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:

_____ _____

Samantha M Iamurri Date

Exploring the Role of Cofactor and Protein Sequence Variability Across the Old Yellow Enzyme Superfamily

By

Samantha M Iamurri
Doctor of Philosophy

Chemistry

_____

Dr. Stefan Lutz
Advisor

_____

Dr. Simon Blakey
Committee Member

_____

Dr. Emily Weinert
Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____

Date

Exploring the Role of Cofactor and Protein Sequence Variability Across the Old Yellow Enzyme Superfamily

By

Samantha M Iamurri
B.S., The College of New Jersey, 2012

Advisor: Stefan Lutz, 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
2019

Abstract

Exploring the Role of Cofactor and Protein Sequence Variability Across the Old Yellow Enzyme Superfamily

By Samantha M Iamurri

Over the past century enzymes have been the center of intense research efforts focused on the discovery and engineering of novel biocatalysts to produce industrially and commercially relevant chemical compounds. Most of these efforts involve enhancing enzyme efficiency, stereospecificity, stability, and substrate scope through alterations to the native protein backbone. Certain enzymes, however, require small molecule chemical cofactors to perform their native functions. Accordingly, engineering of natural cofactors represents parallel strategy for enzyme customization that can be used alongside traditional backbone modification methodologies. On a more fundamental level, cofactor engineering may also be employed to probe the mechanistic and biophysical characteristics of protein catalysts.

This dissertation thus utilizes engineered cofactor analogs to examine the functional role of the native cofactor flavin mononucleotide (FMN) in the folding of Old Yellow Enzyme 1 (OYE1) from *Saccharomyces pastorianus*. Preliminary work with a cell-free protein expression system determined that FMN must be present during OYE1 synthesis to produce catalytically functional enzymes, disputing a longstanding assumption that the cofactor's presence is unnecessary for proper folding. Nascent OYE1 was then supplemented with cofactor analogs containing key elements of the native FMN in order to systematically study the contribution of specific chemical groups towards proper enzyme folding. Activity assay data suggest that the flavin's isoalloxazine and terminal phosphate moieties are essential for productive OYE1 folding.

With the endless industrial demand for new biocatalysts, this work also details our exploration into the OYE superfamily in search of novel enzymes. Using sequence similarity networks (SSNs) to visually guide our sampling efforts, we have doubled the number of characterized OYEs and have begun functionally profiling the entire superfamily in an unprecedented manner. Excitingly, we have identified a number of novel catalysts capable of catalytically out-performing any of the previously reported native or engineered OYE family members.

Exploring the Role of Cofactor and Protein Sequence Variability Across the Old Yellow Enzyme Superfamily

By

Samantha M Iamurri
B.S., The College of New Jersey, 2012

Advisor: Stefan Lutz, 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
2019

Acknowledgements

There have been a number of people throughout my time at Emory who have made my graduate school experience extremely memorable. First, I would like to thank my advisor Dr. Stefan Lutz for giving me an opportunity to not only work in his lab but free rein to manage it. Thank you for your endless support, encouragement, and belief in me. I truly cherish the relationship we have built. I would also like to thank my committee members, Dr. Simon Blakey and Dr. Emily Weinert for their helpful input about my research over the years. Additionally, I would like to thank all of the members of the Atlanta Flavin Meetings for their memorable feedback and discussions.

Next, I would like to thank the people who make coming to work enjoyable, the best labmates anyone could ask for. To all of the Lutz Lab members both past and present which I have had the privilege to work next to and learn from the lab would not have been the same without our unique hypothetical questions, impromptu sing-a-longs, and our exotic lab lunches. I thank you all for the constant support and love #LL4L. I would especially like to thank Matthew Jenkins for our morning coffee, our wonderful conversations, and the endless support since I joined the lab. I would like to thank Dr. Huanyu Zhao for all the synthetic knowledge and help she provided me during her time in the lab. Next, I would like to thank Evy Kimbrough for all the snacks and support she provided. I especially would like to thank David White and Parisa Joud for all the work they did on the OYE Superfamily project and for the friendship we've built outside the lab.

Finally, I would like to thank my family, especially my parents. You have always unconditionally supported, loved, and provided me with anything I needed to help me succeed. I wouldn't be the person or where I am today without you. Thank you for everything. I love you.

List of Frequently Used Abbreviations

Abbreviation	Full Name
OYE	Old Yellow Enzyme
FMN	Flavin Mononucleodtide
FAD	Flavin Adenine Dinucleotide
$NAD(P)^+$	Nicotinamide Adenine Dinucleotide (Phosphate)$^+$
PURE	Protein Synthesis Using Recombinant Elements
ITC	Isothermal Calorimetry
SSN	Sequence Similarity Network

Table of Contents

Table of Figures

Table of Tables

Chapter 1: General Introduction

I. General Introduction

Enzymes are essential biological catalysts which carry out the "chemistry of life" (1). They allow for vital macro and small molecule transformations to be carried out on a biologically relevant timescale with high enatio-, regio-, and chemoselectivity. Due to these inherent properties of enzymes, they have been the crux of biocatalysis research and innovation for over a century (2, 3). Not only are enzymes attractive alternatives to traditional organic synthetic catalysts due to their high specificities and catalytic efficiencies, they are also cost effective and "green" catalysts (3, 4). However, about half of all chemical reactions performed by enzymes require a cofactor (5). Cofactors are either metal ions or small organic molecules that facilitate the chemistry performed by enzymes, and therefore are also critical for function. The role of the cofactor varies depending on the enzyme in which it is bound, but for example it can act in protein stabilization, as a Lewis acid, or to promote group and electron transfer reactions. While cofactors help to expand the catalytic repertoire of enzymes, various engineering efforts designed to further enhance the natural properties and expand the normal capabilities of enzymes are still a major focus of the field (3, 6, 7).

Traditional methods of protein engineering, directed evolution or rational design, rely on site saturated mutagenesis, error prone PCR, DNA shuffling, or rational amino acid replacement to introduce variations into the protein backbone in order to modulate catalytic activity or broaden substrate scope (8-11). Another facet of protein engineering involves modification of the enzyme's cofactor. Cofactor engineering has enormous potential to alter enzymatic substrate specificities and functionalities, especially when applied in conjunction with traditional protein backbone engineering strategies. Such engineering can be accomplished by exchanging native

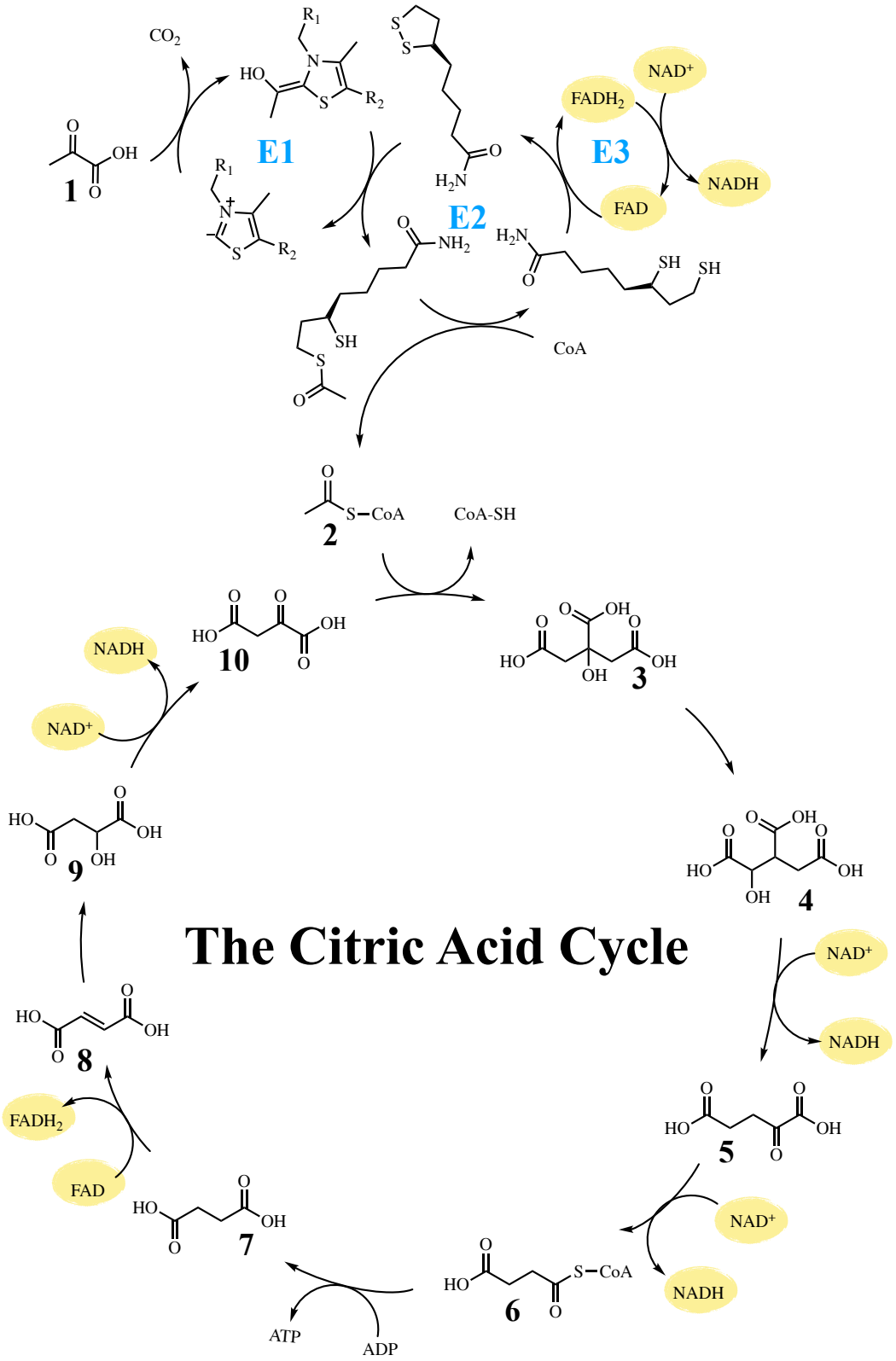
metals for different metals or via synthetic modification to organic small molecules that constitute the active sites. In this chapter the importance of biological cofactors will be discussed while also highlighting the impact of cofactor engineering, specifically within the context of mechanistic studies and attempts to bridge the gap between traditional organic and biological catalysts.

II. Chemical Versatility of Cofactors

There are countless metal and organic cofactors in nature which are responsible for many of life's biochemical processes. It is not in the scope of this dissertation to discuss all of them in detail, but instead to highlight the versatility of select cofactors. To this end, the cofactors needed for life's most important central metabolic pathway, the citric acid cycle will be discussed, starting from pyruvate and concluding with electron transfer in Complex II of oxidative phosphorylation (**Figure 1.1**) (12).

The committing step of the citric acid cycle is the conversion of pyruvate (**1**) to acetyl-CoA (**2**) via the pyruvate dehydrogenase complex (PDC). The PDC is a multi-domain complex made up of repeated copies of three enzymes (E1-3), each of which rely on a different cofactor for catalysis (13). The first enzyme, E1, is thiamin diphosphate (ThDP) dependent and responsible for the decarboxylation of **1** and corresponding formation of a C2α-hydroxyethylidene-ThDP intermediate (**Figure 1.1**). The mechanism of action is referred to as the Umpolung mechanism and is dependent on the deprotonation of the C2H of the thiazolium ring of ThDP. The uniqueness of the thiazolium ring lies in the nitrogen, which acts as an electron sink to stabilize the rate-limiting formation of the carbene (13, 14). The resulting

Figure 1.1: The Citric Acid Cycle. Compounds formed during the cycle are numbered and redox active cofactors are highlighted in yellow. Top: The conversion of pyruvate to acetyl-CoA by the pyruvate dehydrogenase complex with all of the associated cofactors. The PDC subunits are labeled in blue.



reactive carbene can attack a carbonyl moiety on the substrate, which in this case is the ketone of **1**. Decarboxylation of the **1** leads to the common Breslow intermediate that all ThDP enzymes proceed through (15, 16). At this point, catalysis shifts to the second enzyme of PDC, known as E2 or the dihydrolipoamide acetyl transferase. E2 contains a covalently bound lipoic acid cofactor which becomes acetylated as the substrate is transferred from the ThDP to the lipoyl group. The acetyl group is then transferred from the lipoyl group's CoA to form acetyl-CoA (**2)** and dihydrolipoyl-E2. The final enzyme, E3, a dihydrolipomide dehydrogenase, is dependent on two different cofactors to regenerate lipoic acid, flavin adenine dinucleotide (FAD) and nicotinamide adenine dinucleotide (NAD^+). The reduction of FAD to the reduced form, $FADH_2$, oxidizes the dihydrolipoyl moieties to reform the disulfide bond of lipoic acid. The flavin is then oxidized by NAD^+ and the resulting NADH molecule is used in oxidative phosphorlaytion for the productive of ATP.

The PDC itself utilizes chemically diverse cofactors to perform its biological function. While lipoic acid is specific to only E2, E1 and E3 possess cofactors which are widely used throughout the cell. For example ThDP dependent enzymes are used in the pentose phosphate pathway to perform a 2-carbon transfer necessary for the production of fructose-6-phosphate and glyceraldehyde-3-phosphate to feed into glycolysis (17). In plants, ThDP dependent enzymes are used in the Calvin cycle for the degradation of 7-membered ring sugars to pentoses (18). In addition to ThDP, the redox cofactors NAD^+ and FAD will continue to appear throughout the remainder of the citric acid cycle and their importance as electron shuttles will be come apparent.

Following its formation by the PDC, compound **2** binds sequentially after oxaloacetate (**10**) in the active site of citrate synthase for the noncofactor mediated Claisen condensation to

form citrate (**3**) (19). Compound **3** is then converted into isocitrate (**4**) via aconitase, an iron-sulfur [Fe-S] dependent enzyme. [Fe-S] clusters are some of the most versatile and ubiquitous cofactors in nature and can exist as several different core units: [2Fe-2S], [3Fe-4S], [4Fe-4S], and [8Fe-7S] (20). The core units are usually attached via a cysteine residue, which coordinates to one of the iron molecules to complete the tetrahedral coordination. The ability of iron and sulfur to delocalize electron density gives rise to the cluster's primary biological role as a mediator in electron transport (21, 22). However, [Fe-S] proteins serve additional functions beyond electron shuttling. Other important roles include substrate binding and activation, as seen in aconitase. Specifically, the [Fe-S] cluster from aconitase coordinates the hydroxyl group of **3**, which helps simultaneously position and activate the substrate for dehydration (23). The same coordination of substrate by an [Fe-S] cluster is seen in the family of radical S-adenosylmethionine (SAM) dependent enzymes. The [4Fe-4S] binds the SAM molecule to facilitate the reductive cleavage and generate the 5'-deoxyadenosyl radical needed to catalyze the biosynthesis or degradation of DNA precursors, vitamins, cofactors, antibiotics, and herbicides (24, 25). The clusters are also known to play a role in the regulation of gene expression, as seen with redox-sensitive transcriptional activator, SoxR (26). The $[2Fe\text{-}2S]^{2+,+}$ of SoxR functions under a slightly different mechanism in order to regulate gene expression in response to environmental stimuli. At times of oxidative stress the [Fe-S] is oxidized, which stimulates expression of SoxS, and subsequently initiates the oxidative stress response pathway (26). Two new roles for [Fe-S] clusters have more recently emerged, the first of which involves disulfide reductions performed by heterodisulfide reductase in methanogenic archaea as the cluster cleaves the disulfide bond via two single electron donations (27-29). The second example occurs during

biotin synthesis as the [2Fe-2S] was shown to be the source of sulfur needed to convert dethiobiotin to biotin. Following sulfur donation, the cluster is reassembled during each catalytic cycle (30-32). Clearly, [Fe-S] clusters have a diverse set of roles and can fulfill numerous functional metabolic needs within the context of a single pathway, including cellular respiration.

The next step of the citric acid cycle is the conversion of **4** to α-ketoglutarate (**5**) and carbon dioxide by another multi-cofactor dependent enzyme, isocitrate dehydrogenase (IDH). Within the active site of IDH resides a metal ion, either magnesium or manganese, which coordinates the dicarbonyl of **4** while simultaneously being ligated to three aspartate residues of the protein (33). The presence of the magnesium serves two purposes: to help position the substrate and to act as a Lewis acid by deprotonating the alcohol to promote the first step of the mechanism (34). This is followed by a hydride transfer from **4** to NAD^+ to form the oxalosuccinate intermediate. The final step of the mechanism is a decarboxylation, which is believed to be stabilized by the magnesium, to form carbon dioxide and **5** (35). The equivalent of NADH formed acts as an electron shuttle to feed into oxidative phosphorylation.

As evident by the example of IDH above, metal ions are an important class of cofactors which are essential for life, collectively providing the catalytic power to about a third of the proteins found in cells (36). These metals are most commonly involved in either protein stabilization or facilitating chemical reactions that otherwise would be impossible with purely organic molecules, and can be found across all six classes of enzymes (37, 38). Enzymes possessing metal cofactors, referred to as metalloenzymes, are a fundamental part of biology and play roles in photosynthesis, respiration, and nitrogen fixation (37). Metals afford the necessary chemical versatility while the environment of the protein's active site permits different electronic

and structural configurations for the bound metal (38). Magnesium is the most abundant metal found in the cell, existing in millimolar quantities due to its essential role as the coordination partner for phosphate in key biological molecules like DNA, RNA, or ATP. Furthermore, it is most often coordinated to oxygen and plays a role in the polarization of P-O and C-O bonds during catalysis, as seen with the coordination of isocitrate. Due to its natural abundance, it is not surprising that magnesium is found within the active site of IDH; in fact, the majority of metal containing metabolic processes utilize magnesium (37, 39). However, it is surprising that the same level of activity is observed when manganese is bound in IDH as few enzymes exhibit the same catalytic efficiency in the presence of different metal species (33). For example Catechol-O-methyltransferase (COMT), is a magnesium dependent enzyme responsible for the regulation of catecholamine neurotransmitters in the brain and other organs where the metal cofactor is responsible for binding and positioning the substrate in the correct orientation to facilitate catalysis (40, 41). Replacement of magnesium with calcium, which is considered a conservative replacement, completely inactivates the enzyme (39). Specifically, alterations to the active site geometry occur in the presence of calcium which negatively impacts substrate binding. While calcium and magnesium share similar electronic properties, the subtle differences in atomic size can be the difference between an active or an inactive enzyme.

The remaining steps of the citric acid cycle contain cofactors that have been previously described or enzymes that do not require cofactors. The conversion of **5** to succinyl-CoA (**6**) involves the multi-subunit complex α-Ketoglutarate dehydrogenase (KGDH), which possesses numerous copies of subunits E1, E2, and E3 just like PDC (42). In the same manner that **1** is converted to **2**, **5** binds in the active site of E1 and is decarboxylated, then the decarboxylated

intermediate is transferred to E2 where an equivalent of CoA reacts to form **6**. The reduced lipoic acid is reoxidized by E3 via FAD and NAD^+ to complete the catalytic cycle. The following step of the citric acid cycle is unique in that the conversion of **6** to succinate (**7**) uses substrate level phosphorylation to produce a molecule of ATP. The enzyme responsible for promoting this step, succinyl-CoA synthetase, is magnesium dependent for the binding and positioning of the co-substrate, ADP (43). The mechanism includes displacement of CoA by a nucleophilic inorganic phosphate to form succinyl phosphate, followed by the removal of phosphate by an active site histidine to generate **7**. The phosphorylated histidine then transfers the phosphate to ADP (44).

Compound **7** next travels to succinate dehydrogenase (SDH), which is a special enzyme complex that is part of both the citric acid cycle and the electron transport chain (45). SDH is composed of four subunits, two of which are catalytic (Sdh1 and Sdh2). The other two subunits, Sdh3 and Sdh4, are membrane bound units that are responsible for facilitating electron flow from Sdh1 and Sdh2 to the rest of the electron transport chain. Upon substrate **7** binding to Sdh1, the covalently linked FAD facilitates the oxidation of **7** to produce fumarate (**8**) and $FADH_2$. SDH displays an additional feature of FAD that had not yet been described in the pathway. In this case, FAD performs both two electron and single electron transfers as FAD and other flavin cofactors have the ability to participate in two mechanisms. The oxidation of **7** involves a two electron transfer to FAD, then the resulting $FADH_2$ can donate one electron at a time to the [3Fe-4S] cluster of Sdh2 (45). Sdh2 contains 3 different [Fe-S] clusters which aid in the shuttling of electrons from the citric acid cycle to ubiquinone, another redox active cofactor primarily used to transport electrons (46). The utilization of FAD in the citric acid cycle only hints at the versatility of flavin cofactors. Accordingly, these cofactors will be further discussed in section **III**.

Following the complete oxidation of **8**, the three carbons introduced to the citric acid cycle by **1** are converted to malate (**9**) through a simple hydration reaction by the cofactor independent fumarase (47). The final step of the citric acid cycle involving the conversion of **9** to **10** also produces the final equivalent of NADH and allows for the continuation of the cycle. The conversion of **9** to **10** is performed by malate dehydrogenase, which undergoes a conformational change upon binding both the substrate and NAD^+ (48). The conformational change brings other catalytically relevant residues into position, specifically a histidine residue responsible for the deprotonation of the alcohol on **9**, which allows for the direct hydride transfer to NAD^+ and production of **10**. The resulting NADH, as seen previously, feeds into the electron transport chain for oxidative phosphorylation.

Over the course of the pathway, four molecules of NADH are produced. The main role of this redox active cofactor is to shuttle electrons from the citric acid cycle to the electron transport chain for ATP production. Additionally, NAD^+, which is predominantly used in catabolic pathways, has an anabolic counterpart, $NADP^+$. These two cofactors are electronically and structurally identical with the exception that NADPH contains a phosphate group at the 2' position of the adenosine ribose (49). Both are involved in a variety of cellular processes including metabolism, antioxidation, cell death, calcium homeostasis, gene expression, and the reductive synthesis of fatty acids, DNA, and steroids (50-52). NAD^+/$NADP^+$ is heavily utilized as an electron shuttle, as seen with the enoate reductases and some P450s, or in cofactor regeneration, as observed with flavoenzymes. Much like flavins, nicotinamide cofactors can also play more direct catalytic roles. For example, NAD^+ dependent oxidoreducatases utilize the

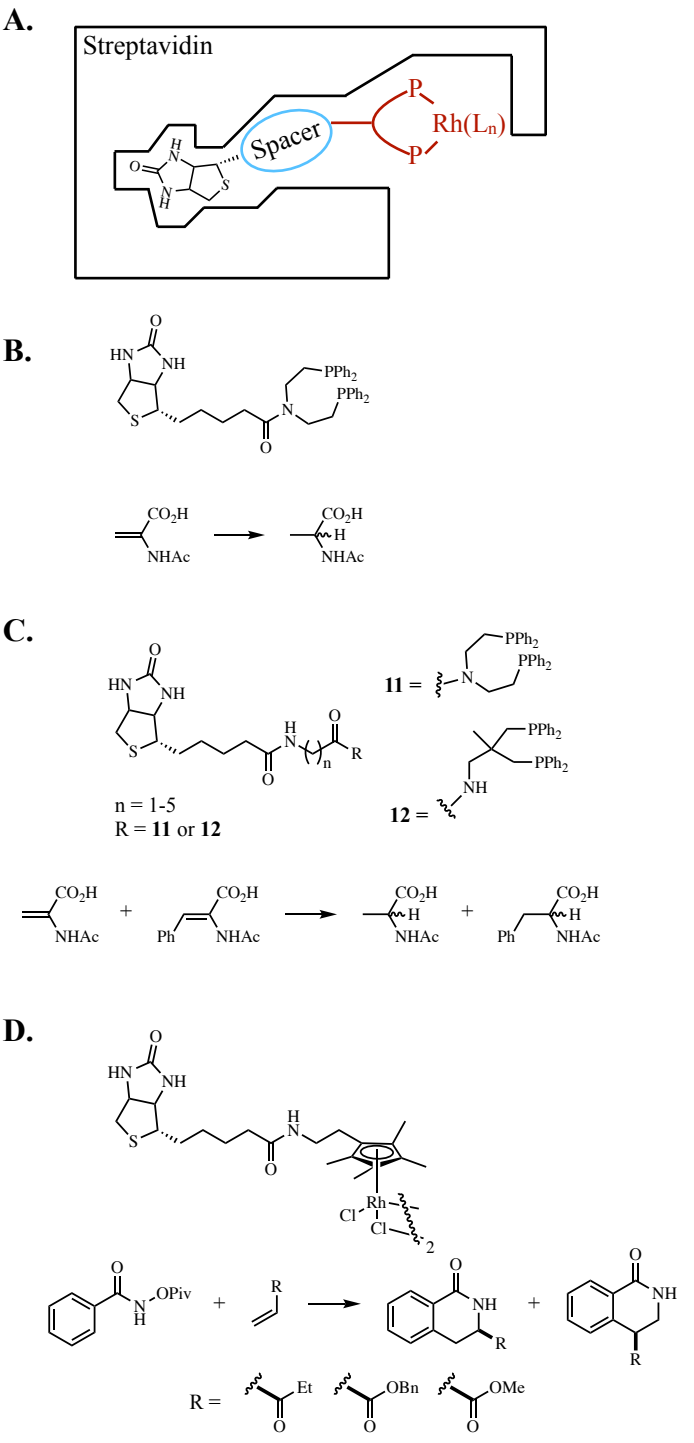
reduced form for the direct oxidation of primary alcohols (53). While the citric acid cycle does not contain all of the cofactors necessary for life, it does exemplify the crucial role of cofactors.

Artificial Metalloenzymes (ArMs)

Cofactor dependent enzymes are still limited to a specific reaction or functioning within the confines of living cells. Deriving inspiration from the versatility of natural cofactors, these enzymes and others have been the subjects of intense research efforts aimed at increasing substrate scope, catalytic efficiency, and stability outside the cell through efforts such as protein and cofactor engineering. Modified enzymes generated from such efforts are referred to as Artificial Metalloenzymes (ArMs), and are usually obtained though either covalent, supramolecular, or dative anchoring of non native metallocofactors, or through metal swapping within metalloproteins (54).

The first account of an ArM was in the 70's by Whitesides *et al.*, where avidin was converted into an artificial hydrogenase by covalently linking a rhodium catalyst to biotin, **Figure 1.2** (55). Placing the rhodium catalyst in the binding pocket of avidin provided a chiral environment for an asymmetric hydrogenation. However, due to limited synthetic and protein engineering tools at the time, further development died off until the early 2000's when the field was revitalized through work done primarily by the Ward group (56-60). Over the last few decades, more groups, such as the Arnold, Lewis, Fasan, and Hyster groups, have expanded the field, producing ArMs for hydrogenation, C-C bond formation, oxygen insertion, and hydration reactions (61-64).

Figure 1.2. Artificial Metalloenzymes. **A.** The general scheme for linking the metal cofactor into the protein scaffold. **B.** The first biotin-based metal cofactor synthesized by Whitesides *et al.* and the reaction it catalyzed. **C.** The next generation of rhodium-biotin metallocofactors used for hydrogenation. **D.** Rhodium-biotin cofactors for carbon-carbon bond formation.



Initially, metal swapping was mainly utilized to study mechanistic details of metalloenzymes, but there are a few cases where native metals were replaced for the sole purpose of either increasing the catalytic efficiency or changing the chemistry of the enzyme (65-68). For instance, when the zinc of carbonic anhydrase is replaced by cobalt the native activity decreases, but there is a concurrent increase in esterase activity (65). Similarly, heme proteins are known for their C-H functionalization and for C-O bond formations, but not for C-C bond formation. Replacement of the native iron in myoglobin with a number of different transition metals, such as, cobalt, copper, manganese, rhodium, iridium, ruthenium, and silver, resulted in a catalytically active ArM that could perform carbene insertion to both β-substituted vinylarenes and unactivated

aliphatic α-olefins (69).

The initial ArM methodology has been greatly expanded on since the revitalization of the field, mainly due to the advancement of directed evolution as a technique for protein engineering. Such engineering strategies allowed for iterative rounds of ArM optimization through mutations to the primary and secondary coordination sphere of the metal cofactor (54). To date, 83 different protein scaffolds have been subjected to the established anchoring methods in hopes of creating ArMs for nonbiological chemistries. A summary table detailing these 83 scaffolds can be found in the 2018 review by Ward *et al.* (54). Described below are some of the more pertinent efforts.

Asymmetric hydrogenation of carbon-carbon double bonds is of high value, as such many traditional organic catalysts are being continuously designed in order to perform this feat. The development of ArMs has furthered such efforts since the incorporation of metal cofactors within protein scaffolds often provides the chiral environment needed to selectively generate desired hydrogenation products. The most common metal used for hydrogenation is rhodium, but ArMs can also contain copper, palladium, and platinum in combination with a variety of ligands, such as biotin, pincer complexes, dipyridine, bisphosphine, and monodentate phosphite (55, 70-79). Of the numerous scaffolds available in nature, some of the more commonly used proteins include avidin, papain, and lipases. One of the more successful ArMs created for asymmetric hydrogenation was developed by the Ward group (56, 76, 80-84). They aimed to improve the ArM initially designed by Whiteside by first switching the protein scaffold from avidin to streptavidin as the lower isoelectric point of streptavidin would, in theory, have a higher affinity for the rhodium complex. A previously identified amino acid, serine 112, within the binding

pocket of streptavidin was determined to have the greatest effect on the catalyst, and thus became the target of site saturated mutagenesis. The biotin-rhodium complex was also mutated both in terms of the length of the linker between the metal and the biotin, as well as the chelating groups surrounding the rhodium (**Figure 1.2C**). The library was tested against two dehydroamino acid derivatives and they discovered that the rhodium-biotin library had more influence over catalytic activity than the streptavidin mutants did; however, S112 did control the enantioselectivity of the ArM and the group did achieve interconversion of enantiomers depending on the amino acid present. The combination of engineering to the protein and the cofactor together allowed for a library of ArMs with broad spectrum activities.

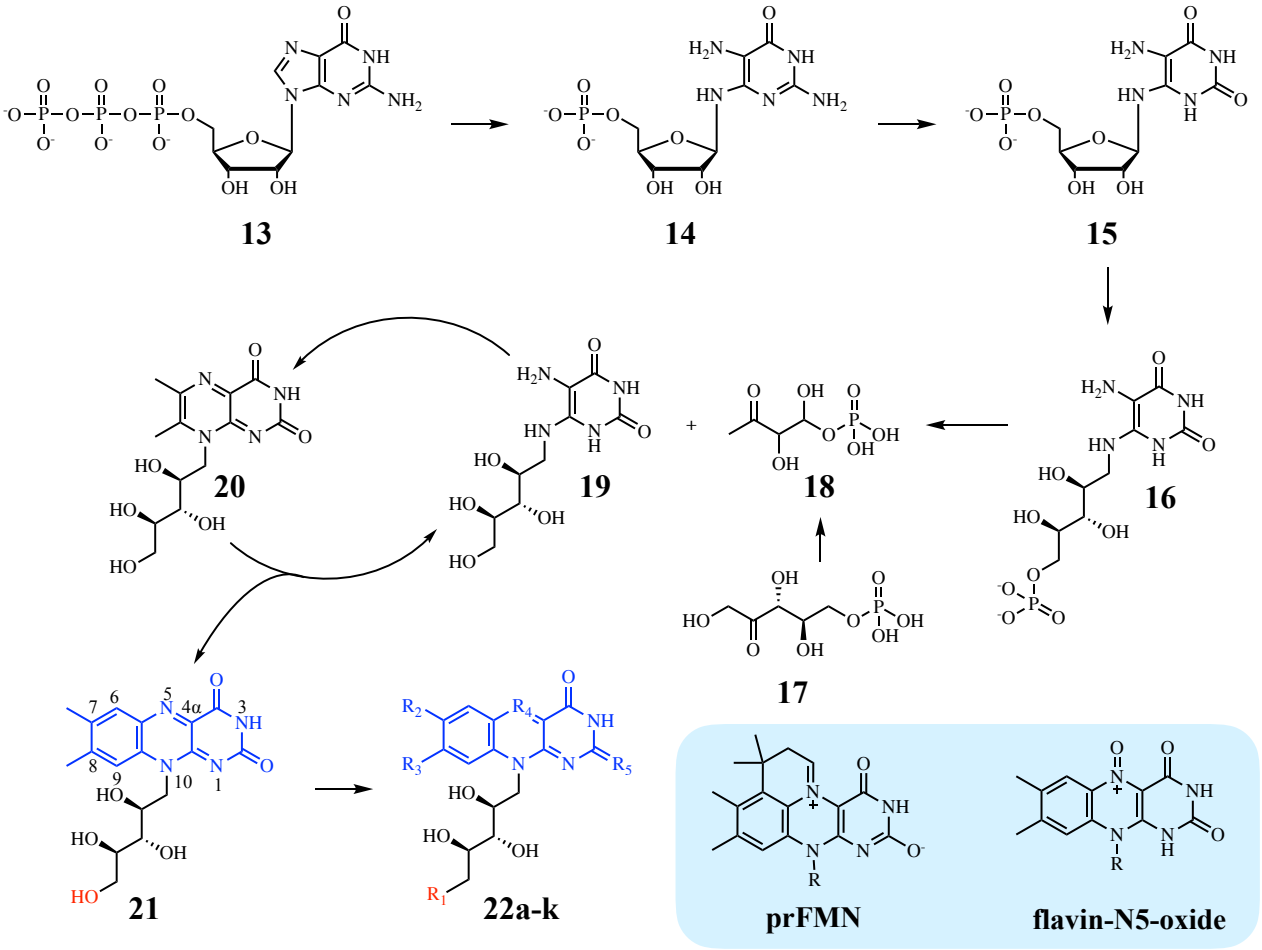
Selective carbon-carbon bond formation is another highly sought after process due to the high value products formed relevant to agrochemical, pharmaceutical, and polymer industries. ArMs have been developed for allylic alkylations, Suzuki cross-coupling, Heck reactions, C-H activation, olefin metathesis, cyclopropanation, polymerization, Diels-Alder reactions, Huisgen [3 + 2] cycloaddition, and Friedel-Crafts reactions (85-93). One notable example of protein and cofactor engineering working symbiotically was demonstrated when Rovis and Ward bound biotinylated [$RhCl_2(H_2O)$] within streptavidin to catalyze the enantioselective hydroarylation of methyl acrylate and pivaloyl-protected benzhydroxamic acid (**Figure 1.2D**) (94, 95). They observed a decreased turn over number (TON) of roughly half compared to free catalyst, but there was a gain in the regioisomeric ratio and enantiomeric excess values. Directed evolution of streptavidin led to the generation of mutants which regained some TON, but more impressively, exhibited further increases in regioisomeric ratio and enantiomeric excess values.

These examples only begin to highlight the efforts taken to create ArMs for nonbiological reactions. While the major drawback of these enzymes is low overall TONs, further cofactor and protein engineering has the potential to overcome this set back. Continued development of ArMs will help to bridge the gap between organic and biological chemistry, which could be combined to conquer a major dilemma: sustainable energy. Specifically, combating the currently inefficient conversion of solar energy.

III. Flavin Cofactors

With the numerous cofactors found in nature, the work of this dissertation is focused around one group of cofactors known as flavins. Riboflavin was first discovered in 1879 as the pigmentation in milk, but it was not until the 1930s that its structure was determined and assigned as vitamin B_2 by Peter Karrer and Richard Kuhn (96, 97). Riboflavin does not usually have a direct metabolic role, but it has been shown to induce disease resistance in plants and can be the apoptosis-inducing factor in mitochondrial protein (98-101). Most importantly, it is the precursor for the biosynthesis of flavin mononucleotide (FMN) and FAD. While riboflavin is an essential vitamin for animals and some bacteria, all plants, fungi, and most bacteria synthesize it *de novo* (96). **Figure 1.3,** shows the synthesis of riboflavin starting with GTP (**13**) and ribulose-5-phosphate (**17**). GTP cyclohydrolase II removes the carbon at the 8 position of the guanidine and the pyrophosphate from GTP to afford compound **14**. Next, there is deamination of the amino group at the 2 position followed by reduction of the ribose to form the ribityl tail in sequential steps by two separate enzymes (compounds **15** and **16**). The next step is the final dephosphorylation, for which the mechanism is unknown, yielding 5-amino-6-

Figure 1.3. Biosynthetic pathway of riboflavin followed by the downstream conversions to FMN and FAD. Compound **22** is a generic compound to showcase different flavin analogs. R groups correspond to different functional groups, as listed in the table. prFMN and flavin-N5-oxide are list separately in the blue inset due to their unique derivatization. They are shown with the ribityl tail represented by R.



22	Compound	R_1	R_2	R_3	R_4	R_5
a	FMN	PO_4^{2-}	$-CH_3$	$-CH_3$	N	O
b	FAD	adenine dinucleotide	$-CH_3$	$-CH_3$	N	O
c	F_0	-OH	—	-OH	-CH	O
d	F_{420}	oligoglutamate	—	-OH	-CH	O
e	Roseoflavin	-OH	$-CH_3$	dimethylamino	N	O
f	8-Chloroflavin	PO_4^{2-}/adenine dinucleotide	$-CH_3$	-Cl	N	O
g	8-Mercaptoflavin	PO_4^{2-}/adenine dinucleotide	$-CH_3$	-SH	N	O
h	2-Thioflavins	PO_4^{2-}/adenine dinucleotide	$-CH_3$	$-CH_3$	N	S
i	8-CyanoFMN	PO_4^{2-}	$-CH_3$	-CN	N	O
j	8-Formylflavin	PO_4^{2-}/adenine dinucleotide	$-CH_3$	-CHO	N	O
k	lampteroflavin	ribose	$-CH_3$	$-CH_3$	N	O

ribitylamin-2,4(1H,3H)-pyrimidinedione (**19**). This compound is then combined with 3,4-dihydroxy-2-butanone-4-phosphate (**18**), which is derived from **17**, to give 6,7-dimethyl-8-ribityllumazine (**20**) in a step mediated by lumazine synthase. In the final step, riboflavin synthase condenses two molecules of **20** to give riboflavin (**21**). From here riboflavin is then transformed to FMN and FAD (**22a** and **b**, respectively) either via a bifunctional riboflavin kinase/FAD synthetase, as in bacteria, or via separate monofunctional riboflavin kinase and FAD synthetase as in eukaryotic organisms.

The completed flavins (compounds **21** and **22a-k**, as shown in **Figure 1.3**) contain three main structural moieties: a terminal recognition group (red), the ribityl tail (black), and the isoalloxazine ring system (blue). The terminal recognition group aids in the binding of FMN and FAD to enzymes while also serving to distinguish the three cofactors from one another. Riboflavin contains a terminal hydroxyl group while FMN and FAD contain a terminal phosphate group or adenine dinucleotide, respectively. The ribityl tail is a 5 carbon chain consisting of 3 secondary hydroxyl groups, which also aid in anchoring the cofactor to the enzyme. The most important part of the cofactors lies in the isoalloxazine ring system, which gives the cofactors their diverse biological and non-biological catalytic activities. The ring system can exist in 3 separate states: oxidized, fully reduced (two electron), and semiquinone (single electron reduced). Due to the fact that it can exist in these three separate states, it therefore can be involved in both one and two electron transfer chemistries. Likewise, the redox potential of the cofactor is extremely tunable, spanning from -400 mV to +60 mV, depending on the enzyme environment surrounding the cofactor. FMN and FAD are the catalytically active

forms, with the exception of the Na^+-pumping NADH:quinine oxidoreductase from *Vibrio cholerae*, which utilizes riboflavin as its cofactor. (96, 102, 103).

Flavoenzymes

Enzymes utilizing either FMN or FAD for catalytic function, referred to as "flavoenzymes," are an extremely versatile class of enzymes in regards to both the types of substrates they accept and the reactions they perform. As stated above, the flavin grants them the expanded ability to perform both one and two electron chemistries compared to other cofactors which exclusively perform either one or two electron transfers (104). More flavoproteins contain FAD over FMN, but both cofactors can be either covalently or noncovalently bound within the active site. While covalent attachment of the cofactor is rare, it does increase the oxidative power (105). The flavin is usually linked either through the 8 position via a histidine or through the 6 position via a cysteine (numbering of flavins in **Figure 1.3**) (106, 107).

Redox active flavoenzymes can be divided into five classes (108). First, transhydrogenases, such as the ene-reductases, involve a two electron transfer from the flavin to the substrate. Second, dehydrogenase-oxides perform a two electron transfer from the substrate to the flavin, as seen in Acyl-CoA dehydrogenases (109). Third, dehydrogenase-monooxygenases where the reduced flavin is first oxidized by molecular oxygen, to form a reactive peroxo-intermediate, followed by oxygen insertion into the substrate, as seen in the Baeyer-Villiger monooxygenases (109, 110). Fourth, dehydrogenase-electron transferases where the flavin is reduced by a two electron transfer and subsequently reoxidized by two one electron transfers to the acceptor, as seen in cytochromes and iron-sulfur proteins (111). Finally, electron transfer

proteins, such as the butyryl-CoA dehydrogenase and flavodoxins, where the flavin is reduced and reoxidized in a single step as it transfers electrons between two redox active proteins (96, 108, 112).

Besides redox reactions, flavoenzymes are involved in light emission, light sensing in plants, and circadian rhythm regulation (113-118). Specifically, FMN is utilized in plant phototropins and chloroplast movement (119, 120). FMN-containing fluorescent proteins are of much interest due to the fact that they can fluoresce in both aerobic and anaerobic conditions, which give them an advantage over green fluorescent proteins when studying anaerobic systems (121-123). FMN is also an important precursor in the synthesis of cobalamin (124, 125). FAD is found in cryptochromes and blue-light using proteins (126, 127). Free flavins can also play a role as extracellular mediators for the reduction of Fe^{3+} and other cations (128-130). Over all, free flavins can act as electron shuttles, be involved in insoluble ion solubilization, and geochemical cycles, displaying some of the extreme versatility of flavins (96).

While FMN and FAD themselves represent a considerable amount of catalytic power and versatility, nature has developed a few riboflavin and FMN analogs which further expand the utility of flavoenzymes. Analogs of note are the 5-deaza flavins, which have a carbon at the 5 position instead of a nitrogen making the cofactors behave more like a nicotinamide cofactor due to the decrease in redox potential (96). Specifically, F_0 (**22c**) and F_{420} (**22d**) both possess 7,8-didemethyl-8-hydroxy-5-deazaflavin isoalloxazine moieties with the only difference between the two being in the terminal recognition group: **22c** has a terminal hydroxyl group while **22d** has an oligoglutamate (**Figure 1.3**). F_0 has been isolated from methanogenic archaea where it is involved in the hydride transfer during the reduction of carbon dioxide and acetate to methane

(131-135). F_0 is also involved in DNA photolysis in cyanobacteria and some eukaryotes (136-139). F_{420} similarly has versatile roles depending on the organism. For example, F_{420} has been found to play a role in the biosynthesis of tetracycline and lincomycin in certain strains of *Streptomycetes* (140-142) while *Mycobacterium* and *Nocardia* use F_{420} as the cofactor for glucose-6-phosphate dehydrogenase (143, 144). Additionally, F_{420} has relevant non-native functionality as an activator for antituberculosis drugs (145, 146).

Another well-studied riboflavin analog is roseoflavin (**22e**) from *Streptomyces davawensis*, which is of particular interest due to its native antimicrobial activity towards Gram-positive bacteria (147). To date, it is one of the few compounds that has shown activity against *Listeria monocytogenes* (148). Roseoflavin contains a dimethyl amino group at the 8 position instead of a methyl group. This substitution causes the isoalloxazine ring system to lose its oxidizing ability due to an intramolecular charge transfer from the dimethylamino group to the pteridine moiety, resulting in a more negative redox potential relative to its FMN and FAD counter parts (149). Roseoflavin is believed to to be an inhibitor to flavoenzymes by functioning as a competitive inhibitor for the native flavin cofactor (149, 150). Another analog, lampteroflavin (**22k**), has been identified as a light emitting chromophore for certain mushrooms (151). It is also known that certain plants secrete riboflavin and analogs during times of iron starvation (152).

Nature has also created catalytically relevant FMN and FAD analogs, for example the prenylated FMN (prFMN, see **Figure 1.3** and **1.4**) and the 8-formyl flavin (**22j**). The prFMN has been reported in conjunction with the two protein system, ubiquinone prenyltransferase and decarboxylase (UbiDX) (153). This complex is known for its oxidative decarboxylation of

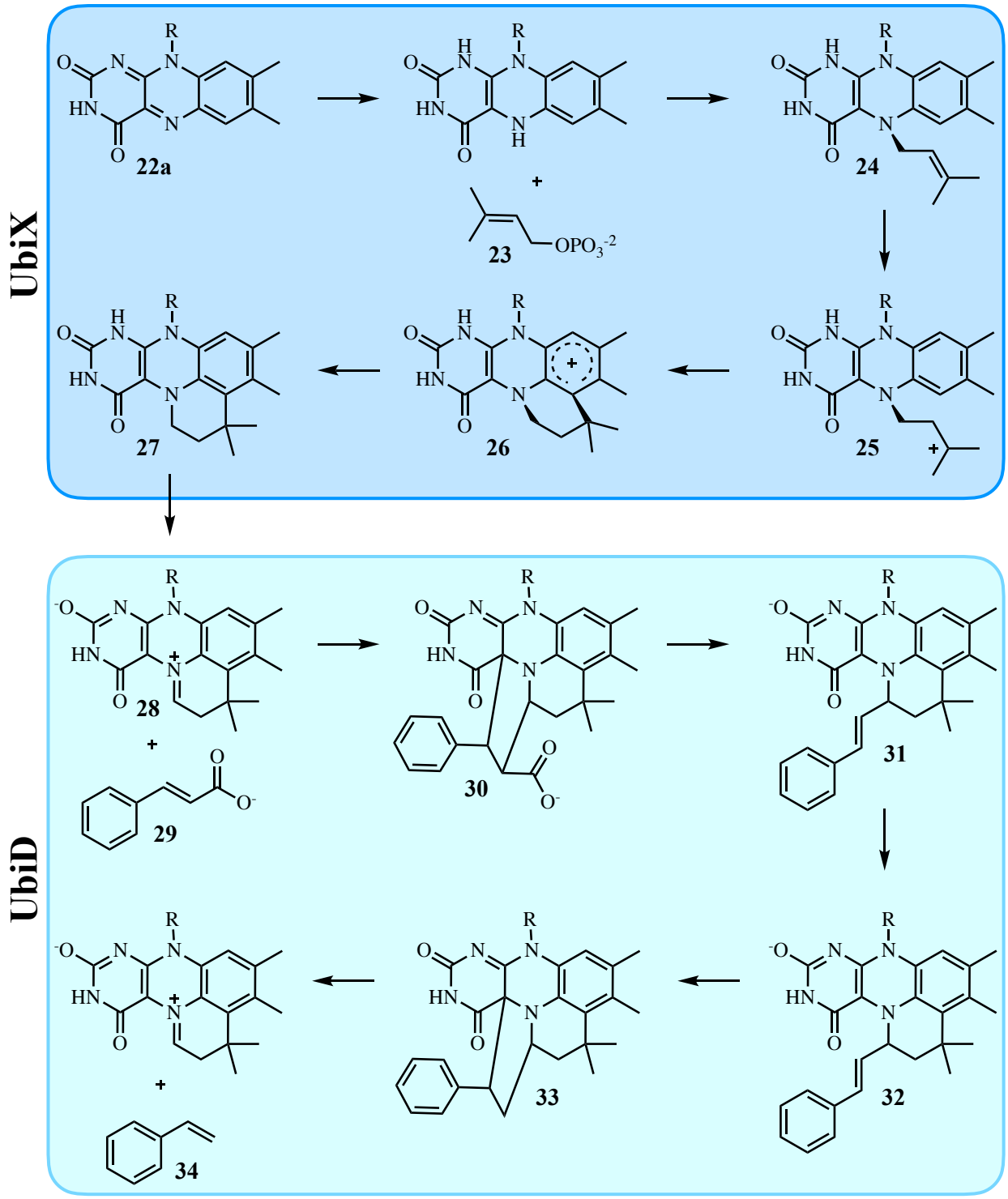


Figure 1.4. Synthesis and catalytic mechanism of prFMN. **Top**: Synthesis of prFMN by UbiX starting with oxidized FMN (**22a**). **Bottom**: Final maturation step of prFMN and the mechanism for decarboxylation of cinnamic acid (**29**) to produce styrene (**34**) by UbiD.

aromatic complexes. Prenyltransferases are known to be metal dependent, but the recently

discovered prenyltransferases (UbiX) from *Psuedomonas aeruginosa* and *Aspergillus niger* are not, and will only use dimethylallylmonophosphate (DMAP) as opposed to dimethylallyl-pyrophosphate (DMAPP) for prenylation (153-155). The analog is synthesized by UbiX, which binds and reduces FMN followed by combination with DMAP (**23)** to generate a fourth non aromatic ring (154). The mechanism, shown in **Figure 1.4** top, proceeds via an unusual sp^3 N5-prenyl intermediate, which forms an N5-C1' dimethyl adduct (**24** and **25**) prior to the formation of the C6-C3' bond (**26**) (154). The maturation of the active cofactor (**28**) is completed in the second enzyme of the system, the decarboxylase (UbiD). This maturation forms the N5-prenyl C1' iminium group, generating an azomethine ylide-like cofactor (154). The prFMN is said to have similar properties to ThDP or pyridoxal phosphate (PLP) and acts as an electron sink to allow for the decarboxylation of α/β unsaturated aromatic and acrylic acids (153, 156, 157) via a 1,3-dipolar cycloaddition. For example, **Figure 1.4 (**bottom) depicts the decarboxylation of cinnamic acid (**29**) to produce styrene (**34**) (158). Therefore, discovery of the prFMN further expands the catalytic toolbox of flavins.

Thus far, prenylation of a flavin has only been observed in the form of FMN. Conversely, the 8-formyl derivatization has been found as a modification to both FMN and FAD. The first account of 8-formyl-FAD was found in *Chromatium* cytochrome C_{552}; the formylation was achieved through a covalent linkage of a cysteine through the 8 position of the FAD molecule (159). Subsequent studies have reported the presence of noncovalently bound 8-formyl-FMN and FAD in glucose-methanol-choline oxidoreductases from *Aspergillus oryzae* RIB40, L-lactate oxidase from *Aerococcus viridans*, and fungal formate oxidase (160-162). Biophysical characterization of either the 8-formyl-FMN or FAD revealed a more positive redox potential of

about 120 mV compared to the native FMN or FAD (159, 161). The increased redox potential leads to a higher turnover rate in the reductive half reaction, and in the case of the formate oxidase, a 10-fold increase of oxidase activity was observed once FAD was converted to 8-formyl-FAD within the active site (162).

While not an analog in the traditional sense, the flavin-N5-oxide (**Figure 1.3** inset) was first discovered as a new catalytic motif during the 1,3-diketone oxidation in enterocin biosynthesis (163-166). Two other instances have subsequently been reported: in the first example, the motif is involved in dibenzothiophene sulfone oxidation, and in the other it is responsible for the oxidative amide cleavage of uracil (167, 168). In all three cases the flavin hydroperoxide is formed at the N5 instead of the normal C4α, as seen in the Baeyer-Villiger monooxygenases (169). The nucleophilicity of the terminal oxygen of the flavin-N5-oxide is more reduced as compared to the terminal oxygen of the C4α-peroxide, which could lead to better selectivity during Baeyer-Villiger oxidations (169). As more natural examples of the flavin-N5-oxide emerge, its potential for cofactor engineering is becoming more apparent, especially when considering advantages over the traditional C4α-peroxide.

In addition to natural flavin derivatives countless synthetic flavin analogs have been synthesized to probe enzyme active sites for qualities such as solvent accessibility, polarity, reaction stereochemistry, and dynamic behavior (170-172). For example, enzymes reconstituted with either 8-chloroflavin (**22f**) or 8-mercaptoflavin (**22g**), which react with small molecule probes such as sodium sulfide or iodoacetamide, have been used to determine substrate and solvent accessibility into the active site (173). 8-mercaptoflavin and the structurally related 2-thioflavins (**22h**) have also been used to probe active site polarity due to the large spectral

changes that occur upon analog binding (174, 175). Mechanistically, deazaflavin analogs have been utilized to assign stereochemistry of the native hydride transfer reaction for a number of flavoenzymes, which was not possible with the native cofactor (176). Additionally, deazaflavins in general have also garnered significant interest due to their reduced redox potential relative to native flavins, which renders them less susceptible to molecular oxygen (177, 178). Lastly, synthetic roseoflavin analogs have been shown to possess potent antibiotic effects and function as inhibitors to native riboflavin synthesis *in vivo* (150). These examples and others collectively illustrate that while native riboflavin, FMN, FAD, and natural cofactor analogs are both prolific and functionally versatile, synthetic flavin analogs are equally worthy of study given their immense potential for the generation of new therapeutics, the elucidation of reaction mechanisms, and the capacity to expand upon the scope of possible chemical transformations in flavoenzymes.

IV. Old Yellow Enzymes

Old Yellow Enzymes (OYEs) are a flavin-dependent family of ene-reductases with activity primarily towards α/β unsaturated ketones, aldehydes, and nitro groups, and characteristically possess high regio-, stereo-, and enatioselectivity (179). These attractive features make them industrially relevant alternatives to traditional organic catalysts, specifically in the realms of biotechnology, pharmaceuticals, fine chemicals and agrochemistry (180-182). The first OYE discovered, named Old Yellow Enzyme 1 (OYE1) after its characteristically yellow color, was isolated from *Sacchromyces pastorianus* in 1932 (183). Subsequently, there have been a number of OYE1 homologues identified across all kingdoms of life, indicating their

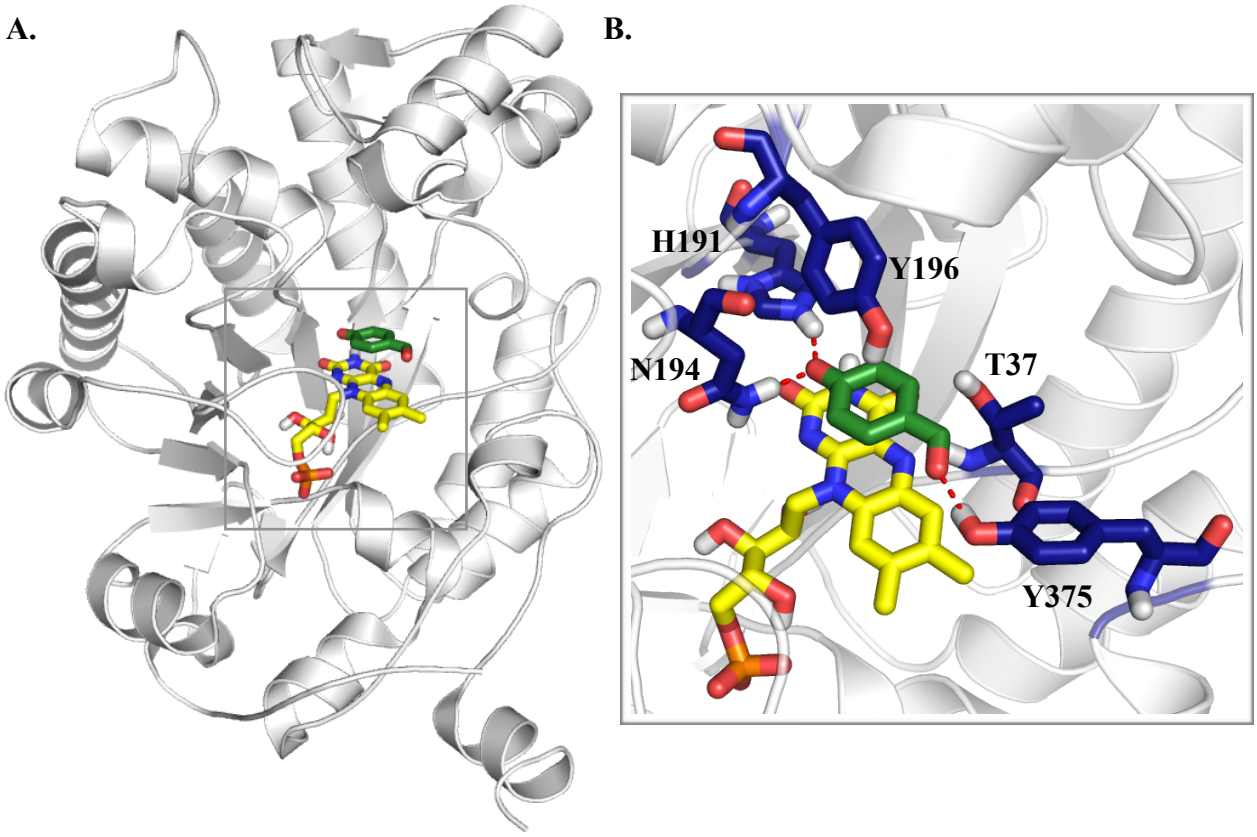


Figure 1.5. **A.** Crystal structure of OYE1 (PDB 1OYB) with FMN (yellow) and *p*-hydroxy-benzaldehyde (green) bound within the active site. **B.** Zoom in of OYE1 active site with relevant active site residues labeled and highlighted in blue. Polar contacts with substrate and active site residues are depicted by red dashes.

ubiquitousness in nature (184-187). Due to their high stabilities and broad substrate scopes, multiple OYE family members have long been the subjects of mechanistic studies and protein engineering efforts (180, 188-194).

Dependency on FMN is presumably conserved throughout the entirety of the superfamily as all OYEs characterized to date contain FMN. The family also shares important structural features. The majority of family members are monomeric and contain an TIM-barrel core domain, which is responsible for binding FMN (**Figure 1.5A**) (195, 196). Within the active site there is high conservation of certain residues, specifically at positions 37, 191, 194, and 196 (numbering based on OYE1 and depicted in **Figure 1.5B**) (197-200). Typically, there is a

threonine at position 37, which is responsible for directly interacting with the N5 of FMN and has been found to modulate the redox potential of the flavin. Another important residue is the tyrosine at position 196, which was originally thought to be the final proton donor for the reduced substrate; however, there is much debate regarding its actual role since mutagenesis studies have shown that it is not necessary for product formation (201-203). The most crucial amino acids for catalysis lie at positions 191 and 194, which directly interact with the oxygen of the substrate's electron withdrawing group to simultaneously coordinate the substrate over the *si*-face of the flavin and activate the β carbon to permit direct hydride transfer (**Figure 1.6**). OYE1 contains a histidine at 191 and an asparagine at 194 while other homologues contain histidines at both positions. Based on this difference in amino acid preference at positions 191 and 194, the superfamily was divided into subgroups, traditionally referred to as the "classical OYEs" and the "YqjM-like OYEs." Despite high structural conversation, there are still slight mechanistic variations and substrate preferences between the two subgroups (203). A more detailed examination of the OYE superfamily groupings and distinctions will be discussed in chapter 4.

Catalytically, OYEs function through a ping pong mechanism (204, 205). The catalytic cycle of the OYEs can be seen in **Figure 1.6**. Briefly, NADPH (in some cases NADH) enters the active site to reduce the oxidized FMN. Following the departure of $NAD(P)^+$, the substrate binds and the reduced flavin can directly transfer a hydride from the N5 position of FMN to the β carbon of the substrate while either tyrosine 196 or a solvent molecule pronates the α carbon in a *trans*-fashion. Pioneering work was performed by Massey in order to elucidate the above mechanism and pinpoint key catalytic residues, providing the ground work for all of the subsequent protein engineering performed on OYE1. Such engineering efforts have been

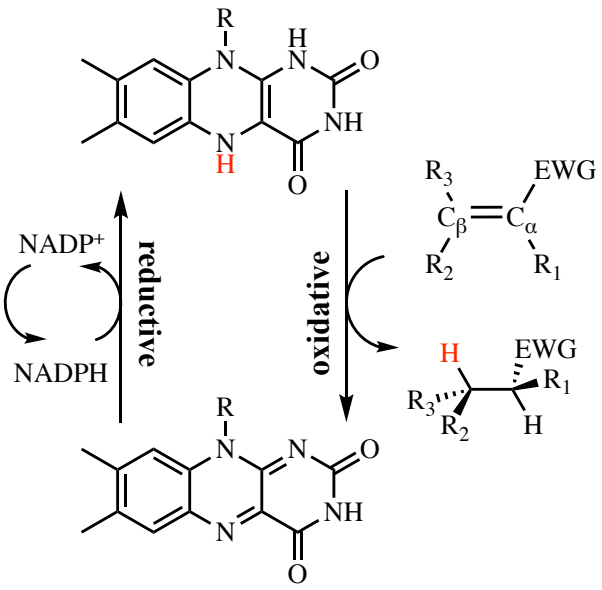


Figure 1.6. The catalytic cycle of OYEs. The oxidative half reaction entails the trans hydrogenation of the activated alkene while the reductive half reaction includes cofactor regeneration by NADPH. The hydride (red) is transferred from the FMN to the Cβ with accompanying protonation at Cα.

undertaken to modulate the enzyme's catalytic properties, such as increase turnover rate, reverse enantioselectivity, increase stability, and broaden substrate scope (179, 181, 200, 206-208).

In addition to traditional protein engineering techniques, cofactor engineering of flavoenzymes has the potential to provide insight into additional mechanistic details, or even fundamentally alter the chemistry the subject enzyme can perform (refer to section III of this chapter). However, cofactor engineering performed in regards to OYE1 has not been prolific. One of the first accounts of cofactor engineering was performed by Massey in order to elucidate the origin of unknown absorption bands observed at higher wavelengths (600 to 700 nm) (209). The absorption at higher wavelengths was proposed to come from a charge transfer complex originating from π-π stacking between the oxidized flavin and the olefinic bond of the substrate, but this assertion was disputed (209-211). To confirm their postulation, Massey *et al.* replaced the native FMN of OYE1 with a number of different FMN analogs possessing varied redox potentials in order to study the charge transfer complex that formed between the FMN and substrate (209, 211). Specifically, they believed there should be a correlation between the redox potential of the enzyme-bound flavin and the energy of the resulting charge transfer complex (211). Experimental data were consistent

with predicted trends that the redox potential of the charge transfer acceptor (FMN or analog) is an indicator of the electron affinity in the corresponding charge transfer complex (209, 211).

The next account of cofactor engineering of OYE1 was also done by Massey. However, rather than of trying to gain insight into the enzymatic mechanism, engineering was conducted in the hopes of changing the type of catalysis performed by OYE1. Massey *et al.* synthesized 8-cyano-FMN (**22i**) as substitutions on the 8 position of flavins have been shown to have the greatest influence on the redox potential of the cofactor (212-214). More specifically, increasing electron withdrawing character of the functional group inserted at the 8 position translates into more positive redox potential of the flavin (215). Increasing the redox potential of the flavin has the possibility to convert OYE1 from a reductase to a desaturase, which is precisely what was observed upon reconstitution of OYE1 with **22i** (215).

With the exception of the two examples detailed above, little work has been done in regards to the cofactor engineering of OYE1. Chapter 3 provides the first example of cofactor engineering of OYE1 utilizing the PURExpress i*n vitro* transcriptional/translation (PURE system) from New England BioLabs (Ipswich, MA) to determine which portions of the flavin may be essential in aiding the proper folding of the enzyme.

Beyond OYE1, the OYE superfamily

Despite the multiple mechanistic and protein engineering works performed previously for OYE1, this single enzyme represents an infinitesimal piece of a much larger enzyme superfamily. With genome sequence data becoming more easily accessible, the annotation of new OYEs has sky rocketed. To date, the number of sequences which have been annotated in the gene

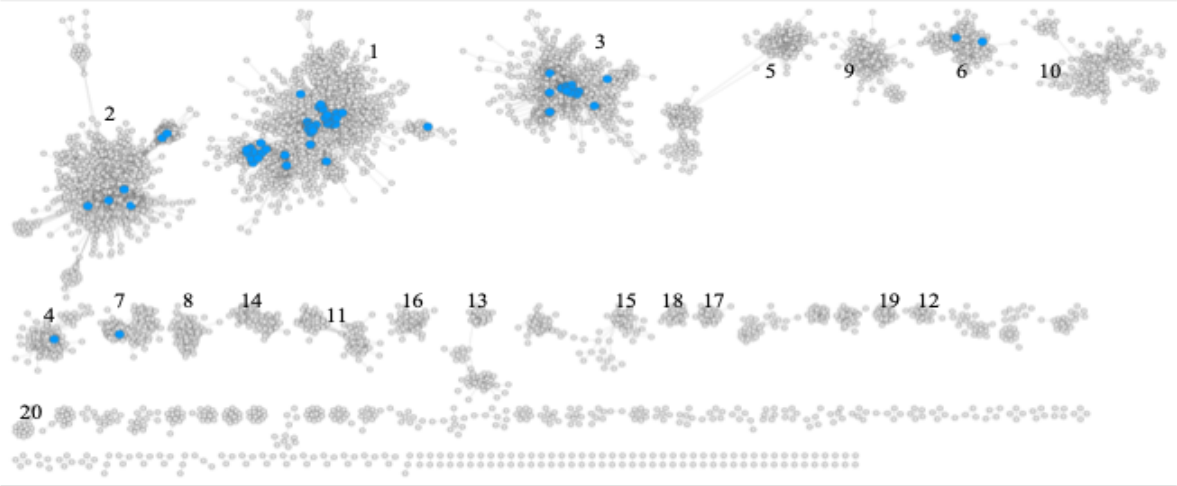


Figure 1.7. Sequence Similarity Network (SSN) of the OYE Superfamily. The 20 most populated clusters are numbered. Node (circles) represent between 1 and 641 sequences with greater than 50% identity. Edges (connecting lines) represent E-value thresholds of 1 x 10^{-85}. Blue nodes represent the current OYEs characterized in the literature. More about the OYE SSN can be found in chapter 4.

databank as members of the OYE superfamily is over 73,000, **Figure 1.7**. Of the 73,000 identified, only 136 have been characterized in the literature (blue nodes in **Figure 1.7** and for the list of sequences see the Supplemental of Chapter 4). Some of the most notable ones are YqjM, PETN reductase, OPR 1 and 3, and NCR (180).

YqjM from *Bacillus subtilis* and was the first Gram-negative bacterial OYE isolated and biocatalytically characterized (216). It shares a 33% sequence identity with OYE1 and natively prefers NADPH over NADH (189, 216). It also has a similar substrate profile to OYE1 but with the added bonus of being able to reduce trinitrotoluene (TNT) (189, 216, 217). YqjM differs from the rest of the family, however, in that its catalytically active form is a homo tetramer while most other family members exist as monomers or dimers. Additionally, crystallization of YqjM revealed key differences in the catalytically relevant amino acids, which caused a distinct split of the superfamily into the “classical OYEs” and “YqjM-like OYEs” (189). YqjM-like OYEs have two histidines at positions 191 and 194, and a cysteine at position 37 in place of the threonine

present in OYE1 (189). The other notable OYE family members, PETN reductase , OPR 1 and 3, and NCR, all fall into the classical OYE grouping. PETN reductase from *Enterobactor cloacae* was isolated from explosive contaminated soil and is of special interest because of the bacteria's ability to use pentaerythritol tetranitrate (PETN) and other environmentally toxic explosives as its sole nitrogen source (188). Accordingly, it is currently being investigated for its bioremediation potential. OPR 1 and 3 from *Solanum lycopersicum* (tomato) shares a very similar substrate profile to YqjM, but demonstrated unusually high oxygen tolerance (190, 192, 217). Finally, NCR from *Zymomonas mobilis* is of note due to its activity towards isomers of citral and its ability to perform the double reduction of an alkyne to an alkane (191, 218). **Table 1.1** provides a list of substrates reduced by OYE1 and the OYE family members described above in greater detail.

Table 1.1. Substrate Scope of Select OYE Family Members. Adapted from (179).

Enzyme	Organism	Substrate
OYE1	*S. pastorianus*	α,β-unsaturated aldehydes, ketones, imides, nitroalkenes, esters, cyclic and acyclic enones, explosives (PETN and TNT), terpenoids, N-ethyl maleimide and codeinone.
YqjM	*B. subtilis*	α,β-unsaturated aldehydes, ketones, maleimides and nitroalkenes, dicarboxylic acids and dimethyl esters and explosives (nitroglycerin and TNT).
PETN Reductase	*E. cloacae*	α,β-disubstituted nitroalkenes, substituted cyclic α,β-unsaturated cyclic and acyclic aldehydes and ketones, terpenoids, aromatic and aliphatic explosives (PETN, N-nitramines, and TNT) and codeinone.
OPR 1 and 3	*S. lycopersicum*	α,β-unsaturated aldehydes, ketones, maleimides and nitroalkenes, dicarboxylates and di- methyl esters and 12-oxophytodienoic acid (OPDA).
NCR	*Z. mobilis*	α,β-unsaturated cyclic and acyclic aldehydes, ketones, imides, carboxylic acids, esters, and terpe- noids.

While a number of industrially useful biocatalysts from the OYE family have been reported, this number accounts for less than 1% of the entire family. Additionally, as shown in the sequence similarity network (SSN) of **Figure 1.7**, the OYEs that have been identified all come from the same clusters, leaving the vast majority of sequence space untapped. Given the vast array of unexplored sequences within the family, the potential to find better catalysts capable of performing novel chemistries or possessing unique substrate scopes is strong. Chapter 4 entails the exploration of the OYE superfamily, essentially doubling the number of characterized OYEs currently known.

V. Cell-Free Protein Expression

Traditional methods of protein engineering involve manipulation of genetic DNA, followed by transformation into a host cell, such as *E. coli*, and then heterologous expression and purification. While this method works, cell-free protein expression systems provide advantageous opportunities, such as allowing the expression of toxic proteins, and limiting both aggregation and degradation of desired products (219, 220). The first cell-free protein expression system utilized cell extracts to identify the biosynthetic pathway of β-galactosidase (221). Expanding upon this previous work, cell free protein synthesis employed either cell extracts in a reaction vessel with a permeable membrane, which allows for constant removal of protein product in conjunction with cofactor or substrate replenishment (222), or the purification and subsequent mixing of each component necessary for transcription and translation (223). While the first method allows for higher protein yields, the second method has more desirable advantages over the utilization of cell lysates. For example, cell lysate contains components from

all of the biosynthetic pathways found within the cell, which could negatively impact the synthesis of the protein of interest (224). Another issue with using cell lysate is the inherent presence of proteases that can prematurely degrade the desired proteins (222, 225). These drawbacks inspired the developed of the second method which has been named, “Protein synthesis Using Recombinant Elements“ (PURE) (223). The PURE system allows for complete control over what components go into the system (223, 226-228). It is important to note some of the drawbacks of the PURE system, the biggest being the cost, which can limit the accessibility to labs (220). However, this drawback has not prevented the PURE system from being at the center of many protein engineering efforts (181, 208, 229-233).

In regards to cofactor engineering with cell-free protein expression systems, similar drawbacks have emerged. The biggest shortcomings arise from the utilization of cell lysate. As described above, the cell lysate is enveloped in a permeable membrane, which allows for the removal of small molecules, including cofactors, found natively in the cell (222). However, this is not a perfect method and trace amounts of cofactors are still present during protein synthesis, prohibiting the complete synthesis of apoprotein (234). While the PURE system has not been utilized previously for cofactor engineering, it has clear advantages over the cell lysate method as the PURE system allows for complete control of both the species and quantity of cofactor added to each reaction mix. Accordingly, the PURE system was chosen in order to discern which parts of FMN are necessary for productive folding of OYE1, as described in chapter 4.

VI. Aims and Scope of the Dissertation

The overall theme of this dissertation focuses on cofactor engineering of FMN in order to elucidate the cofactor dependence in the productive folding of OYE1, while also exploring the versatility of the cofactor within the family OYEs.

Chapter 2, entails the engineering of FAD synthetase from *Corynebacterium ammoniagenes,* the enzyme responsible for converting riboflavin to FMN and then to FAD in nature. Engineering of FAD synthetase serves two purposes: first, it allows for a more streamlined process to synthesize FMN without first having to convert riboflavin all the way to FAD, then back to FMN. Second, FAD synthetase is natively promiscuous and can be used as a tool for the selective phosphorylation of FMN analogs. The use of these analogs will be described in chapter 3.

Chapter 3 includes the synthesis of novel FMN folding analogs, binding studies of the analogs to OYE1, and activity assays with the PURE system to determine which structural elements of the FMN cofactor are necessary for proper folding of OYE1.

Finally, chapter 4 focuses solely on OYE1 and its family of ene-reductases. Specifically, chapter 4 explores the whole OYE superfamily to showcase the versatility of the FMN cofactor in different, but related, OYEs.

VII. References

1. Walsh, C., Enabling the chemistry of life. Nature, 2001. **409**(6817): p. 226.
2. Jaenicke, L., Centenary of the award of a nobel prize to eduard buchner, the father of biochemistry in a test tube and thus of experimental molecular bioscience. Angewandte Chemie International Edition, 2007. **46**(36): p. 6776.
3. Bornscheuer, U.T., et al., Engineering the third wave of biocatalysis. Nature, 2012. **485**: p. 185.
4. Itoh, T. and U. Hanefeld, Enzyme catalysis in organic synthesis. Green Chemistry, 2017. **19**(2): p. 331.
5. Nomenclature committee of the international union of biochemistry and molecular biology, in Enzyme nomenclature. 1992, Academic Press: San Diego. p. ii.
6. Lalonde, J. and A. Margolin, Enzyme catalysis in organic synthesis: A comprehensive handbook, second edition. 2008. p. 163.
7. Lütz, S., Biocatalysts and enzyme technology. 2nd edition. By klaus buchholz, volker kasche and uwe theo bornscheuer. Angewandte Chemie International Edition, 2013. **52**(28): p. 7073.
8. Protein engineering: Methods and applications, in Protein therapeutics. p. 189.
9. Estell, D.A., T.P. Graycar, and J.A. Wells, Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. Journal of Biological Chemistry, 1985. **260**(11): p. 6518.
10. Clark, D.P. and N.J. Pazdernik, *Chapter 11 - protein engineering*, in *Biotechnology (second edition)*, D.P. Clark and N.J. Pazdernik, Editors. 2016, Academic Cell: Boston. p. 365.
11. Dhanjal, J.K., et al., Computational protein engineering approaches for effective design of new molecules, in Encyclopedia of bioinformatics and computational biology, S. Ranganathan, et al., Editors. 2019, Academic Press: Oxford. p. 631.
12. Bender, D.A. and P.A. Mayes, The citric acid cycle: The central pathway of carbohydrate, lipid & amino acid metabolism, in Harper's illustrated biochemistry, 30e, V.W. Rodwell, et al., Editors. 2016, McGraw-Hill Education: New York, NY.

13. Patel, M.S., et al., The pyruvate dehydrogenase complexes: Structure-based function and regulation. Journal of Biological Chemistry, 2014. **289**(24): p. 16615.

14. Meyer, D., et al., Observation of a stable carbene at the active site of a thiamin enzyme. Nature Chemical Biology, 2013. **9**: p. 488.

15. Breslow, R., On the mechanism of thiamine action. Iv.1 evidence from studies on model systems. Journal of the American Chemical Society, 1958. **80**(14): p. 3719.

16. Richter, M., Functional diversity of organic molecule enzyme cofactors. Natural Product Reports, 2013. **30**(10): p. 1324.

17. Zhao, J. and C.-J. Zhong, A review on research progress of transketolase. Neuroscience Bulletin, 2009. **25**(2): p. 94.

18. Bunik, V.I., A. Tylicki, and N.V. Lukashev, Thiamin diphosphate-dependent enzymes: From enzymology to metabolic regulation, drug design and disease models. The FEBS Journal, 2013. **280**(24): p. 6412.

19. Wiegand, G. and S.J. Remington, *Citrate synthase: Structure, control, and mechanism.* Annual Review of Biophysics and Biophysical Chemistry, 1986. **15**(1): p. 97.

20. Johnson, D.C., et al., Structure, function, and formation of biological iron-sulfur clusters. Annual Review of Biochemistry, 2005. **74**(1): p. 247.

21. Noodleman, L. and D.A. Case, Density-functional theory of spin polarization and spin coupling in iron—sulfur clusters, in Advances in inorganic chemistry, R. Cammack, Editor. 1992, Academic Press. p. 423.

22. Glaser, T., et al., Ligand k-edge x-ray absorption spectroscopy: A direct probe of ligand−metal covalency. Accounts of Chemical Research, 2000. **33**(12): p. 859.

23. Lloyd, S.J., et al., The mechanism of aconitase: 1.8 a resolution crystal structure of the s642a:Citrate complex. Protein Sci, 1999. **8**(12): p. 2655.

24. Jarrett, J.T., The generation of 5'-deoxyadenosyl radicals by adenosylmethionine-dependent radical enzymes. Curr Opin Chem Biol, 2003. **7**(2): p. 174.

25. Zhang, Q., et al., Characterization of nocl involved in thiopeptide nocathiacin i biosynthesis: A [4fe-4s] cluster and the catalysis of a radical s-adenosylmethionine enzyme. Journal of Biological Chemistry, 2011. **286**(24): p. 21287.

26. Gorodetsky, A.A., et al., DNA binding shifts the redox potential of the transcription factor soxr. Proc Natl Acad Sci U S A, 2008. **105**(10): p. 3684.

27. Dai, S., et al., Redox signaling in chloroplasts: Cleavage of disulfides by an iron-sulfur cluster. Science, 2000. **287**(5453): p. 655.

28. Duin, E.C., et al., Heterodisulfide reductase from methanothermobacter marburgensis contains an active-site [4fe–4s] cluster that is directly involved in mediating heterodisulfide reduction. FEBS Letters, 2002. **512**(1-3): p. 263.

29. Walters, E.M. and M.K. Johnson, Ferredoxin:Thioredoxin reductase: Disulfide reduction catalyzed via novel site-specific [4fe–4s] cluster chemistry. Photosynthesis Research, 2004. **79**(3): p. 249.

30. Ugulava, N.B., B.R. Gibney, and J.T. Jarrett, Biotin synthase contains two distinct iron−sulfur cluster binding sites: Chemical and spectroelectrochemical analysis of iron−sulfur cluster interconversions. Biochemistry, 2001. **40**(28): p. 8343.

31. Berkovitch, F., et al., Crystal structure of biotin synthase, an s-adenosylmethionine-dependent radical enzyme. Science, 2004. **303**(5654): p. 76.

32. Jameson, G.N.L., et al., Role of the [2fe−2s] cluster in recombinant escherichia coli biotin synthase. Biochemistry, 2004. **43**(7): p. 2022.

33. Siebert, G., M. Carsiotis, and G.W.E. Plaut, The enzymatic properties of isocitric dehydrogenase. Journal of Biological Chemistry, 1957. **226**(2): p. 977.

34. Fedøy, A.-E., et al., Structural and functional properties of isocitrate dehydrogenase from the psychrophilic bacterium desulfotalea psychrophila reveal a cold-active enzyme with an unusual high thermal stability. Journal of Molecular Biology, 2007. **372**(1): p. 130.

35. Cherbavaz, D.B., et al., Active site water molecules revealed in the 2.1 å resolution structure of a site-directed mutant of isocitrate dehydrogenase11edited by i. A. Wilson. Journal of Molecular Biology, 2000. **295**(3): p. 377.

36. Waldron, K.J., et al., Metalloproteins and metal sensing. Nature, 2009. **460**: p. 823.

37. Andreini, C., et al., Metal ions in biological catalysis: From enzyme databases to general principles. JBIC Journal of Biological Inorganic Chemistry, 2008. **13**(8): p. 1205.

38. Hemschemeier, A. and T. Happe, The plasticity of redox cofactors: From metalloenzymes to redox-active DNA. Nature Reviews Chemistry, 2018. **2**(9): p. 231.

39. Valdez, C.E., et al., Mysteries of metals in metalloenzymes. Accounts of Chemical Research, 2014. **47**(10): p. 3110.

40. Guldberg, H.C. and C.A. Marsden, Catechol-o-methyl transferase: Pharmacological aspects and physiological role. Pharmacological Reviews, 1975. **27**(2): p. 135.

41. Männistö, P.T. and S. Kaakkola, Catechol-o-methyltransferase (comt): Biochemistry, molecular biology, pharmacology, and clinical efficacy of the new selective comt inhibitors. Pharmacological Reviews, 1999. **51**(4): p. 593.

42. McLain, A.L., P.A. Szweda, and L.I. Szweda, A-ketoglutarate dehydrogenase: A mitochondrial redox sensor. Free Radic Res, 2011. **45**(1): p. 29.

43. Joyce, M.A., et al., Adp-binding site of escherichia coli succinyl-coa synthetase revealed by x-ray crystallography. Biochemistry, 2000. **39**(1): p. 17.

44. Fraser, M.E., et al., Two glutamate residues, glu 208α and glu 197β, are crucial for phosphorylation and dephosphorylation of the active-site histidine residue in succinyl-coa synthetase. Biochemistry, 2002. **41**(2): p. 537.

45. Rutter, J., D.R. Winge, and J.D. Schiffman, Succinate dehydrogenase - assembly, regulation and role in human disease. Mitochondrion, 2010. **10**(4): p. 393.

46. Swiezewska, E., Ubiquinone and plastoquinone metabolism in plants, in Methods in enzymology. 2004, Academic Press. p. 124.

47. Leshets, M., et al., Fumarase: From the tca cycle to DNA damage response and tumor suppression. Frontiers in Molecular Biosciences, 2018. **5**(68).

48. Goward, C.R. and D.J. Nicholls, Malate dehydrogenase: A model for structure, evolution, and catalysis. Protein Sci, 1994. **3**(10): p. 1883.

49. Ying, W., Nad+/nadh and nadp+/nadph in cellular functions and cell death: Regulation and biological consequences. Antioxidants & Redox Signaling, 2007. **10**(2): p. 179.

50. Berger, F., M.a.H. Ramírez-Hernández, and M. Ziegler, The new life of a centenarian: Signalling functions of nad(p). Trends in Biochemical Sciences, 2004. **29**(3): p. 111.

51. Belenky, P., K.L. Bogan, and C. Brenner, Nad⁺ metabolism in health and disease. Trends in Biochemical Sciences, 2007. **32**(1): p. 12.

52. Pollak, N., C. Dölle, and M. Ziegler, The power to reduce: Pyridine nucleotides--small molecules with a multitude of functions. Biochem J, 2007. **402**(2): p. 205.

53. Sellés Vidal, L., et al., Review of nad(p)h-dependent oxidoreductases: Properties, engineering and application. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2018. **1866**(2): p. 327.

54. Schwizer, F., et al., Artificial metalloenzymes: Reaction scope and optimization strategies. Chemical Reviews, 2018. **118**(1): p. 142.

55. Wilson, M.E. and G.M. Whitesides, Conversion of a protein to a homogeneous asymmetric hydrogenation catalyst by site-specific modification with a diphosphinerhodium(i) moiety. Journal of the American Chemical Society, 1978. **100**(1): p. 306.

56. Collot, J., et al., Artificial metalloenzymes for enantioselective catalysis based on biotin−avidin. Journal of the American Chemical Society, 2003. **125**(30): p. 9030.

57. Thomas, C.M. and T.R. Ward, Artificial metalloenzymes: Proteins as hosts for enantioselective catalysis. Chemical Society Reviews, 2005. **34**(4): p. 337.

58. Thomas, C.M. and T.R. Ward, *Design of artificial metalloenzymes*. Applied Organometallic Chemistry, 2005. **19**(1): p. 35.

59. Creus, M. and T.R. Ward, Designed evolution of artificial metalloenzymes: Protein catalysts made to order. Organic & Biomolecular Chemistry, 2007. **5**(12): p. 1835.

60. Mao, J. and T.R. Ward, Artificial metalloenzymes for enantioselective catalysis based on the biotin-avidin technology. CHIMIA International Journal for Chemistry, 2008. **62**(12): p. 956.

61. Hyster, T.K., et al., Biotinylated rh(iii) complexes in engineered streptavidin for accelerated asymmetric c–h activation. Science, 2012. **338**(6106): p. 500.

62. Wang, Z.J., et al., Cytochrome p450-catalyzed insertion of carbenoids into n–h bonds. Chemical Science, 2014. **5**(2): p. 598.

63. Yang, H., et al., A general method for artificial metalloenzyme formation through strain-promoted azide–alkyne cycloaddition. ChemBioChem, 2014. **15**(2): p. 223.

64. Bordeaux, M., V. Tyagi, and R. Fasan, Highly diastereoselective and enantioselective olefin cyclopropanation using engineered myoglobin-based catalysts. Angewandte Chemie, 2015. **127**(6): p. 1764.

65. Coleman, J.E., Metal ion dependent binding of sulphonamide to carbonic anhydrase. Nature, 1967. **214**(5084): p. 193.

66. Cuatrecasas, P., S. Fuchs, and C.B. Anfinsen, Catalytic properties and specificity of the extracellular nuclease of staphylococcus aureus. Journal of Biological Chemistry, 1967. **242**(7): p. 1541.

67. Gomez, J.E., E.R. Birnbaum, and D.W. Darnall, Metal ion acceleration of the conversion of trypsinogen to trypsin. Lanthanide ions as calcium ion substitutes. Biochemistry, 1974. **13**(18): p. 3745.

68. Tainer, J.A., V.A. Roberts, and E.D. Getzoff, *Metal-binding sites in proteins.* Current Opinion in Biotechnology, 1991. **2**(4): p. 582.

69. Key, H.M., et al., Abiological catalysis by artificial haem proteins containing noble metals in place of iron. Nature, 2016. **534**: p. 534.

70. Lin, C.-C., C.-W. Lin, and A. S.C. Chan, Catalytic hydrogenation of itaconic acid in a biotinylated pyrphos–rhodium(i) system in a protein cavity. Vol. 10. 1999. 1887.

71. Reetz, M.T., Directed evolution of selective enzymes and hybrid catalysts. Tetrahedron, 2002. **58**(32): p. 6595.

72. Reetz, M.T., et al., Towards the directed evolution of hybrid catalysts. CHIMIA International Journal for Chemistry, 2002. **56**(12): p. 721.

73. Collot, J., et al., Artificial metalloenzymes for enantioselective catalysis: The phenomenon of protein accelerated catalysis. Journal of Organometallic Chemistry, 2004. **689**(25): p. 4868.

74. Kruithof, C.A., et al., Lipase active-site-directed anchoring of organometallics: Metallopincer/protein hybrids. Chemistry – A European Journal, 2005. **11**(23): p. 6869.

75. Panella, L., et al., Merging homogeneous catalysis with biocatalysis; papain as hydrogenation catalyst. Chemical Communications, 2005(45): p. 5656.

76. Skander, M., et al., Chemical optimization of artificial metalloenzymes based on the biotin-avidin technology: (s)-selective and solvent-tolerant hydrogenation catalysts via the introduction of chiral amino acid spacers. Chemical Communications, 2005(38): p. 4815.

77. Rutten, L., et al., Solid-state structural characterization of cutinase–ece-pincer–metal hybrids. Chemistry – A European Journal, 2009. **15**(17): p. 4270.

78. den Heeten, R., et al., Synthesis of hybrid transition-metalloproteins via thiol-selective covalent anchoring of rh-phosphine and ru-phenanthroline complexes. Dalton Transactions, 2010. **39**(36): p. 8477.

79. Basauri-Molina, M., et al., Lipase active site covalent anchoring of rh(nhc) catalysts: Towards chemoselective artificial metalloenzymes. Chemical Communications, 2015. **51**(31): p. 6792.

80. Skander, M., et al., Artificial metalloenzymes: (strept)avidin as host for enantioselective hydrogenation by achiral biotinylated rhodium−diphosphine complexes. Journal of the American Chemical Society, 2004. **126**(44): p. 14411.

81. Klein, G., et al., Tailoring the active site of chemzymes by using a chemogenetic-optimization procedure: Towards substrate-specific artificial hydrogenases based on the biotin–avidin technology. Angewandte Chemie International Edition, 2005. **44**(47): p. 7764.

82. Mazurek, S., T.R. Ward, and M. Novič, Counter propagation artificial neural networks modeling of an enantioselectivity of artificial metalloenzymes. Molecular Diversity, 2007. **11**(3): p. 141.

83. Rusbandi, E., et al., Second generation artificial hydrogenases based on the biotin−avidin technology: Improving activity, stability and selectivity by introduction of enantiopure amino acid spacers. Vol. 349. 2007. 1923.

84. Rusbandi, U.E., et al., Second-generation artificial hydrogenases based on the biotin–avidin technology: Improving selectivity and organic solvent tolerance by introduction of an (r)-proline spacer. Comptes Rendus Chimie, 2007. **10**(8): p. 678.

85. Astruc, D., The metathesis reactions: From a historical perspective to recent developments. New Journal of Chemistry, 2005. **29**(1): p. 42.

86. Abe, S., et al., Polymerization of phenylacetylene by rhodium complexes within a discrete space of apo-ferritin. Journal of the American Chemical Society, 2009. **131**(20): p. 6958.

87. Podtetenieff, J., et al., An artificial metalloenzyme: Creation of a designed copper binding site in a thermostable protein. Angewandte Chemie International Edition, 2010. **49**(30): p. 5151.

88. Davies, H.M.L., J. Du Bois, and J.-Q. Yu, *C–h functionalization in organic synthesis.* Chemical Society Reviews, 2011. **40**(4): p. 1855.

89. Suzuki, A., Cross-coupling reactions of organoboranes: An easy way to construct c c bonds (nobel lecture). Angewandte Chemie International Edition, 2011. **50**(30): p. 6722.

90. Ueno, T., H. Tabe, and Y. Tanaka, Artificial metalloenzymes constructed from hierarchically-assembled proteins. Chemistry – An Asian Journal, 2013. **8**(8): p. 1646.

91. Drienovská, I., et al., Novel artificial metalloenzymes by in vivo incorporation of metal-binding unnatural amino acids. Chemical Science, 2015. **6**(1): p. 770.

92. Reynolds, E.W., et al., *An evolved orthogonal enzyme/cofactor pair.* Journal of the American Chemical Society, 2016. **138**(38): p. 12451.

93. Delbecq, F. and C. Len, Chapter 5 - application of heck alkenylation reaction in modified nucleoside synthesis, in Palladium-catalyzed modification of nucleosides, nucleotides and oligonucleotides, A.R. Kapdi, D. Maiti, and Y.S. Sanghvi, Editors. 2018, Elsevier. p. 147.

94. Fagnou, K. and M. Lautens, Rhodium-catalyzed carbon−carbon bond forming reactions of organometallic compounds. Chemical Reviews, 2003. **103**(1): p. 169.

95. Souillart, L. and N. Cramer, Catalytic c–c bond activations via oxidative addition to transition metals. Chemical Reviews, 2015. **115**(17): p. 9410.

96. Abbas, C.A. and A.A. Sibirny, Genetic control of biosynthesis and transport of riboflavin and flavin nucleotides and construction of robust biotechnological producers. Microbiol Mol Biol Rev, 2011. **75**(2): p. 321.

97. Northrop-Clewes, C.A. and D.I. Thurnham, The discovery and characterization of riboflavin. Annals of Nutrition and Metabolism, 2012. **61**(3): p. 224.

98. Dong, H. and S.V. Beer, Riboflavin induces disease resistance in plants by activating a novel signal transduction pathway. Phytopathology, 2000. **90**(8): p. 801.

99. Miramar, M.D., et al., Nadh oxidase activity of mitochondrial apoptosis-inducing factor. Journal of Biological Chemistry, 2001. **276**(19): p. 16391.

100. Zhang, S., et al., Riboflavin-induced priming for pathogen defense in arabidopsis thaliana. Journal of Integrative Plant Biology, 2009. **51**(2): p. 167.

101. Asai, S., K. Mase, and H. Yoshioka, A key enzyme for flavin synthesis is required for nitric oxide and reactive oxygen species production in disease resistance. The Plant Journal, 2010. **62**(6): p. 911.

102. Massey, V., The chemical and biological versatility of riboflavin. Biochemical Society Transactions, 2000. **28**(4): p. 283.

103. Juárez, O., et al., Riboflavin is an active redox cofactor in the na+-pumping nadh:Quinone oxidoreductase (na+-nqr) from vibrio cholerae. Journal of Biological Chemistry, 2008. **283**(48): p. 33162.

104. Mansoorabadi, S.O., C.J. Thibodeaux, and H.-w. Liu, The diverse roles of flavin coenzymes--nature's most versatile thespians. J Org Chem, 2007. **72**(17): p. 6329.

105. Fraaije, M.W., et al., Covalent flavinylation is essential for efficient redox catalysis in vanillyl-alcohol oxidase. Journal of Biological Chemistry, 1999. **274**(50): p. 35514.

106. Winkler, A., et al., Biochemical evidence that berberine bridge enzyme belongs to a novel family of flavoproteins containing a bi-covalently attached fad cofactor. Journal of Biological Chemistry, 2006. **281**(30): p. 21276.

107. Heuts, D.P.H.M., et al., *What's in a covalent bond?* The FEBS Journal, 2009. **276**(13): p. 3405.

108. Palmer, G. and J. Reedijk, Nomenclature of electron−transfer proteins: Recommendations 1989. European Journal of Biochemistry, 1991. **200**(3): p. 599.

109. Miura, R., Versatility and specificity in flavoenzymes: Control mechanisms of flavin reactivity. The Chemical Record, 2001. **1**(3): p. 183.

110. Mattevi, A., To be or not to be an oxidase: Challenging the oxygen reactivity of flavoenzymes. Trends in Biochemical Sciences, 2006. **31**(5): p. 276.

111. Nisimoto, Y., et al., Constitutive nadph-dependent electron transferase activity of the nox4 dehydrogenase domain. Biochemistry, 2010. **49**(11): p. 2433.

112. Sancho, J., Flavodoxins: Sequence, folding, binding, function and beyond. Cellular and Molecular Life Sciences CMLS, 2006. **63**(7): p. 855.

113. Sancar, A., Cryptochrome: The second photoactive pigment in the eye and its role in circadian photoreception. Annual Review of Biochemistry, 2000. **69**(1): p. 31.

114. Christie, J.M. and W.R. Briggs, *Blue light sensing in higher plants.* Journal of Biological Chemistry, 2001. **276**(15): p. 11457.

115. Froehlich, A.C., et al., White collar-1, a circadian blue light photoreceptor, binding to the frequency promoter. Science, 2002. **297**(5582): p. 815.

116. Imaizumi, T., et al., Fkf1 is essential for photoperiodic-specific light signalling in arabidopsis. Nature, 2003. **426**(6964): p. 302.

117. Sancar, A., Photolyase and cryptochrome blue-light photoreceptors, in Advances in protein chemistry. 2004, Academic Press. p. 73.

118. Li, Z., et al., Sensing and responding to excess light. Annual Review of Plant Biology, 2009. **60**(1): p. 239.

119. Haupt, W., Chloroplast movement: From phenomenology to molecular biology, in Progress in botany: Genetics cell biology and physiology systematics and comparative morphology ecology and vegetation science, K. Esser, et al., Editors. 1999, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 3.

120. Briggs, W.R. and J.M. Christie, Phototropins 1 and 2: Versatile plant blue-light receptors. Trends in Plant Science, 2002. **7**(5): p. 204.

121. Drepper, T., et al., Reporter proteins for in vivo fluorescence without oxygen. Nature Biotechnology, 2007. **25**: p. 443.

122. Tielker, D., et al., Flavin mononucleotide-based fluorescent protein as an oxygen-independent reporter in candida albicans and saccharomyces cerevisiae. Eukaryotic Cell, 2009. **8**(6): p. 913.

123. Taheri, P. and S. Tarighi, Riboflavin induces resistance in rice against rhizoctonia solani via jasmonate-mediated priming of phenylpropanoid pathway. Journal of Plant Physiology, 2010. **167**(3): p. 201.

124. Warren, M.J., et al., The biosynthesis of adenosylcobalamin (vitamin b12). Natural Product Reports, 2002. **19**(4): p. 390.

125. Campbell, G.R.O., et al., sinorhizobium meliloti blub is necessary for production of 5,6-dimethylbenzimidazole, the lower ligand of b<sub>12</sub&gt. Proc Natl Acad Sci U S A, 2006. **103**(12): p. 4634.

126. Lin, C., et al., Association of flavin adenine dinucleotide with the arabidopsis blue light receptor cry1. Science, 1995. **269**(5226): p. 968.

127. Gomelsky, M. and G. Klug, Bluf: A novel fad-binding domain involved in sensory transduction in microorganisms. Trends in Biochemical Sciences, 2002. **27**(10): p. 497.

128. Marsili, E., et al., shewanella secretes flavins that mediate extracellular electron transfer. Proceedings of the National Academy of Sciences, 2008. **105**(10): p. 3968.

129. von Canstein, H., et al., Secretion of flavins by shewanella species and their role in extracellular electron transfer. Applied and Environmental Microbiology, 2008. **74**(3): p. 615.

130. Coursolle, D., et al., The mtr respiratory pathway is essential for reducing flavins and electrodes in <em>shewanella oneidensis</em&gt. J Bacteriol, 2010. **192**(2): p. 467.

131. McCormick, J.R.D. and G.O. Morton, Identity of cosynthetic factor i of streptomyces aureofaciens and fragment fo from coenzyme f420 of methanobacterium species. Journal of the American Chemical Society, 1982. **104**(14): p. 4014.

132. Eker, A.P.M., J.K.C. Hessels, and R.H. Dekker, Photoreactivating enzyme from streptomyces griseus—vi. Action spectrum and kinetics of photoreactivation. Photochemistry and Photobiology, 1986. **44**(2): p. 197.

133. Walsh, C., Naturally occurring 5-deazaflavin coenzymes: Biological redox roles. Accounts of Chemical Research, 1986. **19**(7): p. 216.

134. Graham, D.E. and R.H. White, Elucidation of methanogenic coenzyme biosyntheses: From spectroscopy to genomics. Natural Product Reports, 2002. **19**(2): p. 133.

135. Lessner, D.J., et al., An unconventional pathway for reduction of co2 to methane in co-grown methanosarcina acetivorans revealed by proteomics. Proc Natl Acad Sci U S A, 2006. **103**(47): p. 17921.

136. Eker, A.P., et al., DNA photoreactivating enzyme from the cyanobacterium anacystis nidulans. Journal of Biological Chemistry, 1990. **265**(14): p. 8009.

137. Sancar, A., Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. Chemical Reviews, 2003. **103**(6): p. 2203.

138. Eker, A.P.M., et al., *DNA repair in mammalian cells.* Cellular and Molecular Life Sciences, 2009. **66**(6): p. 968.

139. Glas, A.F., et al., The archaeal cofactor f₀ is a light-harvesting antenna chromophore in eukaryotes. Proceedings of the National Academy of Sciences, 2009. **106**(28): p. 11540.

140. M. Rhodes, P., et al., Biochemical and genetic characterization of streptomyces rimosus mutants impaired in oxytetracycline biosynthesis. Vol. 124. 1981. 329.

141. Peschke, U., et al., Molecular characterization of the lincomycin-production gene cluster of streptomyces lincolnensis 78-11. Molecular Microbiology, 1995. **16**(6): p. 1137.

142. Ikeno, S., et al., DNA sequencing and transcriptional analysis of the kasugamycin biosynthetic gene cluster from streptomyces kasugaensis m338-m1. The Journal of Antibiotics, 2006. **59**(1): p. 18.

143. Purwantini, E. and L. Daniels, Purification of a novel coenzyme f420-dependent glucose-6-phosphate dehydrogenase from mycobacterium smegmatis. J Bacteriol, 1996. **178**(10): p. 2861.

144. Purwantini, E. and L. Daniels, Molecular analysis of the gene encoding f420-dependent glucose-6-phosphate dehydrogenase from mycobacterium smegmatis. J Bacteriol, 1998. **180**(8): p. 2212.

145. Stover, C.K., et al., A small-molecule nitroimidazopyran drug candidate for the treatment of tuberculosis. Nature, 2000. **405**(6789): p. 962.

146. Boshoff, H.I. and C.E. Barry, Is the mycobacterial cell wall a hopeless drug target for latent tuberculosis? Drug Discovery Today: Disease Mechanisms, 2006. **3**(2): p. 237.

147. Pedrolli, D.B., et al., The antibiotics roseoflavin and 8-demethyl-8-amino-riboflavin from streptomyces davawensis are metabolized by human flavokinase and human fad synthetase. Biochem Pharmacol, 2011. **82**(12): p. 1853.

148. Matern, A., et al., Uptake and metabolism of antibiotics roseoflavin and 8-demethyl-8-aminoriboflavin in riboflavin-auxotrophic listeria monocytogenes. J Bacteriol, 2016. **198**(23): p. 3233.

149. Shinkai, S., et al., Coenzyme models: 40. Spectral and reactivity studies of roseoflavin analogs: Correlation between reactivity and spectral parameters. Bioorganic Chemistry, 1986. **14**(2): p. 119.

150. Mack, M. and S. Grill, Riboflavin analogs and inhibitors of riboflavin biosynthesis. Applied Microbiology and Biotechnology, 2006. **71**(3): p. 265.

151. Uyakul, D., M. Isobe, and T. Goto, Lampteroflavin, the first riboflavinyl alpha ribofuranoside as light emitter in the luminous mushroom, l. Japonicus. Tetrahedron, 1990. **46**(4): p. 1367.

152. Susín, S., et al., Riboflavin 3'- and 5'-sulfate, two novel flavins accumulating in the roots of iron-deficient sugar beet (beta vulgaris). Journal of Biological Chemistry, 1993. **268**(28): p. 20958.

153. Marshall, S.A., K.A.P. Payne, and D. Leys, The ubix-ubid system: The biosynthesis and use of prenylated flavin (prfmn). Archives of Biochemistry and Biophysics, 2017. **632**: p. 209.

154. White, M.D., et al., Ubix is a flavin prenyltransferase required for bacterial ubiquinone biosynthesis. Nature, 2015. **522**: p. 502.

155. Winkelblech, J., A. Fan, and S.-M. Li, Prenyltransferases as key enzymes in primary and secondary metabolism. Applied Microbiology and Biotechnology, 2015. **99**(18): p. 7379.

156. Coldham, I. and R. Hufton, Intramolecular dipolar cycloaddition reactions of azomethine ylides. Chemical Reviews, 2005. **105**(7): p. 2765.

157. Jordan, F. and H. Patel, Catalysis in enzymatic decarboxylations: Comparison of selected cofactor-dependent and cofactor-independent examples. ACS Catal, 2013. **3**(7): p. 1601.

158. Leys, D., Flavin metamorphosis: Cofactor transformation through prenylation. Curr Opin Chem Biol, 2018. **47**: p. 117.

159. Edmondson, D.E., Intramolecular hemiacetal formation in 8-formylriboflavine. Biochemistry, 1974. **13**(14): p. 2817.

160. Yorita, K., et al., Interaction of two arginine residues in lactate oxidase with the enzyme flavin: Conversion of fmn to 8-formyl-fmn. Proceedings of the National Academy of Sciences, 2000. **97**(24): p. 13039.

161. Doubayashi, D., et al., Formate oxidase, an enzyme of the glucose-methanol-choline oxidoreductase family, has a his-arg pair and 8-formyl-fad at the catalytic site. Bioscience, Biotechnology, and Biochemistry, 2011. **75**(9): p. 1662.

162. Robbins, J.M., et al., Enzyme-mediated conversion of flavin adenine dinucleotide (fad) to 8-formyl fad in formate oxidase results in a modified cofactor with enhanced catalytic properties. Biochemistry, 2017. **56**(29): p. 3800.

163. Teufel, R., et al., Flavin-mediated dual oxidation controls an enzymatic favorskii-type rearrangement. Nature, 2013. **503**: p. 552.

164. Teufel, R., et al., Biochemical establishment and characterization of encm's flavin-n5-oxide cofactor. Journal of the American Chemical Society, 2015. **137**(25): p. 8078.

165. Teufel, R., V. Agarwal, and B.S. Moore, Unusual flavoenzyme catalysis in marine bacteria. Curr Opin Chem Biol, 2016. **31**: p. 31.

166. Piano, V., B.A. Palfey, and A. Mattevi, Flavins as covalent catalysts: New mechanisms emerge. Trends in Biochemical Sciences, 2017. **42**(6): p. 457.

167. Adak, S. and T.P. Begley, Dibenzothiophene catabolism proceeds via a flavin-n5-oxide intermediate. Journal of the American Chemical Society, 2016. **138**(20): p. 6424.

168. Adak, S. and T.P. Begley, Ruta-catalyzed oxidative cleavage of the uracil amide involves formation of a flavin-n5-oxide. Biochemistry, 2017. **56**(29): p. 3708.

169. Adak, S. and T.P. Begley, Flavin-n5-oxide: A new, catalytic motif in flavoenzymology. Archives of Biochemistry and Biophysics, 2017. **632**: p. 4.

170. Ghisla, S. and V. Massey, New flavins for old: Artificial flavins as active site probes of flavoproteins. Biochem J, 1986. **239**(1): p. 1.

171. Murthy, Y.V.S.N. and V. Massey, [41] syntheses and applications of flavin analogs as active site probes for flavoproteins, in Methods in enzymology. 1997, Academic Press. p. 436.

172. Hefti, M.H., J. Vervoort, and W.J.H. van Berkel, *Deflavination and reconstitution of flavoproteins*. European Journal of Biochemistry, 2003. **270**(21): p. 4227.

173. Schopfer, L.M., V. Massey, and A. Claiborne, Active site probes of flavoproteins. Determination of the solvent accessibility of the flavin position 8 for a series of flavoproteins. Journal of Biological Chemistry, 1981. **256**(14): p. 7329.

174. Massey, V., S. Ghisla, and E.G. Moore, 8-mercaptoflavins as active site probes of flavoenzymes. Journal of Biological Chemistry, 1979. **254**(19): p. 9640.

175. Claiborne, A., et al., 2-thioflavins as active site probes of flavoproteins. Journal of Biological Chemistry, 1982. **257**(1): p. 174.

176. Manstein, D.J., et al., Absolute stereochemistry of flavins in enzyme-catalyzed reactions. Biochemistry, 1986. **25**(22): p. 6807.

177. Entsch, B., et al., Oxygen reactivity of p-hydroxybenzoate hydroxylase containing 1-deaza-fad. Journal of Biological Chemistry, 1980. **255**(4): p. 1420.

178. Mansurova, M., M.S. Koay, and W. Gärtner, Synthesis and electrochemical properties of structurally modified flavin compounds. European Journal of Organic Chemistry, 2008. **2008**(32): p. 5401.

179. Amato, E.D. and J.D. Stewart, Applications of protein engineering to members of the old yellow enzyme family. Biotechnology Advances, 2015. **33**(5): p. 624.

180. Toogood, H.S., J.M. Gardiner, and N.S. Scrutton, Biocatalytic reductions and chemical versatility of the old yellow enzyme family of flavoprotein oxidoreductases. ChemCatChem, 2010. **2**(8): p. 892.

181. Daugherty, A.B., S. Govindarajan, and S. Lutz, Improved biocatalysts from a synthetic circular permutation library of the flavin-dependent oxidoreductase old yellow enzyme. Journal of the American Chemical Society, 2013. **135**(38): p. 14425.

182. Scholtissek, A., et al., Old yellow enzyme-catalysed asymmetric hydrogenation: Linking family roots with improved catalysis. Catalysts, 2017. **7**(5): p. 130.

183. Warburg, O. and W. Christian, Ein zweites sauerstoffübertragendes ferment und sein absorptionsspektrum. Naturwissenschaften, 1932. **20**(37): p. 688.

184. Williams, R.E. and N.C. Bruce, 'New uses for an old enzyme' – the old yellow enzyme family of flavoenzymes. Microbiology, 2002. **148**(6): p. 1607.

185. Gao, X., et al., Biochemical characterization and substrate profiling of a new nadh-dependent enoate reductase from lactobacillus casei. Enzyme and Microbial Technology, 2012. **51**(1): p. 26.

186. Fu, Y., K. Castiglione, and D. Weuster-Botz, Comparative characterization of novel ene-reductases from cyanobacteria. Biotechnology and Bioengineering, 2013. **110**(5): p. 1293.

187. Nizam, S., et al., Comprehensive genome-wide analysis reveals different classes of enigmatic old yellow enzyme in fungi. Scientific Reports, 2014. **4**: p. 4013.

188. Binks, P.R., et al., Degradation of pentaerythritol tetranitrate by enterobacter cloacae pb2. Applied and environmental microbiology, 1996. **62**(4): p. 1214.

189. Kitzing, K., et al., The 1.3 å crystal structure of the flavoprotein yqjm reveals a novel class of old yellow enzymes. Journal of Biological Chemistry, 2005. **280**(30): p. 27904.

190. Hall, M., et al., Asymmetric bioreduction of activated alkenes using cloned 12-oxophytodienoate reductase isoenzymes opr-1 and opr-3 from lycopersicon esculentum (tomato): A striking change of stereoselectivity. Angewandte Chemie International Edition, 2007. **46**(21): p. 3934.

191. Müller, A., B. Hauer, and B. Rosche, Asymmetric alkene reduction by yeast old yellow enzymes and by a novel zymomonas mobilis reductase. Biotechnology and Bioengineering, 2007. **98**(1): p. 22.

192. Hall, M., et al., 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): p. 1511.

193. Oberdorfer, G., et al., Stereopreferences of old yellow enzymes: Structure correlations and sequence patterns in enoate reductases. ChemCatChem, 2011. **3**(10): p. 1562.

194. Fryszkowska, A., et al., A surprising observation that oxygen can affect the product enantiopurity of an enzyme-catalysed reaction. The FEBS Journal, 2012. **279**(22): p. 4160.

195. Karplus, P. and K. Fox, Structure-function relations for old yellow enzyme. Vol. 9. 1995.

196. Fraaije, M.W. and A. Mattevi, Flavoenzymes: Diverse catalysts with recurrent features. Trends in Biochemical Sciences, 2000. **25**(3): p. 126.

197. Brown, B.J., et al., On the active site of old yellow enzyme: Role of histidine 191 and asparagine 194. Journal of Biological Chemistry, 1998. **273**(49): p. 32753.

198. Xu, D., R.M. Kohli, and V. Massey, The role of threonine 37 in flavin reactivity of the old yellow enzyme. Proc Natl Acad Sci U S A, 1999. **96**(7): p. 3556.

199. Brown, B.J., et al., The role of glutamine 114 in old yellow enzyme. Journal of Biological Chemistry, 2002. **277**(3): p. 2138.

200. Padhi, S.K., D.J. Bougioukou, and J.D. Stewart, Site-saturation mutagenesis of tryptophan 116 of saccharomyces pastorianus old yellow enzyme uncovers stereocomplementary variants. Journal of the American Chemical Society, 2009. **131**(9): p. 3271.

201. Kohli, R.M. and V. Massey, The oxidative half-reaction of old yellow enzyme: The role of tyrosine 196. Journal of Biological Chemistry, 1998. **273**(49): p. 32763.

202. Khan, H., et al., Proton transfer in the oxidative half-reaction of pentaerythritol tetranitrate reductase. The FEBS Journal, 2005. **272**(18): p. 4660.

203. Opperman, D.J., et al., Crystal structure of a thermostable old yellow enzyme from thermus scotoductus sa-01. Biochemical and Biophysical Research Communications, 2010. **393**(3): p. 426.

204. Massey, V. and L.M. Schopfer, Reactivity of old yellow enzyme with alpha-nadph and other pyridine nucleotide derivatives. Journal of Biological Chemistry, 1986. **261**(3): p. 1215.

205. Breithaupt, C., et al., X-ray structure of 12-oxophytodienoate reductase 1 provides structural insight into substrate binding and specificity within the family of oye. Structure, 2001. **9**(5): p. 419.

206. Vaz, A.D.N., S. Chakraborty, and V. Massey, Old yellow enzyme: Aromatization of cyclic enones and the mechanism of a novel dismutation reaction. Biochemistry, 1995. **34**(13): p. 4246.

207. Walton, A.Z., et al., Residues controlling facial selectivity in an alkene reductase and semirational alterations to create stereocomplementary variants. ACS Catal, 2014. **4**(7): p. 2307.

208. Quertinmont, L.T. and S. Lutz, Cell-free protein engineering of old yellow enzyme 1 from saccharomyces pastorianus. Tetrahedron, 2016. **72**(46): p. 7282.

209. Abramovitz, A.S. and V. Massey, Interaction of phenols with old yellow enzyme. Physical evidence for charge-transfer complexes. Journal of Biological Chemistry, 1976. **251**(17): p. 5327.

210. Eweg, J.K., F. MÜLler, and W.J.H. Van Berkel, On the enigma of old yellow enzyme's spectral properties. European Journal of Biochemistry, 1982. **129**(2): p. 303.

211. Stewart, R.C. and V. Massey, Potentiometric studies of native and flavin-substituted old yellow enzyme. Journal of Biological Chemistry, 1985. **260**(25): p. 13639.

212. Schopfer, L.M., A. Wessiak, and V. Massey, Interpretation of the spectra observed during oxidation of p-hydroxybenzoate hydroxylase reconstituted with modified flavins. Journal of Biological Chemistry, 1991. **266**(20): p. 13080.

213. Hasford, J.J. and C.J. Rizzo, Linear free energy substituent effect on flavin redox chemistry. Journal of the American Chemical Society, 1998. **120**(10): p. 2251.

214. Murthy, Y.V.S.N. and V. Massey, Synthesis and properties of 8-cn-flavin nucleotide analogs and studies with flavoproteins. Journal of Biological Chemistry, 1998. **273**(15): p. 8975.

215. Murthy, Y.V.S.N., Y. Meah, and V. Massey, Conversion of a flavoprotein reductase to a desaturase by manipulation of the flavin redox potential. Journal of the American Chemical Society, 1999. **121**(22): p. 5344.

216. Fitzpatrick, T.B., N. Amrhein, and P. Macheroux, Characterization of yqjm, an old yellow enzyme homolog from bacillus subtilis involved in the oxidative stress response. Journal of Biological Chemistry, 2003. **278**(22): p. 19891.

217. Hall, M., et al., Asymmetric bioreduction of c⌋c bonds using enoate reductases opr1, opr3 and yqjm: Enzyme-based stereocontrol. Advanced Synthesis & Catalysis, 2008. **350**(3): p. 411.

218. Müller, A., et al., Stereospecific alkyne reduction: Novel activity of old yellow enzymes. Angewandte Chemie International Edition, 2007. **46**(18): p. 3316.

219. Stiege, W. and V.A. Erdmann, The potentials of the in vitro protein biosynthesis system. Journal of Biotechnology, 1995. **41**(2): p. 81.

220. Carlson, E.D., et al., Cell-free protein synthesis: Applications come of age. Biotechnology advances, 2012. **30**(5): p. 1185.

221. Kung, H.F., et al., DNA-directed in vitro synthesis of beta-galactosidase. Studies with purified factors. Journal of Biological Chemistry, 1977. **252**(19): p. 6889.

222. Spirin, A., et al., A continuous cell-free translation system capable of producing polypeptides in high yield. Science, 1988. **242**(4882): p. 1162.

223. Shimizu, Y., et al., Cell-free translation reconstituted with purified components. Nature Biotechnology, 2001. **19**(8): p. 751.

224. Jewett, M.C., et al., An integrated cell-free metabolic platform for protein production and synthetic biology. Mol Syst Biol, 2008. **4**: p. 220.

225. Moore, S.D. and R.T. Sauer, The tmrna system for translational surveillance and ribosome rescue. Annual Review of Biochemistry, 2007. **76**(1): p. 101.

226. Shimizu, Y., T. Kanamori, and T. Ueda, Protein synthesis by pure translation systems. Methods, 2005. **36**(3): p. 299.

227. Li, J., et al., Improved cell-free rna and protein synthesis system. PLOS ONE, 2014. **9**(9): p. e106232.

228. Li, J., et al., Dissecting limiting factors of the protein synthesis using recombinant elements (pure) system. Translation (Austin), 2017. **5**(1): p. e1327006.

229. Forster, A.C., et al., Programming peptidomimetic syntheses by translating genetic codes designed de novo. Proceedings of the National Academy of Sciences, 2003. **100**(11): p. 6353.

230. Lipovsek, D. and A. Plückthun, In-vitro protein evolution by ribosome display and mrna display. Journal of Immunological Methods, 2004. **290**(1): p. 51.

231. Josephson, K., M.C.T. Hartman, and J.W. Szostak, *Ribosomal synthesis of unnatural peptides.* Journal of the American Chemical Society, 2005. **127**(33): p. 11727.

232. Fen, C.X., et al., Directed evolution of p53 variants with altered DNA-binding specificities by in vitro compartmentalization. Journal of Molecular Biology, 2007. **371**(5): p. 1238.

233. Quertinmont, L.T., R. Orru, and S. Lutz, Rapid parallel protein evaluator (rapper), from gene to enzyme function in one day. Chemical Communications, 2015. **51**(1): p. 122.

234. Abe, M., et al., Detection of structural changes in a cofactor binding protein by using a wheat germ cell-free protein synthesis system coupled with unnatural amino acid probing. Proteins: Structure, Function, and Bioinformatics, 2007. **67**(3): p. 643.

Chapter 2: Truncated FAD synthetase for direct biocatalytic conversion of riboflavin and analogs to their corresponding flavin mononucleotides

Previously published in: PEDS, 2013, 26, 791

Abstract

The preparation of flavin mononucleotide (FMN) and FMN analogs from their corresponding riboflavin precursors is traditionally performed in a two-step procedure. After initial enzymatic conversion of riboflavin to flavin adenine dinucleotide (FAD) by a bifunctional FAD synthetase, the adenyl moiety of FAD is hydrolyzed with snake venom phosphodiesterase to yield FMN. To simplify the protocol, we have engineered the FAD synthetase from *Corynebacterium ammoniagenes* by deleting its N-terminal adenylation domain. The newly created biocatalyst is stable and efficient for direct and quantitative phosphorylation of riboflavin and riboflavin analogs to their corresponding FMN cofactors at preparative-scale.

I. Introduction

Flavin monophosphate (FMN) and flavin adenine dinucleotide (FAD) constitute two important natural cofactors for one and two-electron transfer reactions. In their protein-bound forms, FMN and FAD are involved in numerous critical cellular functions including oxidative phosphorylation and photosynthesis, fatty acid oxidation, DNA metabolism, detoxification and biosynthesis of other cofactors (Joosten and van Berkel, 2007) where they facilitate a diverse set of chemistries including dehydrogenation, mono and dioxygenation, halogenation and reductions (Miura, 2001; Macheroux *et al.*, 2011; de Gonzalo and Fraaije, 2013;Walsh and Wencewicz, 2013). Both cofactors are synthesized from GTP and ribulose-5-phosphate via the common intermediate riboflavin (RF) (Bacher *et al*., 2000). Subsequently, RF is converted to FMN and FAD through two consecutive reactions, an initial phosphorylation catalyzed by ATP:riboflavin 5’-phosphotransferase (EC 2.7.1.26), followed by adenylation facilitated by ATP:FMN

adenyltransferase (E.C. 2.7.7.2) (**Figure 2.1**). Interestingly, eukaryotes possess separate riboflavin kinases (RFKs) and FMN adenyltransferases to catalyze these two reactions (Merrill and McCormick, 1980). In contrast, prokaryotes utilize a bi-functional fusion protein (FADS) with an N-terminal adenylation domain and a C-terminal RFK domain (Hagihara *et al.*, 1995). A series of crystal structures of eukaryotic and prokaryotic representatives have been solved. For the bi-functional FADSs, the structures of two enzymes from Thermotoga maritima (PDB access#: 1MRZ, 1T6X, 1T6Y, and 1T6Z (Wang *et al.*, 2005; Shin *et al.*, 2013)) and Corynebacterium ammoniagenes (PDB access#: 2X0K (Herguedas *et al.*, 2010); **Figure 2.1**) were reported.

A comparison of their N-terminal adenyl-transfer domain with its eukaryotic homologs indicates only low structural and sequence similarity. In contrast, the core structures of the RFKs have substantial overlap although differences exist in loop regions. Although there are a number of hydrogen bonding and hydrophobic interactions at the domain interface, Medina and coworkers demonstrated that each individual domain could be expressed in soluble form (Frago *et al.*, 2008). However, only the C-terminal RFK was catalytically active based on tests with crude cell extracts.

Prokaryotic FADS and eukaryotic RFKs have both been applied for the phosphorylation of RF to FMN in the laboratory. These enzymatic routes represent attractive alternatives to organic synthetic methods that require extra protection/deprotection steps of the ribityl moiety to ensure regioselectivity (Hersh and Walsh, 1980). In early studies, researchers employed either rat liver RFK (Merrill and McCormick, 1980) or FADS from Corynebacterium ammoniagenes (CaFADS) (Spencer *et al.*, 1976). The rat enzyme has the advantage of directly producing the

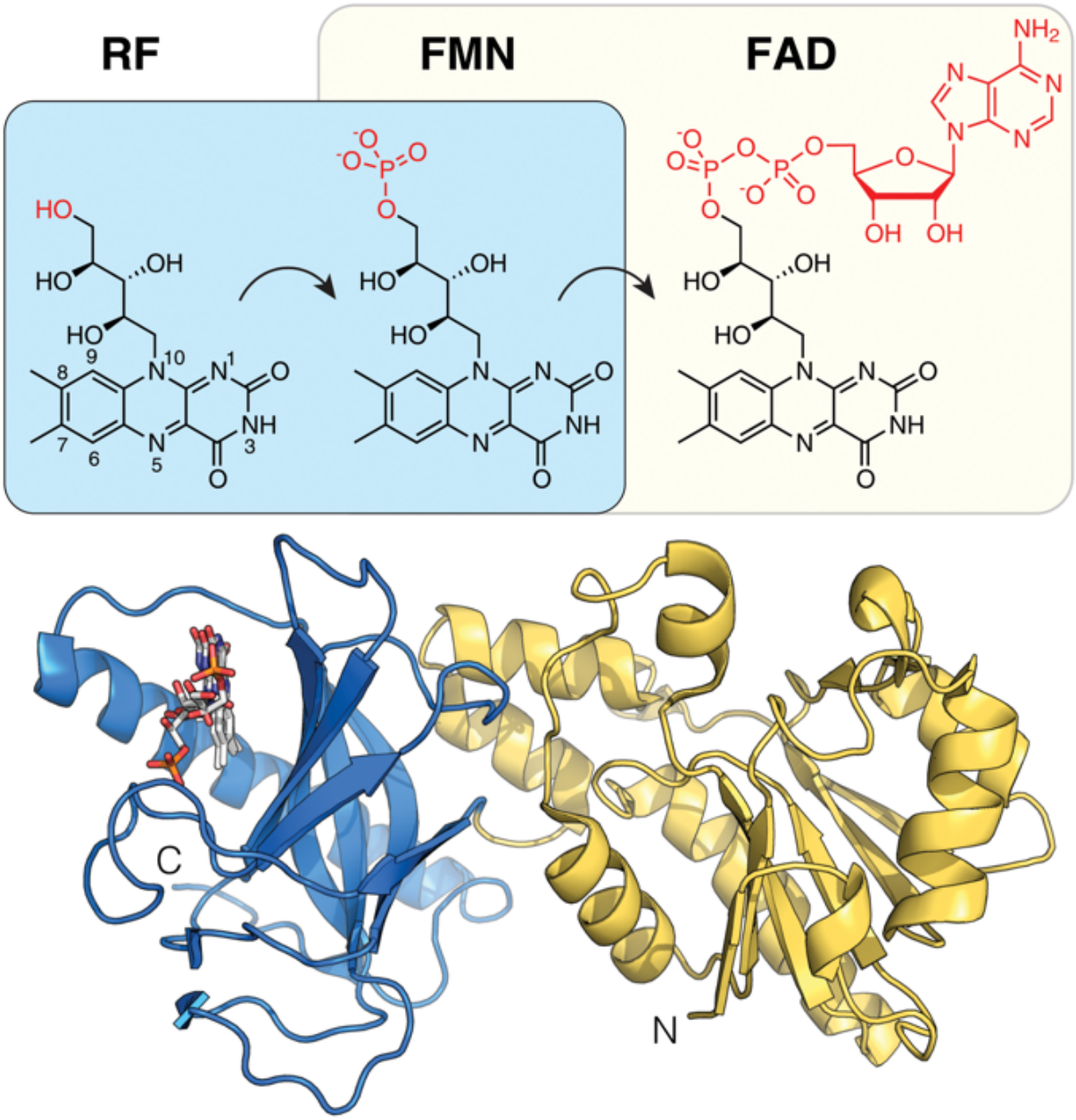


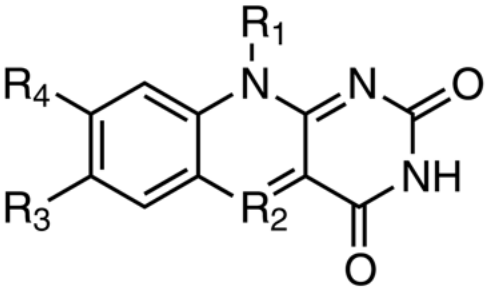
Figure 2.1. Structures of naturally occurring RF, FMN and FAD, as well as FADS from *Corynebacterium ammoniagenes* (PDB 2X0K; Herguedas *et al.*, 2010). With the help of ATP, the C-terminal phosphoryl-transfer domain (shown in blue) initially converts RF to FMN. The presumed active site is marked by two superimposed FMN structures (sticks) from aligned Thermotoga and human kinase homologs. PDB code: 1P4M (Karthikeyan *et al.*, 2003) and 1T6Y (Shin *et al.*, 2013). Next, the N-terminal adenylation domain (highlighted in yellow) facilitates the ATP-driven conversion of FMN to flavin adenine dinucleotide (FAD).

desired mononucleotidic flavin product, yet requires a lengthy enzyme purification process and lacks stability which greatly limits its practical use as a biocatalyst. In contrast, the bifunctional

CaFADS is a robust biocatalyst but complicates the synthesis as it directly converts RF to the corresponding FAD. Furthermore, technical difficulties arise with CaFADS when used at preparative scales since the adenylation step is subject to substrate inhibition and the reaction is reversible, producing a mixture of FMN and FAD (Efimov *et al*., 1998). The preparation of FMN therefore requires extra purification and a second reaction step involving treatment of FAD with snake venom phosphodiesterase to cleave off the adenyl moiety. Despite these problems, the conversion of RF with CaFADS remains the standard procedure for *in vitro* synthesis of FMN (and FAD) today. A number of *in vivo* strategies have also been developed for the intracellular biosynthesis of FMN (Yatsyshyn *et al.*, 2009; Krauss *et al.*, 2011), yet the *in vitro* enzymatic approach remain popular as it offers greater experimental versatility. More specifically, researchers can expand the scope of FMN synthesis beyond RF by exploiting the enzymes' substrate promiscuity which allows for effective conversion of naturally occurring and synthetic RF analogs (Murthy and Massey, 1997; Ghisla and Edmondson, 1999; Miller and Edmondson, 1999; Mack and Grill, 2006; Taylor *et al*., 2013). Such flavin analogs are important functional probes. Substitutions in the isoalloxazine moiety as shown in **Figure 2.2** can, for example, modulate the redox potential of cofactors for alternate chemical reactivity or provide novel sites for covalent crosslinking of cofactor and enzyme (Walsh *et al*., 1978; Ghisla and Massey, 1986).

Motivated by a desire to create a stable RFK for the direct enzymatic phosphorylation of RF and its analogs, we herein describe the expression, purification and functional evaluation of a truncated C-terminal RFK domain from CaFADS. The engineered flavokinase was expressed heterologously in Escherichia coli in high yields and purified in one step via metal affinity chromatography. More importantly, we demonstrate that the isolated enzyme is an effective

biocatalyst for the phosphorylation of RF and three flavin analogs 7,8-dichloro-riboflavin (**1**), 8-amino-riboflavin (**2**) and 5-deaza-riboflavin (**3**) to their corresponding FMN derivatives.



compound	R_1	R_2	R_3	R_4
RF	ribityl	N	$-CH_3$	$-CH_3$
FMN	ribityl-PO_4^{2-}	N	$-CH_3$	$-CH_3$
1	ribityl	N	–Cl	–Cl
1a	ribityl-PO_4^{2-}	N	–Cl	–Cl
2	ribityl	N	$-CH_3$	$-NH_2$
2a	ribityl-PO_4^{2-}	N	$-CH_3$	$-NH_2$
3	ribityl	CH	$-CH_3$	$-CH_3$
3a	ribityl-PO_4^{2-}	CH	$-CH_3$	$-CH_3$

Figure 2.2. Overview of modifications in the isoalloxazine moiety of flavin analogs used in this study. Besides the native substrate (RF), the engineered FADSs were tested with 7,8-dichloro riboflavin (**1**), 8-amino riboflavin (**2**) and 5-deaza riboflavin (**3**) to produce the corresponding monophosphate derivates (**1a–3a**).

II. Results and discussion

Heterologous expression and purification of full-length and truncated FADSs

As part of our efforts to create an effective biocatalytic system for the direct phosphorylation of RF and analogs to their corresponding monophosphates, the two genes encoding for full length Ca FADS (amino acids 1 – 338) and the truncated C-terminal RF kinase domain (tcRFK; amino acids 187– 338) were cloned into a standard pET vector. Upon induction of heterologous protein expression, both enzymes could be produced in soluble form with high yields of typically 70 mg per liter of culture broth in *E. coli*. The subsequent protein purification involved a two-step procedure with initial ion-exchange chromatography, followed by gel filtration for final clean-up. The gel filtration step also doubled for the analysis of the proteins' quaternary structures. As previously reported, the elution profile of full-length FADS showed two

characteristic peaks at molecular weights corresponding to its monomeric and trimeric forms (Herguedas *et al*., 2010). In contrast, the tcRFK eluted as a single peak at an estimated molecular weight of 16.6 kDa, consistent with monomeric enzyme.

To assist in protein purification and simplify preparative-scale FMN synthesis by immobilization of the biocatalyst, we also tested the impact of an N or C-terminal poly-His tag on the stability and catalytic activity of tcRFK. Heterologous protein expression of both variants shows that the N-terminal tag was not tolerated by the truncated enzyme, resulting in protein aggregates with no detectable catalytic activity. In contrast, tcRFK with the C-terminal affinity tag (tcRFKhis) exhibits robust expression of soluble and active enzyme at levels comparable with full-length FADS.

While the deletion of the N-terminal domain of CaFADS might seem relatively straightforward, the principle cannot be applied to all bi-functional enzymes in this subfamily. We were unsuccessful in using the same strategy for the N-terminal truncation of the FADS from *T. maritima*. The crystal structures of the two FADSs are superimposable with an RMSD of 2.07Å (204 out of 266 amino acids; PDB access#: 1MRZ (Wang *et al*., 2005) & 2X0K (Herguedas *et al*., 2010), even though they only share about 30% protein sequence identity. The high structural and functional similarity suggested that deletion of the N-terminal adenylation domain to TmFADS could yield a more thermostable RFK. While heterologous expression of the full length enzyme produced a functional biocatalyst with moderate catalytic activity for RF relative to CaFADS, attempts to generate a truncated version of the *T. maritima* protein resulted in mostly insoluble protein aggregates (data not shown).

Characterization of secondary structure and stability by circular dichroism spectroscopy

Although high expression levels of soluble protein is a good indicator for the structural integrity of tcRFK and tcRFKhis, we studied their secondary structure composition by far-UV circular dichroism (CD) spectroscopy and quantitatively assessed their stability in thermodenaturation experiments. The far-UV spectra of the truncated variants showed a significant shift toward greater b-strand content compared with the typical mixed α/β structure spectrum of full-length FADS (**Figure 2.3**). The change is consistent with an analysis of the secondary structure elements assigned in the crystal structure of CaFADS. The bi-functional enzyme contains 11 α-helices and 22 β-strands, making up 28 and 29% of the 338-amino acid sequence, respectively. In contrast, the C-terminal domain of FADS which doubles as a model

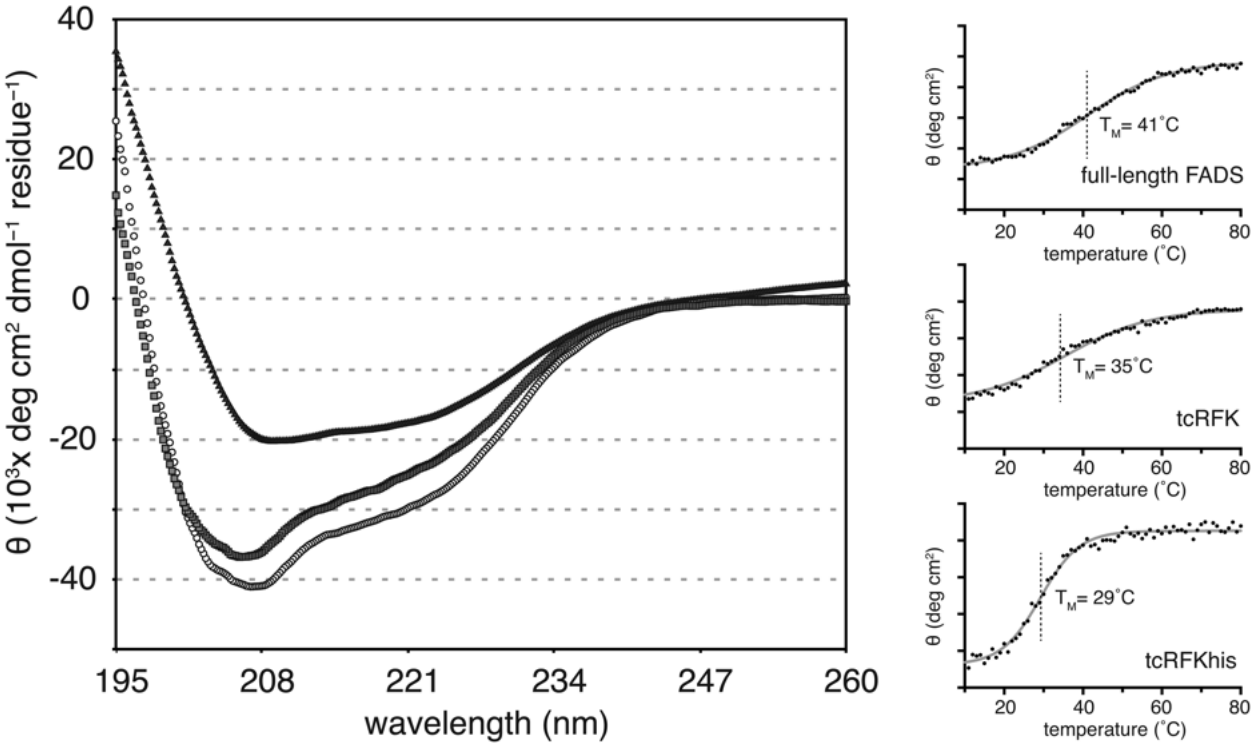


Figure 2.3. Far UV CD spectra and thermodenaturation data for full-length FADS (filled triangle), C-terminally truncated FADS (tcRFK, filled square) and His-tagged C-terminally truncated FADS (tcRFKhis, empty circle). For the thermodenaturation experiments, CD signal was measured at 222 nm from 10–80°C and T_M data was calculated based on nonlinear curve fit.

for tcRFK is considerably more b-strand rich, consisting of only one α-helix but eleven β-strands which shifts the percentages to 16 and 39% for the 151-residue protein, respectively. The experimentally observed changes in the far-UV CD spectra are consistent with such a predicted shift in secondary structure content.

Separately, we evaluated the stability of the full-length and truncated proteins via thermodenaturation in the CD spectrophotometer. Deletion of the N-terminal adenosyl transfer domain in the tcRFK variants results in a drop in the TM value from 40°C (parental CaFADS) to 35°C for tcRFK. The decline could at least in part be caused by the disruption of hydrogen bonding interactions and exposure of hydrophobic surface areas at the domain interface. Addition of the C-terminal poly-His tag further reduces the TM to 30°C, presumably due to the conformationally flexible tail formed by the extra amino acids. From a practical perspective, the lower temperature of unfolding for the tcRFK variants did not, however, impact the enzyme's application for FMN (analog) synthesis. Extended incubation of the biocatalyst in the Tris-HCl/ $MgCl_2$ reaction buffer at temperatures from 20 to 37°C did not result in a notable decline in activity. The discrepancy between CD and activity data can be explained by the switch in buffer salts from phosphate to Tris which is necessary due to limiting spectral properties and strong pH shifts associated with temperature change of Tris-HCl buffer.

Catalytic activity with RF and flavin analogs

Beyond the structural integrity of tcRFK, the catalytic properties of the truncated enzymes were first tested with RF. Product formation could easily be followed qualitatively by t.l.c. analysis while high performance liquid chromatography (HPLC) analysis of quenched reaction aliquots provided a quantitative measure for enzyme activity (**Figure 2.4**). Based on the

HPLC data, the rate of conversion for RF phosphorylation to FMN by tcRFKhis under standard reaction conditions (see Materials and Methods section) was determined as 0.9 μM/min/μg of enzyme and reached >95% completion (**Table 2.1**). The measured rate compares well with the previously reported kinetic parameter for the full-length CaFADS (Herguedas *et al*., 2010), suggesting that the deletion of the N-terminal domain does not alter the catalytic performance of our engineered RF kinase.

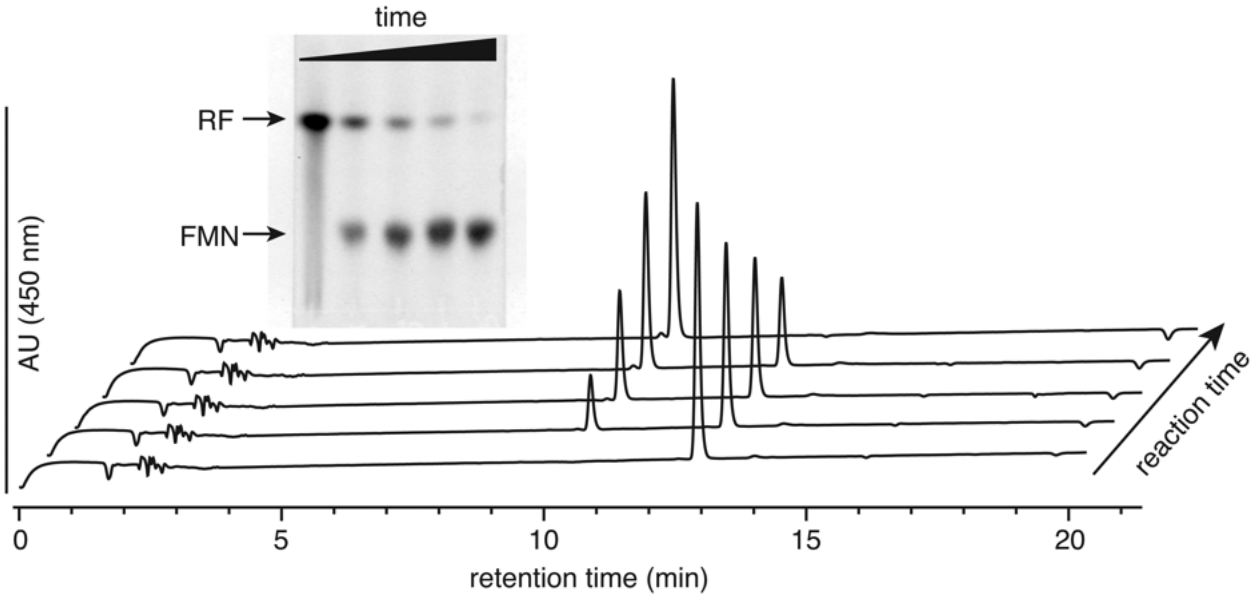


Figure 2.4. Enzymatic conversion of RF to FMN by tcRFKhis. The progress of the preparative-scale reaction was monitored by t.l.c. and HPLC analysis.

Besides phosphorylation of the natural substrate, we were particularly interested in the performance of tcRFKhis with RF analogs in order to establish a direct enzymatic route for preparing the corresponding FMN derivatives. To demonstrate the promiscuity of the engineered RF kinase, we tested three flavin analogs 7,8-dichloro-riboflavin (**1**), 8-amino-riboflavin (**2**) and 5-deaza-riboflavin (**3**) (**Figure 2.2**) as substrates for tcRFKhis. As for the native substrate, the reactions were monitored by t.l.c. and HPLC and the identity of products was confirmed by spectrophotometric analysis and mass spectrometry (data not shown). The flavin analogs were

Table 2.1. Biochemical/biophysical parameters for flavin analogs

Compound	rt (min)[a]	λ_{max} (nm)	ε (M^{-1} cm^{-1})	Conversion[b]	% conv.
RF	13.0	450	12,200	0.9 min^{-1}	>95%
FMN	10.4	450	12,200		
1	15.5	442	12,200	2.7 min^{-1}	>95%
1a	12.0	445	11,700		
2	9.6	478	46,600	5.9 min^{-1}	>95%
2a	7.8	482	44,000		
3	16.3	398	12,000	1.3 min^{-1}	~95%
3a	13.3	398	12,000		

[a]rt = Retention time based on HPLC system reported in Materials and Methods (see Supplementary Materials).
[b]Initial rate of conversion. Rate of product formation was linear for 30 min and reaction (for conditions see Supplementary Materials) reached 95% conversion after 60 min. Standard error ± 20%.

effectively and completely converted (>95%) except for **3** where the reaction reach 90–95%. The reason for the incomplete reaction of 3 is unclear but could be caused by small amounts of contaminants from its chemical synthesis. The measured rates of conversion for all analogs are similar to or better than that for RF. The dichloro (**1**) and 8-amino (**2**) derivatives were turned over approximately three- and six-fold faster, respectively, while the 5-deaza analog (**3**) was phosphorylated at a rate comparable to the native substrate (**Table 2.1**). We associate these rate changes with actual gains in catalytic turnover. The high substrate concentration (50 μM to 0.5 mM) in our experiments relative to the reported KM value for RF of <1 mM (Efimov *et al.*, 1998) should more than compensate for possible differences in substate binding affinity.

Finally, the tcRFK variants were shown to function as robust biocatalysts for preparative-scale synthesis of FMN (analogs). Although substrate concentrations of >0.5 mM are problematic due to limited solubility of the RFs, the reaction was successfully scaled-up to 100

ml to enable conversion of milligram quantities typically required for laboratory experiments. To simplify product isolation and assist in the recovery of the biocatalyst, we chose to immobilize the tcRFKhis variant on Ni-NTA agarose. In experiments with the native RF and our three RF analogs, the performance of the immobilized enzyme was indistinguishable (monitored by t.l.c.) from the homogenous reactions at comparable enzyme loads.

In summary, tcRFK and tcRFKhis are two stable variants of the C-terminal domain from CaFADS. The truncated enzymes represent an efficient biocatalytic system for the phosphorylation of RFs, benefiting from the robustness and simple expression/purification of the bacterial kinases while eliminating the need for two separate enzymatic reactions for the synthesis of FMN from RF via FAD. Furthermore, the substrate promiscuity of tcRFK makes these biocatalysts highly suitable for the phosphorylation of RF analogs to their corresponding FMN anabolites.

III. Supplemental Information

Materials and Methods

Materials: Reagents and chemicals were purchased from Sigma-Aldrich (St. Louis, MO) unless indicated otherwise. Ligase and restriction endonucleases were obtained from New England Biolabs (NEB, Ipswich, MA). Pfu DNA polymerase (Stratagene, LA Jolla, CA) was used for the PCR. Plasmid DNA was isolated using the QIAprep Miniprep Kit and PCR products were purified with QIAquick PCR Purification Kit (Qiagen, Valencia, CA).

Riboflavin analog synthesis: The RF derivatives were synthesized as previously reported in the literature. The syntheses of 7,8-dichloro-riboflavin (**1**) and 8-amino-riboflavin (**2**) were

performed as reported (Lambooy *et al*., 1961, Kasai *et al*., 1987). The preparation of 5-deaza-riboflavin (**3**) followed the report by Yoneda (Yoneda, 1980) although Vilsmeier reagent was substituted with triphosgene (Murthy and Massey, 1997).

Gene cloning, protein expression and purification: The wild type gene for Corynebacterium ammoniagenes FADS (NCBI#: D37967.1) in pET-23a was originally obtained from Dr. W.S. McIntire (Efimov *et al*., 1998). The plasmid was used as template for isolating the DNA sequence that encodes for the FMN synthetase activity in the C-terminal fragment of the two-domain enzyme (corresponding to amino acid residues 187-338; numbering based on full-length Ca FADS; PDB access number: 2X0K (Herguedas *et al*., 2010). Following the PCR with sequence-specific flanking oligonucleotide primers (forward: 5'-CGC-CAT-ATG-TTT-TAT-GTC-ACA-GGT-3'; reverse: 5'-CGC-CTC-GAG-TTA-GCT-TTCTGC-TTG-3'; restriction sites marked in italics, STOP codon is underlined), the gene fragment was cloned into pET-23a using restriction sites NdeI and XhoI.

The same gene fragment including a poly His-tag at the C-terminus was prepared by PCR using the above forward primer and a reverse primer in which the STOP codon was mutated to a glycine codon (underlined) (5'-CGC-CTC-GAG-CCC-GCT-TTC-TGC-TTG-3'). Following restriction digestion, ligation of the PCR product into pET-23a fused the vector's affinity tag to the C-terminus of our target gene. All gene sequences were confirmed by DNA sequencing (Genewiz, South Plainfield, NJ).

For protein overexpression, plasmids were transformed into chemically competent *E. coli* BL21(DE3)-RIL and cultured on 2-YT-agar plates containing ampicillin (100 mg/mL) and chloramphenicol (34 mg/mL) at 37°C. Individual colonies were picked to inoculate overnight

cultures, using 2 mL 2-YT media supplemented with antibiotics at 37°C. An aliquot of the overnight culture was used to inoculate a fresh 250 mL culture of 2-YT supplemented with antibiotics. Cultures were grown to an OD600 of ~0.5, followed by induction with a final concentration of 0.3 mM IPTG and incubation at 20°C overnight. Cells were spun down at 4˚C and 4000 *g* for 20 min and resulting pellets stored at -20˚C until further purification.

For protein purification, cell pellets were resuspended in buffer A (40 mM Tris-HCl pH 8.0, 20 mM NaCl) supplemented with 5 μL benzonase (Novagen), 50 μL protease inhibitor cocktail, 1 mM EDTA, 1 mM DTT, and 1 μM PMSF. The mixture was incubated on ice for 30 min. Cells were lysed by sonication (10-sec pulse/20-sec pause for 4 min), followed by centrifugation at 4°C and 8000 *g* for 30 min. Wild type (full-length) Ca FAD synthetase was purified by ion exchanged chromatography. Briefly, a 1-mL HiTrap DEAE FF column (GE Healthcare) was equilibrated with 10 column volumes (CVs) of buffer A, 10 CVs of buffer B (40 mM Tris-HCl pH 8.0, 1 M NaCl) , and again with 5 CVs of buffer A. After loading of the clear cell lysate, protein was eluted by applying a linear gradient to 100% buffer B over 20 min. Ca FAD synthetase was recovered between 15-50% buffer B. The combined product fractions were further purified via size exclusion chromatography (Superdex 200, 10/300 GL column, equilibrated with buffer C (50 mM Tris-HCl pH 8.0, 200 mM NaCl)) and target protein was concentrated by Millipore Centricon spin columns (MWCO: 10K). For the purification of the C-terminal fragment of Ca FAD synthetase, the clear cell lysate was loaded on a 5 mL-HiTrap Q FF column (GE Healthcare) pre-equilibrated in buffer A. Upon applying a linear gradient of buffer B, target protein eluted between 28-36% buffer B. The protein was further purified by size exclusion chromatography as described above. Product fractions were analyzed by SDS-PAGE and found to be of >80% purity. Purified proteins were flash frozen in liquid nitrogen and stored

at -80°C. For purification of the His-tagged C-terminal fragment of Ca FAD synthetase, cell pellets were resuspended in buffer D (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM imidazole). Following cell lysis by sonication and centrifugation as described above, the clear supernatant was incubated with 1 mL Ni-NTA agarose resin (Qiagen) pre-equilibrated with buffer D at 4°C for 1 h. The suspension was loaded on a PolyPrep-column (BioRad, Carlsbad CA) and the resin washed with 5 CVs of buffer D. Target protein was then recovered with elution buffer (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 250 mM imidazole). The combined product fractions were dialyzed (MWCO: 10-kDa cellulose membrane; Sigma-Aldrich) against 3 L of storage buffer (50 mM Tris-HCl pH 8.0, 200 mM NaCl) at 4°C for 4 h. Purified protein was flash frozen in liquid nitrogen and stored at -80°C. In contrast to previous reports by Efimov regarding the co-purification of FAD with full-length CaFADS (Efimov *et al*., 1998), we did not detect an absorbance band at 450 nm in our tcRFK variants, eliminating the need for additional purification steps to remove native flavin cofactor which otherwise would contaminate our experiments with RF analogs.

Enzyme activity assay by t.l.c.: To detect FMN synthetase activity, 1 μM of enzyme was added to 150 μL reaction mixture containing 50 μM riboflavin (or corresponding substrate analogs), 200 μM ATP, 10 mM MgCl2 in 50 mM Tris-HCl pH 8.0. The reaction mixture was incubated at 37°C and aliquots were taken after 0, 30, and 60 min. Product formation was monitored by spotting reaction samples on t.l.c. plates (Silicagel60/F254 on aluminum; EMD Chemicals) and separating the reaction components in butanol:acetic acid:water (3:1:1) (Frago *et al*., 2008). Starting material and products were visualized at 254 nm with the help of a handheld UV light.

Enzyme activity assay by HPLC: HPLC was used as a quantitative method to monitor the enzymatic FMN (analog) synthesis. For activity assays at 37˚C, 0.1 to 1 μM enzyme was added to 1 mL reaction mixture containing 50 μM riboflavin (or corresponding substrate analogs), 1 mM ATP, 10 mM MgCl2 in 50 mM Tris-HCl pH 8.0. Aliquots (150 μL) were removed from the reaction mixture after 0, 5, 10, 15, and 30 min and immediately quench with 30 uL of 100% TCA, followed by storage on ice for 5 min. After centrifugation at 8000 g for 10 min, the clear supernatant was neutralized with 1.6 mM NaOH and 100 μL solution was used for HPLC analysis (Agilent 1200 series; Eclipse Plus C18: 5μm, 4.6 x 250mm, equipped with guard column; Agilent Technologies). Samples were separated by running a linear gradient of (A) 50 mM ammonium acetate (pH 6) and (B) 70% acetonitrile (in solvent A) at: t = 0 min / 95:5 (A:B); t = 20 min / 40:60 (A:B); t = 22 min / 40:60 (A:B); t 29 min / 95:5 (A:B) with detection at 450 nm and 262 nm (for RF/FMN; wavelenght adjusted according to spectral properties of samples based on UV analysis (**Figure 2.2**) (Cao *et al*., 2008, Krauss *et al*., 2011). Assays were performed in triplicate and data were analyzed in Prism6 (GraphPad, La Jolla, CA).

Preparative-scale conversion of riboflavin (analogs): For preparative-scale synthesis of FMN (analogs), 0.2 mg of substrate ([final] 0.5 mM, dissolved in DMF) was added to a 1 mL reaction volume containing enzyme (10 μM) in reaction buffer (50 mM Tris-HCl pH 8, 100 mM MgCl2, 10 mM ATP). The reaction mixture was incubated at ambient temperature and product formation was followed by t.l.c. as described above. Under the given reaction conditions, substrate conversion was typically completed within 60 min. FMN (analogs) was purified via gravity-flow over Sep-Pak C18 columns (Waters, Milford MA). Product identity was confirmed by ESI-MS analysis (Zhang *et al*., 2003).

In addition to substrate conversion in solution, FMN synthesis could also be performed with immobilized artificial FMN synthetase. Purified enzyme was diluted to 1 μM in 5 mL reaction buffer (50 mM Tris-HCl pH 8.0) and bound to 1 mL Ni-NTA agarose resin (pre-equilibrated with same buffer) at 4°C for 1 h. Excess buffer was discarded and replaced with fresh reaction buffer containing 50 μM riboflavin, 0.2 mM ATP, 10 mM MgCl2, followed by incubation of the reaction mixture at room temperature. Product conversion was monitored by t.l.c as outlined above.

Circular dichroism spectroscopy: The secondary structure and stability of all enzyme variants were assessed using a Jasco J-810 CD spectropolarimeter plus a Jasco PFD-4252 water bath to regulate temperature. All samples had a concentration of 0.5 mg/mL in 50 mM potassium phosphate pH 8.0. The enzymes were monitored from 195-260 nm in a 1 mm quartz cuvette in triplicate with a band width of 2 nm and a scanning speed of 200 nm/min. The raw data was normalized to give mean molar absorptivity (deg·cm2·dmol-1). To obtain the melting temperature for the enzymes the unfolding events were monitored at 222 nm from 4 to 85°C with a temperature slope of 1°C per minute and a band width of 4 nm.

IV. References

1. Bacher,A., Eberhardt,S., Fischer,M., Kis,K. and Richter,G. (2000) Annu. Rev. Nutrit., 20, 153–167.
2. de Gonzalo,G. and Fraaije,M.W. (2013) Chem. Cat. Chem., 5, 403–415. Efimov,I., Kuusk,V., Zhang,X. and McIntire,W.S. (1998) Biochemistry, 37, 9716–9723.
3. Frago,S., Martı´nez-Ju´lvez,M., Serrano,A. and Medina,M. (2008) BMC Microbiol., 8, 1 – 16.
4. Ghisla,S. and Edmondson,D.E. (1999) In Ghisla,S., Kroneck,P., Macheroux,P. and Sund,H. (eds), Flavins and Flavoproteins. Konstanz, Germany: Agency for Scientific Publications, pp. 71– 76.
5. Ghisla,S. and Massey,V. (1986) Biochem. J., 239, 1 –12.
6. Hagihara,T., Fujio,T. and Aisaka,K. (1995) Appl. Microbiol. Biotechnol., 42, 724–729.
7. Herguedas,B., Martinez-Julvez,M., Frago,S., Medina,M. and Hermoso,J.A. (2010) J. Mol. Biol., 400, 218–230.
8. Hersh,L.B. and Walsh,C. (1980) Meth. Enzymol,, 66, 277–287.
9. Joosten,V. and van Berkel,W.J.H. (2007) Curr. Opin. Chem. Biol., 11, 195–202.
10. Karthikeyan,S., Zhou,Q., Mseeh,F., Grishin,N.V., Osterman,A.L. and Zhang,H. (2003) Structure, 11, 265– 273.
11. Krauss,U., Svensson,V., Wirtz,A., Knieps-Grunhagen,E. and Jaeger,K.E. (2011) Appl. Environ. Microb., 77, 1097–1100.
12. Macheroux,P., Kappes,B. and Ealick,S.E. (2011) FEBS J., 278, 2625–2634.
13. Mack,M. and Grill,S. (2006) Appl. Microbiol. Biotechnol., 71, 265– 275.
14. Merrill,A.H. and McCormick,D.B. (1980) J. Biol. Chem., 255, 1335–1338.
15. Miller,J.R. and Edmondson,D.E. (1999) J. Biol. Chem., 274, 23515–23525.
16. Miura,R. (2001) Chem. Rec., 1, 183– 194.
17. Murthy,Y.V. and Massey,V. (1997) Meth. Enzymol., 280, 436–460.
18. Shin,D.H., Wang,W., Kim,R., Yokota,H. and Kim,S.H. (2013) Crystal structure of ADP bound FAD synthetase. Berkeley Structural Genomics Center.
19. Spencer,R., Fisher,J. and Walsh,C. (1976) Biochemistry, 15, 1043–1053.
20. Taylor,M., Scott,C. and Grogan,G. (2013) Trends Biotechnol., 31, 63–64.

21. Walsh,C.T., Fisher,J., Spencer,R., Graham,D.W., Ashton,W.T., Brown,J.E., Brown,R.D. and Rogers,E.F. (1978) Biochemistry, 17, 1942–1951.
22. Walsh,C.T. and Wencewicz,T.A. (2013) Nat. Prod. Rep., 30, 175–200.
23. Wang,W., Kim,R., Yokota,H. and Kim,S.H. (2005) Proteins, 58, 246– 248.
24. Yatsyshyn,V.Y., Ishchuk,O.P., Voronovsky,A.Y., Fedorovych,D.V. and Sibirny,A.A. (2009) Metabol. Eng., 11, 163– 167.

Chapter 3: Elucidating the Role of FMN in the Productive Folding of Old Yellow Enzyme 1

I. Introduction

Old Yellow Enzymes (OYEs) are a family of FMN-depenent ene-reductases which perform *trans*-hydrogenations with high regio-, enantio-, and stereospecificity. OYEs derive their catalytic power from a non-covalently bound FMN cofactor (**Figure 3.1**). Generally speaking, FMN and related flavin cofactors are catalytically versatile (see Chapter 1, Section III) and have been shown to be tunable for a variety of biotechnological applications. For example, modifications to the 8 position of the flavin are known to modulate redox potential (1-3) while modifications to the 5 position can alter reactivity of the flavin (4-6). As a result of this mutability, cofactor replacement in flavoenzymes with chemically-altered cofactor analogs has proven a useful strategy for studying mechanistic details and elucidating specific aspects of the active site environment. While the roles flavin cofactors play with respect to catalysis have been extensively studied, the involvement of flavins in the productive folding of nascent flavoenzymes *in vivo* is much less clearly understood as there are currently only a few literature accounts describing attempts to study this phenomenon (7-9).

Given that flavins are essential cofactors and are thus ubiquitously dispersed in all living organisms, the folding of apoproteins has been immensely difficult to study. Consequently, most published studies to date relied on a strategy involving stripping the native flavin cofactor out of a flavoprotein and then reintroducing it back into the scaffold to draw inferences regarding the respective protein folding pathways. Bollen *et al.* were the first to propose any correlation between the binding of FMN and the role of the cofactor in the folding of the flavodoxin from *Azotobacter vinelandii* (7). The group coined the phrase “Last In, First Out,” believing that since they were able to remove the FMN and then reconstitute it back into the apoprotein with no

apparent difference in the recorded stabilities of the apo- and holoproteins, the flavodoxin must therefore incorporate its FMN cofactor after the protein has assumed its folded tertiary structure. While this ability to remove and reconstitute flavins is common amongst FMN-dependent proteins, there remains insufficient evidence to assert that FMN plays no role in the folding of the flavodoxin given that the flavodoxin was synthesized in the presence of FMN prior to the removal of the cofactor. A second group, Abe *et al.*, utilized a wheat germ cell-free protein synthesis system in order to determine the role FMN plays in the folding of a FMN-binding protein from *Desulfovibrio vulgaris* (8). The primary advantage of this method over the study performed by Bollen *et al.* is that the utilization of the cell-free system permitted the synthesis of the FMN-binding protein in the absence of FMN. After synthesis of the apoprotein, exogenous FMN was then introduced into the apoFMN-binding protein, resulting in successful reconstitution of native protein functions. The authors concluded that FMN is thus not involved in the productive folding of the FMN-binding protein. However, upon closer inspection of the absorbance spectra collected for the apoFMN-binding protein, absorbance peaks corresponding to the characteristic FMN absorbance signature can be observed in the protein sample. These spectra indicate that trace amounts of FMN exist in the wheat germ cell extract used for cell-free "apoprotein" synthesis, and subsequently instill doubts surrounding the authors' claim that the binding protein was synthesized in the total absence of cofactor.

In general, the involvement of FMN in the productive folding of flavoproteins remains ambiguous due to the inability to definitively synthesize flavoproteins in their apo-forms. Fortunately, recent advancements in cell-free protein expression technologies have allowed us to address this unanswered question more thoroughly using the cell-free PURE (Protein synthesis

Using Recombinant Elements) *in vitro* transcription/translation system (10, 11). The major advantage of the PURE system over the previous wheat germ cell-free system utilized by Abe *et al.* lies with the fact that all of the components within the PURE system are chemically defined, effectively minimizing the possibility of free FMN contamination during apoprotein synthesis. Accordingly, the PURE system allowed us to not only synthesize apoOYE1, but also to synthesize OYE1 in the presence of various FMN analogs. In so doing, we have defined a systematic approach for probing which structural aspects of the cofactor are necessary to facilitate proper folding of the nascent enzyme. This approach allowed us to repeat the experiment performed by Abe *et al.* with OYE1, leading us to conclude from the resulting data that the cofactor is involved in the productive folding of OYE1. Furthermore, our methodology

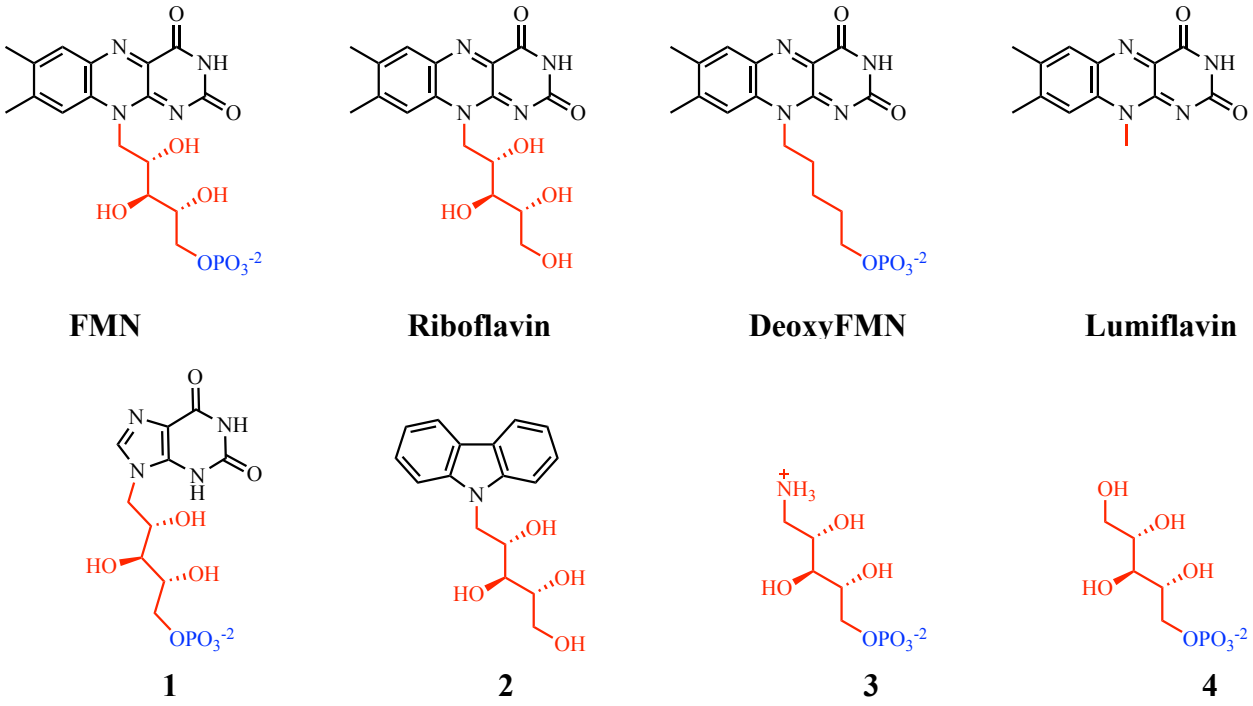
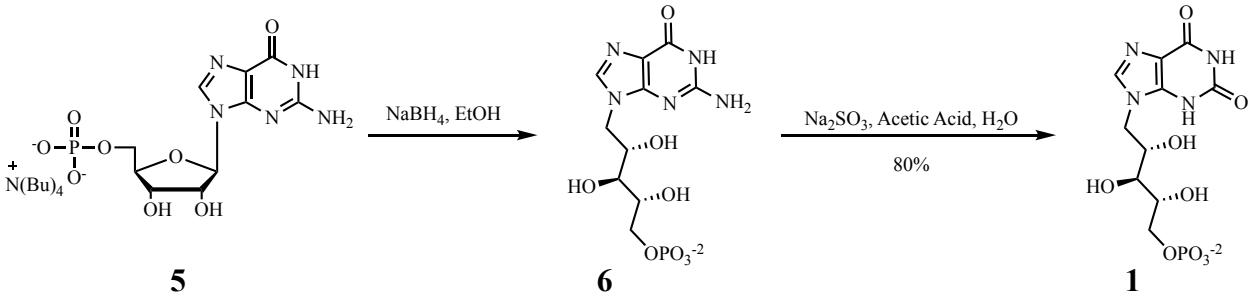


Figure 3.1. Riboflavin and FMN (Analogs). The different moieties of the cofactors are colored to highlight the differences: isoalloxazine (black), ribityl tail (red), and phosphate group (blue).These cofactors were used in the folding studies of OYE1. Cofactors **1-4** are novel analogs.

allowed us to determine the exact structural components of FMN that directly contribute to folding.

To probe the folding of OYE1, we have developed four novel FMN analogs (**Figure 3.1**, **1**-**4**), which each contain a specific structural feature derived from FMN. First, (2*R*,3*S*,4*S*)-5-(2,6-dioxo-1,2,3,6-tetrahydro-9H-purin-9-yl)-2,3,4-trihydroxypentyl phosphate (**1**), the reduced form of xanthosine monophosphate lacking the hydrophobic portion of the isoalloxazine moiety, was chemically synthesized to probe the importance of the isoalloxazine's hydrogen bonding capacity. Next, (2*R*,3*S*,4*S*)-5-(9H-carbazol-9-yl)pentane-1,2,3,4-tetraol (**2**), a carbazole derivative lacking any of the hydrogen bonding groups of the isoalloxazine moiety, was synthesized to elucidate the impact of native FMN's hydrophobic components. Lastly, analogs **3** and **4**, (2*R*,3*S*, 4*S*)-5-ammonio-2,3,4-trihydroxypentyl phosphate and (2*R*,3*S*,4*S*)-2,3,4,5-tetrahydroxypentyl phosphate, respectively, completely lack the isoalloxazine moiety and allow us to monitor the contributions of the flavin's ribityl tail during protein folding. Given that the primary amine of **3** will be protonated at physiological pH, which could negatively interfere with binding, analog **4** was designed to present no formal charges at the same pH conditions. Three additional cofactor analogs, riboflavin, deoxyFMN, and lumiflavin, were also tested (**Figure 3.1**) as these analogs have traditionally been utilized in similar studies designed to determine the structural aspects of FMN that contribute to cofactor binding within apoproteins (12 13). We believe that our panel of FMN analogs effectively samples the various functional groups of FMN in order to elucidate which component(s) are necessary for the proper folding of OYE1.

Figure 3.2. Synthesis of Analog **1**.

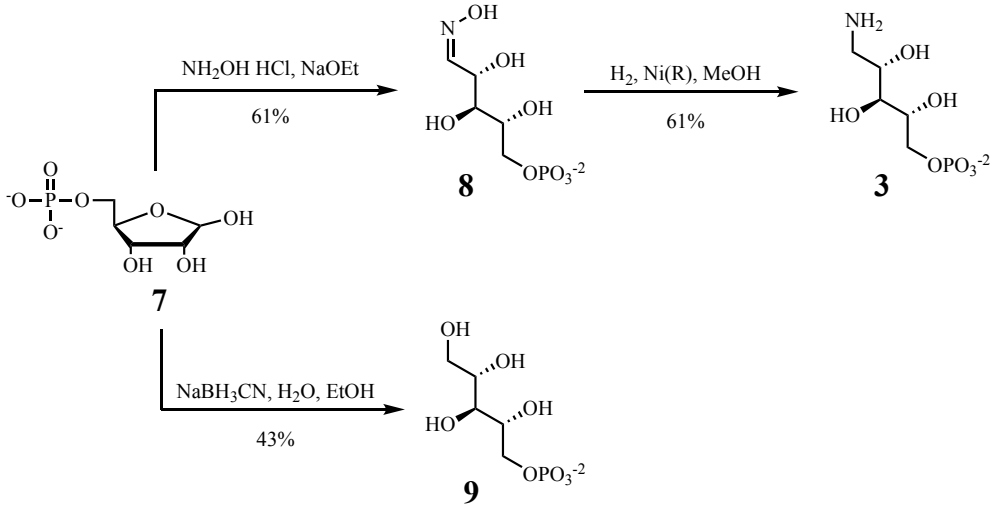


II. Synthesis of FMN Analogs

*Cofactor Analogs **1, 3, 4***

Analogs **1**, **3**, and **4** were synthesized according to literature procedures with modifications to fit our particular compounds. In all cases creation of the phosphorylated ribityl tail originated from a monophosphorylated sugar. Specifically, for analog **1**, synthesis started from guanosine monophosphate (**5**, **Figure 3.2**) while for **3** and **4** synthesis started from ribose-5-phosphate (**7, Figure 3.4**). These cyclic sugars were then reduced to their linear forms based on literature procedures (14-16). For **1**, deamination of the guanosine base is unique in that it requires the generation of nitrous acid *in situ* as previously reported by Van Aerschot *et al*. (17).

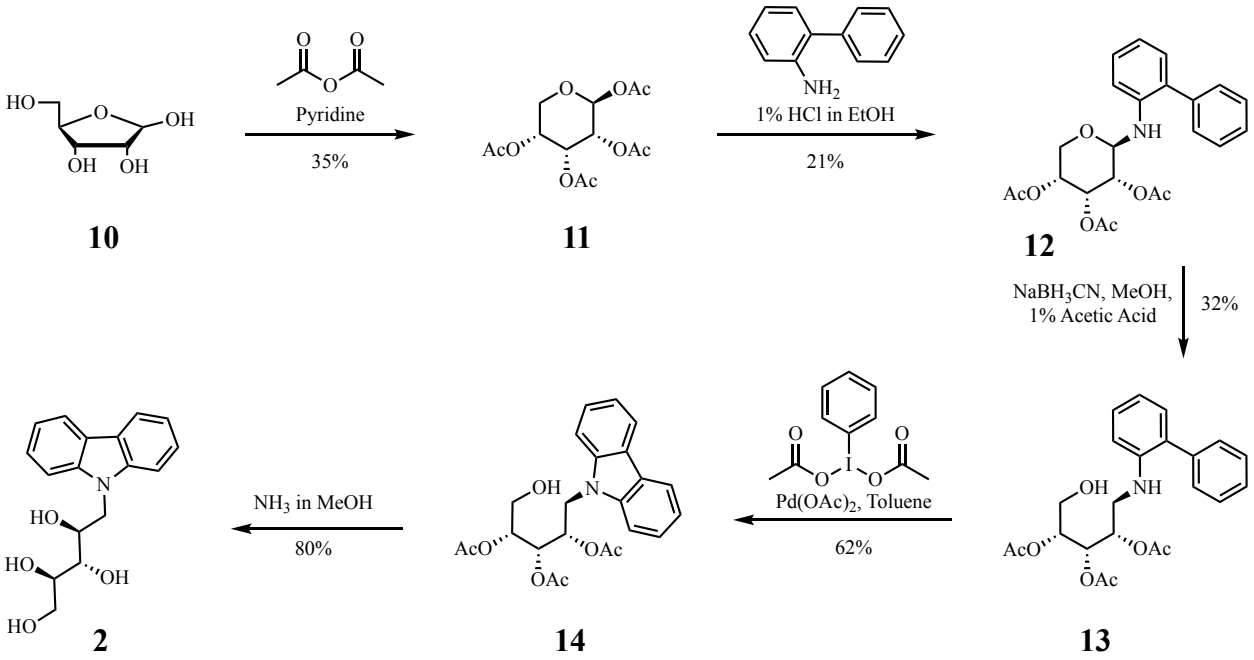
Figure 3.3. Synthesis of Analogs **3** and **4**.



*Cofactor Analog **2***

Unlike the other cofactor analogs analogs, **2** does not start from a phosphorylated sugar due to the incompatibility of the phosphate group with many of the organic solvents needed to couple the carbazole moiety to the sugar. Therefore, synthesis of **2** required a synthetic route with a combination of protection and deprotection steps intended to ensure selective phosphorylation (**Figure 3.4**). Acetate groups were chosen to protect the secondary hydroxyls for their relatively small size and ease of installation and removal. Fortuitously, reaction conditions for the acetate protection led to an isomerization of the sugar from a furanose to pyranose. This isomerization simultaneously protected the secondary hydroxyls with acetate groups and the 5'-hydroxyl was now bound to the C1. The original synthetic route involved a tin(II) chloride-mediated coupling of the carbazole moiety to compound **11**; however, the carbazole moiety hindered the reduction of the cyclic sugar in the subsequent step. Multiple reducing agents were employed albeit with

Figure 3.4. Synthesis of Analog **2**.



no success. We speculated that the lone pair on the nitrogen of the carbazole is part of the aromatic system, which hinders the necessary formation of the iminium intermediate needed for reduction of the sugar. To overcome this issue we instead coupled 2-aminobiphenyl, the precursor to carbazole, to compound **11** through an acid catalyzed reaction to yield **12**. Mild reducing conditions allowed for the linearization of the sugar to afford **13**. Unfortunately, these conditions resulted in a mixture of products which we suspected were isomers of the protected product, **13**, due to migration of acetate groups as there were three sets of acetate groups seen in by ^{1}H-NMR. We confirmed our theory by deprotecting a portion of **13** and observing convergence of peaks via ^{1}H-NMR over time. Efforts were made to prevent the migration of the protecting groups by replacing the acetate groups with benzoyl groups, as well as syilated protecting groups. Unfortunately, the bulkiness of the benzoyl groups prevented the complete protection of the ribose and the presence of the syilated groups prevented the ring closure of the 2-aminobiphenyl to carbazole. In the end, we decided to move forward with the acetate protecting groups, utilizing the mixture of **13** in the next step. The carbazole ring system was formed according to literature procedures to afford **14** (18). At this step chemical phosphorylation was attempted under the impression that only the isomer with the unprotected 5’-hydroxyl would react. However, we never saw formation of product to confirm this theory. Subsequently, we decided to fully deprotect compound **14** and attempt phosphorylation of the resulting compound, **2**. Attempts to phosphorylate **2** were unsuccessful as we were unable to achieve selective phosphorylation of the 5’-hydroxyl group in either high purity or yield. We were, however, able to separate the three phosphorylated products with roughly 80% purity, allowing by NOESY that one of the products was the desired phosphorylated **2** compound while

the other two products had their secondary hydroxyl groups phosphorylated. After multiple attempts to increase both yield and purity of the desired phosphorylated version of analog **2**, we eventually settled on using the unphosphorylated compound for our binding studies. Ultimately, with the help of Dr. Huanyu Zhao (Horizon Therapeutics, Chicago, IL), synthesis of analog **2** was performed in five steps with a 1.2% overall yield.

III. Results and Discussion

ITC Binding Studies of FMN (analogs)

The binding affinity and stoichiometry of the respective analogs for the OYE1 scaffold were determined by isothermal calorimetry (ITC) studies. To perform these studies, OYE1 was heterologously expressed and purified in the presence of the native FMN, which was subsequently removed via dialysis with 5 M potassium bromide to produce apoOYE. Individual cofactors or analogs were titrated into a solution of apoOYE and thermal changes associated with binding were monitored (**Figure 3.5**). Native FMN exhibited a K_D of 0.047 μM with a 0.7 binding stoichiometry while the deoxyFMN had a 28-fold higher K_D of 1.3 μM with a 1.3 binding stoichiometry. The large difference in K_D between FMN and deoxyFMN suggests that while the secondary hydroxyls do contribute to binding affinity, they are not essential for cofactor binding. This result is consistent with previous studies in which apoOYE1 displayed a similarly decreased binding affinity for deoxyFMN (19). It is well known that OYE1 binds one molecule of FMN (20); however, the discrepancy in binding stoichiometry for both FMN and deoxyFMN, could be due to the error associated with determining apoOYE1 concentration. For the remainder of the analogs (**Figure 3.5**, **B**, **D-H**), none of them bound to the apoOYE1 scaffold

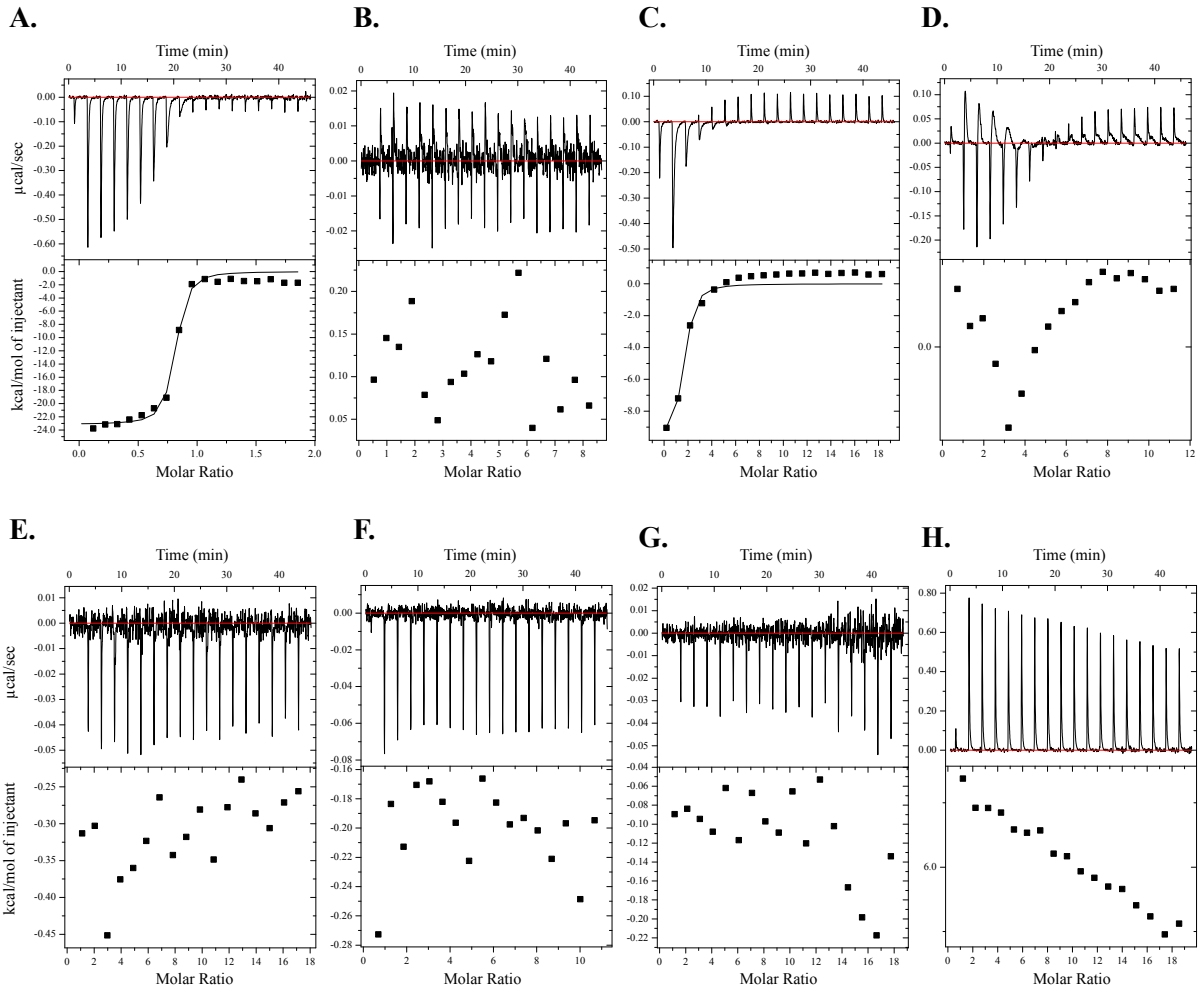


Figure 3.5. FMN (analog) binding to apoOYE1 probed by ITC. **A.** 173.8 µM FMN titrated against 17.4 µM apoOYE1, **B.** 463 µM riboflavin titrated against 10.4 µM apoOYE1, **C.** 852 µM deoxyFMN titrated against 8.6 µM apoOYE1, **D.** 1.05 mM lumiflavin titrated against 17.4 µM apoOYE1, **E.** 556 µM **1** titrated against 6 µM apoOYE1, **F.** 1.73 mM **2** titrated against 17.4 µM apoOYE1, **G.** 1.0 mM **3** titrated against 10.4 µM apoOYE1, **H.** 523 µM **4** titrated against 5.23 µM apoOYE1. In each case, FMN (analog) was in the syringe and apoOYE1 was in the cell. The upper panels include the heat changed observed upon injections of 2 µL into the cell spaced by 2.5 min intervals. The red line represents the baseline. The lower panels show the resulting binding isotherms. When applicable solid lines represent fits to a single set of binding sites.

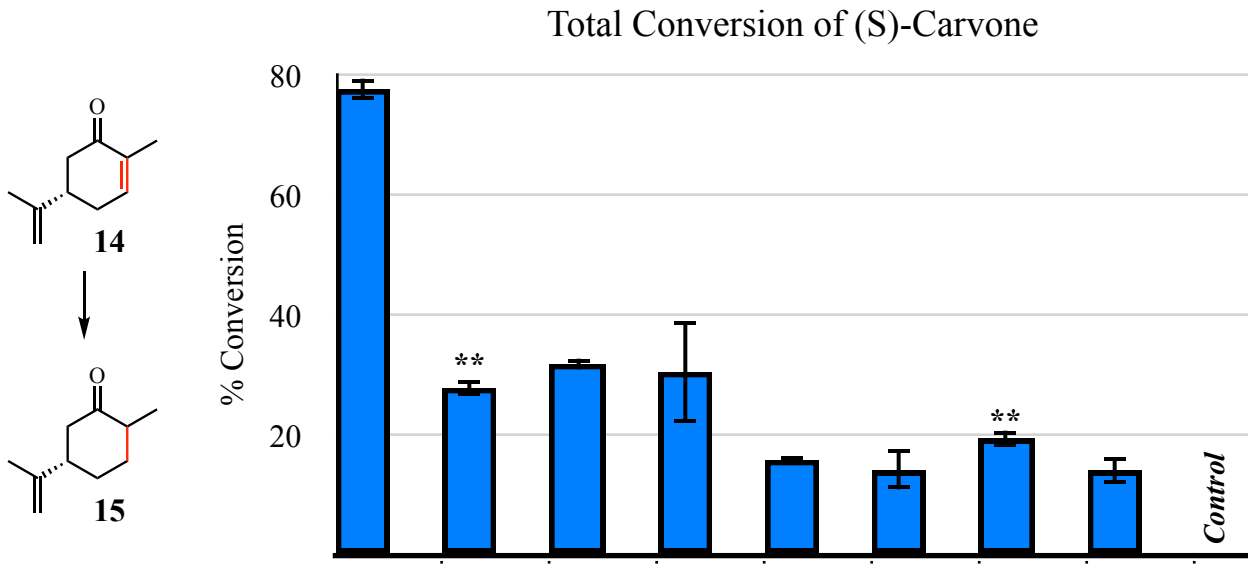
when added at a concentration 10-fold to 100-fold greater than the amount of apoenzyme. This range was chosen to ensure full coverage of possible binding affinities. Based on these results it would appear that the complete isoalloxazine moiety is a major contributor to binding as analogs **1**-**4** lack this functional group.

A lack of any measurable binding for riboflavin in particular is surprising given the similarity in structure between riboflavin, FMN, and deoxyFMN, suggesting that the terminal phosphate group of FMN is crucial both for recognition and for binding of the cofactor to the pre-folded apoprotein. Further, the lack of riboflavin binding indicates that the isoalloxazine and the ribityl hydroxyls are insufficient to promote binding in the absence of the phosphate group. Collectively, the combination of the isoalloxazine and phosphate moieties are necessary for efficient cofactor anchoring within the enzyme. This result is in contrast to previous studies. For example, Lostao *et al.* studied the binding interactions of FMN, riboflavin, and lumiflavin within the apoflavodoxin from *Anabaena* PCC 7119 in order to determine the individual binding contributions of the isoalloxazine, ribityl tail, and the phosphate moieties (7). The authors determined that the isoalloxazine and phosphate moieties are the greatest contributors to binding while the hydroxyl groups of the ribityl tail contribute minimally. This claim is consistent with previous studies performed by Tollin *et al.* in which the authors measured the respective dissociation constants for FMN, deoxyFMN, riboflavin, and lumiflavin following reconstitution of the apoflavodoxin from *Azotobacter vinelandii* with these cofactors (6). These differential binding affinities are unsurprising since variations to active sites residues interacting with the isoalloxazine moiety or direct modifications to the isoalloxazine moiety can impact binding of the cofactor (12, 13, 21). Furthermore, while all of these flavoproteins share FMN as their cofactor, these results clearly establish that the binding mechanism for all flavoproteins is not the same.

PURE System Folding Studies of OYE1

Utilizing the information gleaned from the ITC studies, we sought to more thoroughly investigate which portions of the FMN cofactor are necessary for the folding of catalytically competent nascent OYE1. Given that FMN is an essential cofactor and cannot be entirely eliminated during protein overexpression *in vivo*, we turned to the cell-free PURE system to probe the cofactor-dependence of OYE1 folding (10, 11). In particular, the PURE system allows for complete control over the identity and concentration of cofactor present during protein synthesis. For our studies, OYE1 was synthesized in the presence of 100 μM of FMN, 100 μM cofactor analog, or in the absence of cofactor. However, due to the limited volume (25 μL) of the PURE reaction samples, the amount of synthesized enzyme was too low to perform traditional biophysical characterizations designed to assess the extent of protein folding. We thus assessed proper OYE1 folding using activity assays as unfolded or improperly folded enzyme should not be able to perform the native ene-reductase chemistry.

In order to properly assess enzymatic activity, cofactor analogs present during OYE1 synthesis, particularly non-catalytically active analogs such as lumiflavin and cofactors **1**-**4**, needed to be removed and reconstituted with the native FMN cofactor. Given our ITC results (see above) indicating that FMN has the highest binding affinity among the cofactor analogs tested with OYE1, we believed that the native cofactor would successfully outcompete the analogs for active site occupancy if FMN was introduced into the PURE reaction samples following protein synthesis. Accordingly, OYE1 was synthesized within the PURE system in the presence of various analogs, and then a 1.5-fold excess of FMN was added into each sample to displace any active site-bound analogs. FMN was allowed to incubate with synthesized protein



Enzyme	OYE1	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
	DHFR	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
IVTT Cofactor (100 µM)	FMN	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Riboflavin	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-
	DeoxyFMN	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-
	Lumiflavin	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-
	1	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-
	3	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-
Competition Cofactor	1.5x FMN	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Figure 3.6. Activity assay of OYE1 synthesized in the presence of FMN or analogs. Reaction of interest shown on left. DHFR (NEB) was used as a negative control to monitor if the presence of either cofactor analog or free FMN contributes to observed activity. The left-most column represents a positive control of OYE1 synthesized in the presence of FMN while the right-most column depicts a negative control of OYE1 synthesized in the absence of cofactor. All reactions were run in triplicate with the exception of those marked **, which were run five times and subsequently had outlier datapoints removed.

for 4 hours before activity was analyzed. Activity of the FMN-reconsituted OYE1 was then assessed by monitoring the conversion of (5*S*)-carvone (**14**) to (2*R*,5*S*)-dihydrocarvone (**15**). All samples were run in triplicate except for the samples containing riboflavin and analog **3**, which

were run 5 times. Individual trials from each were identified as outliers and were subsequently removed according to literature procedures during data processing (22). Upon repeating the study performed by Abe *et al.* mentioned above, synthesis of apoOYE1 and subsequent reconstitution with FMN, we did not observe any active enzyme, suggesting that the cofactor needs to be present during protein synthesis. These results contradict the postulate of "Last In, First Out" initially proposed by Bollen *et al*., indicating that the ability to remove FMN from heterologously expressed and purified OYE1 does not mean that the cofactor plays no role in the folding of the nascent enzyme.

Surprisingly, it appears that all of the cofactor analogs tested produced some level of properly folded protein with the samples synthesized in the presence of riboflavin, deoxyFMN, and lumiflavin displaying the greatest substrate turnover values in our activity assay (**Figure 3.6**). This result was unexpected as the ITC data suggest that none of the analogs bind to the apoenzyme. However, this discrepancy can be rationalized given that both our own ITC assay and previous accounts from the literature relied on reconstitution of apoOYE1 with cofactor analogs. In contrast the PURE system integrates the cofactor (analog) during protein synthesis. It appears that the combination of the ribityl phosphate group and the isoalloxazine moiety play the biggest role in the binding of cofactors to apoprotein while the activity assay data suggest that the isoalloxazine moiety alone appears to have the largest influence over the proper folding of nascent OYE1. Specifically, it appears that the amphipathic nature of the isoalloxazine moiety is critical for folding as cofactor analogs **1** and **2**, which each solely possess either the hydrophobic or the hydrophilic portions of the native isoalloxazine moiety, only minimally produce active

enzyme. However, it should be noted that analog **2** is missing the terminal phosphate group, which could have negatively impacted its incorporation during folding of the nascent protein.

IV. Conclusion

This is the first account in which OYE1 has been synthesized in the complete absence of FMN. Our results indicate that apoOYE1 synthesized in a cofactor-free environment does not yield active enzyme, allowing us to assert that the presence of the cofactor is essential during the synthesis of the enzyme and plays a role in the proper folding of catalytically active OYE1. These results contradict conclusions derived from previous studies with flavodoxin and FMN-binding protein systems. However, the cofactor binding site can vary greatly between flavoproteins, as evidenced by comparisons of various flavoenzymes, such as OYE1, NADPH cytochrome c, NADPH oxidase, pyridoxine phosphate oxidase, and glycolate oxidase, which each display unique affinities for FMN and FMN analogs (19). Specifically, OYE1 and NADPH cytochrome c seem to only weakly bind FMN analogs with modifications to the isoalloxazine moieties while NADPH oxidase, pyridoxine phosphate oxidase, and glycolate oxidase are particularly hindered by FMN analogs with modifications to the ribityl tail.

From our studies, we found that for apoOYE1 specifically, the phosphate group seems to be a crucial element for binding given the lack of riboflavin and lumiflavin binding during our ITC studies. Additionally, deoxyFMN bound weakly to apoOYE1 compared to FMN, indicating that the while the secondary hydroxyls of the ribityl tail do contribute to the high binding affinity of FMN, they are not obligately necessary. This result is somewhat surprising as a majority of the polar contacts between the FMN and the protein observed in crystallographic structures reside

between binding site residues and the combined phosphate group and ribityl tail. However, while the phosphate group is a major contributor to binding, the lack of binding for analogs **1**, **3**, and **4**, which each contain the ribityl tail and terminal phosphate group but lack the isoalloxazine moiety, strongly suggests that it is not the only structural aspect of FMN that influences binding. It would appear that the combination of a terminal phosphate group and the isoalloxazine moiety together are needed for effective cofactor incorporation. Further studies may also be necessary to address if the physical distance between the isoalloxazine moiety and the phosphate group also influences cofactor binding.

While we can garner a great deal or information pertaining to the environment of the cofactor binding site within a protein, this data does not always equate to an effective understanding of the role the cofactor plays in facilitating the folding of the nascent protein chain. With the use of the PURE system, we were able to synthesize OYE1 in the presence of various FMN analogs and collect what we believe to be sufficient evidence to conclude that FMN must be present during protein synthesis in order to produce catalytically active OYE1. Furthermore, we assert that the complete isoalloxazine moiety represents the greatest contributor to proper folding. The ribityl tail and phosphate group, though essential for cofactor binding, are not essential for folding as the OYE1 synthesized in the presence of lumiflavin produced nearly the same level of active enzyme as OYE1 synthesized in the presence of either riboflavin or deoxyFMN. In conclusion, the PURE express *in vitro* transcription/translation system represents an invaluable tool for the study of protein/cofactor interactions. The methodologies developed in this work can be broadly applied to both other members of the OYE family and other

flavoenzymes in general as a pathway to determine key structural differences between flavoproteins.

V. Experimental

General Information

The PURExpress kit was purchased from New England Biolabs (Ipswich, MA). Deoxyriboflavin and lumiflavin were gifts from Dr. Dale Edmondson (Emory University). All reagents were purchased from Sigma Aldrich (St. Louis, MO) unless otherwise specified. ITC spectra were obtained on a Malvern Panalytical (United Kingdom) MicroCal Auto-iTC200. Data acquired from the Auto-iTC200 was analyzed with MicroCal Analysis Launcher and Origin 7 (OriginLab, Northampton, MA). 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 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 200°C, MS mode, EI; detector voltage, 0.2 kV; mass range, 12-250 u; scan speed, 833 u/s. All GC/MS data were acquired by a GC/MS Lab Solutions software (Shimadzu).

*Synthesis of (2R,3S,4S)-5-(2,6-dioxo-1,2,3,6-tetrahydro-9H-purin-9-yl)-2,3,4-trihydroxypentyl phosphate (**1**)*

Guanosine monophosphate (320 mg, 786 μmole) was dissolved in water and loaded onto 5 mL of Dowex 50WX8-100 (H^+) ion-exchange resin. Column was washed with DI water until

the flow through was a neutral pH. Flow through was collected into 10 mL of ethanol containing tetrabutylammonium hydroxide (786 µL of 1.0 M solution, 786 µmole). Solution was concentrated down and lyophilized to afford **5** in 67% yield. ^{1}H NMR (399 MHz, Deuterium Oxide) δ 8.14 (s, 1H), 5.88 (d, *J* = 6.3 Hz, 1H), 4.74 (dd, *J* = 6.3, 5.1 Hz, 1H), 4.69 (s, 0H), 4.44 (dd, *J* = 5.2, 3.2 Hz, 1H), 4.27 (qd, *J* = 3.4, 1.4 Hz, 1H), 3.96 (dd, *J* = 4.9, 3.5 Hz, 2H), 3.32 – 2.95 (m, 19H), 1.60 (p, *J* = 7.7 Hz, 20H), 1.31 (h, *J* = 7.4 Hz, 18H), 0.90 (t, *J* = 7.4 Hz, 26H). ^{31}P NMR (162 MHz, Deuterium Oxide) δ 2.84.

Compound **5** (100 mg, 92 µmole) and $NaBH_4$ (14 mg, 368 µmole) were dissolved in ethanol (184 µL 500 mM) and refluxed for 1.5 h. Reaction was quenched with 1 M HCl and concentrated to remove ethanol and afford **6**. Without purification **6** was carried forward. To the solution of **6** in water 1 µL acetic acid was added followed by the addition of sodium sulfite (55 mg, 440 µmole). Reaction was stirred at room temperature for 3 h. Reaction was quenched with 1 M Tris HCl pH 7 and diluted to 50 mL with DI water to obtain a concentration of 2 mg/mL. Final product was purified by reverse phase semi-prepative HPLC (Supelco, Discovery C-18, 250 x 21.2 mm, 5 µm). Linear gradient from 0% to 10% B (Buffer A: 100 mM TEAA pH 7 Buffer B: 100% acetonitrile). HPLC purification yielded the desired product, **1**, in 80% yield. ^{1}H NMR (399 MHz, Deuterium Oxide) δ 8.06 (s, 1H), 5.88 (d, *J* = 6.1 Hz, 1H), 4.74 (d, *J* = 2.0 Hz, 1H), 4.70 (t, *J* = 5.6 Hz, 1H), 4.43 (dd, *J* = 5.2, 3.3 Hz, 1H), 4.29 (dd, *J* = 3.8, 2.0 Hz, 1H), 4.05 (t, *J* = 3.8 Hz, 2H). ^{31}P NMR (162 MHz, Deuterium Oxide) δ 0.53. ESI-MS: *m*/*z* 365.06904 $[M+H]^{+}$.

*Synthesis of (2R,3S,4S)-5-(9H-carbazol-9-yl)pentane-1,2,3,4-tetraol (**2**)*

D-ribose (**9**, 12 g, 80 mmole) was dissolved in pyridine (80 mL) followed by the addition of acetic anhydride (75.6 mL, 800 mmole). Reaction was stirred overnight at room temperature. Reaction was concentrated down and the product (**10**) was purified via crystallization in 35% yield. ^{1}H NMR (399 MHz, Chloroform-*d*) δ 6.03 (d, *J* = 4.7 Hz, 1H), 5.49 (t, *J* = 3.4 Hz, 1H), 5.15 (dt, *J* = 6.5, 3.4 Hz, 1H), 5.04 (ddd, *J* = 4.8, 2.7, 0.8 Hz, 1H), 4.11 – 3.79 (m, 3H), 2.14 (d, *J* = 0.9 Hz, 3H), 2.11 (d, *J* = 0.9 Hz, 3H), 2.10 (d, *J* = 0.9 Hz, 4H), 2.09 (d, *J* = 1.0 Hz, 4H).

Compound **10** (3.18 g, 10 mmole) and 2-aminobiphenyl (2.54 g, 15 mmole) were dissolved in ethanol (20 mL) along with catalytic amounts of concentrated HCl. The 2-aminobiphenyl was added in 3 portions over 3 days. The mixture was stirred at reflux for the entity of the reaction. Reaction was cooled to room temperature before solvent was evaporated off. Product was purified via flash chromatography and eluted with 20:80 diethyl ether/hexanes to produce **11** in 21% yield. ^{1}H NMR (399 MHz, Chloroform-*d*) δ 7.50 – 7.33 (m, 1H), 7.26 (s, 1H), 7.12 (tt, *J* = 6.4, 1.5 Hz, 0H), 7.02 – 6.85 (m, 1H), 5.71 – 5.56 (m, 0H), 5.50 (s, 0H), 5.29 (dd, *J* = 10.4, 5.7 Hz, 0H), 5.19 – 5.09 (m, 0H), 5.04 – 4.92 (m, 0H), 4.86 (ddd, *J* = 9.0, 2.9, 1.3 Hz, 0H), 4.65 (d, *J* = 8.3 Hz, 0H), 4.12 (qd, *J* = 7.1, 1.3 Hz, 0H), 3.98 – 3.72 (m, 0H), 3.54 (dd, *J* = 11.6, 4.7 Hz, 0H), 2.18 (d, *J* = 1.3 Hz, 0H), 2.04 (d, *J* = 1.3 Hz, 1H), 2.01 (d, *J* = 1.4 Hz, 0H), 1.96 (d, *J* = 1.3 Hz, 0H), 1.90 (d, *J* = 1.2 Hz, 0H).

Compound **11** (0.427 g, 1 mmole) was combined with $NaBH_3CN$ (0.1 g, 1.5 mmole) in methanol (10 mL) and a drop of glacial acetic acid was added at 0°C. Reaction was allowed to slowly warm up to room temperature and stirred overnight. Product was extracted 3 times into

ethyl acetate, washed with brine, and dried over Na_2SO_4. Product was purified via flash chromatography and eluted with 40:60 ethyl aceteate/hexanes to yield **12** in 32% yield. ^{1}H NMR (399 MHz, Chloroform-*d*) δ 7.50 – 7.39 (m, 6H), 7.39 – 7.32 (m, 6H), 7.14 – 7.06 (m, 2H), 6.89 – 6.76 (m, 5H), 5.35 (dt, *J* = 7.4, 4.0 Hz, 1H), 5.14 (dd, *J* = 7.9, 3.3 Hz, 1H), 5.05 (dq, *J* = 10.7, 5.7, 5.2 Hz, 2H), 4.59 – 4.03 (m, 6H), 4.04 – 3.90 (m, 3H), 3.65 – 3.27 (m, 6H), 2.09 (d, *J* = 0.7 Hz, 3H), 2.06 (dq, *J* = 3.4, 1.1 Hz, 8H), 2.01 (d, *J* = 0.7 Hz, 3H), 1.97 (d, *J* = 0.8 Hz, 2H), 1.92 (d, *J* = 0.8 Hz, 3H).

The mixture of compound **12** (20 mg, 47 μmole) was combined with $Pd(OAc)_2$ (2.5 mg, 10 μmole) and stirred at 50°C for 1 h. (Diacetoxyiodo)benzene (18.4 mg, 57 μmole) was added and stirred at 50°C for 24 h. The reaction was concentrated down, purified via flash chromatography, and eluted with 40:60 ethyl aceteate/hexanes to yield **13** in 62% yield. ^{1}H NMR (399 MHz, Chloroform-*d*) δ 8.07 (dddd, *J* = 7.8, 3.4, 1.5, 0.7 Hz, 1H), 7.59 – 7.32 (m, 2H), 7.26 – 7.20 (m, 1H), 5.80 – 5.62 (m, 0H), 5.47 – 5.33 (m, 0H), 5.27 (ddd, *J* = 7.8, 2.7, 0.7 Hz, 0H), 5.11 – 4.96 (m, 0H), 4.78 – 4.07 (m, 3H), 4.02 (s, 0H), 3.81 – 3.68 (m, 0H), 2.17 (d, *J* = 0.7 Hz, 1H), 2.12 (d, *J* = 0.7 Hz, 1H), 2.04 (dd, *J* = 2.8, 0.6 Hz, 1H), 1.79 (d, *J* = 0.6 Hz, 1H), 1.67 (d, *J* = 0.7 Hz, 1H). LCMS: *m*/*z* 450.0 $[M+Na]^+$.

Compound **13** (44 mg, 103 μmole) was dissolved in methanol (400 μL) and 7 M ammonia in methanol (60 μL) was added at room temperature. The reaction was stirred overnight. The reaction was concentrated down and purified via flash chromatography. The product was eluted with 10:90 methanol/dichloromethane to afford **2** in 80% yield. ^{1}H NMR (399 MHz, Methanol-d_4) δ 8.02 (dq, *J* = 7.7, 0.9 Hz, 2H), 7.65 – 7.55 (m, 2H), 7.43 – 7.32 (m,

2H), 7.19 – 7.07 (m, 2H), 4.59 (dt, *J* = 14.9, 2.1 Hz, 1H), 4.43 (ddd, *J* = 15.0, 9.1, 1.6 Hz, 1H), 4.29 – 4.19 (m, 1H), 3.88 – 3.59 (m, 5H). LCMS: *m/z* 302.0 $[M+H]^+$ and 324.0 $[M+Na]^+$.

*Synthesis of (2R,3S,4S)-5-ammonio-2,3,4-trihydroxypentyl phosphate (**3**)*

Hydroxylamine hydrochloride (147 mg, 2.11 mmol) was dissolved in 2.5 mL of absolute ethanol. Two drops of a 1% phenolphthalein were added. Sodium ethoxide (165 µL, 2.11 mmol) is added drop wise until the solution stays pink for ~1 min. The white precipitant was filtered off leaving a yellow filtrate. To the filtrate, **7** (500 mg, 1.05 mmol) was slowly added. The reaction was placed in a cold oil bath and gently heated to 60°C. The yellow precipitant is collected and recrystalized in ethanol to yield the ribityloxime-5-phosphate, **8** in 99% yield. ^{1}HNMR (400 MHz, CD_3OD) (E)-oxime: δ 7.54 (d, 1H), 4.44 (dd, 1H), 3.85-3.60 (m, 5H), (Z)-oxime: δ 6.90 (d, 1H), 5.09 (dd, 1H), 3.85-3.60 (m, 5H). ^{31}P NMR (162 MHz, CD_3OD) δ 5.34.

Compound **8** was then hydrogenated in the presence of 5% palladium on carbon (112 mg, 52.6 µmole) in methanol. Catalyst was filtered off and filtrate was concentrated to yield **3** in 73% yield. ^{1}HNMR (400 MHz, CD_3OD) δ 3.9-3.5 (m, 5H), 2.9-2.7 (m, 2H). ^{31}P NMR (162 MHz, CD_3OD) δ 5.24.

*Synthesis of (2R,3S,4S)-2,3,4,5-tetrahydroxypentyl phosphate (**4**)*

Ribose-5-phosphate (**7**, 25 mg, 81 µmole) was dissolved in 30 µL of DI water and 500 µL of ethanol (152 mM). Sodium cyanoborohydride (15 mg, 243 µmole) was added to the mixture and the reaction was stirred at room temperature over night. Reaction was quenched and filtered to remove sodium cyanoborohydride. Product was purified via borate column. Borate column

was made by washing Dowex 50WX8 hydrogen form with 0.1 M ammonium borate pH 9.2. Product (**4**) was eluted with 0.1 M ammonium borate pH 9.2 in 43% yield. ^{1}H NMR (500 MHz, Deuterium Oxide) δ 4.05 (ddd, *J* = 10.9, 5.8, 2.9 Hz, 1H), 3.96 (dt, *J* = 11.0, 6.3 Hz, 1H), 3.90 (td, *J* = 6.3, 2.8 Hz, 1H), 3.83 (t, *J* = 5.7 Hz, 1H), 3.77 (dd, *J* = 12.0, 3.0 Hz, 1H), 3.71 (t, *J* = 6.5 Hz, 1H), 3.63 (t, *J* = 9.7 Hz, 1H). ^{31}P NMR (162 MHz, Deuterium Oxide) δ 0.54. ESI-MS: *m*/*z* 231.02761 $[M+H]^{+}$.

Conversion of Deoxyriboflavin to DeoxyFMN

For synthesis of deoxyFMN, the procedure described in Chapter 2 was followed. Briefly, 0.2 mg of deoxyriboflavin was added to a 1 mL reaction volume containing CaRFK (10 μM) in reaction buffer (50 mM Tris-HCl pH 8, 100 mM $MgCl_2$, 10 mM ATP). The reaction mixture was incubated at 37°C temperature for 18 h and product formation was monitored by t.l.c.. DeoxyFMN was purified via gravity-flow over Sep-Pak C18 columns (Waters, Milford MA) in 40% yield. Product identity was confirmed by ESI-MS.

Expression and Purification of OYE1

For protein overexpression, OYE1 in pET14b was transformed into *E. coli* BL21(DE3) cells. Individual colonies were grown overnight in 5 mL of 2YT supplemented with ampicillin (100 µg/mL) at 37°C. An aliquot of the overnight culture was used to inoculate 250 mL of fresh 2YT containing ampicillin. The culture was grow at 37°C until and OD_{600} of 0.6 was reached, then expression was induced by the addition of 0.3 mM ITPG, and carried out at 20°C for 18 h.

Cells were pelleted by centrifugation at 4°C, 4000 *g* for 20 min, the clear supernatant removed and pellets stored at -20°C until further usage.

For protein purification, cell pellets were resuspended in buffer A (40 mM Tris-HCl pH 8.0, 10 mM NaCl) supplemented with 2.5 µL/gram of pellet DNase I (EMD Millipore), and 50 µL/gram of pellet protease inhibitor cocktail. The mixture was incubated on ice for 20 min. Cells were lysed by sonication (10-sec pulse/20-sec pause for 4 min) in two batches, followed by centrifugation at 4°C, 4000 *g* for 30 min. The clarified lysate was loaded on a HiTrap Q FF (5 mL) anion-exchange column, pre-equilibrated in buffer A. The column was washed with 2 column volumes (CV) buffer A, followed by a linear gradient over 20 CVs to 100% Buffer B (40 mM Tris-HCl pH 8.0, 1 M NaCl). Protein elution was monitored by UV absorption (280 and 450 nm), product fractions were collected and concentrated using a Millipore filter unit (MWCO: 10 kDa). The protein was further purified by size exclusion chromatography (Superdex 200, 10/300 GL column) in buffer C (40 mM Tris HCl pH 8.0, 300 mM NaCl). Protein elution was monitored as described above. Fractions containing the protein of interests were collected and assessed for purity by SDS-PAGE.

Removal of native FMN cofactor was achieved by dialyzing pooled protein fractions against buffer D (100 mM potassium phosphate pH 5.0, 2 M postassium bromide, 6.7 mM ETDA) for 4 h at 4°C producing apoOYE1. Followed by dialysis against buffer E (50 mM Tris HCl pH 7.5, 20 mM NaCl) for 4 h at 4°C. Apoprotein was immediately used in ITC studies.

Isothermal Calorimetry Binding Studies

Calorimetric studies were performed on a MicroCal Auto-iTC200 at 25°C following standard procedures. ApoOYE1 was dialyzed against buffer F (50 mM Tris HCl pH 7.5). ApoOYE1 was added to the calorimetric reaction cell at a concentration between 10-15 μM and titrated with FMN (10x the concentration of apoOYE1), riboflavin (10x, 44x, and 22x), deoxyFMN (10x, 20x, 44x, 79x, 99x), lumiflavin (10x and 60x), **1** (10x, 90x, 110x, 150x), **2** (10x and 100x), **3** (10x and 100x), and **4** (10x and 100x) in the same buffer. Each titration experiment was performed with 18 injections of 2 μL at 150 s equilibration intervals. Data were fit to a single-site binding model by nonlinear least-squares regression with the Origin software package. The fit of data yields the binding affinity, enthalpy change, entropy change, and binding stoichometry for titration. Concentration of apoOYE1 was determined through a fluorescence titration of FMN on a FluoroMax-3 (Horiba Scientific, Piscataway, NJ) and a breakpoint analysis (SegReg (23)).

PURExpress In vitro Transcription/Translation of OYE1 with FMN (analogs)

OYE1 in pET-14b was used as template DNA for *in vitro* transcription/translation (IVTT). IVTT reactions were assembled according to the PURExpress in vitro protein synthesis kit with the following modifications optimized for our system. Reactions were assembled on a 25 μL scale containing PURExpress Solution A, PURExpress Solution B, 20 U of Murine RNAse Inhibitor (NEB), 250 ng of DNA, and OmniPur RNAse Free Water (EMD). Reactions contained either 100 μM FMN, 100 μM cofactor analog, or no cofactor. All reactions were run simultaneously with dihydrofolate reductase (DHFR, NEB) as a negative control. Reaction

mixtures were incubated at 37°C for 2.5 h. Following synthesis, reactions were quenched by cooling to 4°C for 5 min before being directly utilized for FMN competition assays.

Cofactor Competition and Enzymatic Activity Assay

For competition reactions, cofactor analogs were replaced by native FMN by adding 150 μM of FMN into each IVTT reaction. IVTT/competition reactions were incubated at 4°C for 4 h before being used directly in activity assay. Ene-reductase activity for enzymes was measured in triplicate unless noted under anaerobic conditions at room temperature using GDH from *Thermoplasma acidophilum* for the regeneration of NADPH. Substrate was dissolved in DMSO (0.5% final concentration) to overcome poor water solubility. Reaction stock was prepared in 50 mM Tris HCl pH 7.5 containing 100 mM glucose, 2 U per reaction GDH, 200 μM NADPH, and 250 μM of **14**. 35 μL of reaction stock was added to 25 μL of IVTT/competition reactions. After a 24 h period, reactions were quenched and products were extracted twice by adding 30 μL of ethyl acetate containing 0.5 mM cyclohexanone as an internal standard. 2 μL of the organic phase was injected and analyzed via chiral GC-MS. Product was confirmed by a commercial standard and mass spectrometry.

*(5*S*)-Carvone GC/MS Protocol*: 110°C to 150 (5 min) at 2°C/min; linear velocity 52.8 cm/sec; injection temperature 250°C. (retention times for **14**: 15.8 min and **15**: 12.8 min)

VI. References

1. Schopfer, L.M., A. Wessiak, and V. Massey, Interpretation of the spectra observed during oxidation of p-hydroxybenzoate hydroxylase reconstituted with modified flavins. Journal of Biological Chemistry, 1991. **266**(20): p. 13080.

2. Hasford, J.J. and C.J. Rizzo, Linear free energy substituent effect on flavin redox chemistry. Journal of the American Chemical Society, 1998. **120**(10): p. 2251.

3. Murthy, Y.V.S.N. and V. Massey, Synthesis and properties of 8-cn-flavin nucleotide analogs and studies with flavoproteins. Journal of Biological Chemistry, 1998. **273**(15): p. 8975.

4. Entsch, B., et al., Oxygen reactivity of p-hydroxybenzoate hydroxylase containing 1-deaza-fad. Journal of Biological Chemistry, 1980. **255**(4): p. 1420.

5. Manstein, D.J., et al., Absolute stereochemistry of flavins in enzyme-catalyzed reactions. Biochemistry, 1986. **25**(22): p. 6807.

6. Mansurova, M., M.S. Koay, and W. Gärtner, Synthesis and electrochemical properties of structurally modified flavin compounds. European Journal of Organic Chemistry, 2008. **2008**(32): p. 5401.

7. Bollen, Y.J.M., et al., Last in, first out: The role of cofactor binding in flavodoxin folding. Journal of Biological Chemistry, 2005. **280**(9): p. 7836.

8. Abe, M., et al., Detection of structural changes in a cofactor binding protein by using a wheat germ cell-free protein synthesis system coupled with unnatural amino acid probing. Proteins: Structure, Function, and Bioinformatics, 2007. **67**(3): p. 643.

9. Huijbers, M.M.E., et al., Proline dehydrogenase from thermus thermophilus does not discriminate between fad and fmn as cofactor. Scientific Reports, 2017. **7**: p. 43880.

10. Shimizu, Y., et al., Cell-free translation reconstituted with purified components. Nature Biotechnology, 2001. **19**(8): p. 751.

11. Shimizu, Y., T. Kanamori, and T. Ueda, Protein synthesis by pure translation systems. Methods, 2005. **36**(3): p. 299.

12. Tollin, G. and D.E. Edmondson, *[37] purification and properties of flavodoxins*, in *Methods in enzymology*, A. San Pietro, Editor. 1980, Academic Press. p. 392.

13. Lostao, A., et al., Dissecting the energetics of the apoflavodoxin-fmn complex. Journal of Biological Chemistry, 2000. **275**(13): p. 9518.

14. Abdek-Akher, M., J.K. Hamilton, and F. Smith, *The reduction of sugars with sodium borohydride1*. Journal of the American Chemical Society, 1951. **73**(10): p. 4691.

15. Taylor, E.J. and C. Djerassi, Mechanism of the sodium cyanoborohydride reduction of .Alpha.,.Beta.-unsaturated tosylhydrazones. Journal of the American Chemical Society, 1976. **98**(8): p. 2275.

16. Peterson, K.E., et al., Alcohol-, diol-, and carbohydrate-substituted indenoisoquinolines as topoisomerase i inhibitors: Investigating the relationships involving stereochemistry, hydrogen bonding, and biological activity. Journal of Medicinal Chemistry, 2011. **54**(14): p. 4937.

17. Aerschot, A.V., et al., Double protection of the heterocyclic base of xanthosine and 2′-deoxyxanthosine. Nucleosides and Nucleotides, 1989. **8**(2): p. 159.

18. Jordan-Hore, J.A., et al., Oxidative pd(ii)-catalyzed c−h bond amination to carbazole at ambient temperature. Journal of the American Chemical Society, 2008. **130**(48): p. 16184.

19. Tsibris, J.C.M., D.B. McCormick, and L.D. Wright, Studies on the binding and function of flavin phosphates with flavin mononucleotide-dependent enzymes. Journal of Biological Chemistry, 1966. **241**(5): p. 1138.

20. Theorell, H. and Å. Åkeson, Molecular weight and fmn11fmn = riboflavine mononucleotide. Content of crystalline “old yellow enzyme”. Archives of Biochemistry and Biophysics, 1956. **65**(1): p. 439.

21. Lostao, A., et al., Differential stabilization of the three fmn redox forms by tyrosine 94 and tryptophan 57 in flavodoxin from anabaena and its influence on the redox potentials. Biochemistry, 1997. **36**(47): p. 14334.

22. Schwertman, N.C., M.A. Owens, and R. Adnan, A simple more general boxplot method for identifying outliers. Computational Statistics & Data Analysis, 2004. **47**(1): p. 165.

23. Oosterbaan, R., et al., Crop production and soil salinity: Evaluation of field data from india by segmented linear regression with breakpoint. 2019.

Chapter 4: Exploration of the Old Yellow Enzyme Superfamily Utilizing Sequence Similarity Networks

This work is done in collaboration with Dr. Janine Copp from the University of British Columbia as well as with other lab members, David White, Parisa Keshavarz-Joud, and Tamra Blue.

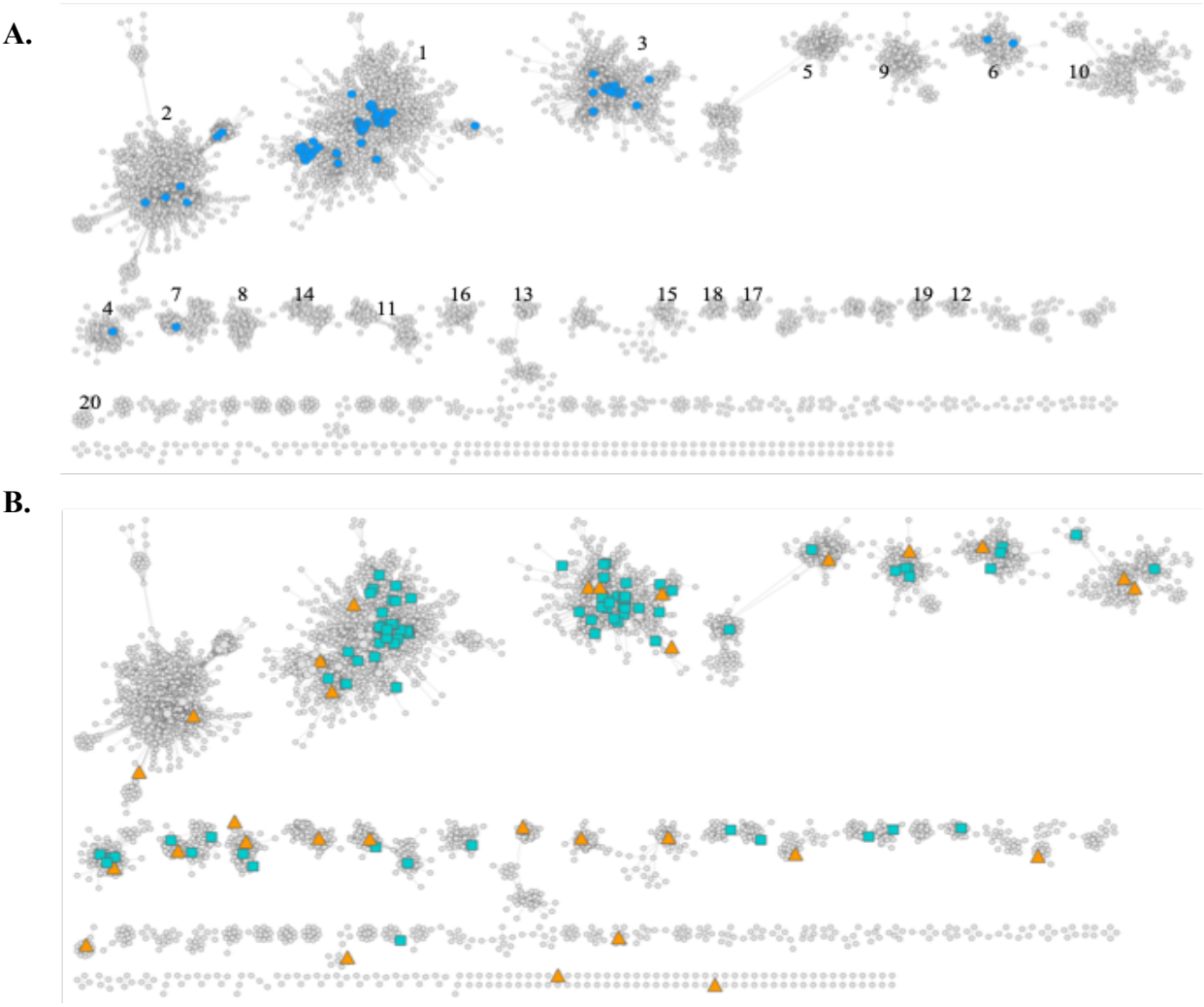
I. Introduction

Over the last century, the discovery and isolation of new enzymes has provided powerful catalytic tools for the sustainable synthesis of fine chemicals, flavors, and pharmaceutical intermediates (1). Enzymes provide an attractive, green alternative to traditional chemocatalysts as they are highly efficient under mild conditions. Of particular interest are the FMN dependent Old Yellow Enzymes (OYEs) for their asymmetric *trans*-hydrogenation of activated alkenes with increased regio- and entioselectivity compared to their transition metal counterparts (2). *Trans*-hydrogenations have the potential to add up to two stereo centers, which is a desirable process in organic synthesis and has garnered much interest in OYEs as biocatalysts. Accordingly, OYEs have been utilized in the synthesis of anti-inflammatory drugs, macrocyclic antibiotics, and anti convulsants (2-4). However, the ability to engineer novel biocatalysts cannot currently meet industrial demands, and there remains a need for more efficient methods of discovering new biocatalysts. Protein superfamilies, defined by a group of proteins which share a common ancestor, potentially harbor a plethora of biocatalytic opportunities, though most have gone unexplored (5). For example, the OYE superfamily contains over 70,000 sequences, but only 136 enzymes have been characterized. Despite OYEs being heavily studied over the last 85 years, the characterization of these enzymes has been limited to a handful of substrates, and has been hampered by the fact that the biological function is currently only known for 8 OYEs (6-12). Considering the vastness of the family and our limited understanding of OYEs in general, the potential for isolating novel biocatalysts of industrial or commercial relevance within the depths of the superfamily is promising.

No previous work has attempted to visualize the OYE superfamily in its entirely. The closest account was the theoretical work performed by Nizam *et al*. in which a genome-wide search of 60 fungal species was performed, and 424 putative OYEs were identified in order to study the evolutionary significance of OYEs throughout the fungal kingdom (13). All other studies have relied on previously characterized OYEs in order to draw conclusions and make inferences about the family. Accordingly, grouping and classification of family members is often based on crystal structures and homology models of known OYEs. Originally OYEs were categorized into two classes, the 'classical' and 'thermophillic-like,' based on structural features, but more recently there has been an expansion to five classes (14).

Until the early 2000's all the OYEs studied were thought to be structurally similar. It was not until the discovery and crystallization of YqjM from *Bacillus subtilis* that another structural variant of OYE was identified, splitting the family into the 'classical' OYEs and 'YqjM-like' OYEs (15). Subsequently, the class of YqjM-like OYEs were found to contain the thermophilic OYEs as well, and was thus renamed as the 'thermophilic-like' OYEs (10, 16). During a follow-up study performed by Nizam *et al*., both a novel class of OYEs and a new naming system were proposed (17). Class I, previously referred to as the 'classical' OYEs, uses OYE1 as a reference sequence and is identified by conserved active site residues Thr37, His191, Asn194, Tyr196, Phe250, and Tyr375 (numbering is based on OYE1). The thermophilic-like OYEs were recategorized as Class II with the hallmark active site resides Cys26, Tyr28, Lys109, His164, His167, Tyr169, and Arg 336 (numbering based on YqjM). Another major structural distinction of the Class II OYEs is that they are catalytically active as dimers or tetramers and possess "arginine-fingers" at the C-termini of each monomer which protrude into the active site of the

Figure 4.1. Sequence Similarity Network for OYE Superfamily. There are 9,706 nodes which each represent between 1 and 641 sequences with greater than 50% identity. Edges represent E-value threshold set at 1 x 10^{-85}. Largest 20 clusters are numbered. **A.** Distribution of the currently characterized OYEs (enlarged blue nodes). **B.** Distribution of OYEs from first (orange triangles) and second (cyan squares) generation library.



other monomer in the dimer (2). Lastly, Nizam *et al*. added a new class, Class III, containing a conservation YGGS motif along with either set of active site residues belonging to Class I or Class II. To date, no Class III OYEs have been fully characterized, but rather members have only been identified through bioinformatics (17). Recently, the three classes have been reordered by Scholtissek *et al*. according to the phylogeny of 63 previously characterized OYEs (2). They

broke the original ‘classical’ OYEs into two different groups, Class I and Class II. According to Scholtissek *et al*., Class I contains the classical OYEs from plants and bacteria while Class II contains all the classical fungal OYEs. Class III contains the thermophilic-like OYEs from bacteria. In early 2019, the number of classes was yet again expanded from three to five based on data collected from NCBI’s Conserved Domain Database (14). These five classes resemble the classes proposed by Nizam *et al*. with Class I and Class II containing the classical and thermophilic-like OYEs, respectively, while the fungal OYEs have been designated as Class V. Peters *et al*. then characterized two enzymes from each of their newly proposed classes, III and IV (14). However, upon comparison of the important active site residues from enzymes in Class III and IV, these OYEs closely resemble the Class I and II OYEs.

Collectively, there has been much debate over the classification of OYEs, both putative and characterized, over the past five years. However, there is no defined standardization for how OYEs are grouped, leading to much confusion in the field. We propose a more systematic method for comparing and grouping the whole superfamily. Using sequence similarity networks (SSN), we can take large datasets like the OYE superfamily and unbiasedly compare the sequences in order to clearly classify them.

Recent advances in sequencing technologies have led to dramatic increases in the volume of available sequencing data with more than 150 million protein sequences currently deposited in the Uniprot database (18). However, the ability to process sufficient levels of sequence information in order to assign function and effectively explore the unknown sequence space of protein families remains an unmet challenge for multiple biological disciplines (5). Over the past decade numerous bioinformatics tools have been developed in an attempt to address this

dilemma. With the development of Cytoscape, which allows for the integration of biomolecular interactions with expression data, phenotypes, and activity profiles, computational SSNs have emerged as an enhanced method for visualizing large groups of closely related proteins, known as “superfamilies” (19). SSNs are a collection of independent pairwise alignments between sequences meant to provide an effective bioinformatics tool for the exploration of a protein family’s sequence space (20). Furthermore, SSNs are an attractive alternative to traditional multiple sequence alignments and phylogenic trees for a few key reasons. First, SSNs provide a rapid method of viewing evolutionarily-related proteins based on threshold set by the user. Phylogenetic trees, by comparison, are limited in the number of sequences which can be aligned and appropriately visualized. Additionally, the networks allow for an overlay of similarity with biologically useful data such as activity, size, origin, structure, etc (20). With over 70,000 members in the OYE family and only a limited subset of the family that has been explored, SSNs visualization lends itself as a viable strategy to systematically sort through the unexplored sequence space in order to find novel biocatalysts.

Using a SSN, we systematically sampled the OYE superfamily by selecting representatives from a number of different clusters, groups of closely related proteins within the superfamily, to find enzymes with increased activity, a broader or different substrate scope, and novel chemistries. Since the native function of OYEs is unknown, it is possible that there is functional diversity within the family to be discovered. Our initial library, the first generation, was designed to broadly sample the superfamily in order to learn about the sequence-function relationship. Utilizing the insights gained from the first generation, we more narrowly selected additional family members (i.e. the second generation) to further explore specific regions of the

superfamily in greater detail. Between the two libraries, we were able to screen 120 novel OYEs by sampling from 27 clusters, 21 of which were unexplored, and essentially doubling the number of knowns OYEs.

Selected OYEs were profiled against a total of 16 substrates in mixes: the 'good' (I), the 'bad' (II), the 'ugly' (III), and the 'crazy' (IV) (**Figure 4.2**). The 'good' mix contains (*S*)-carvone (**1**), ketoisophorone (**2**), cinnamaldehyde (**3**), and 4-phenyl-3-butyn-2-one (**4**) representing benchmark substrates used in previous OYE studies. Conversion of **4** is unique in that it proceeds through two consecutive reductions (21). The 'bad' mix (**5-8**) is made up of furanones with increasingly bulky substituents. Such furanone derivatives represent precursors to desirable synthons, though OYEs have difficulty reducing them (22). The 'ugly' mix consists of methyl-2-(hydroxymethyl)acrylate (**9**), a precursor to the valuable 'Roche ester' for which OYEs show little to no activity (23). Substrates **10** and **11** are substituted cyano-alkenes which are sought after synthons and were included to assess pour substrate conversion by known OYEs. Additionally, nitro-alkene (**12**) was included to probe for possible nitro reductase activity.

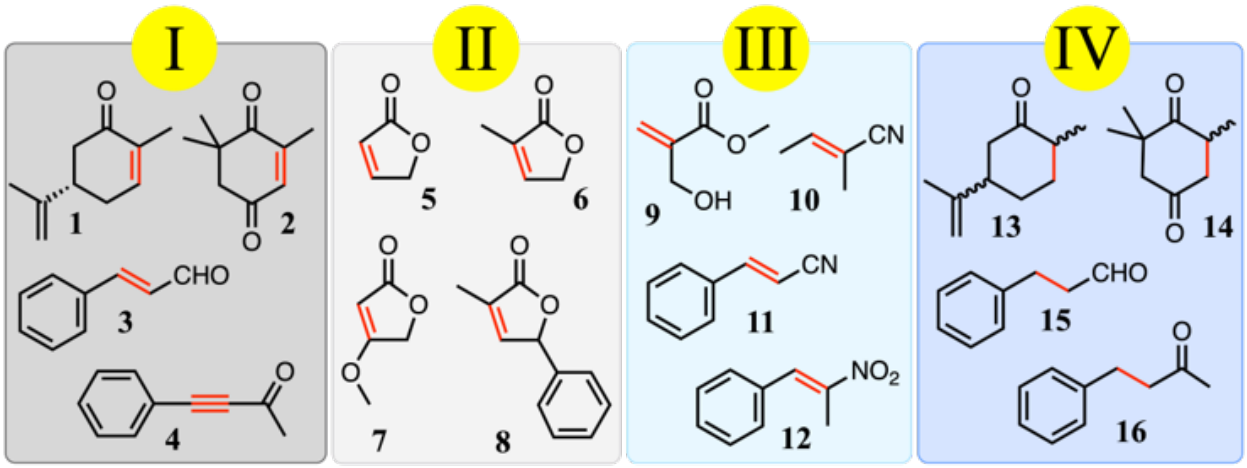


Figure 4.2: Substrate Mixes I-IV. The substrates are grouped based on known OYE activity. The bond which is reduced or oxidized is highlighted in red.

Finally, the ‘crazy’ mix (**13-16**) consists of the reduced products of the ‘good’ substrates. This mix was designed to evaluate library members for desaturase activity. Such oxidative activity has only been observed at very low levels at high temperatures with one OYE family member (24), yet hints at the catalytic versatility of this superfamily.

II. Results and Discussion

Sequence Similarity Network (SSN)

The network was generated by collecting roughly 70,000 sequences from the UniProtKB database identified with the PF00724 ‘Oxidored FMN’ domain (**Figure 4.1**). Subsequently, sequences were clustered based on a similarity factor which was calculated by an “all versus all” BLAST pairwise alignment and visualized by a SSN. The resulting network was made up of 9,705 nodes which each contain anywhere from 1 to 641 sequences that have greater than 50% identity. The nodes were connected by edges if the BLAST e-value between them was greater than 1×10^{-85}, creating more than 20 distinct clusters of sequences. This edge threshold best divides the superfamily into clusters with different known functions or structures. For example, the ‘classical’ and ‘thermophilic’ OYEs fall into two separate clusters: 1 and 3, respectively. A third group of OYEs, referred to as the 2-enoate reductases, make up cluster 2. The 2-enoate reductase OYEs are typically 600-700 amino acids in length and consist of two domains, in the middle lies an iron-sulfur cluster used to mediate electron transfers between a FAD binding domain and the catalytic FMN binding domain (25). Collectively, these top 3 clusters contain over 15,000 sequences each and account for roughly 75% of the superfamily. Upon creation of the SSN we were able to map out previously known and characterized OYEs fall (**Figure 4.1A**).

The majority of these OYEs are dispersed among clusters 1-3. Beyond clusters 1-3, there are six additional clusters which contain at least one known OYEs. Lastly, there are 20 clusters possessing greater than 100 sequences each, 11 of which have been completely unexplored. Given that the majority of previously characterized OYEs originate from the first three clusters, the superfamily as a whole remains largely underrepresented and could contain multiple enzymes with unique catalytic profiles.

The SSN not only allows for the complete and unbiased classification of all the known family members, but also generates a comprehensive roadmap for the exploration of OYES. Interestingly, a hallmark feature distinguishing the 'classical' and 'thermophilic' OYEs was the difference in active site residues at positions 191 and 194 (all numbering is based on OYE1). Traditionally, the 'classical' OYEs were known to contain a histidine at 191 (H191) and an asparagine at 194 (N194) while the 'thermophilic' OYEs contained histidines in both positions. While the 'classical' and 'thermophilic' OYEs fall into different clusters, cluster 1 contains sequences with H191/N194 as well as sequences with H191/H194. The class II, 'thermophilic,' OYEs can be found in cluster 3. Interestingly, this entire cluster contains a conserved cysteine residue at position 37, which is known to interact with the flavin and act as a redox modulator. This conservation of residues is not seen to this extent with any other cluster, as structural motifs can usually be seen across multiple clusters. After the first two classes, the traditional classification of OYEs breaks down further. The OYEs that were identified by Peters *et al*. as class III are actually found within 3 different clusters, (3, 6, and 13) while the class IV OYEs reside in cluster 4. The class V OYEs, also identified by Nizam *et al*. as the fungal OYEs, are also scattered amongst clusters, 5, 9, and 16. Clearly, considering the family as a whole instead

of focusing solely on the characterized OYEs offers a more comprehensive view of the OYE superfamily.

Visualization of the superfamily in the form of a SSN allowed us to observe certain additional trends. While OYEs are ubiquitous in nature, nearly 70% of the sequences identified in the superfamily are bacterial. Some of the interesting clusters in regards to the distribution of sequences amongst the biological kingdoms are 1-3, 5, and 9. For example, cluster 2 is unique in that it is made up of 95% bacterial OYEs. Conversely, cluster 9 is solely composed of eukaryotic OYEs with a majority of the sequences originating from the phylum nematoda. Clusters 1 and 3 are very similar in that they contain OYEs from fungi, bacteria, and archea, though cluster 1 also contains a number of OYEs from plants while cluster 3 does not. Finally, cluster 5 is identified as the cluster with 'islands,' or highly related sub-clusters where each island contains either fungal or bacteria OYEs. From this analysis the clustering of the superfamily does not appear to be dictated by any particular organism of origin, but instead could be grouped based on native function. Further studies into each cluster is needed to verify this assertion.

In addition to these particular organismal groupings, looking at the superfamily as a whole allowed us to identify a common structural motif which may have previously been overlooked. The 2-enoate reductases of cluster 2 presumably all contain an iron-sulfur clusters, which are known to be coordinated by a conserved $CxxCxxCx_nC$ sequence in proteins utilizing this cofactor. Within the top 20 clusters, five other clusters besides cluster 2 contain this conserved sequence motif. It is unclear at present whether this motif could be an evolutionary artifact or whether it could hint at the native function of these uncharacterized OYEs.

Selection of Novel OYE Family Members

To validate the network the first generation library made up of 32 sequences, from bacteria, fungi, archaea, protista, and plantae over 22 different clusters were selected to explore the untapped sequence space of the family (**Figure 4.1B**). As a positive control, four enzymes from cluster 1 that had previously been characterized were included within this test library. Insights gained from the first generation guided the selection of sequences for the second round of testing. In the second generation library, consisting of 93 novel sequences from bacteria, fungi, archaea, protista, metazoa, and plantae, two-thirds of the sequences came from either cluster 1 or 3 as these clusters make up 50% of the total family. Sequences from cluster 2 were not selected because of the lack of activity for enzymes from this cluster in the first generation screen. Given that cluster 2 contains the 2-enoate reductases, the observed lack of activity is likely due to an inability to properly charge these enzymes with the necessary iron-sulfur clusters in our current screening system. The remaining sequences came from 7 new clusters to further probe the sequence space of the family and 8 clusters that were sampled in the first generation. The additional sequences chosen from these 8 clusters were selected either to further explore the cluster or to look for functional alternatives to first generation member exhibiting poor solubility.

Evaluation of Novel OYE Family Members

To efficiently evaluate two libraries of novel OYEs, we utilized the cell-free PURE (Protein synthesis Using Recombinant Elements) *in vitro* transcription/translation system (26, 27). The PURE system allows for a much more streamline expression of genes compared to traditional cell lysates or protein purification methods. Additionally, the chemically-defined

nature of the PURE system drastically reduced the occurrence of undesired background reactions while also creating a tunable environment to suit individual needs of enzymes of interest. For example, the system was supplemented with both FMN and FAD as previous research has hinted at the necessity of the cofactor being present during synthesis of OYE proteins (see chapter 3). To date, the OYEs which have been characterized are all FMN dependent, though FAD was included in the PURE reactions as we cannot rule out the possibility of FAD dependent enzymes being present within the superfamily (28). Based on previously reported literature procedures, purified chaperonin GroEL/GroES from *E. coli* were also added to increase the chances of soluble protein expression (29, 30).

Following *in vitro* synthesis, each family member was tested for activity against the four substrate mixes. For mixes I-III, assays were performed anaerobically to prevent side reactions between the reduced flavin and molecular oxygen. Additionally, reactions were supplemented with both NADH and NADPH as electron donors since characterized OYEs will preferentially use one of these reductants over the other for catalysis (2). The previously reported glucose/ glucose dehydrogenase (GDH) system was added to regenerate the nicotinamide cofactors (31). Assays for mix IV were performed aerobically and without the addition of nicotinamide or the regeneration system. Activity will be confirmed with traditional heterologous expression and purification for candidates with greater activity towards the substrates or with a unique substrate profile compared to our positive control, OYE1.

Within the first generation library, 14 OYEs had activity towards at least one of the substrate mixes, two showed activity towards all four mixes, and five showed novel desaturase activity (**Figure 4.3A**). The active enzymes originate from eight different clusters, five of which

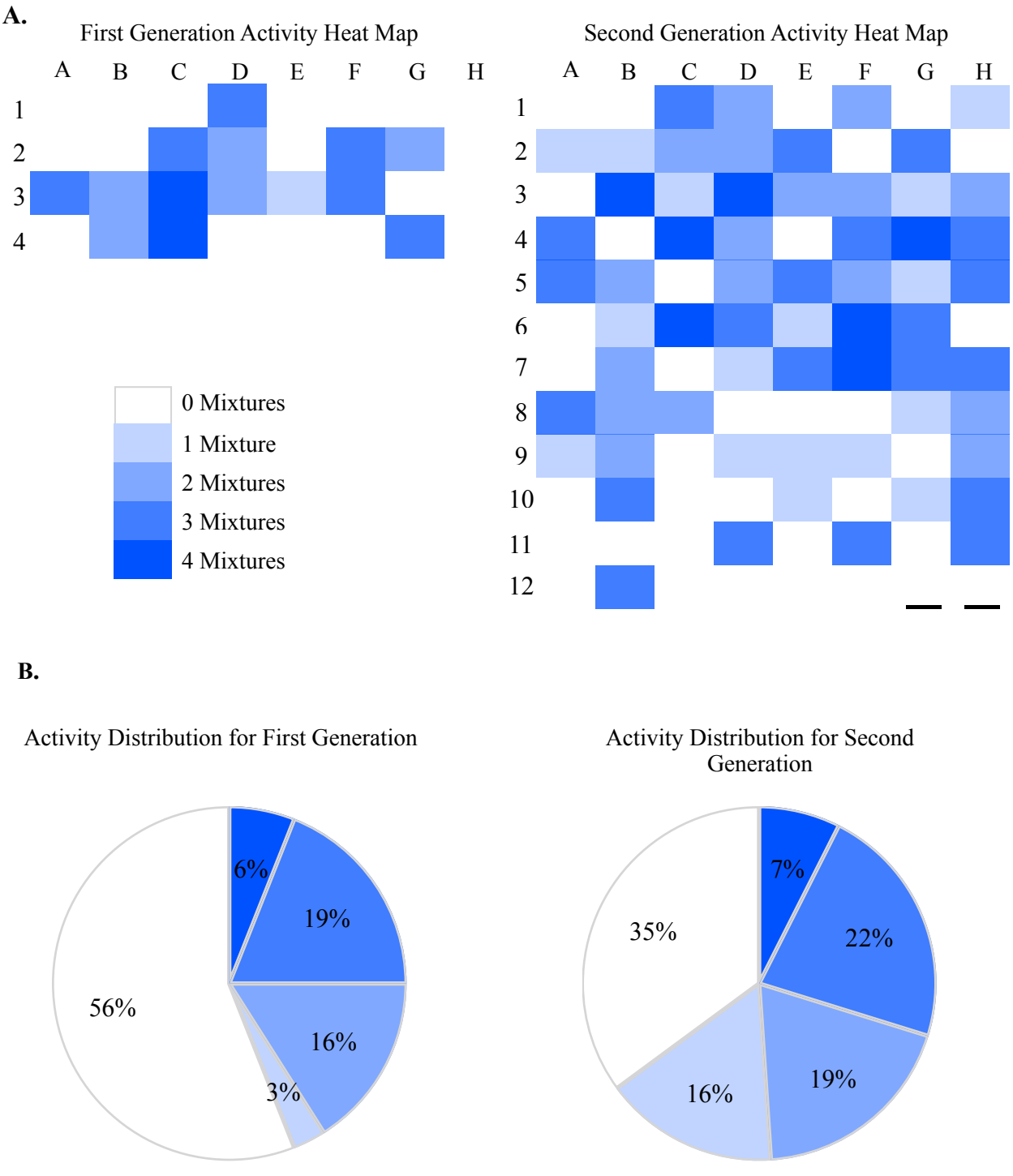


Figure 4.3. Activity Data for First and Second Generation Libraries. Uniprot codes for each well can be found in the supplemental. **A.** Heat maps show number of mixes each library member was active towards from 0 (white) to 4 (blue) for first (top left) and second (top right) generation. **B.** Pie charts show percent distribution of enzymes with each amount of activity.

have not been explored previously. Examination of the 18 inactive library members revealed that nine of them contain putative iron-sulfur binding motifs. The PURE system does not contain the accessory proteins necessary for proper charging of the iron-sulfur cluster, which could account for their lack of activity. The reason for the inactivity of the remaining nine proteins is unclear and could be due to lack of proper folding, post-translational modifications, or lack of proper substrates.

Unsurprisingly, our five benchmark OYEs showed activity towards mix I, II, and III; however, three of these enzymes unexpectedly showed oxidase activity, a phenomenon which has not been reported previously. **Figure 4.4** displays the various levels of activity of each first generation library member towards specific substrates. Of note are Q2TJB8 from *Trypanosoma cruzi* and B9T8J4 from *Ricinus communis* for their broad substrate profiles, exhibiting activity in all four mixes. Found in clusters 1 and 3 respectively, these unique enzymes illustrate that even though both clusters contain the majority of previously characterized OYEs, both clusters potentially harbor a significant number of novel catalysts within the unexplored sequence space. Additionally, novel OYEs A0A017SC09 (cluster 9) from *Aspergillus ruber* CBS 135680 and A0YFJ6 (cluster 11) from marine gamma proteobacterium HTCC2143 show activity towards the bulky substrate **8**. A0YFJ6 is unique in that it also shows activity towards **6**, which has canonically been a difficult substrate due to the methyl group at the 2 position yet, it does not have activity towards substrate **5,** as more commonly seen with characterized OYEs. This activity profile would suggest that the active site of A0YFJ6 could be larger compared to OYE1, allowing for the accommodation of bulky substrates. Further characterization of A0YFJ6 is necessary to more throughly elucidate the differences in substrate profile.

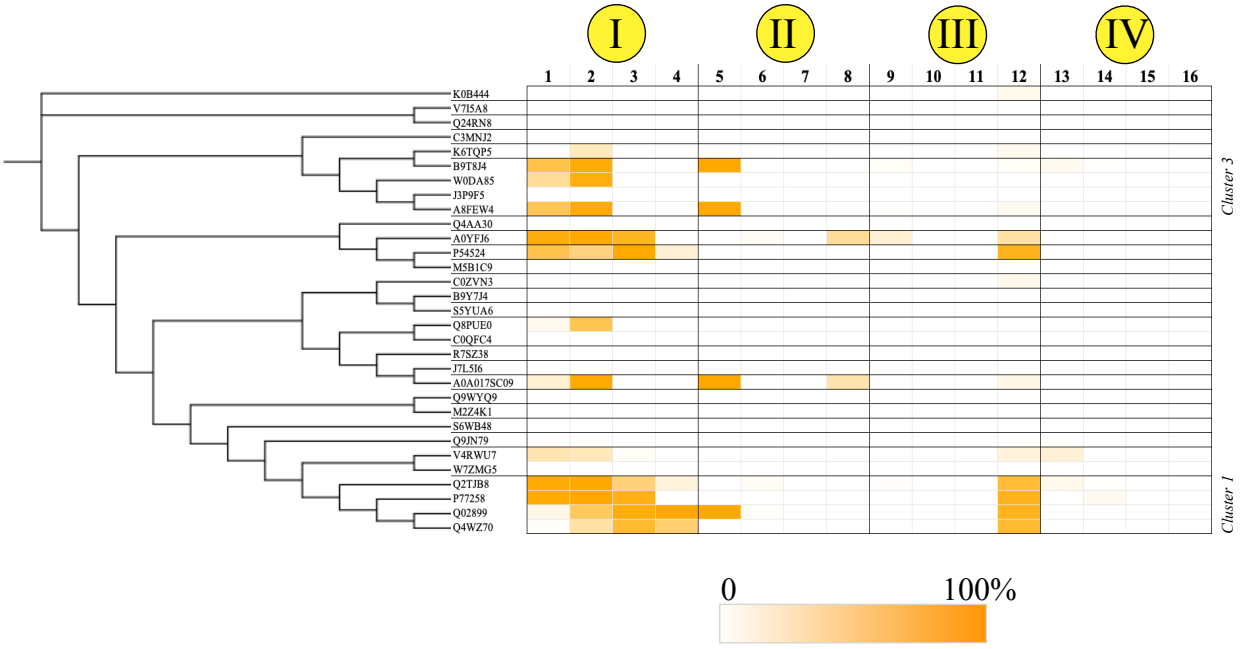


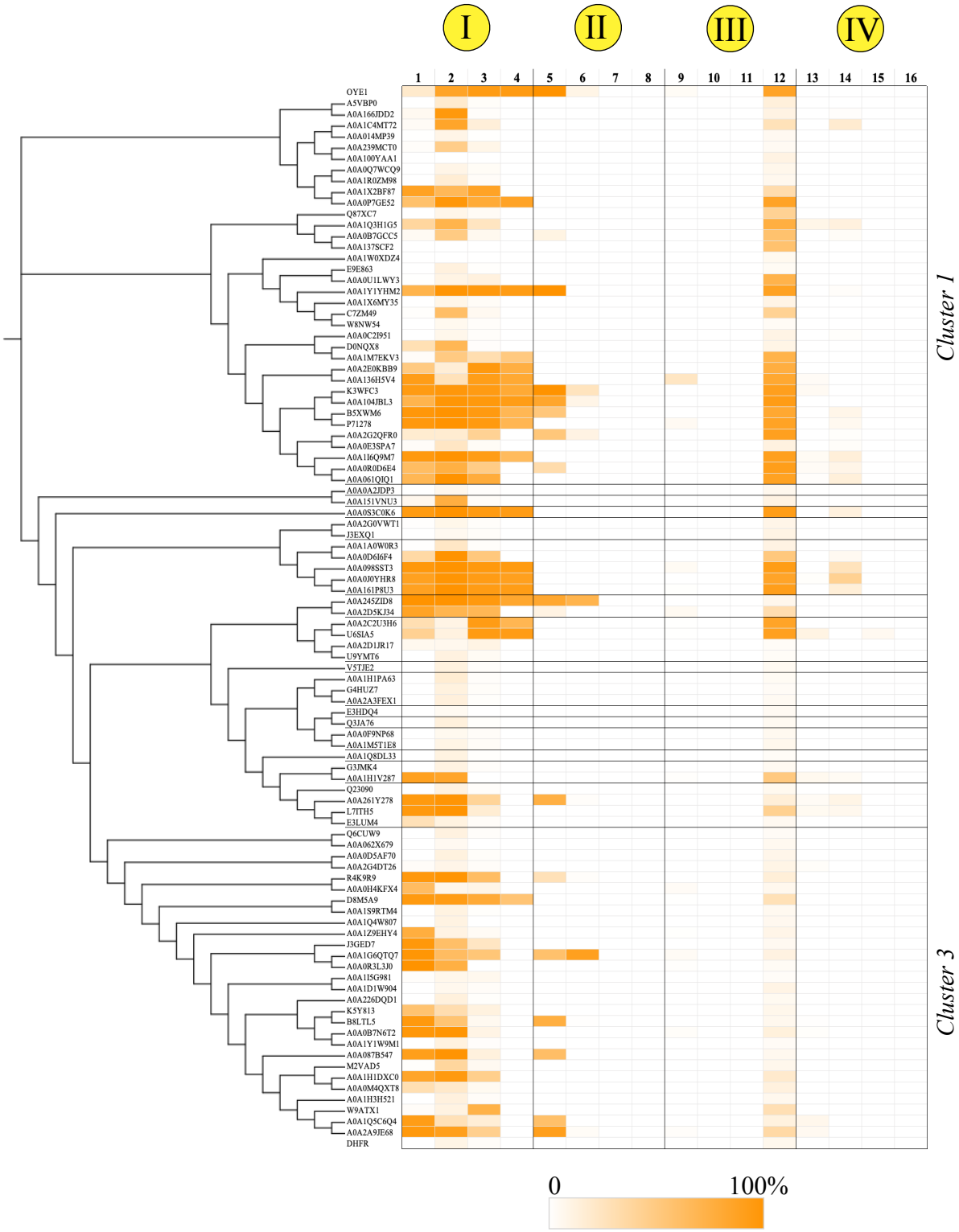
Figure 4.4. Phylogenic Tree and Substrate Consumption for First Generation Library. Each enzyme is represented by their Uniprot code, grouped by cluster, and ordered based on ancestry. Clusters 1 and 3 are labeled. As validation Cluster 1 enzymes have been previously characterized.

By comparison to the first library the second generation library saw an overall increase in the total percentage of active enzymes which originate from nine different clusters (**Figure 4.3B**), three of which are novel and were not sampled in the first generation library. Of the 61 active enzymes, seven of them show activity towards all four substrate mixes. Cluster 1 contains six of these seven enzymes while the remaining enzyme is from cluster 3. Based on the current data set we have these results suggest that OYEs with the broadest substrate profiles come from clusters 1 and 3. Additionally, K3WFC3 (cluster 1) from *Pythium ultimum* DAOM BR144 and A0A2A9JE68 (cluster 3) from *Saccharopolyspora erythraea* have the broadest activity, converting 8 substrates in total. **Figure 4.5** displays the percent consumption of the second generation library members, grouped by clusters, towards each substrate (OYE1 is at the top for reference). In general, there are some apparent initial trends linking the observed catalytic

activities for candidate enzymes and their respective clusters of origin. Clusters 1, 3, and 4 are the most active specifically for mix I with cluster 1 containing the highest number of active enzymes overall. Interestingly, cluster 3 enzymes show little to no activity towards **4** or **12** which seem to be good substrates for cluster 1 and 4. These small differences in substrate profiles could be indicative of unique structural variations within the active sites of enzymes from each of these clusters.

Of the activity profiles observed across both libraries, there are specific OYEs of note. For example, in reference to mix I, OYE1 is known for its high (>85%) conversion of substrates **2**, **3**, and **4,** though it has only shown modest (~20%) activity towards substrate **1**. We have identified several family members with comparable activities towards **2**, **3**, and **4** while also displaying greater than 85% activity towards substrate **1,** which represents a 4-fold increase in activity relative to OYE1. As for mix II, we, unfortunately, do not see any activity toward substrate **8,** as was observed in the first generation library, but there are a number of enzymes with activity towards **6**. Specifically, A0A1G6QTQ7 (cluster 3) from *Bradyrhizobium* sp. R5 and A0A245ZID8 (cluster 11) from *Sphingomonas dokdonensis* display the highest activities reported to date for **6**: 7.0 and 8.8-fold greater than OYE1, respectively. A0A245ZID8 is also a highly efficient enzyme for substrates in mix I (> 80% conversion for **1**-**4**), but shows no activity for mixes III or IV. Another interesting OYE is A0A136H5V4 (cluster 1) from *Neptuniibacter* sp. Phe_28 for its 6.5-fold increase in activity towards **9** compared to OYE1. It also shows high (>80%) activity towards **1**, **3** and **4,** but comparatively low activity towards **2** (27%). With the exception of A0A151VNU3 (cluster 16) from *Hypsizygus marmoreus*, OYEs from clusters 13 to 24 do not show any activity, which could be do to low solubility, insufficient substrates, or lack

Figure 4.5. Phylogenic Tree and Substrate Consumption for Second Generation Library. Each enzyme is represented by their Uniprot code, grouped by cluster, and ordered based on ancestry. OYE1 is listed at the top as a reference and DHFR (negative control) is at the bottom.



of possible protein binding partners necessary for activity or stability. Overall, the majority of the activity observed for mix III was for substrate **12**, though enzymes capable of reducing **12** produced largely racemic products with the exceptions of A0A0R0D6E4 (cluster 1) from *Stenotrophomonas chelatiphaga*, A0A2G2QFR0 (cluster 1) from *Methylophaga* sp., and A0A061QIQ1 (cluster 1) from alpha proteobacterium Q-1. Each of these three enzymes exhibit approximately 90% conversion of **12** with at least 70% enantiomeric excess towards the same enantiomer (due to lack of a commercially available standard at this time, we cannot determine which isomer is produced in excess). While A0A0R0D6E4 and A0A2G2QFR0 have activity for mixes I and II as well, A0A061QIQ1 does not. Interestingly, however, A0A061QIQ1 does have around 15% desaturase activity towards **14**. As a whole enzymes from cluster 4 show extremely high levels of activity for mix I while also having the highest levels of desaturase activity among both the first and second generation libraries. For example, A0A0J0YHR8 (cluster 4) from *Flavobacterium* sp. ABG exhibits at least 87% conversion for all substrates in mix I while also yielding 42% conversion of **14**.

An additional observation of the active enzymes from the second generation is that half of them (28 in total) exhibited desaturase activity towards substrates **13** and **14**, as seen with the first generation, as well as substrate **15**. A majority of these enzymes come from cluster 1, though clusters 3, 4, 5, 6, 9, and 12 also contain enzymes with desaturase activity, suggesting that it is inherent across the entire superfamily. Further studies will need to be performed in order to deduce the mechanistic cause underpinning this novel activity. Our ongoing efforts are focusing on A0A0C2I951 from *Pseudomonas batumici* as it is the only enzyme which solely displays desaturase activity. Previous works have determined that the amino acid at position 37 modulates

the redox potential of the flavin, which in theory would affect the enzyme's ability to oxidize or reduce a substrate. However, position 37 is not conserved across the clusters; for example, cluster 1 contains a threonine while cluster 3 has a cysteine. Future investigations will need to explore whether other amino acids may influence the redox potential of the FMN. Additionally, A0A0J0YHR8 (cluster 4) displays the highest desaturase activity observed to date with 42% conversion of **14** in our system. Just like A0A0C2I951, A0A0J0YHR8 also contains a threonine at position 37 further proving that the total protein environment likely influences redox potential. Our laboratory is currently seeking to determine the redox potential of library members with desaturase activity to help elucidate the cause of this novel activity. Furthermore, U6SIA5 (cluster 6) from *Bacillus marmarensis* DSM 21297 displays desaturase activity towards **15**. This could serve as another potential candidate for further biochemical characterizations.

Solubility of Novel OYE Family Members

The genes for both the first and second generation libraries were codon optimized for *E. coli* expression and chemically synthesized. OYEs were heterologously expressed and their solubility was assessed by SDS-PAGE based on a five point scale. **Figure 4.4A** and **B** show the solubility profile and statistical distribution across the first and second generation enzymes. For the first generation, 6 enzymes fell into either the 'mostly soluble' or '50% soluble' categories, which we consider to be purifiable. The largest percentage, which corresponds to 11 enzymes, were 'mostly insoluble' and there are another six enzymes which were completely 'insoluble.' Lastly, there was one enzyme that did not express in our system. Ten enzymes from the first generation contain an iron-sulfur binding motif, eight of which were either mostly or completely

insoluble. The lack of iron-sulfur charging machinery in *E. coli* could account for the poor solubility for these enzymes. In comparison, the second generation library had 38 ‘mostly soluble,’ 10 ’50% soluble,’ 15 ‘mostly insoluble,’ 24 ‘insoluble’ enzymes, and six enzymes which did not express. Between the two generations, there was an increase of purifiable enzymes from 38% to 53%. Unfortunately, the percentage of insoluble OYEs remained the same across the two generations, and there was an increase in the number of enzymes that did not express in the second generation.

In terms of soluble enzymes which were active, all of the ‘mostly soluble’ enzymes from the first generation were active whereas there were 7 ‘mostly soluble’ enzymes of the second generation which were not active. Interestingly, in both libraries there are enzymes which have activity when screened in our cell-free assay, yet they are insoluble in our current heterologous expression system. For example, two OYEs (B9T8J4 and A0YFJ6) from the first generation, originating from clusters 3 and 11, respectively, were of particular interest due to their activity profiles. However, both entities are considered insoluble proteins based on our solubility testings. Upon examination, the cluster 3 enzyme, B9T8J4 from *Ricinus communis*, was found to possess an 81 amino acid N-terminal extension which is predicted to be a transmembrane anchor domain. Upon removal of this putative anchor domain, we observed a shift from insoluble to ‘mostly soluble,’ allowing for further biochemical characterization. Conversely, there was no obvious N- or C-terminal extension available for truncation on the cluster 11 enzyme, identified as A0YFJ6 from marine gamma proteobacterium. Instead, the enzyme was recombinantly fused to maltose binding protein (MBP) to produce soluble protein. However, these strategies are not general. Other insoluble proteins, such as cluster 9’s A0A017SC09 from *Aspergillus ruber* CBS 135680,

Figure 4.6. Solubility Data for First and Second Generation Libraries. Uniprot codes for each well can be found in the supplemental. **A.** Heat maps show level of solubility from no expression (white) to mostly soluble (green) for first (top left) and second (top right) generation. **B.** Pie charts show percent distribution of enzymes at each level of solubility.

were subjected to both approaches without any increase solubility. Of the 6 enzymes in the second generation that did not express in *E. coli*, an archeal enzyme from cluster 3, A0A1Z9EHY4, did have activity in the PURE system. The second generation is still being analyzed for candidates whose solubility could potentially be improved.

III. Conclusion

OYEs are an extensively studied family of enzymes, but there is still much that we can learn about this very large group of potential biocatalysts. Previously, researchers have focused on a relatively limited subset of enzymes. Modern sequencing technologies and advances in bioinformatic tools are now enabling us to explore the superfamily more thoroughly. Thus far our analysis has allowed us to observe differences in activity and substrate profiles which have not yet been reported. Our analyses has relied heavily on the co-development of methods for the rapid and systematic exploration of the OYE superfamily in silico (SSN) and at the bench (PURE) granting us the opportunity to screen 120 novel OYEs. The results from these initial studies of the first and second generation libraries have been prolific and have proven how expansive this superfamily truly is as we have discovered biocatalysts possessing increased catalytic activity towards industrially relevant substrates and enzymes which display novel reactivity. For example, we found promising new biocatalysts, K3WFC3 and A0A2A9JE68, which have broad specificities, as well as A0A1G6QTQ7, A0A245ZID8, A0A136H5V4, and A0A017SC09, which show conversion of previously challenging substrates, **6**, **8**, and **9**. Additionally, we identified OYEs with increased enantioselectivity for substrate **12** (A0A0R0D6E4, A0A2G2QFR0, and A0A061QIQ1). Lastly, we have observed 34 enzymes

possessing various levels of desaturase activity with the most active one (A0A0J0YHR8) converting 42% of **14**.

Since the first literature report of OYE1 in 1935, these enzymes have been subjected to various types of mutagenesis in order to improve catalytic activity, stability, alter substrate scope, and enhance enantioselectivity. While these efforts have been successful, our initial dive into the superfamily has uncovered OYEs which can outcompete currently published native and engineered OYE biocatalysts without the need for any mutagenic alterations. This further provides evidence that superfamilies in general are harboring an enormous amount of untapped potential which could meet currently unfulfilled industrial demands.

Utilization of SSNs for visualization of the superfamily has also provided insight into how these enzymes cluster together, laying a foundational framework for classifying related enzymes within the superfamily. Furthermore, our characterization experiments are beginning to grant us empirical evidence underpinning the unique functional aspects inherent to each individual cluster. For example, there is a strong argument that the structural differences of the ‘classical’ and ‘thermophilic’ OYEs are significantly distinctive based not only on how these enzymes separate into clusters 1 and 3, respectively, but also based on their affinities toward different substrates. However, for the other clusters within the superfamily the distinctions are not necessarily as clear, yet. The manner in which these enzymes are grouped into clusters is more complex than simply relying on the typical active site residues that are traditionally considered when classifying OYEs. Further biochemical characterization and protein crystallization will be necessary to discern these mechanistic and structural differences.

IV. Experimental

General Information

The PURExpress kit was purchased from New England Biolabs (Ipswich, MA). Full length genes of library members were codon optimized for *E. coli* expression and then chemically synthesized and ligated into pET28a by the DOE Joint Genome Institute. Substrate **8** was a gift from Melanie Hall at the University of Graz. All other reagents were purchased from Sigma Aldrich (St. Louis, MO) unless otherwise specified. 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 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 200°C, MS mode, EI; detector voltage, 0.2 kV; mass range, 12-250 u; scan speed, 833 u/s. All GC/MS data were acquired by a GC/MS Lab Solutions software (Shimadzu).

PURExpress In vitro Transcription/Translation of OYE Library Members

Library members in pET-28a were used as template DNA for *in vitro* transcription/ translation (IVTT). IVTT reactions were assembled according to the PURExpress *in vitro* protein synthesis kit with the following modifications optimized for our system. Reactions were assembled on a 20 uL scale containing 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, 200 ng of DNA, and OmniPur RNAse Free Water (EMD). All reactions were run simultaneously with OYE1 in pET14b as a positive control and dihydrofolate reductase (DHFR, NEB) as a

negative control. Reaction mixtures were incubated at 37°C for 2.5 hours. Following synthesis, reactions were quenched by cooling to 4°C for 5 minutes before being directly adding into the activity assay.

Enzymatic Activity Assay for OYE Library Members

Ene-reductase activity for enzymes was measured in duplicate under anaerobic conditions at room temperature using GDH from *Thermoplasma acidophilum* for the regeneration of NADPH and NADH. All substrates were dissolved in DMSO (0.5% final concentration) to overcome poor water solubility. Three separate reaction stocks were prepared, one for each mix I-III. The stocks were prepared in 50 mM Tris HCl pH 7.5 containing 100 mM glucose, 2 U per reaction GDH, 200 μM NADPH, 200 μM NADH, and either 250 μM of substrates **1**-**4** for Mix I, 250 μM of substrates **4**-**8** for Mix II, or 1 mM of **9**-**12** for Mix III. Desaturase activity was measured under aerobic conditions at room temperature. Reaction stocks were prepared in 20 mM pyrophosphate buffer pH 8.5 with 250 μM of substrates **13**-**16**. 4 μL of the IVTT reaction was added to 26 μL of each mix I-IV. After a 24 h period, reactions were quenched and products were extracted twice by adding 30 μL of ethyl acetate containing 0.5 mM cyclohexanone as an internal standard. 2 μL of the organic phase was injected and analyzed via chiral GC-MS. Products were confirmed by commercial standards when available or mass spectrometry.

Mix I and IV GC/MS Protocol: 110°C to 150 (5 min) at 2°C/min; linear velocity 52.8 cm/sec; injection temperature 250°C. (retention times for **1**: 15.8 min, **2**: 14.1 min, **3**: 20.4 min, **4**: 18.0 min, (*R*)-**13**: 13.6 min, (*R*)-**14**: 14.9 min, **15**: 13.8 min, **16**: 16.9 min.

Mix II GC/MS Protocol: 80°C (1 min) to 100°C (1 min) at 1°C/min to 185°C (10 min) at 2.5 °C/min; linear velocity 53.9 cm/sec; injection temperature 220°C. **5**: 8.4 min, **6**: 10.4 min, **7**: 22.4 min, **8**a and **8**b: 34.9 and 35 min.

Mix III GC/MS Protocol: 75°C (5 min) to 100°C at 2°C/min to 200°C (5 min) at 10°C/min; linear velocity 73.1 cm/sec; injection temperature 210°C. **9**: 15.2 min, **10**: 3 min, **11**: 25.3 min, **12**: 26.7 min.

Solubility of OYE Library Members

Plasmid DNA was transformed into *E. coli* BL21(DE3) cells for expression and plated on LB plates supplemented with kanamycin (30 mg/mL) at 37°C. Individual colonies were used to inoculate 5 mL overnight cultures of LB with kanamycin at 37°C. 100 μL of the saturated overnight culture was used to inoculate a fresh 5 mL of 2YT with antibiotics. Cultures were grown at 37°C to an OD_{600} of 0.6, then protein expression was initiated by adding 0.3 mM IPTG, and expression was carried out overnight at 20°C for 18 h. Cultures were centrifuged for 3 min at 3220 g and 4 °C. Pellets were lysed with 250 μL of BugBuster (EMD) and rocked at room temperature for 20 min. Lysed samples were then centrifuged at 8000 g for 10 min at 4°C. Soluble fraction was removed and the insoluble pellet was resuspended in 250 μL of DI water. Soluble and insoluble fractions were analyzed for protein content by SDS-PAGE.

V. Supplemental

Table 4.S1. Uniprot Codes for First and Second Generation Library. Each table corresponds to Figures 4.3A and 4.6A.

Uniprot Codes for First Generation Members

	A	B	C	D	E	F	G	H
1	M5B1C9	V7I5A8	W7ZMG5	A0YFJ6	B9Y7J4	Q9JN79	Q9WYQ9	R7SZ38
2	S6WB48	Q24RN8	D7ZLJ8	P54524	C0ZVN3	A0A017SC09	V4RWU7	Q4AA30
3	A8FEW4	Q4WZ70	B9T8J4	W0DA85	K6TQP5	P77258	M2Z4K1	C3MNJ2
4	K0B444	Q8PUE0	Q2TJB8	S5YUA6	J7L5I6	C0QFC4	Q02899	J3P9F5

Uniprot Codes for Second Generation Members

	A	B	C	D	E	F	G	H
1	E3HDQ4	A0A062X679	R4K9R9	A0A0H4KFX4	A0A1Q4W807	A0A166JDD2	A0A014MP39	A0A1I5G981
2	A0A1H1DXC0	A5VBP0	A0A0P7GE52	A0A2E0KBB9	A0A1Q5C6Q4	A0A0D5AF70	A0A136H5V4	A0A1R0ZM98
3	A0A1H3H521	A0A0R0D6E4	A0A0M4QXT8	A0A2G2QFR0	W9ATX1	A0A2D5KJ34	A0A0E3SPA7	A0A1M7EKV3
4	A0A1Q3H1G5	A0A1X6MY35	A0A1Y1YHM2	A0A245ZID8	A0A239MCT0	A0A1I6Q9M7	B5XWM6	P71278
5	A0A104JBL3	C7ZM49	A0A0Q7WCQ9	A0A087B547	A0A098SST3	A0A1X2BF87	Q87XC7	A0A0J0YHR8
6	U9YMT6	A0A2G0VWT1	A0A0B7GCC5	A0A0D6I6F4	A0A0C2I951	A0A2A9JE68	A0A161P8U3	A0A1M5T1E8
7	M2VAD5	A0A2C2U3H6	A0A0F9NP68	A0A137SCF2	U6SIA5	K3WFC3	A0A0S3C0K6	A0A061QIQ1
8	A0A1C4MT72	A0A1Z9EHY4	D8M5A9	A0A1D1W904	J3EXQ1	A0A100YAA1	A0A151VNU3	A0A2G4DT26
9	A0A2D1JR17	A0A0U1LWY3	W8NW54	J3GED7	A0A1A0W0R3	D0NQX8	A0A0A2JDP3	A0A0R3L3J0
10	E9E863	A0A1G6QTQ7	G4HUZ7	A0A226DQD1	K5Y813	A0A2A3FEX1	E3LUM4	A0A1H1V287
11	A0A1Q8DL33	A0A1Y1W9M1	Q6CUW9	A0A0B7N6T2	A0A1H1PA63	L7ITH5	G3JMK4	A0A261Y278
12	A0A1W0XDZ4	B8LTL5	A0A1S9RTM4	V5TJE2	Q3JA76	Q23090		

Table 4.S2. Current List of OYEs in the Literature.

UniProt #	Name	Organism	Cluster	Ref.
B0C3H7	AcaryoER1	Acaryochloris marina	1	(76)
A8ZP20	AcaryoER3	Acaryochloris marina	1	(76)
I3V5V6	Achr-OYE4	Achromobacter sp. JA82	1	(83)
Q8LAH7	AtOPR1	Arabidopsis thaliana	1	(65)
Q8GYB8	AtOPR2	Arabidopsis thaliana	1	(42)
Q9FUP0	AtOPR3	Arabidopsis thaliana	1	(44)
C5H429	DBR2	Artemisia annua	1	(8)
Q4WZ70	EasA	Aspergillus fumigatus	1	(78)
P43084	EBP1	Candida albicans	1	(36)
A0A0U2HAF1	Chr-OYE2	Chryseobacterium	1	(91)
C4Y835	ClER	Clavispora lusitaniae	1	(80)
B7K0Q6	CyanothER1	Cyanothece	1	(76)
B7JW34	CyanothER2	Cyanothece	1	(76)
Q6JL81		Enterobacter cloacae	1	(7)
P71278	PETNR	Enterobacter cloacae	1	(39)
P77258	NemR	Escherichia coli	1	(41)
A1E8I9	GoOYE	Gluconobacter oxydans	1	(68)
Q5FTL6	Gox0502	Gluconobacter oxydans	1	(63)
Q5HXK5	Gox2684	Gluconobacter oxydans	1	(63)
P40952	KYE1	Kluyveromyces lactis	1	(59)
Q6I7B7	CYE	Kluyveromyces marxianus	1	(47)
W0T7Z3	KYE2	Kluyveromyces marxianus CBS4857	1	(98)
E9AGH7	NTR2	Leishmania infantum	1	(93)
A0YMI2	LyngbyaER1	Lyngbya	1	(76)
A5DR62	MgER	Meyerozyma guilliermondii	1	(8)
Q8YVV8	NostocER1	Nostoc sp. PCC 7120	1	(76)

Table 4.S2. Continued

UniProt #	Name	Organism	Cluster	Ref.
B2IX00	NospuncER1	Nostoc punctiforme	1	(76)
Q84QK0	OsOPR1	Oryza sativa subsp. japonica	1	(67)
Q69TH8	OsOPR4	Oryza sativa subsp. japonica	1	(56)
Q69TI0	OsOPR5	Oryza sativa subsp. japonica	1	(56)
Q6Z965	OsOPR7	Oryza sativa subsp. japonica	1	(56)
Q0E0C6	OsOPR8	Oryza sativa subsp. japonica	1	(56)
W6Q9S9	fgaOx3	Penicillium roqueforti	1	(12)
Q8J293	HYE1	Pichia angusta	1	(48)
Q8J292	HYE2	Pichia angusta	1	(48)
Q8J291	HYE3	Pichia angusta	1	(48)
Q9RPM1	XenB	Pseudomonas fluorescens	1	(45)
Q51990	MR	Pseudomonas putida	1	(35)
A0A0A0VDJ9	XpdB	Rhizobium radiobacter	1	(101)
O31246	NerA	Rhizobium radiobacter	1	(97)
Q03558	OYE2	Saccharomyces cerevisiae	1	(52)
P41816	OYE3	Saccharomyces cerevisiae	1	(60)
Q02899	OYE1	Saccharomyces pastorianus	1	(34)
A3LT82	OYE2.6	Scheffersomyces stipitis	1	(81)
Q3V799	SL41	Scytosiphon lomentaria	1	(54)
Q8EEC8	SYE1	Shewanella oneidensis	1	(58)
Q8E9V9	SYE3	Shewanella oneidensis	1	(57)
Q8EBV3	SYE4	Shewanella oneidensis	1	(94)
Q9XG54	OPR1	Solanum lycopersicum	1	(49)
Q9FEX0	OPR2	Solanum lycopersicum	1	(49)
Q9FEW9	OPR3	Solanum lycopersicum	1	(49)
Q31R14	Syn7942ER	Synechococcus elongatus	1	(76)

Table 4.S2. Continued

UniProt #	Name	Organism	Cluster	Ref.
Q2TJB8	TcOYE	Trypanosoma cruzi	1	(70)
A0A2G4TY30	Yers-ER	Yersinia bercovieri	1	(71)
Q5NLA1	NCR	Zymomonas mobilis	1	(60)
A4J778	Dred_2421	Desulfotomaculum reducens	2	(88)
P42593	DCR	Escherichia coli	2	(40)
E1CIA4		Lactococcus garvieae	2	(69)
P16099	TMADH	Methylophilus methylotrophus	2	(32)
O52935		Moorella thermoacetica	2	(46)
Q6IWJ5	HADH	Nocardioides simplex	2	(9)
I3V5V5	Achr-OYE3	Achromobacter sp. JA81	3	(75)
Q3M1A7	AnabaenaER3	Anabaena variabilis	3	(76)
H9N833	PfvC	Arthrobacter sp. JBH1	3	(74)
Q2UCA4	AspER	Aspergillus oryzae	3	(55)
X2KU66	BcOYE	Bacillus coagulans	3	(86)
P54550	YqjM	Bacillus subtilis (strain 168)	3	(50)
A0A2M9W051	Bac-OYE1	Bacillus vallismortis	3	(8)
A0A023MJL1	Chr-OYE3	Chryseobacterium sp. CA49	3	(85)
Q9RSD4	DrER	Deinococcus radiodurans	3	(76)
A0A0R0LNE8	FOYE-1	Ferrovum sp. JA12	3	(95)
Q5KXG9	GkOYE	Geobacillus kaustophilus	3	(24)
W0T2Q3	GeoER	Geobacillus sp.	3	(82)
Q74DE5	GSU1371	Geobacter sulfurreducens	3	(87)
Q7NEQ0	GloeoER	Gloeobacter violaceus	3	(76)
Q9R9V9	XenA	Pseudomonas putida	3	(43)
Q1LDQ5	RmER	Ralstonia metallidurans	3	(73)
A0A1B1KA86	OYERo2	Rhodococcus opacus	3	(90)

Table 4.S2. Continued

UniProt #	Name	Organism	Cluster	Ref.
Q883F8		Pseudomonas syringae pv. tomato str. DC3000	3	(25)
Q87X32		Pseudomonas syringae pv. tomato str. DC3000	1	(25)
Q48MP2		Pseudomonas syringae pv. phaseolicola 1448A	4	(25)
Q48LU5		Pseudomonas syringae pv. phaseolicola 1448A	3	(25)
Q48EK4		Pseudomonas syringae pv. phaseolicola 1448A	1	(25)
Q0SBS7		Rhodococcus sp. RHAI	8	(25)
Q0RXM6		Rhodococcus sp. RHAI	N/A	(25)
A0A0H3JQS4		Staphylococcus auerus	6	(25)
Q82MZ3		Streptomyces sp. MA-4680	3	(25)
Q92YK2		Sinorhizobium meliloti 1021	1	(25)
Q8PDM8		Xanthomonas campestris	3	(25)
Q97DV0	CaER	Clostridium acetobutylicum	2	(25)
A5N665		Clostridium kluyveri	2	(46)
D8GRM6		Clostridium ljungdahlii	2	(25)
O52933	ER-CT	Clostridium tyrobutyricum	2	(46)
G2TQU6		Bacillus coagulans	2	(25)
P32382		Thermoanaerobacter brockii	2	(33)
O29794		Archaeoglobus fulgidus	2	(51)
P32370	baiH	Clostridium scindens	2	(37)
P19410	baiCD	Clostridium scindens	2	(61)
V9P074		Eubacterium ramulus	2	(79)
G9F1Y9	FldZ	Clostridium sporogenes	2	(99)

Table 4.S2. Continued

UniProt #	Name	Organism	Cluster	Ref.
A0A125RY22	RhrER 2178	Rhodococcus rhodochrous	3	(76)
B0KAH1	TOYE	Thermoanaerobacter pseudethanolicus E39	3	(16)
B0JDW3	TsOYE	Thermus scotoductus	3	(62)
A0A0U2H4S5	Chr-OYE1	Chryseobacterium	4	(91)
P54524	YqiG	Bacillus subtilis	6	(92)
S4ZS69	LacER	Lactobacillus casei	6	(72)
A0A0U3I6M0	Nox	Rhodococcus erythropolis	7	(11)
Q747D7	GSU3330	Geobacter sulfurreducens	10	(89)
A0A2A9I9R7	Lla-ER	Lactococcus lactis	13	(14)
C0ZMK6	Rer-ER7	Rhodococcus erythropolis	4	(14)
A0A378Y5I0	Ppo-ER3	Paenibacillus polymyxa	4	(14)
A0A370VYJ4	ChrR	Streptomyces sp. M7	3	(96)
E1YD54	NCR	metagenomic Desulfobacterium sp.	2	(66)
Q9KCT8		Bacillus halodurans C-125	3	(25)
Q7NSC5		Chromobacterium violaceum	1	(25)
Q8Y4H1		Listeria monocytogenes	3	(25)
Q82S49		Nitrosomonas europaea	1	(25)
Q9HZR5		Pseudomonas aeruginosa PAOI	1	(25)
Q9HW45		Pseudomonas aeruginosa PAOI	1	(25)
Q88PD0		Pseudomonas putida KT2440	1	(25)
Q88NF7		Pseudomonas putida KT2440	3	(25)
Q88MU0		Pseudomonas putida KT2440	3	(25)
Q887W0		Pseudomonas syringae pv. tomato str. DC3000	4	(25)

Table 4.S2. Continued

UniProt #	Name	Organism	Cluster	Ref.
C5H429	DBR2	Artemisia annua	1	(8)
Q48303	dmd	Hyphomicrobium sp. (strain x)	2	(38)
Q0JZ81	ReuER	Ralstonia eutropha	1	(100)
Q60I59	HADH	Rhizobium sp.	2	(53)
M9NZ71	DZNR	Slackia isoflavoniconvertens	2	(77)
C1KYN8	LmoER	Listeria monocytogenes	3	(100)
Q8NLA7	CglER	Corynebacterium glutamicum	3	(100)
V0AYG0	KpnER	Klebsiella pneumoniae	3	(100)
U5KKF9	XenB	Pseudomonas putida ATCC 17453	1	(14)
W5RUC0	NemA	Pseudomonas putida ATCC 17453	1	(14)
Q88K07	XenC	Pseudomonas putida KT2440	1	(64)
Q88K04	XenD	Pseudomonas putida KT2440	3	(64)
Q88I29	XenF	Pseudomonas putida KT2440	1	(64)

VI. References

1. Bornscheuer Uwe, T., *The fourth wave of biocatalysis is approaching.* Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2018. **376**(2110): p. 20170063.

2. Scholtissek, A., et al., Old yellow enzyme-catalysed asymmetric hydrogenation: Linking family roots with improved catalysis. Catalysts, 2017. **7**(5): p. 130.

3. Korpak, M. and J. Pietruszka, Chemoenzymatic one-pot synthesis of γ-butyrolactones. Advanced Synthesis & Catalysis, 2011. **353**(9): p. 1420.

4. Pietruszka, J. and M. Schölzel, Ene reductase-catalysed synthesis of (r)-profen derivatives. Advanced Synthesis & Catalysis, 2012. **354**(4): p. 751.

5. Copp, J.N., et al., Revealing unexplored sequence-function space using sequence similarity networks. Biochemistry, 2018. **57**(31): p. 4651.

6. Breithaupt, C., et al., X-ray structure of 12-oxophytodienoate reductase 1 provides structural insight into substrate binding and specificity within the family of oye. Structure, 2001. **9**(5): p. 419.

7. Williams, R.E., et al., Biotransformation of explosives by the old yellow enzyme family of flavoproteins. Applied and Environmental Microbiology, 2004. **70**(6): p. 3566.

8. Zhang, Y., et al., The molecular cloning of artemisinic aldehyde δ11(13) reductase and its role in glandular trichome-dependent biosynthesis of artemisinin in artemisia annua. Journal of Biological Chemistry, 2008. **283**(31): p. 21501.

9. Reed, T., et al., Crystal structure of histamine dehydrogenase from nocardioides simplex. Journal of Biological Chemistry, 2010. **285**(33): p. 25782.

10. Toogood, H.S., J.M. Gardiner, and N.S. Scrutton, Biocatalytic reductions and chemical versatility of the old yellow enzyme family of flavoprotein oxidoreductases. ChemCatChem, 2010. **2**(8): p. 892.

11. Khairy, H., J.H. Wübbeler, and A. Steinbüchel, The nadh:Flavin oxidoreductase nox from rhodococcus erythropolis mi2 is the key enzyme of 4,4′-dithiodibutyric acid degradation. Letters in Applied Microbiology, 2016. **63**(6): p. 434.

12. Gerhards, N. and S.-M. Li, A bifunctional old yellow enzyme from penicillium roqueforti is involved in ergot alkaloid biosynthesis. Organic & Biomolecular Chemistry, 2017. **15**(38): p. 8059.

13. Nizam, S., et al., Comprehensive genome-wide analysis reveals different classes of enigmatic old yellow enzyme in fungi. Scientific Reports, 2014. **4**: p. 4013.

14. Peters, C., et al., Novel old yellow enzyme subclasses. ChemBioChem. **0**(0).

15. Kitzing, K., et al., The 1.3 å crystal structure of the flavoprotein yqjm reveals a novel class of old yellow enzymes. Journal of Biological Chemistry, 2005. **280**(30): p. 27904.

16. Adalbjörnsson, B.V., et al., Biocatalysis with thermostable enzymes: Structure and properties of a thermophilic 'ene'-reductase related to old yellow enzyme. ChemBioChem, 2010. **11**(2): p. 197.

17. Nizam, S., et al., Comparative structural modeling of six old yellow enzymes (oyes) from the necrotrophic fungus ascochyta rabiei : Insight into novel oye classes with differences in cofactor binding, organization of active site residues and stereopreferences. PLOS ONE, 2014. **9**(4): p. e95989.

18. Apweiler, R., et al., Uniprot: The universal protein knowledgebase. Nucleic Acids Research, 2004. **32**(suppl_1): p. D115.

19. Shannon, P., et al., Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res, 2003. **13**(11): p. 2498.

20. Atkinson, H.J., et al., Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. PLOS ONE, 2009. **4**(2): p. e4345.

21. Müller, A., et al., Stereospecific alkyne reduction: Novel activity of old yellow enzymes. Angewandte Chemie International Edition, 2007. **46**(18): p. 3316.

22. Turrini, N.G., M. Hall, and K. Faber, Enzymatic synthesis of optically active lactones via asymmetric bioreduction using ene-reductases from the old yellow enzyme family. Advanced Synthesis & Catalysis, 2015. **357**(8): p. 1861.

23. Winkler, C.K., et al., Asymmetric bioreduction of activated alkenes to industrially relevant optically active compounds. Journal of Biotechnology, 2012. **162**(4): p. 381.

24. Schittmayer, M., et al., Old yellow enzyme-catalyzed dehydrogenation of saturated ketones. Advanced Synthesis & Catalysis, 2011. **353**(2−3): p. 268.

25. Joo, J.C., et al., Alkene hydrogenation activity of enoate reductases for an environmentally benign biosynthesis of adipic acid. Chemical Science, 2017. **8**(2): p. 1406.

26. Shimizu, Y., et al., Cell-free translation reconstituted with purified components. Nature Biotechnology, 2001. **19**(8): p. 751.

27. Shimizu, Y., T. Kanamori, and T. Ueda, Protein synthesis by pure translation systems. Methods, 2005. **36**(3): p. 299.

28. Forster, A.C., et al., Programming peptidomimetic syntheses by translating genetic codes designed de novo. Proceedings of the National Academy of Sciences, 2003. **100**(11): p. 6353.

29. Inobe, T., et al., Asymmetry of the groel-groes complex under physiological conditions as revealed by small-angle x-ray scattering. Biophysical Journal, 2008. **94**(4): p. 1392.

30. Li, J., et al., Improved cell-free rna and protein synthesis system. PLOS ONE, 2014. **9**(9): p. e106232.

31. Weckbecker, A. and W. Hummel, *Glucose dehydrogenase for the regeneration of nadph and nadh*, in *Microbial enzymes and biotransformations*, J.L. Barredo, Editor. 2005, Humana Press: Totowa, NJ. p. 225.

32. Boyd, G., et al., Trimethylamine dehydrogenase of bacterium w3a1 molecular cloning, sequence determination and over-expression of the gene. FEBS Letters, 1992. **308**(3): p. 271.

33. Liu, X.-L. and R.K. Scopes, Cloning, sequencing and expression of the gene encoding nadh oxidase from the extreme anaerobic thermophile thermoanaerobium brockii. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression, 1993. **1174**(2): p. 187.

34. Fox, K.M. and P.A. Karplus, Old yellow enzyme at 2 Å resolution: Overall structure, ligand binding, and comparison with related flavoproteins. Structure, 1994. **2**(11): p. 1089.

35. French, C.E. and N.C. Bruce, Purification and characterization of morphinone reductase from pseudomonas putida m10. Biochem J, 1994. **301 (Pt 1)**(Pt 1): p. 97.

36. Madani, N.D., et al., Candida albicans estrogen-binding protein gene encodes an oxidoreductase that is inhibited by estradiol. Proc Natl Acad Sci U S A, 1994. **91**(3): p. 922.

37. Baron, S.F. and P.B. Hylemon, Expression of the bile acid-inducible nadh:Flavin oxidoreductase gene of eubacterium sp. Vpi 12708 in escherichia coli. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology, 1995. **1249**(2): p. 145.

38. Yang, C.-C., L.C. Packman, and N.S. Scrutton, The primary structure of hyphomicrobium x dimethylamine dehydrogenase. European Journal of Biochemistry, 1995. **232**(1): p. 264.

39. Binks, P.R., et al., Degradation of pentaerythritol tetranitrate by enterobacter cloacae pb2. Applied and environmental microbiology, 1996. **62**(4): p. 1214.

40. He, X.-Y., S.-Y. Yang, and H. Schulz, Cloning and expression of the fadh gene and characterization of the gene product 2,4-dienoyl coenzyme a reductase from escherichia coli. European Journal of Biochemistry, 1997. **248**(2): p. 516.

41. Miura, K., et al., Molecular cloning of the nema gene encoding n-ethylmaleimide reductase from escherichia coli. Biological & Pharmaceutical Bulletin, 1997. **20**(1): p. 110.

42. Biesgen, C. and E.W. Weiler, Structure and regulation of opr1 and opr2, two closely related genes encoding 12-oxophytodienoic acid-10,11-reductases from arabidopsis thaliana. Planta, 1999. **208**(2): p. 155.

43. Blehert, D.S., B.G. Fox, and G.H. Chambliss, Cloning and sequence analysis of two pseudomonas flavoprotein xenobiotic reductases. J Bacteriol, 1999. **181**(20): p. 6254.

44. Costa, C.L., P. Arruda, and C.E. Benedetti, An arabidopsis gene induced by wounding functionally homologous to flavoprotein oxidoreductases. Plant Molecular Biology, 2000. **44**(1): p. 61.

45. Pak, J.W., et al., Transformation of 2,4,6-trinitrotoluene by purified xenobiotic reductase b from pseudomonas fluorescens i-c. Applied and environmental microbiology, 2000. **66**(11): p. 4742.

46. Rohdich, F., et al., Enoate reductases of clostridia : Cloning, sequencing, and expression. Journal of Biological Chemistry, 2001. **276**(8): p. 5779.

47. Kataoka, M., et al., Old yellow enzyme from candida macedoniensis catalyzes the stereospecific reduction of the c=c bond of ketoisophorone. Bioscience, Biotechnology, and Biochemistry, 2002. **66**(12): p. 2651.

48. Komduur, J.A., et al., Old yellow enzyme confers resistance of hansenula polymorpha towards allyl alcohol. Current Genetics, 2002. **41**(6): p. 401.

49. Strassner, J., et al., Characterization and cdna-microarray expression analysis of 12-oxophytodienoate reductases reveals differential roles for octadecanoid biosynthesis in the local versus the systemic wound response. The Plant Journal, 2002. **32**(4): p. 585.

50. Fitzpatrick, T.B., N. Amrhein, and P. Macheroux, Characterization of yqjm, an old yellow enzyme homolog from bacillus subtilis involved in the oxidative stress response. Journal of Biological Chemistry, 2003. **278**(22): p. 19891.

51. Kengen, S.W.M., J. van der Oost, and W.M.d. Vos, Molecular characterization of h2o2-forming nadh oxidases from archaeoglobus fulgidus. European Journal of Biochemistry, 2003. **270**(13): p. 2885.

52. Wada, M., et al., Production of a doubly chiral compound, (4r,6r)-4-hydroxy-2,2,6-trimethylcyclohexanone, by two-step enzymatic asymmetric reduction. Applied and Environmental Microbiology, 2003. **69**(2): p. 933.

53. Bakke, M., et al., Histamine dehydrogenase from rhizobium sp.: Gene cloning, expression in escherichia coli, characterization and application to histamine determination. Journal of Biotechnology, 2005. **119**(3): p. 260.

54. Fujita, S., et al., Identification and characterization of a fluorescent flagellar protein from the brown alga scytosiphon lomentaria (scytosiphonales, phaeophyceae): A flavoprotein homologous to old yellow enzyme. European Journal of Phycology, 2005. **40**(2): p. 159.

55. Machida, M., et al., Genome sequencing and analysis of aspergillus oryzae. Nature, 2005. **438**(7071): p. 1157.

56. Sasaki, T. and P. International Rice Genome Sequencing, *The map-based sequence of the rice genome.* Nature, 2005. **436**(7052): p. 793.

57. Brigé, A., et al., Comparative characterization and expression analysis of the four old yellow enzyme homologues from shewanella oneidensis indicate differences in physiological function. Biochemical Journal, 2006. **394**(1): p. 335.

58. van den Hemel, D., et al., Ligand-induced conformational changes in the capping subdomain of a bacterial old yellow enzyme homologue and conserved sequence fingerprints provide new insights into substrate binding. Journal of Biological Chemistry, 2006. **281**(38): p. 28152.

59. Chaparro-Riggers, J.F., et al., Comparison of three enoate reductases and their potential use for biotransformations. Advanced Synthesis & Catalysis, 2007. **349**(8−9): p. 1521.

60. Müller, A., B. Hauer, and B. Rosche, Asymmetric alkene reduction by yeast old yellow enzymes and by a novel zymomonas mobilis reductase. Biotechnology and Bioengineering, 2007. **98**(1): p. 22.

61. Kang, D.-J., et al., Clostridium scindens baicd and baih genes encode stereo-specific 7α/7β-hydroxy-3-oxo-δ4-cholenoic acid oxidoreductases. Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids, 2008. **1781**(1): p. 16.

62. Opperman, D.J., L.A. Piater, and E. van Heerden, A novel chromate reductase from thermus scotoductus sa-01 related to old yellow enzyme. J Bacteriol, 2008. **190**(8): p. 3076.

63. Schweiger, P., et al., Vinyl ketone reduction by three distinct gluconobacter oxydans 621h enzymes. Applied Microbiology and Biotechnology, 2008. **80**(6): p. 995.

64. van Dillewijn, P., et al., Subfunctionality of hydride transferases of the old yellow enzyme family of flavoproteins of pseudomonas putida. Applied and Environmental Microbiology, 2008. **74**(21): p. 6703.

65. Beynon, E.R., et al., The role of oxophytodienoate reductases in the detoxification of the explosive 2,4,6-trinitrotoluene by arabidopsis. Plant Physiology, 2009. **151**(1): p. 253.

66. Bergmann, F., et al., Genomic insights into the metabolic potential of the polycyclic aromatic hydrocarbon degrading sulfate-reducing deltaproteobacterium n47. Environmental microbiology, 2011. **13**(5): p. 1125.

67. Li, W., et al., Comparative characterization, expression pattern and function analysis of the 12-oxo-phytodienoic acid reductase gene family in rice. Plant Cell Reports, 2011. **30**(6): p. 981.

68. Richter, N., H. Gröger, and W. Hummel, Asymmetric reduction of activated alkenes using an enoate reductase from gluconobacter oxydans. Applied Microbiology and Biotechnology, 2011. **89**(1): p. 79.

69. Shimada, Y., et al., Identification of two novel reductases involved in equol biosynthesis in lactococcus strain 20–92. Journal of Molecular Microbiology and Biotechnology, 2011. **21**(3-4): p. 160.

70. Yamaguchi, K., et al., Structure of the inhibitor complex of old yellow enzyme from trypanosoma cruzi. Journal of Synchrotron Radiation, 2011. **18**(1): p. 66.

71. Yanto, Y., et al., Asymmetric bioreduction of alkenes using ene–reductases yerser and kye1 and effects of organic solvents. Organic Letters, 2011. **13**(10): p. 2540.

72. Gao, X., et al., Biochemical characterization and substrate profiling of a new nadh-dependent enoate reductase from lactobacillus casei. Enzyme and Microbial Technology, 2012. **51**(1): p. 26.

73. Grazu, V., F. López-Gallego, and J.M. Guisán, Tailor-made design of penicillin g acylase surface enables its site-directed immobilization and stabilization onto commercial mono-functional epoxy supports. Process Biochemistry, 2012. **47**(12): p. 2538.

74. Husserl, J., J.B. Hughes, and J.C. Spain, Key enzymes enabling the growth of arthrobacter sp. Strain jbh1 with nitroglycerin as the sole source of carbon and nitrogen. Applied and Environmental Microbiology, 2012. **78**(10): p. 3649.

75. Liu, Y.-J., et al., Asymmetric bioreduction of activated alkenes by a novel isolate of achromobacter species producing enoate reductase. Applied Microbiology and Biotechnology, 2012. **95**(3): p. 635.

76. Fu, Y., K. Castiglione, and D. Weuster-Botz, Comparative characterization of novel ene-reductases from cyanobacteria. Biotechnology and Bioengineering, 2013. **110**(5): p. 1293.

77. Schröder, C., et al., Identification and expression of genes involved in the conversion of daidzein and genistein by the equol-forming bacterium <span class="named-content genus-species" id="named-content-1">slackia isoflavoniconvertens</ span&gt. Applied and Environmental Microbiology, 2013. **79**(11): p. 3494.

78. Chilton, A.S., A.L. Ellis, and A.L. Lamb, Structure of an aspergillus fumigatus old yellow enzyme (easa) involved in ergot alkaloid biosynthesis. Acta Crystallogr F Struct Biol Commun, 2014. **70**(Pt 10): p. 1328.

79. Gall, M., et al., Enzymatic conversion of flavonoids using bacterial chalcone isomerase and enoate reductase. Angewandte Chemie International Edition, 2014. **53**(5): p. 1439.

80. Ni, Y., et al., An ene reductase from clavispora lusitaniae for asymmetric reduction of activated alkenes. Enzyme and Microbial Technology, 2014. **56**: p. 40.

81. Patterson-Orazem, A., B. Sullivan, and J.D. Stewart, Pichia stipitis oye 2.6 variants with improved catalytic efficiencies from site-saturation mutagenesis libraries. Bioorganic & Medicinal Chemistry, 2014. **22**(20): p. 5628.

82. Tsuji, N., et al., Isolation and characterization of a thermotolerant ene reductase from geobacillus sp. 30 and its heterologous expression in rhodococcus opacus. Applied Microbiology and Biotechnology, 2014. **98**(13): p. 5925.

83. Wang, H.-B., X.-Q. Pei, and Z.-L. Wu, An enoate reductase achr-oye4 from achromobacter sp. Ja81: Characterization and application in asymmetric bioreduction of c=c bonds. Applied Microbiology and Biotechnology, 2014. **98**(2): p. 705.

84. Wang, Z.J., et al., Cytochrome p450-catalyzed insertion of carbenoids into n–h bonds. Chemical Science, 2014. **5**(2): p. 598.

85. Xu, M.-Y., X.-Q. Pei, and Z.-L. Wu, Identification and characterization of a novel “thermophilic-like” old yellow enzyme from the genome of chryseobacterium sp. Ca49. Journal of Molecular Catalysis B: Enzymatic, 2014. **108**: p. 64.

86. Zhou, X., H.L. Chow, and J.C. Wu, Bioreduction of activated alkenes by a novel “ene”-reductase from the thermophilic strain bacillus coagulans wcp10-4. Biocatalysis and Biotransformation, 2014. **32**(5-6): p. 267.

87. Ali, M., et al., Physiological characterization of anaerobic ammonium oxidizing bacterium ‘candidatus jettenia caeni’. Environmental Microbiology, 2015. **17**(6): p. 2172.

88. Otwell, A.E., et al., Identification of proteins capable of metal reduction from the proteome of the gram-positive bacterium desulfotomaculum reducens mi-1 using an nadh-based activity assay. Environmental Microbiology, 2015. **17**(6): p. 1977.

89. Pump, J., J. Pratscher, and R. Conrad, Colonization of rice roots with methanogenic archaea controls photosynthesis-derived methane emission. Environmental Microbiology, 2015. **17**(7): p. 2254.

90. Riedel, A., et al., Functional characterization and stability improvement of a ‘thermophilic-like’ ene-reductase from rhodococcus opacus 1cp. Frontiers in Microbiology, 2015. **6**(1073).

91. Pei, X.-Q., M.-Y. Xu, and Z.-L. Wu, Two “classical” old yellow enzymes from chryseobacterium sp. Ca49: Broad substrate specificity of chr-oye1 and limited activity of chr-oye2. Journal of Molecular Catalysis B: Enzymatic, 2016. **123**: p. 91.

92. Sheng, X., et al., Identification and characterization of a novel old yellow enzyme from bacillus subtilis str.168. Journal of Molecular Catalysis B: Enzymatic, 2016. **130**: p. 18.

93. Wyllie, S., et al., Activation of bicyclic nitro-drugs by a novel nitroreductase (ntr2) in leishmania. PLOS Pathogens, 2016. **12**(11): p. e1005971.

94. Elegheert, J., et al., Structural dissection of shewanella oneidensis old yellow enzyme 4 bound to a meisenheimer complex and (nitro)phenolic ligands. FEBS Letters, 2017. **591**(20): p. 3391.

95. Scholtissek, A., et al., A thermophilic-like ene-reductase originating from an acidophilic iron oxidizer. Applied Microbiology and Biotechnology, 2017. **101**(2): p. 609.

96. Shan, W., et al., Draft genome sequence of streptomyces sp. Xy006, an endophyte isolated from tea (camellia sinensis). Genome Announc, 2017. **5**(37): p. e00971.

97. Waller, J., et al., Structural insights into the ene-reductase synthesis of profens. Organic & Biomolecular Chemistry, 2017. **15**(20): p. 4440.

98. Li, Z., et al., Identification of an ene reductase from yeast kluyveromyces marxianus and application in the asymmetric synthesis of (r)-profen esters. Asian Journal of Organic Chemistry, 2018. **7**(4): p. 763.

99. Mordaka, P.M., et al., Recombinant expression and characterisation of the oxygen-sensitive 2-enoate reductase from clostridium sporogenes. Microbiology, 2018. **164**(2): p. 122.

100. Wang, J., et al., Microbial production of branched-chain dicarboxylate 2-methylsuccinic acid via enoate reductase-mediated bioreduction. Metabolic Engineering, 2018. **45**: p. 1.

101. Zahradník, J., et al., The crystal structure of xdpb, the bacterial old yellow enzyme, in an fmn-free form. PLOS ONE, 2018. **13**(4): p. e0195299.

Chapter 5: Conclusions and Future Work

I. General Conclusions

Over the last century researchers have strived to understand enzyme mechanisms, as well as to develop biocatalysts for industrial, therapeutic, diagnostic, and environmental purposes (1-4). One of the ways enzymes have been engineered is through modifications to the cofactor housed within the active site. Inherently, cofactors provide catalytic power and versatility to enzymes to facilitate reactions which otherwise would not be feasible (5, 6). The most recent advances in cofactor engineering have focused on the development of artificial metalloenzymes to help bridge the gap between organic chemistry and biocatalysis (7, 8). Traditionally, cofactor engineering has been used in order to elucidate enzyme mechanisms and study active site environments (9-13). In this dissertation, I focused on the application of cofactor engineering on Old Yellow Enzyme 1 (OYE1) and the versatile role of the FMN cofactor within the OYE superfamily. In the cofactor engineering studies, I developed a number of FMN analogs to determine which structural aspects of the cofactor contribute in the folding of the catalytically active enzyme. Through these efforts I was able to conclude that FMN is indeed necessary for productive folding of the nascent enzyme; specifically, the isoalloxazine moiety is the driving force behind folding of OYE1 (chapter 3). Interestingly, the mechanism underpinning binding of the cofactor to pre-folded enzyme is different than the mechanism of the nascent protein folding as shown by the isothermal calorimetry (ITC) studies performed. For binding of the cofactor to apoenzyme, both the isoalloxazine and the phosphate moieties are necessary for efficient cofactor recognition. It is important to note that the folding and binding mechanisms are not universal across flavoenzymes and the role of the cofactor is dependent on the protein environment. My work involving the OYE superfamily (chapter 4) has showcased this point: the

activity profile across the OYE superfamily is diverse, which we can attribute to the variety of backbone scaffolds in which the FMN is housed. We have only just begun to sample the unexplored sequence space within the superfamily. So far, we have successfully identified several novel enzymes displaying superior catalytic efficiency compared to previously reported OYE family members. Future studies will investigate the subtle functional and structural differences between the family members in order to further identify biocatalysts of interest.

II. Continued Use of Cofactor Engineering

In chapter 3, I determined that the "Last In, First Out" theory pertaining to the involvement of FMN in protein folding does not apply to OYE1. The presence of FMN during protein synthesis is necessary for the production of catalytically active enzyme, which contradicts previous literature accounts. Based on these results it would appear that the combination of hydrophobic character along with the hydrogen bonding ability of the isoalloxazine moiety are necessary as analogs **1** and **2**, which each contain either the hydrogen bonding or hydrophobic aspect, respectively, did not produce high levels of active enzyme. However, it is important to note that analog **2** lacks the terminal phosphate group, which could have negatively impacted the analog's role during the folding of the nascent protein. Further attempts would be necessary to confirm this assertion. First, the synthesis of the phosphorylated **2** would need to be further optimized. Our current route does not produce **2** in high enough yield to successfully and selectively phosphorylate the 5'-hydroxyl of the ribitly tail. The largest detriment to yield occurs during the coupling of the 2-aminobiphenyl to the acetylated ribose. Optimization of this step would help with overall yield and would provide more material to

attempt phosphorylation. The current chemical phosphorylation method, which includes the use of phosphoryl chloride, are not selective enough, as shown by the mixture of products we obtain. In order to successfully phosphorylate **2**, the synthetic route may require an additional round of protection and deprotection steps. For example, once **2** is formed there would need to be a selective protection of the 5'-hydroxyl with a trityl protecting group followed by protection of the secondary hydroxyls with an additional protection group. The challenge lies in finding a second protecting group small enough to ensure that all three hydroxyls are protected while also being insensitive to the acid used in the deprotection of the trityl group. Once the hydroxyls are protected, the remaining steps of the synthetic route would include removal of the trityl group, followed by phosphorylation of the terminal hydroxyl, and finally deprotection of the secondary hydroxyls. While, this procedure adds five additional steps to the synthetic route, having the phosphorylated version of analog **2** would further elucidate the contributions of the structural aspects of FMN. Additionally, folding of OYE1 in the presence of just the carbazole moiety, representing the hydrophobic equivalent to lumiflavin, could be monitored allowing us to conclude if the aromatic aspect of FMN is enough to produce catalytically productive enzyme.

One of the major conclusions we drew from this work is that not all flavoenzymes follow the same folding or cofactor binding mechanism. Therefore, further studies could integrate the work performed in chapter 3 and apply it to the work performed in chapter 4. Expanding upon our ongoing exploration of other members of the OYE superfamily offers a fantastic opportunity to more broadly investigate the role of FMN for the functional assembly of these biocatalysts. Given that all OYE family members share the same FMN cofactor, the functional diversity cannot solely be linked to FMN but is more likely associated with the differences in protein

sequence. Utilizing the cofactor analogs developed in chapter 3, we could study the folding mechanism of these unique family members in order to probe the active site environment. This would allow us to observe similarities or differences in how these enzymes fold and provide further insight into structure-function relationships within the superfamily. This, of course, would be supplemented by crystallographic and kinetic studies of the OYE family members of interest.

III. Further Exploration of the OYE Superfamily

Continuing the ongoing exploration of OYE sequence space, my work has focussed on the systematic exploration of the OYE superfamily. Previous research involving the OYE superfamily has only focused on a small subset of the family, leaving the majority of the family unexplored. Since OYE1 was first discovered in 1935, it has been extensively studied for its catalytic potential as a multi-application biocatalyst due to its high regio- enantio-, and stereospecificity. As a result of these inherent attributes, OYE1 and other family members have been extensively engineered in order to increase efficiency, stability, or substrate scope. While these efforts have been successful, protein engineering has certain limitations dictated by the characteristics of the parent sequence and the our ability to predict specific amino acid contributions to overall performance (14). Rather than engineering previously reported OYEs, we set out to discover novel enzymes within this large superfamily possessing enhanced catalytic efficiencies, broader substrate scopes, or the ability to perform novel chemistries. As part of our systematic search for new OYEs, we employed sequence similarity networks (SSN) to help visualize and sort this superfamily. Through our screening efforts we have identified OYEs with increased catalytic efficiencies and distinct stereoselectivity for some substrates compared to our

benchmark enzyme, OYE1. The most exciting discovery has been the detection of high levels of desaturase activity among 34 of our screened OYE family members, which is unprecedented for this family. Due to the nature of this project and the amount of information we have gathered thus far, the future directions of this project are virtually endless. In the immediate future, the nexts steps we are currently undertaking include the selection of 10-12 novel OYEs for further kinetic, biophysical, and structural characterization in order to parse out the origins of observed differences in the catalytic profiles.

In conjunction with the heterologous characterization of select candidates we can explore the family from a different angle. Within the second generation library there are seven OYE family members which solubly expressed in our *E. coli* system yet lack any activity when screened with the PURE system. Of these seven enzymes five are eukaryotic and the remaining two are bacterial. It is possible that these enzyme require some post translational modification or require a different redox partner for activity which our PURE system and activity assay do not contain. While the substrates we have chosen cover a variety of structural motifs it is possible that the lack of activity we see for these seven OYEs is due to not having the desired substrate in our panel. Therefore, we plan to utilize genomic context analysis to examine the functions of the adjacent genes flanking our target proteins in the hopes of elucidating possible biological functions which could provide better substrates for these enzymes. In addition to the genomic context analysis, we plan to heterologously express and purify these enzymes to see if they retain the FMN and if the cofactor can be reduced. Elucidation of the lack of activity could further help profile and explain functional differences within the family.

Additionally, our selection of OYEs for screening consciously avoided enzymes from cluster 2 due to their larger sizes and the fact that this cluster contains the iron-sulfur containing 2-enoate reductases. Our current system does not support these particular enzymes; however, they have the potential to be powerful biocatalysts as they have the ability to reduce additional substrates such as α/β unsaturated carboxylic acids and esters (15). Interestingly, other clusters seem to conserve the same iron-sulfur cluster sequence motif, but have the added benefit of being single-domain proteins. Future work includes studying these single-domain putative iron-sulfur cluster containing OYEs in order to determine if an iron-sulfur cluster contributes to catalytic activity, which could expand the catalytic repertoire of the OYEs and give them an advantage over their two domain counterparts.

With the increase in genomic sequencing tools we now have a vast array of unexplored sequence information potentially housing numerous biocatalysts of interest. Considering there are 80,000 putative family members just within the context of the OYE superfamily, the probability of identifying novel OYEs with desirable functional traits is promising. While we have essentially doubled the number of known OYEs, we have not even begun to truly dive deeply into this superfamily. Clusters 1 and 3 consist of roughly 50% of the superfamily, and at this current time between our work and the literature they have been the most extensively sampled clusters. Interestingly, sequences within each respective cluster share a high degree of conservation (>50% ID) in their active site environments, but seemingly possess a substantial functional diversity. This would suggest that important functional contributions are made by residues and domains bordering the active site. These subtle structural differences are difficult to relate to function by traditional crystallographic or kinetic methods given the complexity of the

relationship between structure and function of proteins. Therefore, machine learning lends itself as a powerful tool for identifying correlations between the function we observed and the distal structure of OYEs. Development of this predictive framework would increase our ability to more accurately identify promising biocatalysts to meet our specific functional demands.

IV. References

1. Farinas, E.T., T. Bulter, and F.H. Arnold, *Directed enzyme evolution.* Current Opinion in Biotechnology, 2001. **12**(6): p. 545.

2. Brannigan, J.A. and A.J. Wilkinson, *Protein engineering 20 years on.* Nature Reviews Molecular Cell Biology, 2002. **3**(12): p. 964.

3. Renata, H., Z.J. Wang, and F.H. Arnold, Expanding the enzyme universe: Accessing non-natural reactions by mechanism-guided directed evolution. Angewandte Chemie International Edition, 2015. **54**(11): p. 3351.

4. Porter, J.L., R.A. Rusli, and D.L. Ollis, Directed evolution of enzymes for industrial biocatalysis. ChemBioChem, 2016. **17**(3): p. 197.

5. Nomenclature committee of the international union of biochemistry and molecular biology, in Enzyme nomenclature. 1992, Academic Press: San Diego. p. ii.

6. Prier, C.K. and F.H. Arnold, Chemomimetic biocatalysis: Exploiting the synthetic potential of cofactor-dependent enzymes to create new catalysts. Journal of the American Chemical Society, 2015. **137**(44): p. 13992.

7. Rosati, F. and G. Roelfes, *Artificial metalloenzymes.* ChemCatChem, 2010. **2**(8): p. 916.

8. Schwizer, F., et al., Artificial metalloenzymes: Reaction scope and optimization strategies. Chemical Reviews, 2018. **118**(1): p. 142.

9. Ghisla, S. and V. Massey, New flavins for old: Artificial flavins as active site probes of flavoproteins. Biochem J, 1986. **239**(1): p. 1.

10. Manstein, D.J., et al., Absolute stereochemistry of flavins in enzyme-catalyzed reactions. Biochemistry, 1986. **25**(22): p. 6807.

11. Duine, J.A., *Cofactor engineering.* Trends in Biotechnology, 1991. **9**(1): p. 343.

12. Murthy, Y.V.S.N. and V. Massey, [41] syntheses and applications of flavin analogs as active site probes for flavoproteins, in Methods in enzymology. 1997, Academic Press. p. 436.

13. Hefti, M.H., J. Vervoort, and W.J.H. van Berkel, *Deflavination and reconstitution of flavoproteins.* European Journal of Biochemistry, 2003. **270**(21): p. 4227.

14. Lutz, S. and S.M. Iamurri, *Protein engineering: Past, present, and future*, in *Protein engineering: Methods and protocols*, U.T. Bornscheuer and M. Höhne, Editors. 2018, Springer New York: New York, NY. p. 1.

15. Caldeira, J., et al., Epr and mössbauer spectroscopic studies on enoate reductase. Journal of Biological Chemistry, 1996. **271**(31): p. 18743.