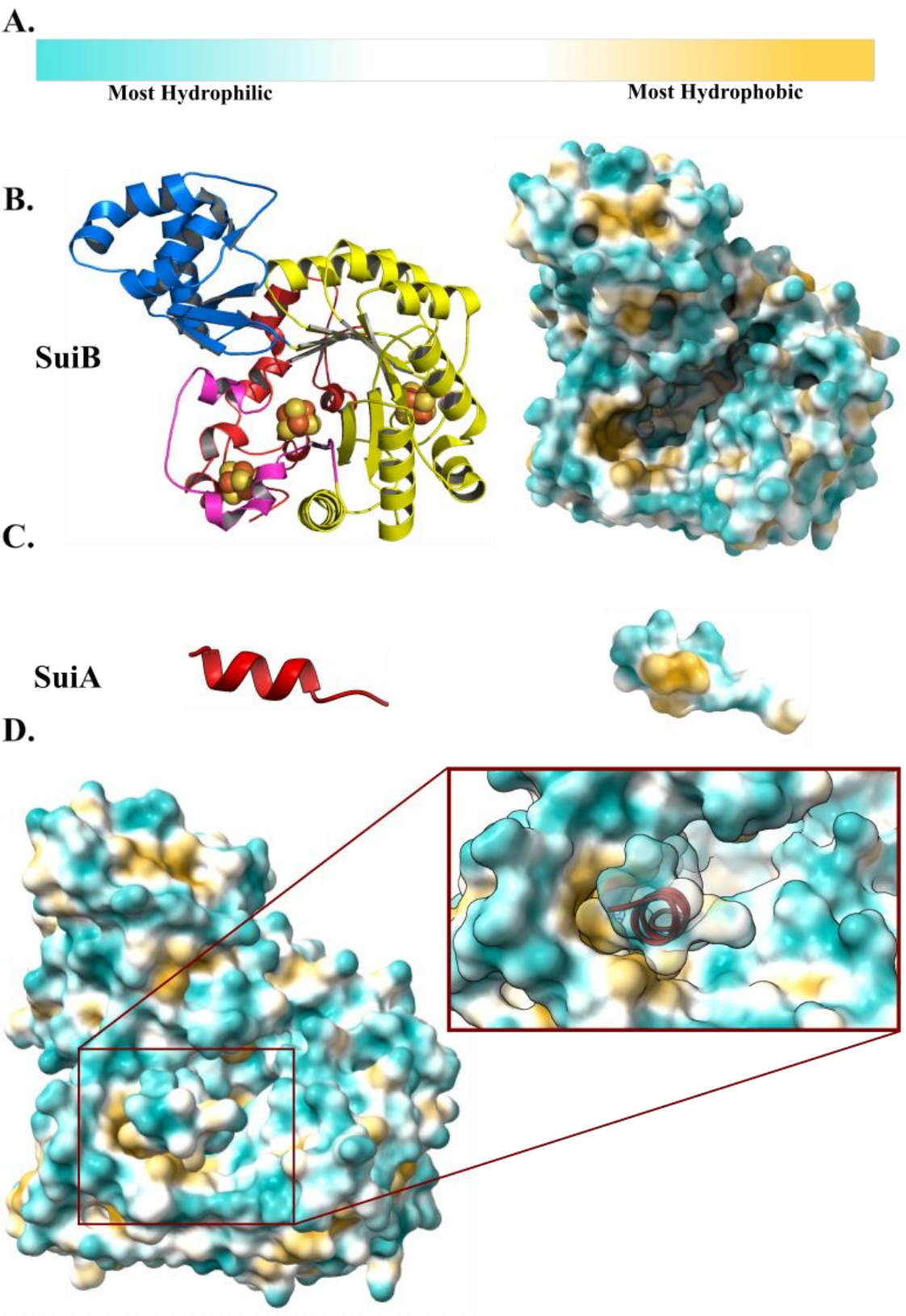



*Figure 4.11. Hydrophobicity Analysis of SuiB and SuiA*

**A.**

```
CLUSTAL O(1.2.4) multiple sequence alignment

SuiB(A0A0Z8EWX1)    MRTISEDILFRLEKFGGILINKTNFERIELDETEAFFLYLVQNHGIEIATSFFKKEIEMG    60
CteB                -----MAMIHKFSMMGTNIVVDVNSGAVHV----------VDDISFDI--LDYYKNFTAG    43
                          ::.::. :*  :: ..*   :.:          *:: .::*    : *::  *

SuiB(A0A0Z8EWX1)    KLERALSLNIYSDNNIEDSL--------------NNPYETLQNARKHVAKLKKHNILSFP    106
CteB                EIKNKLAH-KYNADEIDEALREIESLEAEGLLFSEDPYKEYVS------SMD---RKSVV    93
                    :::. *:   *. ::*:::*              ::**:   .      .:.     *.

SuiB(A0A0Z8EWX1)    LELVIYPSMYCDLKCGFCFLANREDRNAKPA-------KDWERILRQAKDNGVLSVSILG    159
CteB                KALCLHISHDCNLRCKYCFASTGNFGGQRNMMSLEVGKKAIDFLISESGNRKNLEIDFFG    153
                      * :: *  *:*:* :** :. :  . :          *  : :: :: :.  *.:.::*

SuiB(A0A0Z8EWX1)    GEPTRYFDIDNLLIACEELKIK-------TTITTNAQLIKKSTVEILAKSKYITPVLSLQ    212
CteB                GEPMMNFDVVKGIIEYARQKEKEHNKNFRFTLTTNGLLLNDENIKYINEN-MQNIVLSID    212
                    ***   **: : :*   . * *        *:***. *::...:: : :.   . ***::

SuiB(A0A0Z8EWX1)    ---TLDSKLNFELMGVRPDR-QIKLAKYFNEVGKKCRINAVYTKQSYEQIIELVDFCIEN    268
CteB                GRKEVNDRMRIRIDGSGCYDDILPKFKYVAE---SRNQDNYYVRGTFTR--ENMDFS--N    265
                        ::.::.:.: *       :   **. *   . . :  *.: :: :  * :**.  *

SuiB(A0A0Z8EWX1)    KIDRFSVANYSEVTG---YTKIKKKYDLA--DLRRLN----EYVTDYITQREANLNFATE    319
CteB                DVLHLADEGFRQISVEPVVAAKDSGYDLREEDLPRLFEEYEKLAYEYVKRRKEGN-----    320
                    .: :::  .: :::     :  .. ***   ** **     : . :*:.:*: .

SuiB(A0A0Z8EWX1)    GCHLFTAYPELINNSIEFSEFDEMYYGCRAKYTKMEIMSNGDILPCIAFLGVNQTKQNAF    379
CteB                ---WFNFFHFMI-DLTQGPCIVKRLTGCGSGHEYLAVTPEGDIYPCHQFVGNEKFKMGNV    376
                        *. :  :* :  :   : :  ** : :  : :  :*** **  *:* :: * . .

SuiB(A0A0Z8EWX1)    EKDLLD-----VWYDDPLYGGIRSFRTKNSKCLSCQLKICEGGCYVNLIKEKSP-EYFR    433
CteB                KEGVLNRDIQNYFKNSNV--------YTKKECDSCWAKFYCSGGCAANSYNFHKDINTVY    428
                    ::.:*:      : :. :         .:.:* **     *.*** .*  : :.  : .

SuiB(A0A0Z8EWX1)    DSVCQL----------------    439
CteB                KVGCELEKKRVECALWIKAQEM    450
                    .  *:*
```

**B.**

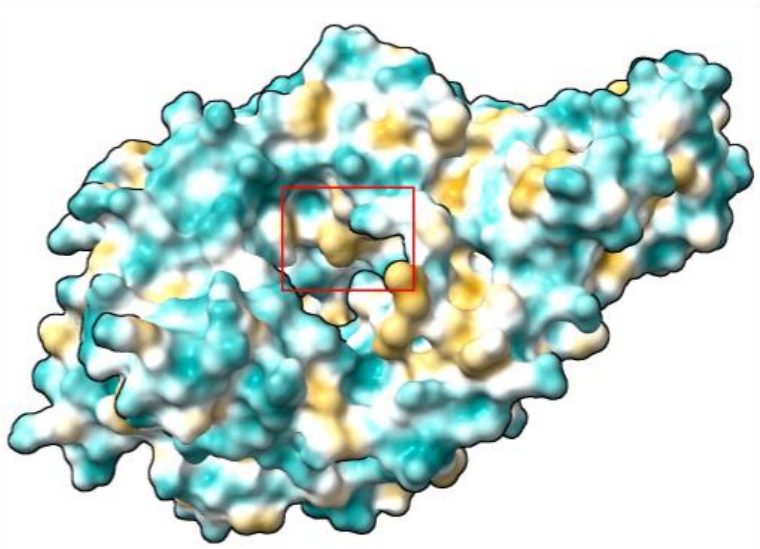


*Figure 4.12. A. Alignment of CteB and SuiB RRE Region (blue), rSAM Domain (yellow), SPASM Domain (red), and Bridging Region (pink). Residues in the SuiB hydrophobic bridging and rSAM pockets B. CteB hydrophobicity analysis, red box depicts hydrophobic pocket in briding domain.*

Interestingly, the SuiB-RRE region lacks a hydrophobic region or pocket at β3 spatially compatible with the hydrophobic nodes on SuiA in helical conformation, supporting our hypothesis that the peptide cannot interact with the RRE in the ordered conformation (Figure 4.13B). CteB-RRE has two compatible hydrophobic regions with its peptide, CteA, at β3 residues and between β3 and α3 helixes (Figure 4.13C). SuiB does display a hydrophobic pocket between α2 and α3 helixes and β1 and β2 (residues Ileu8 and Leu19); however, unlike CteB-RRE there is no hydrophobic pocket at β1 which could result in weaker binding to the RRE or an alternative binding strategy.

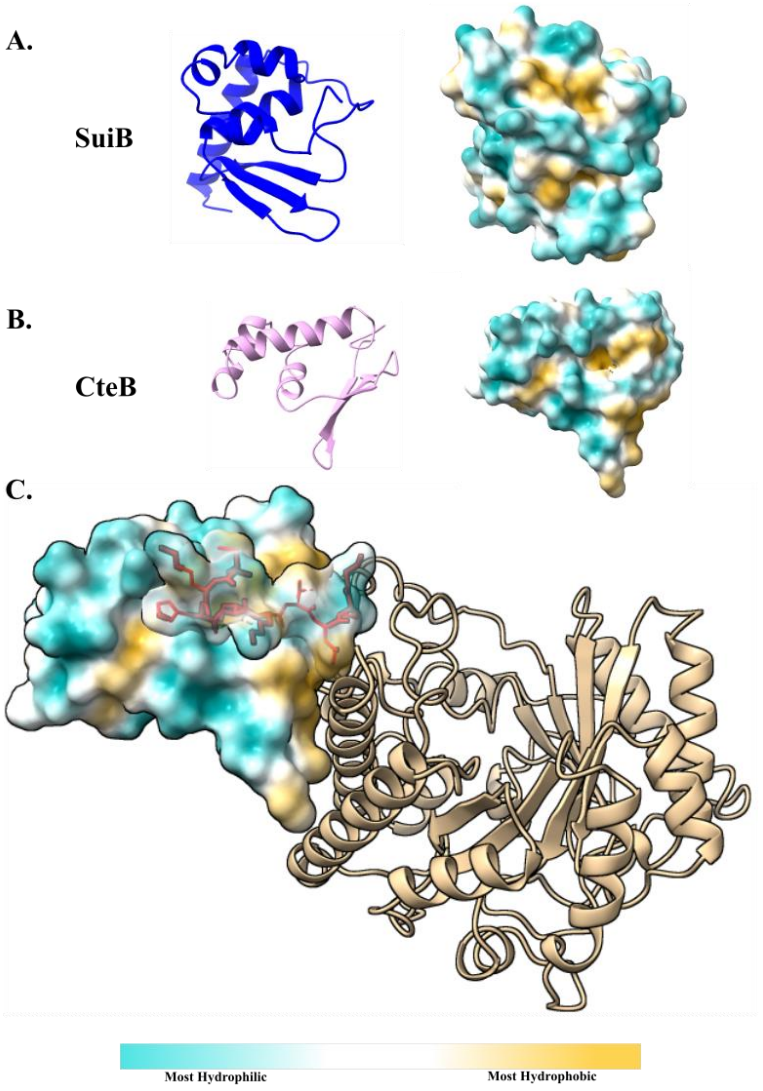



*Figure 4.13. A. Hydrophobicity analysis of CteB. B. Hydrophobicity analysis of SuiB. C. Hydrophobicity analysis of CteB binding to CteA by β3 and the gap between β3 and α3*

### 4.2.9. Comparison to YydG from *Bacillus subtilis*

The lack of data to support interaction between the RRE and SuiA prompted our exploration of alternative mechanisms for RiPP recognition. This lead us to the exploration of YydG from *Bacillus subtilis* (UniProt accession code: A0A6M3ZHV5) is a RiPP modifying rSAM that lacks both an embedded RRE-like domain and a discrete RRE in the *yydFGHIJ* BGC (Figure 4.11). Despite this absence of RRE, YydG is able to catalyzes the epimerization of non-polar residues (valine, leucine, and isoleucine) on its precursor peptide, YydF from their L- into their D-counterparts.[66] YydG is activity raises questions as to how the tailoring enzyme recognizes YydF. By investigating structural similarities between YydG and SuiB, it may be possible to elucidate novel alternative recognition domains.

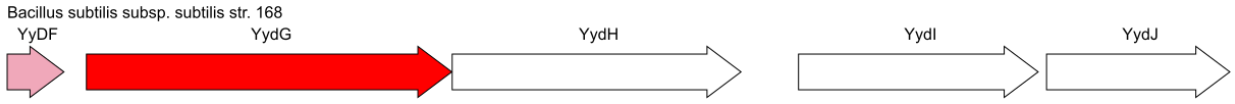



*Figure 4.14. The yydFGHIJ BGC. BGC lacks an RRE-like domain*

It has been previously reported that both SuiB and YydG (sequence identity: 13.2%, sequence similarity 22.1%) contain SPASM domains.[26, 67] However, sequence alignment identifies that YydG has a bridging domain similar to SuiB's (Figure 4.15). Within that bridging region are three of the four crystographically supported SuiA binding resides: Gly320, Phe324, and Tyr344. This similarity may indicate hydrophobic contacts in the bridging domain may be significant in RiPP recognition. However further data is needed to support such a hypothesis.

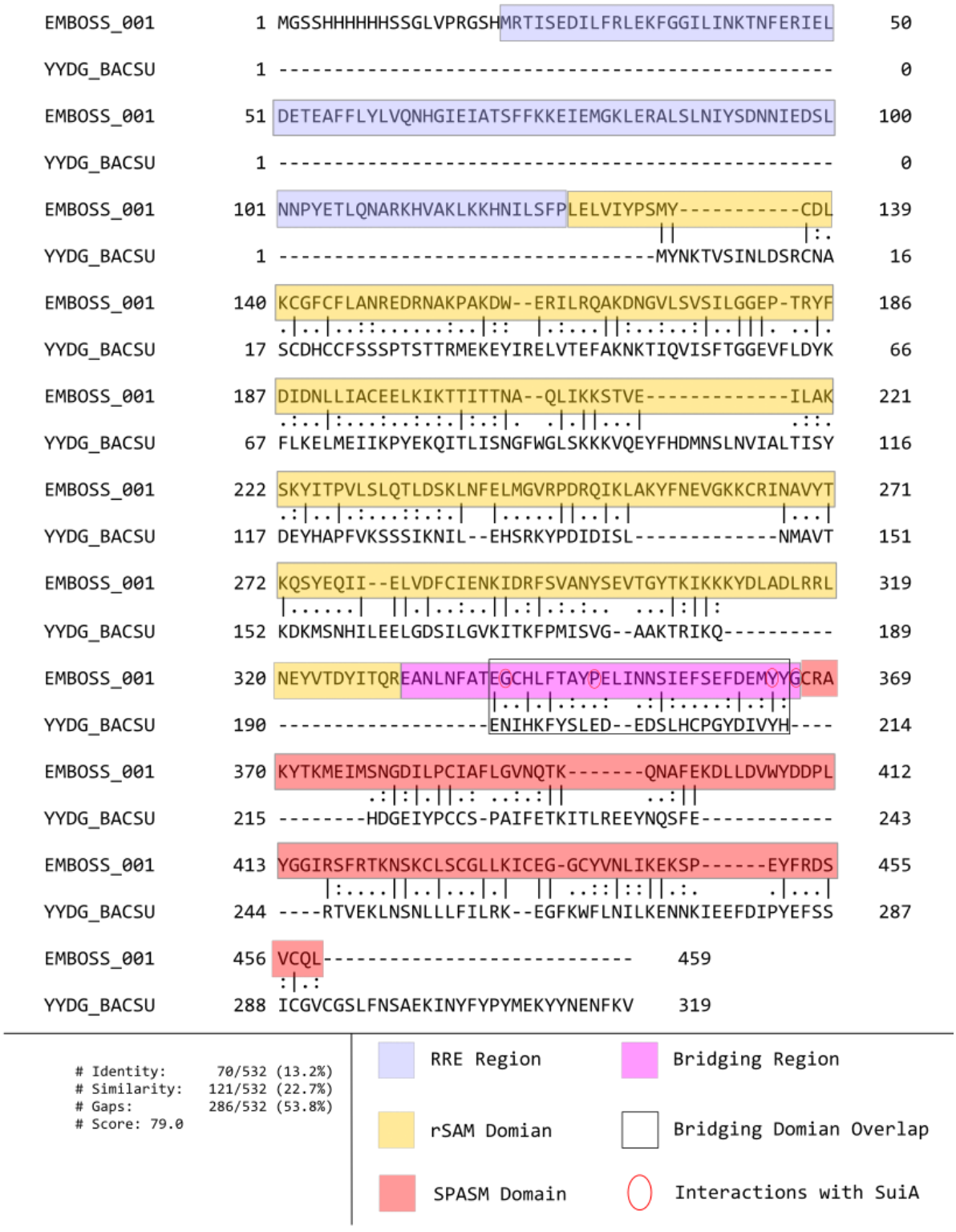



*Figure 4.15. Sequence alignment of YydG and SuiB(EMBOSS_001). RRE Region (blue), rSAM Domain (yellow), SPASM Domain (red), and Bridging Region (pink). Residues in the SuiB Bridging Domain that align in YydG are boxed, and residues that interact with SuiA in the bridging region are circled in red on EMBOSS_001 line 320.*

There are no crystal structures of YydF or YydG. Thus, to visualize if the YydG: YydF complex mimics SuiB: SuiA bridging domain interactions, JPred and AlphaFold models of YydF and YydG were generated (Figure 4.16).[57, 68] YydF is double the size of SuiA and displays 22.4 % sequence identity and 28.6% sequence similarity. Similar to the leader sequence of SuiA, the N-terminal of YydF is predicted to be helical, however, the core region is predicted to form a β-hairpin (Figure 4.16A). After *ab initio* docking of the YydF and YydG models (Figure 4.16B), YydF displays significant hydrogen bond interactions with the bridging domain at YydF-Lys37:YydG-Asp202 (1.6 Å), YydF-Arg42:YydG-Asp210 (1.6 Å), and YydF-Asn41:YydG-Asp202 (2.2 Å). YydG-Asp210 is identical and YydG-Asp202 is has high similarity to the bridging domain of SuiB, identifying another two bridging domain residues potentially involved in RiPP recognition. YydF forms interactions with the rSAM domain partial TIM barrel between YydF-Leu45:YydG-Lys174 (2.4 Å) and YydF-Asp31:YydG-Asn157 (2.3 Å). The docking results of YydG with YydF hints at a novel binding mode for RiPP recognition, where the RiPP potentially completes the partial TIM barrel of the rSAM (Figure 4.16D).

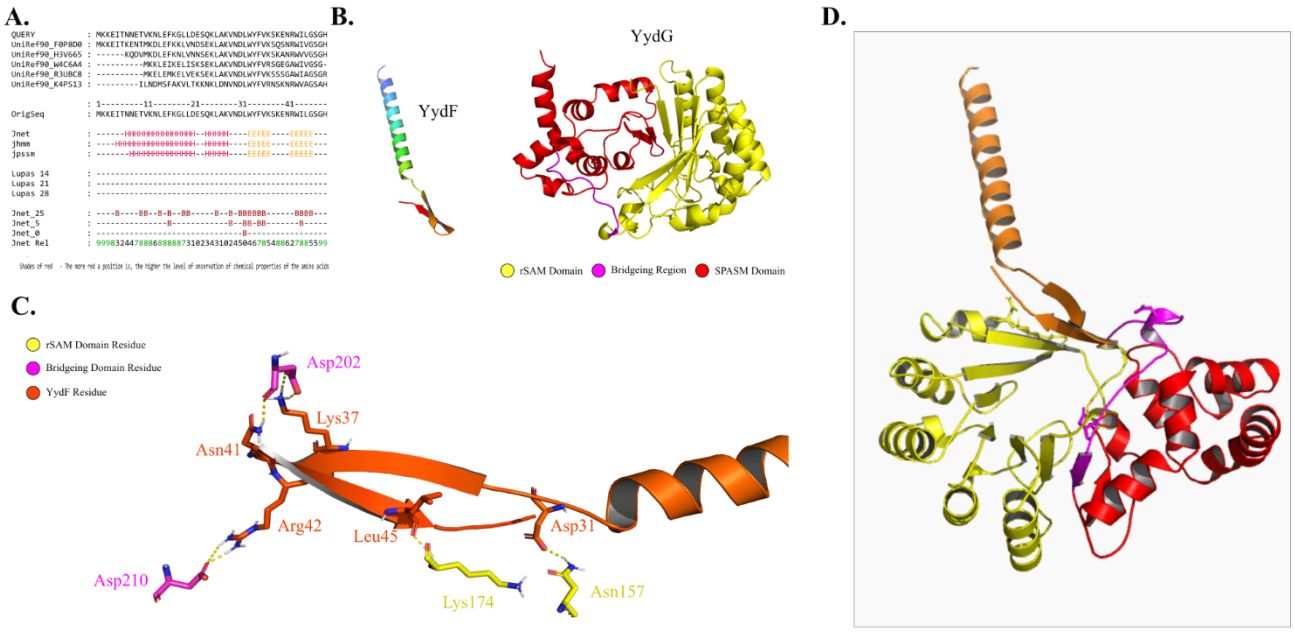



*Figure 4.16. Structural prediction of YydF and YydG. (A) Jpred4 secondary structure prediction of YydF. (B) Alphafold model of YydF and YydG. (C) Docking of YydG with YydF(orange) showing hydrogen bonding with residues in the bridging domain (pink) and rSAM domain (yellow) (D) Docking results of YydG and YydF depicting RiPP completion of the rSAM partial TIM barrel.*

### 4.3.Conclusion

RiPPs are a class of structurally and functionally diverse natural products that are of great interest in developing novel antimicrobials. In Chapter 4, we explored the recognition mechanism of RiPPs by tailoring enzymes within their BGCs. It is widely accepted that RiPP recognition is governed by the RRE-like domain located discretely or embedded with the tailoring enzyme. However, this chapter has evidence that an additional protein domain may significantly impact RiPP recognition by rSAM enzymes: the bridging domain. Through *in silico* protein binding and dynamics studies, we explore the SuiB-RRE-dependent recognition of SuiA. Fluorescence assays display detectable binding interactions of SuiB-RRE to SuiA in solution ($K_D$ = 0.82 ± 0.3 μM). *Ab initio* docking of the SuiA leader sequence in a helical conformation, as represented in PDB: 5V1T, yielded no

binding interactions with the RRE; however, it reproduced crystallographic binding interactions with the bridging region.

MD simulations of SuiB with and without SuiA implicate the role of dynamic movement in SuiA binding interactions. MD simulations were able to reproduce crystallographically observed dynamics giving us confidence in our procedure. These parameters were used to classify conformational changes of SuiA. Over 80 ns, SuiA was in a random coil conformation 70% of the simulation leading us to classify SuiA as an intrinsically disordered peptide. These results were further confirmed by CD analysis. Intrinsically disordered peptides lack a structured secondary structure until interaction with a binding partner.[69] As the crystal structure of SuiB with SuiA represents SuiA in a helical conformation, we conclude that SuiB orders SuiA upon binding to the catalytic barrel.

*Ab initio* docking of the partially disordered constructs of SuiA resulted in limited interactions with the RRE supported by significant bridging domain contacts within 2.0 Å. During MD simulations where SuiA does not dissociate within the first 200 ns, the peptide gains additional interactions with the bridging domain. After 300 ns, the peptide appears angels into the catalytic barrel in a similar coordinates as bound SuiA. The studies in Ch. 4 demonstrate that while SuiA interacts with the RRE, the bridging domain is more significant for positioning the RRE for catalysis. We hypothesize that recognition in SuiA is accomplished by weak interaction with the RRE that allows for stronger interactions and guidance into the catalytic barrel by the bridging domain, where the peptide gains order due to hydrophobic contacts in the barrel.

The role of the bridging domain In peptide binding is further supported by *ab intio* docking studies of YydG with YydF. The RRE absent tailoring enzyme was able to recognize the peptide via

hydrogen bonding interactions at the bridging domain. YydF To further support this conclusion, YydG has been transformed and solubility expressed. The crystal structure of YydG coupled to YydF is being obtained.

Additional studies are needed to verify the extent of RRE recognition in SuiB. While a different disordered conformation of SuiA may yield better RRE binding, manual random structural sampling is time-consuming and computationally expensive. As an alternative route, we have begun growing protein crystals of SuiB-RRE with SuiA protein to gain a physiologically relevant starting conformation of SuiA.

## 4.4.Experimental

### 4.4.1. SuiB:RRE-SuiA Fractional Saturation Assay

A fluorescently labeled SuiA derivative (SuiA-Fl) was generated with 6-FAM ligated to the N-terminus of SuiA. Aerobically, SuiA-Fl was dissolved and diluted in 50 mM HEPES buffer, pH 7.6, and serially diluted in a black 96-well plate to a final volume of 100 µL and final concentration of 2 µM. The isolated SuiB:RRE was diluted to 10 µM and added to each well in increasing concentrations from 0 µM to 5 µM. This experiment was monitored from 510-600 nm and run in duplicate.

### 4.4.2. *suiA* genomic analysis

Genomic analysis was conducted as previously described in Section 3.4.12 (Genomic Neighborhood Analysis of BurkOYE) utilizing the Protein IDs of the tailoring enzyme detailed in Table 4.1.

### 4.4.3. HADDOCK 2.2 *ab initio* Docking

Structures for SuiB: SuiA docking were extracted from PDB: 5V1T and prepared for docking using COOTs rotamer validation tool, while alanine stubs were manually corrected using the Simple Mutate function.[70] The protein, protein fragment, and peptide structure coordinates were extracted into individual files using Pymol. The unbound structures, the docking complexes of **(1)** SuiBwt (residues 1-439), **(2)** RRE (residues 1-101), and **(3)** Catalytic-barrel (residues 107-439), each with the SuiA construct, were generated. Ten thousand models for each system were generated during the rigid-body docking (it0) step. The leading 400 models (i.e., the 400 most negative HADDOCK scores) were further refined via the flexible interface (it1) stage and a water step. The top 400 structures from the water step are reduced to the top 200 structures and clustered within 2.0 Å by backbone RMSD. Structures that most resembled physicologically relevant binding were further analyzed by MD and hydrogen bond analysis.

The structures of CteB and the leader stand of CteA were extracted from PDB: 5WGG. Docking runs were performed under the same parameters as SuiB: SuiA docking.

### 4.4.4. MD with GROMACS

To achieve the most physiologically relevant SuiB system, parameterization of the metal cluster was manually supplemented into the AMBER99SB-ILDN protein, nucleic AMBER94 (Lindorff-Larsen et al., Proteins 78, 1950-58, 2010) forcefield in GROMACS 2020.1. Amber compatible charge parameterization originated from quantum mechanical calculations described in Banci 1992[71] and further optimized in Vesper 2012.[72] In both studies, the cysteinyl coordination of the [4Fe-4S] cluster was accounted for by installing a Cys-Fe bond represented by a metallic cysteine residue (CYF). As the SuiB model used in this system is post SAM cleavage, the literature charges

were modified so the sum of the cysteine and rSAM cluster charge results in the net charge of +2 (Figure 4.1D). Under these conditions, all three metal clusters remained stable throughout all MD simulations. The final [4Fe-4S] parameters used in this study are listed in Table 4.3. FeS-Cluster Parameter Modifications to amber99sb-ildn.ff Post successful parameterization the SuiB models were further improved by incorporating the ligand topologies of SAM built in xtleap.

Simulations for each system were carried out under periodic boundary conditions (PBS) using GROMACS version 2020.1. All the systems were solvated with the default solvent model Simple Point Charge water (SPC) (with coordinates from $GMXLIB/spc216.gro) in a cubic box. Excess charges were neutralized with $Na^+$ and $Cl^-$ ions. Prior to energy minimization (EM), volume and temperature (NVT), and pressure and temperature (NPT) equilibrium, the complex was centered and placed at least 1.0 nm from the box edge. The simulation boxes underwent steepest descent EM until maximum force ($F_{max}$) acting on each atom is below 1000 kJ/mol/nm. The Canonical ensemble (aka NVT equilibration) was performed on the EM system for 100 ps at a time-step of 2 fs using the leap-frog integrator and modified Berendsen thermostat for equilibrating the system to a temperature of 300 K. Initial velocities were assigned from a Maxwellian distribution. The NVT equilibration was followed by a NPT equilibration, for an additional 100 ps with 2 fs time-step to stabilize the system's pressure at 1 bar using Parrinello-Rahman barostats with the compressibility of $4.5 \times 10^{-5}$ $bar^{-1}$. The isotropic position scaling protocol was used to control the pressure. Long-range electrostatic interactions were modeled using the Particle Mesh Ewald (PME) model, and the length of all covalent bonds constrained with the linear constraint solver (LINCS) algorithm. A 10 Å cut-off was used for short-range interactions. The molecular dynamics simulation was performed for 5, 80, or 100 ns at a time-step of 2 fs. Simulations longer than 100

ns utilized checkpoint files continue runs from previous 100 ns and concatenate post total time step.

Generalizable commands can be viewed under Table 4.4.

### 4.4.5. Structure and Simulation Analysis

The RMSD analysis was performed with by positional comparison of the protein α-C backbone of the energy minimized structure over time using the gmx_rms command in GROMACS and plotting time (ns) vs RMSD (Å). The RMSF analysis of the final step was performed with of the protein α-C backbone of the energy minimized structure using the gmx_rmsf command in GROMACS and plotting AA residue vs RMSD (Å).

Hydrogen bond analysis was performed in Chimera using the h-bond analysis tool to detect interactions less than 4 Å. The hydrophobicity analysis was performed in ChimeraX with pdb 5v1t.

### 4.4.6. Hydrophobicity Analysis

The “hydrophobicity surface” preset in ChimeraX was used to display the molecular surface colored by AA hydrophobicity. SuiB, SuiA, and fragments of SuiB were extracted from 5V1T in pymol.

### 4.4.7. SuiA Conformation Change Analysis

1ns time points of SuiA were extracted from an 80ns MD simulations performed as describe above. Each time point was colormetrically categorized as α-helical (purple) or disordered (white) or a combination of both using Visual Molecular Dynamics (VMD).

### 4.4.8. Circular Dichroism

CD experiments were performed on a J-1500 Circular Dichroism Spectrophotometer with a with a PM-539 detector. Continuous scans of each sample were performed from 260 - 190 nm at 19.99 °C. Samples were prepared at 3mg/mL of full length SuiA or purified RRE.

## 4.5.Supporting Information

Table 4.1. Role in RRE:Tailoring enzyme pairs and Protein IDs

| **Enyzme** | **Protein ID** | **RRE-Containing** | **Tailoring Enzyme** |
|---|---|---|---|
| BalhC | ABK83869.1 | X | |
| BalhD | ABK83870.1 | | X |
| LarC | BAL72548.1 | X | |
| LarD | BAL72549.1 | | X |
| PqqD | CAA41582.1 | X | |
| PqqE | CAA41583.1 | | X |
| SuiB | AWL26605.1 | X | |
| YydG | QJP90684.1 | | X |

Table 4.2. Conformation changes in SuiA

| **Time(ns)** | **Step** | **Majority AH** | **50:50 AH:LP** | **Majority LP** |
|---|---|---|---|---|
| 0 | 0 | x | | |
| 0.9999 | 100 | x | | |
| 1.9998 | 200 | x | | |
| 2.9997 | 300 | x | | |
| 3.9996 | 400 | x | | |
| 4.9995 | 500 | x | | |
| 5.9994 | 600 | | x | |
| 6.9993 | 700 | | x | |
| 7.9992 | 800 | | x | |
| 8.9991 | 900 | | | x |
| 9.999 | 1000 | | | x |

| | | | | |
|---|---|---|---|---|
| 10.9989 | 1100 | | | x |
| 11.9988 | 1200 | | x | |
| 12.9987 | 1300 | | x | |
| 13.9986 | 1400 | | | x |
| 14.9985 | 1500 | x | | |
| 15.9984 | 1600 | | x | |
| 16.9983 | 1700 | | x | |
| 17.9982 | 1800 | | x | |
| 18.9981 | 1900 | | | x |
| 19.998 | 2000 | | | x |
| 20.9979 | 2100 | | | x |
| 21.9978 | 2200 | | | x |
| 22.9977 | 2300 | | | x |
| 23.9976 | 2400 | | | x |
| 24.9975 | 2500 | | | x |

| | | | | |
|---|---|---|---|---|
| 25.9974 | 2600 | | | x |
| 26.9973 | 2700 | | | x |
| 27.9972 | 2800 | | | x |
| 28.9971 | 2900 | | | x |
| 29.997 | 3000 | | | x |
| 30.9969 | 3100 | | | x |
| 31.9968 | 3200 | | | x |
| 32.9967 | 3300 | | | x |
| 33.9966 | 3400 | | x | |
| 34.9965 | 3500 | | x | |
| 35.9964 | 3600 | | x | |
| 36.9963 | 3700 | | | x |
| 37.9962 | 3800 | | | x |
| 38.9961 | 3900 | | | x |
| 39.996 | 4000 | | | x |

| | | | | |
|---|---|---|---|---|
| 40.9959 | 4100 | | | x |
| 41.9958 | 4200 | | | x |
| 42.9957 | 4300 | | | x |
| 43.9956 | 4400 | | | x |
| 44.9955 | 4500 | | | x |
| 45.9954 | 4600 | | | x |
| 46.9953 | 4700 | | | x |
| 47.9952 | 4800 | | | x |
| 48.9951 | 4900 | | | x |
| 49.995 | 5000 | | | x |
| 50.9949 | 5100 | | | x |
| 51.9948 | 5200 | | | x |
| 52.9947 | 5300 | | | x |
| 53.9946 | 5400 | | | x |
| 54.9945 | 5500 | | | x |

| | | | | |
|---|---|---|---|---|
| 55.9944 | 5600 | | | x |
| 56.9943 | 5700 | | | x |
| 57.9942 | 5800 | | | x |
| 58.9941 | 5900 | | x | |
| 59.994 | 6000 | | | x |
| 60.9939 | 6100 | | x | |
| 61.9938 | 6200 | | | x |
| 62.9937 | 6300 | | x | |
| 63.9936 | 6400 | | x | |
| 64.9935 | 6500 | | x | |
| 65.9934 | 6600 | | x | |
| 66.9933 | 6700 | | x | |
| 67.9932 | 6800 | | | x |
| 68.9931 | 6900 | | | x |
| 69.993 | 7000 | | | x |

| | | | | |
|---|---|---|---|---|
| 70.9929 | 7100 | | | x |
| 71.9928 | 7200 | | | x |
| 72.9927 | 7300 | | | x |
| 73.9926 | 7400 | | | x |
| 74.9925 | 7500 | | | x |
| 75.9924 | 7600 | | | x |
| 76.9923 | 7700 | | | x |
| 77.9922 | 7800 | | | x |
| 78.9921 | 7900 | | | x |
| 79.992 | 8000 | | | x |
| 80.9919 | 8001 | | | x |

AH = Helical leader and core sequence

LP= Fully disordered leader and core sequence

50:50 AH:LP = Leader sequence α-Helix, while core sequence was disordered

Table 4.3. FeS-Cluster Parameter Modifications to amber99sb-ildn.ff

| **aminoacids.rtp** | **aminoacids.hdb** | **ffbonded.itp** |
|---|---|---|
| [ SF4 ] | CYF 3 | [ bondtypes ] |
| [ atoms ] | 1 1 H N -C CA | FS SH 1 0.226 120000.0 |
| FE1 FS 1.465 1 | 1 5 HA CA N CB C | [ angletypes ] |
| FE2 FS 1.465 2 | 2 6 HB CB CA SG | SH FS S 1 180.000 0.000 |
| FE3 FS 1.465 3 | | CT SH FS 1 180.000 0.000 |
| FE4 FS 1.465 4 | | [ dihedraltypes ] |
| S1 S -0.9817 5 | | CT SH FS S 9 180.0 0.00000 1 |
| S2 S -0.9817 6 | | FS S FS X 9 180.0 0.00000 1 |
| S3 S -0.9817 7 | | |
| S4 S -0.9817 8 | | **atomtypes.atp** |
| [ bonds ] | | FS 55.00000 |
| FE1 S2 0.2298 100000 | | |
| FE1 S3 0.2316 100000 | | |

FE1 S4 0.2285 100000

FE2 S1 0.2323 100000

FE2 S3 0.2247 100000

FE2 S4 0.2183 100000

FE3 S1 0.2306 100000

FE3 S2 0.2228 100000

FE3 S4 0.2230 100000

FE4 S1 0.2138 100000

FE4 S2 0.2143 100000

FE4 S3 0.2103 100000

[ angles ]

S2 FE1 S3 92.970 460.2

S2 FE1 S4 103.009 460.2

S2 FE3 S1 100.043 460.2

S2 FE3 S4 107.137 460.2

S2 FE4 S3 104.018 460.2

S2 FE4 S1 108.531 460.2

S3 FE1 S4 102.404 460.2

S3 FE2 S1 95.786 460.2

S3 FE2 S4 108.102 460.2

S3 FE4 S1 106.155 460.2

S4 FE2 S1 101.824 460.2

S4 FE3 S1 100.939 460.2

FE2 S1 FE3 73.974 460.2

FE2 S1 FE4 75.172 460.2

FE2 S3 FE1 73.534 460.2

FE2 S3 FE4 77.488 460.2

FE2 S4 FE3 78.249 460.2

FE2 S4 FE1 75.356 460.2

FE3 S1 FE4 73.083 460.2

FE3 S2 FE1 74.155 460.2

FE3 S2 FE4 74.565 460.2

FE3 S4 FE1 74.399 460.2

FE4 S2 FE1 78.152 460.2

FE4 S3 FE1 78.532 460.2

**[ CYF]**

[ atoms ]

N N -0.41570 1

H H 0.27190 2

CA CT 0.02130 3

HA H1 0.11240 4

CB CT -0.33740 5

HB1 H1 0.00405 6

HB2 H1 0.00405 7

SG SH -0.49000 8

C C 0.59730 9

O O -0.56790 10

[ bonds ]

| | | |
|---|---|---|
| N H | | |
| N CA | | |
| CA HA | | |
| CA CB | | |
| CA C | | |
| CB HB1 | | |
| CB HB2 | | |
| CB SG | | |
| C O | | |
| -C N | | |
| [ impropers ] | | |
| -C CA N H | | |
| CA +N C O | | |

Table 4.4. Commands for MD simulations

1. gmx editconf -f filename.gro -o filenamebox.gro -c -d 1.0 -bt cubic
2. gmx solvate -cp Filenamebox.gro -cs spc216.gro -o Filenamesolv.gro -p topol.top
3. gmx grompp -f ions.mdp -c Filenamesolv.gro -p topol.top -o ions.tpr

4. gmx genion -s ions.tpr -o Filenameions.gro -p topol.top -pname NA -nname CL -neutral
5. gmx grompp -f minim.mdp -c Filenameions.gro -p topol.top -o em.tpr
6. gmx mdrun -v -deffnm em
7. gmx grompp -f nvt.mdp -c em.gro -r em.gro -p topol.top -o nvt.tpr
8. gmx mdrun -deffnm nvt
9. gmx grompp -f npt.mdp -c nvt.gro -r nvt.gro -t nvt.cpt -p topol.top -o npt.tpr
10. gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o md80ns.tpr
11. gmx mdrun -deffnm md80ns

*Table 4.5. FASTA Sequences for SuiB, CteB, and YydG*

>SuiB(A0A0Z8EWX1)

MRTISEDILFRLEKFGGILINKTNFERIELDETEAFFLYLVQNHGIEIATSFFKKEIEMGKL
ERALSLNIYSDNNIEDSLNNPYETLQNARKHVAKLKKHNILSFPLELVIYPSMYCDLKC
GFCFLANREDRNAKPAKDWERILRQAKDNGVLSVSILGGEPTRYFDIDNLLIACEELKI
KTTITTNAQLIKKSTVEILAKSKYITPVLSLQTLDSKLNFELMGVRPDRQIKLAKYFNE
GKKCRINAVYTKQSYEQIIELVDFCIENKIDRFSVANYSEVTGYTKIKKKYDLADLRRL
EYVTDYITQREANLNFATEGCHLFTAYPELINNSIEFSEFDEMYYGCRAKYTKMEIMSN
DILPCIAFLGVNQTKQNAFEKDLLDVWYDDPLYGGIRSFRTKNSKCLSCGLLKICEGG
YVNLIKEKSPEYFRDSVCQL

>CteB(A3DDW1)

MAMIHKFSMMGTNIVVDVNSGAVHVVDDISFDILDYYKNFTAGEIKNKLAHKYNADE
IDEALREIESLEAEGLLFSEDPYKEYVSSMDRKSVVKALCLHISHDCNLRCKYCFASTG
NFGGQRNMMSLEVGKKAIDFLISESGNRKNLEIDFFGGEPMMNFDVVKGIIEYARQKE
KEHNKNFRFTLTTNGLLLNDENIKYINENMQNIVLSIDGRKEVNDRMRIRIDGSGCYD

DILPKFKYVAESRNQDNYYVRGTFTRENMDFSNDVLHLADEGFRQISVEPVVAAKDS

GYDLREEDLPRLFEEYEKLAYEYVKRRKEGNWFNFFHFMIDLTQGPCIVKRLTGCGSG

HEYLAVTPEGDIYPCHQFVGNEKFKMGNVKEGVLNRDIQNYFKNSNVYTKKECDSC

WAKFYCSGGCAANSYNFHKDINTVYKVGCELEKKRVECALWIKAQEM

>YydG(Q45595)

MYNKTVSINLDSRCNASCDHCCFSSSPTSTTRMEKEYIRELVTEFAKNKTIQVISFTGG

EVFLDYKFLKELMEIIKPYEKQITLISNGFWGLSKKKVQEYFHDMNSLNVIALTISYDE

YHAPFVKSSSIKNILEHSRKYPDIDISLNMAVTKDKMSNHILEELGDSILGVKITKFPMI

SVGAAKTRIKQENIHKFYSLEDEDSLHCPGYDIVYHHDGEIYPCCSPAIFETKITLREEY

NQSFERTVEKLNSNLLLFILRKEGFKWFLNILKENNKIEEFDIPYEFSSICGVCGSLFNSA

EKINYFYPYMEKYYNENFKV

**4.6.Bibliography**

(1) Tan, S. Y.; Tatsumura, Y. Alexander Fleming (1881-1955): Discoverer Of Penicillin. *Singapore Medical Journal* **2015**, *56* (7), 366-367.

(2) Dias, D. A.; Urban, S.; Roessner, U. A Historical Overview Of Natural Products In Drug Discovery. *Metabolites* **2012**, *2* (2), 303-336.

(3) Williams, D. H.; Stone, M. J.; Hauck, P. R.; Rahman, S. K. Why Are Secondary Metabolites (Natural Products) Biosynthesized? *Journal Of Natural Products* **1989**, *52* (6), 1189-1208.

(4) Maplestone, R. A.; Stone, M. J.; Williams, D. H. The Evolutionary Role Of Secondary Metabolites — A Review. *Gene* **1992**, *115* (1), 151-157.

(5) Houbraken, J.; Frisvad, J. C.; Samson, R. A. Fleming's Penicillin Producing Strain Is Not Penicillium Chrysogenum But P. Rubens. *IMA Fungus* **2011**, *2* (1), 87-95.

(6) Lobanovska, M.; Pilla, G. Penicillin's Discovery And Antibiotic Resistance: Lessons For The Future? *Yale Journal of Biology and Medicine* **2017**, *90* (1), 135-145.

(7) Gordon, E.; Mouz, N.; Duée, E.; Dideberg, O. The Crystal Structure Of The Penicillin-Binding Protein 2x From Streptococcus Pneumoniae And Its Acyl-Enzyme Form: Implication In Drug Resistance11edited By R. Huber. *Journal Of Molecular Biology* **2000**, *299* (2), 477-485.

(8) Fleming, A. The Discovery Of Penicillin. *British Medical Bulletin* **1944**, *2* (1), 4-5.

(9) Fleming, S. A. Penicillin. Nobel Prize Outreach, 1945. *Nobelprize.Org* (Accessed 2023 May 17th ).

(10) Abraham, E. P.; Chain, E. An Enzyme From Bacteria Able To Destroy Penicillin. *Nature* **1940**, *146* (3713), 837-837.

(11) Munita, J. M.; Arias, C. A. Mechanisms Of Antibiotic Resistance. *Microbiol Spectrum* **2016**, *4* (2).

(12) Ventola, C. L. The Antibiotic Resistance Crisis: Part 1: Causes And Threats. *Pharmacy and Therapeutics* **2015**, *40* (4), 277-283.

(13) Terreni, M.; Taccani, M.; Pregnolato, M. New Antibiotics For Multidrug-Resistant Bacterial Strains: Latest Research Developments And Future Perspectives. *Molecules* **2021**, *26* (9).

(14) Murray, C. J. L.; Ikuta, K. S.; Sharara, F.; Swetschinski, L.; Robles Aguilar, G.; Gray, A.; Han, C.; Bisignano, C.; Rao, P.; Wool, E.; Et Al. Global Burden Of Bacterial Antimicrobial Resistance In 2019: A Systematic Analysis. *The Lancet* **2022**, *399* (10325), 629-655.

(15) COVID-19: U.S. Impact On Antimicrobial Resistance, Special Report 2022. Centers For Disease, C., Prevention, National Center For, E., Zoonotic Infectious, D., Division Of Healthcare Quality, P., Eds.: Hyattsville, MD, 2022.

(16) *National Infection & Death Estimates For Antimicrobial Resistance*. CDC, 2021.. (Accessed 2023.

(17) Gould, I. M.; Bal, A. M. New Antibiotic Agents In The Pipeline And How They Can Help Overcome Microbial Resistance. *Virulence* **2013**, *4* (2), 185-191.

(18) Do, T.; Link, A. J. Protein Engineering In Ribosomally Synthesized And Post-Translationally Modified Peptides (Ripps). *Biochemistry* **2023**, *62* (2), 201-209.

(19) Zhong, G.; Wang, Z. J.; Yan, F.; Zhang, Y.; Huo, L. Recent Advances In Discovery, Bioengineering, And Bioactivity-Evaluation Of Ribosomally Synthesized And Post-Translationally Modified Peptides. *ACS Bio Med Chem Au* **2023**, *3* (1), 1-31.

(20) Li, Y.; Rebuffat, S. The Manifold Roles Of Microbial Ribosomal Peptide–Based Natural Products In Physiology And Ecology. *Journal Of Biological Chemistry* **2020**, *295* (1), 34-54.

(21) Walker, J. A.; Hamlish, N.; Tytla, A.; Brauer, D. D.; Francis, M. B.; Schepartz, A. Redirecting Ripp Biosynthetic Enzymes To Proteins And Backbone-Modified Substrates. *ACS Central Science* **2022**, *8* (4), 473-482.

(22) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; Et Al. Ribosomally Synthesized And Post-Translationally Modified Peptide Natural Products: Overview And Recommendations For A Universal Nomenclature. *Natural Product Reports* **2013**, *30* (1), 108-160.

(23) Atanasov, A. G.; Zotchev, S. B.; Dirsch, V. M.; Orhan, I. E.; Banach, M.; Rollinger, J. M.; Barreca, D.; Weckwerth, W.; Bauer, R.; Bayer, E. A.; Et Al. Natural Products In Drug Discovery: Advances And Opportunities. *Nature Reviews Drug Discovery* **2021**, *20* (3), 200-216.

(24) Montalbán-López, M.; Scott, T. A.; Ramesh, S.; Rahman, I. R.; Van Heel, A. J.; Viel, J. H.; Bandarian, V.; Dittmann, E.; Genilloud, O.; Goto, Y. New Developments In Ripp Discovery, Enzymology And Engineering. *Natural Product Reports* **2021**, *38* (1), 130-239.

(25) Burkhart, B. J.; Hudson, G. A.; Dunbar, K. L.; Mitchell, D. A. A Prevalent Peptide-Binding Domain Guides Ribosomal Natural Product Biosynthesis. *Nature Chemical Biology* **2015**, *11* (8), 564-570.

(26) Davis, K. M.; Schramma, K. R.; Hansen, W. A.; Bacik, J. P.; Khare, S. D.; Seyedsayamdost, M. R.; Ando, N. Structures Of The Peptide-Modifying Radical SAM Enzyme Suib Elucidate The Basis Of Substrate Recognition. *Proceedings Of The National Academy Of Sciences* **2017**, *114* (39), 10420-10425.

(27) Brennan, R. G. The Winged-Helix DNA-Binding Motif: Another Helix-Turn-Helix Takeoff. *Cell* **1993**, *74* (5), 773-776.

(28) Shelton, K. E.; Mitchell, D. A. Bioinformatic Prediction And Experimental Validation Of Ripp Recognition Elements. *Methods Enzymology* **2023**, *679*, 191-233.

(29) Dicaprio, A. J.; Firouzbakht, A.; Hudson, G. A.; Mitchell, D. A. Enzymatic Reconstitution And Biosynthetic Investigation Of The Lasso Peptide Fusilassin. *Journal Of The American Chemical Society* **2019**, *141* (1), 290-297.

(30) Kloosterman, A. M.; Shelton, K. E.; Wezel, G. P. V.; Medema, M. H.; Mitchell, D. A. RRE-Finder: A Genome-Mining Tool For Class-Independent Ripp Discovery. *Msystems* **2020**, *5* (5), E00267-00220.

(31) Kloosterman, A. M.; Medema, M. H.; Van Wezel, G. P. Omics-Based Strategies To Discover Novel Classes Of Ripp Natural Products. *Current Opinion In Biotechnology* **2021**, *69*, 60-67.

(32) Benjdia, A.; Berteau, O. Radical SAM Enzymes And Ribosomally-Synthesized And Post-Translationally Modified Peptides: A Growing Importance In The Microbiomes. *Frontiers Of Chemistry* **2021**, *9*, 678068.

(33) Hudson, G. A.; Burkhart, B. J.; Dicaprio, A. J.; Schwalen, C. J.; Kille, B.; Pogorelov, T. V.; Mitchell, D. A. Bioinformatic Mapping Of Radical S-Adenosylmethionine-Dependent Ribosomally Synthesized And Post-Translationally Modified Peptides Identifies New Cα, Cβ, And Cγ-Linked Thioether-Containing Peptides. *Journal of the American Chemical Society* **2019**, *141* (20), 8228-8238.

(34) Broderick, J. B.; Broderick, W. E.; Hoffman, B. M. Radical SAM Enzymes: Nature's Choice For Radical Reactions. *FEBS Letters* **2023**, *597* (1), 92-101.

(35) Kelly, W. L.; Pan, L.; Li, C. Thiostrepton Biosynthesis: Prototype For A New Family Of Bacteriocins. *Journal of The American Chemical Society* **2009**, *131* (12), 4327-4334.

(36) Freeman, M. F.; Gurgui, C.; Helf, M. J.; Morinaka, B. I.; Uria, A. R.; Oldham, N. J.; Sahl, H.-G.; Matsunaga, S.; Piel, J. Metagenome Mining Reveals Polytheonamides As Posttranslationally Modified Ribosomal Peptides. *Science* **2012**, *338* (6105), 387-390.

(37) Huo, L.; Rachid, S.; Stadler, M.; Wenzel, S. C.; Müller, R. Synthetic Biotechnology To Study And Engineer Ribosomal Bottromycin Biosynthesis. *Chemistry & Biology* **2012**, *19* (10), 1278-1287.

(38) Sofia, H. J.; Chen, G.; Hetzler, B. G.; Reyes-Spindola, J. F.; Miller, N. E. Radical SAM, A Novel Protein Superfamily Linking Unresolved Steps In Familiar Biosynthetic Pathways With Radical Mechanisms: Functional Characterization Using New Analysis And Information Visualization Methods. *Nucleic Acids Research* **2001**, *29* (5), 1097-1106.

(39) Vey, J. L.; Drennan, C. L. Structural Insights Into Radical Generation By The Radical SAM Superfamily. *Chemical Reviews* **2011**, *111* (4), 2487-2506.

(40) Wang, J.; Woldring, R. P.; Román-Meléndez, G. D.; Mcclain, A. M.; Alzua, B. R.; Marsh, E. N. Recent Advances In Radical SAM Enzymology: New Structures And Mechanisms. *ACS Chemical Biology* **2014**, *9* (9), 1929-1938.

(41) Shisler, K. A.; Broderick, J. B. Emerging Themes In Radical SAM Chemistry. *Current Opinion In Structural Biology* **2012**, *22* (6), 701-710.

(42) Mendauletova, A.; Kostenko, A.; Lien, Y.; Latham, J. How A Subfamily Of Radical S-Adenosylmethionine Enzymes Became A Mainstay Of Ribosomally Synthesized And Post-Translationally Modified Peptide Discovery. *ACS Bio & Med Chem Au* **2022**, *2* (1), 53-59.

(43) Goldman, P. J.; Grove, T. L.; Sites, L. A.; Mclaughlin, M. I.; Booker, S. J.; Drennan, C. L. X-Ray Structure Of An Adomet Radical Activase Reveals An Anaerobic Solution For

Formylglycine Posttranslational Modification. *Proceedings Of The National Academy Of Sciences* **2013**, *110* (21), 8519-8524.

(44) Schramma, K. R.; Bushin, L. B.; Seyedsayamdost, M. R. Structure And Biosynthesis Of A Macrocyclic Peptide Containing An Unprecedented Lysine-To-Tryptophan Crosslink. *Nature Chemistry* **2015**, *7* (5), 431-437.

(45) Caruso, A. Uncovering New Metalloenzyme-Mediated Chemistry From Nature's Toolkit: Radical S-Adenosylmethionine Reactions In Natural Product Biosynthesis. Princeton University, 2021.

(46) Breil, B. T.; Ludden, P. W.; Triplett, E. W. DNA Sequence And Mutational Analysis Of Genes Involved In The Production And Resistance Of The Antibiotic Peptide Trifolitoxin. *Journal Of Bacteriol* **1993**, *175* (12), 3693-3702.

(47) Scupham, A. J.; Dong, Y.; Triplett, E. W. Role Of Tfxe, But Not Tfxg, In Trifolitoxin Resistance. *Applied Environmental Microbiology* **2002**, *68* (9), 4334-4340.

(48) Wecksler, S. R.; Stoll, S.; Tran, H.; Magnusson, O. T.; Wu, S. P.; King, D.; Britt, R. D.; Klinman, J. P. Pyrroloquinoline Quinone Biogenesis: Demonstration That Pqqe From Klebsiella Pneumoniae Is A Radical S-Adenosyl-L-Methionine Enzyme. *Biochemistry* **2009**, *48* (42), 10151-10161.

(49) Harrison, K. J.; Crécy-Lagard, V.; Zallot, R. Gene Graphics: A Genomic Neighborhood Data Visualization Web Application. *Bioinformatics* **2018**, *34* (8), 1406-1408.

(50) Inokoshi, J.; Matsuhama, M.; Miyake, M.; Ikeda, H.; Tomoda, H. Molecular Cloning Of The Gene Cluster For Lariatin Biosynthesis Of Rhodococcus Jostii K01-B0171. *Applied Microbiology and Biotechnology* **2012**, *95* (2), 451-460.

(51) Melby, J. O.; Dunbar, K. L.; Trinh, N. Q.; Mitchell, D. A. Selectivity, Directionality, And Promiscuity In Peptide Processing From A Bacillus Sp. Al Hakam Cyclodehydratase. *Journal Of The American Chemical Society* **2012**, *134* (11), 5309-5316.

(52) Latham, J. A.; Iavarone, A. T.; Barr, I.; Juthani, P. V.; Klinman, J. P. Pqqd Is A Novel Peptide Chaperone That Forms A Ternary Complex With The Radical S-Adenosylmethionine Protein Pqqe In The Pyrroloquinoline Quinone Biosynthetic Pathway. *Journal Of Biological Chemistry* **2015**, *290* (20), 12908-12918.

(53) Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. HADDOCK: A Protein−Protein Docking Approach Based On Biochemical Or Biophysical Information. *Journal Of The American Chemical Society* **2003**, *125* (7), 1731-1737.

(54) Henzler-Wildman, K.; Kern, D. Dynamic Personalities Of Proteins. *Nature* **2007**, *450* (7172), 964-972.

(55) Grove, T. L.; Himes, P. M.; Hwang, S.; Yumerefendi, H.; Bonanno, J. B.; Kuhlman, B.; Almo, S. C.; Bowers, A. A. Structural Insights Into Thioether Bond Formation In The Biosynthesis Of Sactipeptides. *Journal Of The American Chemical Society* **2017**, *139* (34), 11734-11744.

(56) Blue, T. C.; Davis, K. M. Computational Approaches: An Underutilized Tool In The Quest To Elucidate Radical SAM Dynamics. *Molecules* **2021**, *26* (9), 2590.

(57) Cuff, J. A.; Barton, G. J. Application Of Multiple Sequence Alignment Profiles To Improve Protein Secondary Structure Prediction. *Proteins* **2000**, *40* (3), 502-511.

(58) Johnson, M.; Zaretskaya, I.; Raytselis, Y.; Merezhuk, Y.; Mcginnis, S.; Madden, T. L. NCBI BLAST: A Better Web Interface. *Nucleic Acids Research* **2008**, *36* (Suppl_2), W5-W9.

(59) Lupas, A.; Van Dyke, M.; Stock, J. Predicting Coiled Coils From Protein Sequences. *Science* **1991**, *252* (5009), 1162-1164.

(60) Chemes, L. B.; Alonso, L. G.; Noval, M. G.; De Prat-Gay, G. Circular Dichroism Techniques For The Analysis Of Intrinsically Disordered Proteins And Domains. In *Intrinsically Disordered Protein Analysis: Volume 1, Methods And Experimental Tools*, Uversky, V. N., Dunker, A. K. Eds.; Humana Press, 2012; Pp 387-404.

(61) Dill, K. A. Dominant Forces In Protein Folding. *Biochemistry* **1990**, *29* (31), 7133-7155.

(62) Malleshappa Gowder, S.; Chatterjee, J.; Chaudhuri, T.; Paul, K. Prediction And Analysis Of Surface Hydrophobic Residues In Tertiary Structure Of Proteins. *Scientificworldjournal* **2014**, *2014*, 971258.

(63) Young, L.; Jernigan, R. L.; Covell, D. G. A Role For Surface Hydrophobicity In Protein-Protein Recognition. *Protein Science* **1994**, *3* (5), 717-729.

(64) Chekan, J. R.; Ongpipattanakul, C.; Nair, S. K. Steric Complementarity Directs Sequence Promiscuous Leader Binding In Ripp Biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **2019**, *116* (48), 24049-24055.

(65) Kyte, J.; Doolittle, R. F. A Simple Method For Displaying The Hydropathic Character Of A Protein. *J Mol Biol* **1982**, *157* (1), 105-132.

(66) Benjdia, A.; Guillot, A.; Ruffié, P.; Leprince, J.; Berteau, O. Post-Translational Modification Of Ribosomally Synthesized Peptides By A Radical SAM Epimerase In Bacillus Subtilis. *Nature Chemistry* **2017**, *9* (7), 698-707.

(67) Popp, P. F.; Benjdia, A.; Strahl, H.; Berteau, O.; Mascher, T. The Epipeptide Yydf Intrinsically Triggers The Cell Envelope Stress Response Of Bacillus Subtilis And Causes Severe Membrane Perturbations. *Frontiers In Microbiology* **2020**, *11*, Original Research.

(68) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A. Highly Accurate Protein Structure Prediction With Alphafold. *Nature* **2021**, *596* (7873), 583-589.

(69) Ezerski, J. C.; Zhang, P.; Jennings, N. C.; Waxham, M. N.; Cheung, M. S. Molecular Dynamics Ensemble Refinement Of Intrinsically Disordered Peptides According To Deconvoluted Spectra From Circular Dichroism. *Biophysical Journal* **2020**, *118* (7), 1665-1678.

(70) Casañal, A.; Lohkamp, B.; Emsley, P. Current Developments In Coot For Macromolecular Model Building Of Electron Cryo-Microscopy And Crystallographic Data. *Protein Science* **2020**, *29* (4), 1069-1078.

(71) Banci, L.; Bertini, I.; Carloni, P.; Luchinat, C.; Orioli, P. L. Molecular Dynamics Simulations On Hipip From Chromatium Vinosum And Comparison With NMR Data. *Journal Of The American Chemical Society* **1992**, *114* (27), 10683-10689.

(72) Vesper, M. D. Collective Dynamics In Allosteric Transitions: A Molecular Dynamics Study. Georg-August University School Of Science (GAUSS), 2012.

# **Chapter 5.** Conclusion and Future Works

## 5.1.General Conclusions

Biocatalysts are macromolecules of biological origin that mediate chemical reactions resulting in enantiospecific compounds. Throughout human history, the development, identification, and utilization of protein biocatalysts have progressed with technological advancements. These have been marked as four major waves of biocatalysis innovation.[1] The fourth wave of biocatalysis, in particular, has accelerated the field by coupling computational and traditional biochemical approaches allowing the analysis of enzyme systems that are difficult to reproduce in a laboratory environment. The previous chapters have demonstrated how fourth-wave techniques have broadened the understanding of metalloenzymes within the rSAM and OYE superfamilies.

## 5.2.Conclusions and Further Exploration of the OYE Superfamily

Chapter 2 details my lab's use of an SSN to provide a classification system for the OYE superfamily that evolves with the family over time. This work was supported by a longitudinal study in Chapter 3. We generated a protein library to represent the sequence diversity in the superfamily, which I utilized in Chapter 2 to identify OYE homologs with the ability to create decalone chiral building blocks. Unfortunately, under our standard operating procedure, no member of our protein library could generate monocarboxylic acid chiral building blocks. This lack of activity was disappointing, however not surprising, as such activity is attributed to the 2-ER metalloenzyme family within the OYE superfamily.[2] This family was bioinformatically explored in Chapter 3. This chapter presents a novel sub-family of 2-ERs with a unique wild-type mutation in the FeS binding domain. The representative of this family (BurkOYE) displayed distinct *N*-methyl-proline demethylase activity while retaining its ability to generate monocarboxylic acid chiral building blocks through reductive chemistry. This oxidative pathway

is significant as the ability of OYEs to display oxidative activity is not well explored in the OYE superfamily. One of the earliest accounts exploring oxidation activity in OYEs by Massey et al. reported that increasing the redox potential via artificial flavin incorporation converted OYE1 into a desaturase enzyme.[3] While these results instigated the belief that the redox potential of an OYE may be responsible for desaturase activity, our large-scale probe of the superfamily found a lack of correlation between the redox potential and activity.[4] This, however, does not exclude the redox potential of the FeS cluster as inconsequential in both pathways. Further studies are needed to uncover the role of the FeS cluster in demethlayse activity and why generating monocarboxylic acid chiral building blocks is exclusive to this family of OYEs.

Additionally, 2-ERs have always had issues with soluble and active expression under aerobic conditions.[5] This work details a method to obtain 2-ERs aerobic, opening the door to 2-ER application in industrial uses. More studies are needed to fully understand the native role of BurkOYE and its wild type of mutation. Obtaining soluble alanine mutate, where all the Cys in the FeS binding domain, as well as the Alanine to Cysteine mutant, are needed to understand the role of the variant position better. Both will be reexpressed under fully anaerobic conditions, and biotransformation and kinetic studies will be employed to detect how the constructs affect the activity and oxygen sensitivity.

Interestingly, there are other clusters in the OYE Superfamily that we have bioinformatically predicted to contain predominantly metalloenzyme members.[4] These clusters are unexplored and contain an FMN-like domain with a [4Fe-4S] cluster binding motif without any additional binding domain. Future works will involve the biochemical and computational characterization of these families to determine activity, along with the exploration of other clusters within the 2-ER family to elucidate the full catalytic activity of the family.

## 5.3.Conclusions and Further Exploration of non-RRE mediated RiPP Recognition

Advancements in modern sequencing approaches have resulted in the exponential growth of raw protein biocatalyst data. That is seen no more impactfully than the bioinformatic-driven discovery of the rSAM superfamily. While rSAM enzymes were known prior to Sofia et al., it is undeniable that bioinformatics techniques have advanced the rSAM superfamily to one of the largest known to date. [6] In Chapter 4, I explore how RiPP modifying rSAMs identify the percussor peptides through in silico analysis supported by circular dichroism and fluorescence analysis. Through this study, I identify hydrophobic pockets within the bridging region of SuiB that may aid in peptide recognition as well as perform docking experiments that consistently result in preferential interaction of the leader sequence with the bridging region.

Further work is needed to determine if the RRE is required for peptide recognition as well as the role of the bridging domain in RiPP recognition. Mutagenic studies coupled with biotransformation would be useful to determine this. Mutating out the RRE and performing fluorescence-based Michaelis–Menten kinetics studies with the construct.[7] Additional studies should be conducted with a mutant that removes the hydrophobic pocket in the bridging domain. Both mutations may result in insoluble protein, so in tandem, *in silico* binding studies could be performed to computationally derive Gibbs free energy of the mutants with SuiA.

In Chapter 4, I also identify that SuiA undergoes dramatic conformational changes as a result of interactions with SuiB. It is still unclear when the conformational change from disordered loop to α-helix occurs. It is possible that order is gained upon weak interactions with the RRE prior to shuttling into the hydrophobic pockets of the catalytic barrel or solely by the barrel. However, additional MD and Metadynamics simulations are needed to predict a wider conformational

landscape. Metadynamics allows for an estimate of the free energy of the system while changing the energy landscape during an MD simulation to afford uncommon macromolecule conformational changes.[8]

Docking was unable to visualize SuiB-RRE:SuiA leader strand interaction. At the moment, the only data supporting SuiB-RRE:SuiA leader interaction is our fluorescence assays. Unfortunately, these assays could not be repeated with the full SuiB construct, so fluorescence studies will be repeated via microscale thermophoresis. Additionally, it is possible that the RRE has a preference for a specific disordered conformation of SuiA to facilitate binding. A random sampling of SuiA conformations from an 80ns MD simulation did not result in SuiA leader strand RRE interaction. So as an alternative approach, the crystallization of the RRE with SuiA may provide a more optimal starting structure for additional computational analysis.

**5.4.Bibliography**

(1) Bornscheuer, U. T. The Fourth Wave Of Biocatalysis Is Approaching. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2018**, *376* (2110).

(2) Toogood, H. S.; Gardiner, J. M.; Scrutton, N. S. Biocatalytic Reductions And Chemical Versatility Of The Old Yellow Enzyme Family Of Flavoprotein Oxidoreductases. *Chemcatchem* **2010**, *2* (8), 892-914.

(3) Murthy, Y. V. S. N.; Meah, Y.; Massey, V. Conversion Of A Flavoprotein Reductase To A Desaturase By Manipulation Of The Flavin Redox Potential. *Journal Of The American Chemical Society* **1999**, *121* (22), 5344-5345.

(4) White, D. W., Iamuri, S., Joud, P., Blue, T.C., Copp, J., Lutz, S. The Hidden Biocatalytic Potential Of The Old Yellow Enzyme Family. *Biorxiv* **2023**.

(5) Mordaka, P. M.; Hall, S. J.; Minton, N.; Stephens, G. Recombinant Expression And Characterisation Of The Oxygen-Sensitive 2-Enoate Reductase From Clostridium Sporogenes. *Microbiology (Reading)* **2018**, *164* (2), 122-132.

(6) Sofia, H. J.; Chen, G.; Hetzler, B. G.; Reyes-Spindola, J. F.; Miller, N. E. Radical SAM, A Novel Protein Superfamily Linking Unresolved Steps In Familiar Biosynthetic Pathways With Radical Mechanisms: Functional Characterization Using New Analysis And Information Visualization Methods. *Nucleic Acids Research* **2001**, *29* (5), 1097-1106.

(7) Palmier, M. O.; Van Doren, S. R. Rapid Determination Of Enzyme Kinetics From Fluorescence: Overcoming The Inner Filter Effect. *Analytical Biochemistry* **2007**, *371* (1), 43-51.

(8) Capelli, R.; Menke, A. J.; Pan, H.; Janesko, B. G.; Simanek, E. E.; Pavan, G. M. Well-Tempered Metadynamics Simulations Predict The Structural And Dynamic Properties Of A Chiral 24-Atom Macrocycle In Solution. *ACS Omega* **2022**, *7* (34), 30291-30296.