AN ABSTRACT OF THE DISSERTATION OF

Kelsey Marie Kean for the degree of Doctor of Philosophy in Biochemistry and Biophysics presented on May 14, 2018.

Title: Structural Insights into Novel Sugar Phosphate Cyclases and Flavoenzymes.

Abstract approved: _____

P. Andrew Karplus

The function of a protein is defined by its three-dimensional structure, and in understanding the three-dimensional structure of a protein, we gain an understanding of its function and mechanism. Protein structures, especially at high resolution, can provide detailed insights into many elements of enzyme function and catalysis – identifying residues directly involved in binding ligand or in carrying out catalysis, illuminating factors involved in promoting catalysis, and revealing subtleties which allow functionally and/or structurally similar enzymes to carry out distinct chemistries.

This dissertation presents work aimed at the functional and structural characterization of two types of proteins: sedoheptulose 7-phosphate cyclases (SH7PCs) and flavoenzymes. Six chapters of original work are presented in this dissertation and include one review and five primary research reports. The first three chapters (Chapters 2-4) focus on SH7PCs, and the second three chapters (Chapters 5-7) focus on flavoenzymes. All but the last of these chapters are published; the final original research chapter on lactate monooxygenase is at the stage of a manuscript in preparation for submission. These central chapters are bookended by a brief introduction to SH7PCs and flavoenzymes (Chapter 1) and concluding remarks on the main highlights and impacts of this work as well as future directions (Chapter 8).

With regard to the work on SH7PCs, I report the first structures of any SH7PC (Chapters 3 and 4). The crystal structures show these enzymes are structurally similar

to each other and to other, closely related sugar phosphate cyclases, in particular dehydroquinate synthase (DHQS). These structures reveal subtle but informative differences between the three members of the SH7PC family – 2-epi-5-epi-valiolone synthase (EEVS), desmethyl-4-deoxygadusol synthase (DDGS), and 2-epi-valiolone synthase (EVS) – and suggest how these enzymes all utilize the same substrate to specifically generate one of three distinct products. The review (Chapter 2) provides an overview of the state of our understanding of SH7PCs, including the structural work reported in this thesis, and gives a broader context for the roles, evolution, and prevalence of these enzymes in nature. Based on structural insights, I propose a novel anomer selection hypothesis to differentiate two of these enzymes, EEVS and DDGS, from the third, EVS, based on which anomer of the substrate, sedoheptulose 7-phosphate, the enzyme selectively binds (Chapter 3). Furthermore, these structures allowed for corrections and additions to the "fingerprint" used to identify and differentiate SH7PCs and DHQS, which further guided genome mining (Chapter 4).

With regard to the work on flavoenzymes, enzymes utilizing either FAD or FMN to carry out their unique chemistries, I report studies that have advanced our understanding of enzyme families represented by three different proteins: glycerol 3-phosphate oxidase from *Mycoplasma pneumoniae* (*Mp*GlpO), ferredoxin-$NADP^+$ reductase from corn root (FNR), and lactate monooxygenase from *Mycobacterium smegmatis* (LMO).

In facilitating glycerol metabolism and producing hydrogen peroxide as a byproduct, glycerol 3-phosphate oxidase (GlpO), which converts glycerol 3-phosphate to dihydroxyacetone phosphate, is implicated as a pathogenicity factor and a potential drug target in pathogens such as *Mycoplasma pneuomniae*. I report the structure of *Mp*GlpO (Chapter 5). Sequence and structural comparisons of *Mp*GlpO to other structurally known glycerol 3-phosphate oxidases and dehydrogenases (GlpO/DHs) reveal there are two distinct types of GlpO/DHs: Type I GlpO/DHs, including mitochondrial GlpDHs and *Streptococcus* GlpO, and Type II GlpO/DHs, of which *Mp*GlpO is representative. Guided by a liganded structure of a close homolog for which the structure was solved as part of the Protein Structure Initiative, I proposed the first plausible binding mode of the glycerol 3-phosphate substrate and a

detailed catalytic mechanism that arguably will apply to all GlpO/DHs, despite the distinct differences in the composition of their active sites.

The FAD-dependent enzyme FNR catalyzes the transfer of electrons from photoreduced ferredoxin to $NADP^+$ during photosynthesis and serves as a model for a broad superfamily of enzymes including NO synthase, cytochrome P450 reductase, and NADPH oxidases. Using a variant replacing a conserved aromatic amino acid to capture the productive binding mode of the nicotinamide portion of NADP(H), we analyzed a suite of high resolution structures (~1.5 Å) in complex with nicotinamide, $NADP^+$, and NADPH. A reinterpretation of previous kinetic data also supports the relevance of these complexes to catalysis. Based on these high resolution structures, we further report insights into factors promoting hydride transfer in FNR and other FNR-like superfamily members. Specifically, we infer that higher anisotropic mobility of the C4 atom of $NADP^+$ compared to NADPH, distortion of FAD geometry from planarity, and a tightly packed active site implicate significant active site compression as a factor promoting hydride transfer. A broadly relevant conclusion of this work is the recognition of active site compression as an important and general – although often overlooked and underappreciated – factor that can promote catalysis.

LMO is an FMN-dependent enzyme that catalyzes the conversion of lactate to acetate, carbon dioxide, and water. LMO is part of a family of α-hydroxy acid oxidases, all of which carry out the same oxidation chemistry, but unlike LMO, proceed through an uncoupled pathway. I report the first structure of an LMO (Chapter 7). This structure reveals details of the LMO active site and provides new insights as to how LMO kinetically and functionally deviates from the other family members by proceeding along a coupled reaction pathway. A highly mobile, variable loop ("loop 4") known to seal the active site in these α-hydroxy acid oxidases is significantly larger in LMO than in the other enzymes in the family, and has both more compact folding and greater buried surface area. We suggest it is the dynamics of this loop that governs the kinetics of intermediate release (or lack thereof) in these enzymes.

Structural Insights into Novel Sugar Phosphate Cyclases and Flavoenzymes

by
Kelsey Marie Kean

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented May 14, 2018
Commencement June 2018

Doctor of Philosophy dissertation of Kelsey Marie Kean presented on May 14, 2018

APPROVED:

Major Professor, representing Biochemistry and Biophysics

Head of the Department of Biochemistry and Biophysics

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Kelsey Marie Kean, Author

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere appreciation to my advisor, Dr. Andy Karplus. I feel like I could not have received better and more well-rounded training as a graduate student – in methodology, scientific communication, writing, experimental design and analysis, and managing challenging professional interactions – and I am especially grateful for the opportunities you allowed me to pursue outside of the laboratory. Teaching, outreach, and leadership have been a huge part of my development over the last five years, and I so appreciate your unwavering support in all of these endeavors. Also, thank you for allowing to pursue many different projects, although sometimes covertly and often without funding, guided by my own interests.

I have to thank past members of the Karplus Lab: Camden Driggers, Arden Perkins, and Andrew Brereton. I thank Camden for teaching me so much of what I know as my mentor in the lab as a rotation student and an early graduate student, for continuing to answer my occasional questions via email, and for introducing me to the Tarntip lunch special. Thank you to Arden Perkins and Andrew Brereton for their friendship and for making the Karplus Lab an enjoyable place to work, and especially to Andrew for helping me clean up the flooded lab. I also need to thank Dale Tronrud, who is mostly accurately described as the "Sage of the Willamette Valley" and contains a wealth of information both science-related and not and who has been incredibly generous with his time and knowledge even after technically leaving the Karplus Lab, especially in keeping our lab computers running and allowing my work to continue smoothly. Thank you for always inviting me on your adventures and getting me to see parts of Oregon I would not have made it to otherwise- in particular Belknap Crater, one of the more unique and memorable hikes I have been on. Thank you to the undergraduates I worked with, trained, and gotten to known over the years in the lab, especially Taylor Roemelt (now a dentist) and Jenna Beyer.

I also thank Ryan Mehl and Kari van Zee – two of my greatest advocates, cheerleaders, and mentors at OSU who have provided excellent examples of

mentorship and teaching to me. Thank you for everything and especially for letting me take on the BB494 Teaching Lab as one of my projects.

I would like to thank the Mehl Lab, current and former, who have taken me in as one of their own, especially as the Karplus Lab has dwindled to one, and have always been willing to answer my questions (especially Linda Benson, Joey Porter, and Robby Blizzard). I would also like to thank all of my collaborators, both near and far, especially Taifo Mahmud, Andrew Osborn, and Alessandro Aliverti. Thank you to Karplus Lab members before me, who often didn't quite finish things up and left interesting projects to me: Calia Elkhal, Russell Carpenter, and Andrea Hall (and especially Andrea for her thorough and clear documentation which has made my life much easier). I also thank Viviana Perez and Shelley Su, my committee members not already mentioned, for their guidance over the last five years

At the heart of my work is crystallography so much of my work was made possible by regular, free access to the Advanced Light Source in Berkeley, California. I would like to thank Stacey Ortega and Marc Allaire at ALS for always being incredibly helpful and knowledgeable and for even coming in on weekends to help me with unique data collection strategies.

I would like to thank the great friends I have made during my time at OSU. In particular, I have to thank my dear friends Chelsea and Ben Wolk, who have been terrific friends and have often treated me like family (even letting me live with them), as well as Steve and Annie Friedman, Joey Porter, Robby Blizzard, Allison Gicking, and Heather Forsythe for many great hikes, family dinners, floats, large personal pizzas, and trips to the Beanery. I would also like to acknowledge all of the great Oregon dogs I have gotten to know: Piper, Jeff Dominguez (J.D.), Bear, Arya, the late Bishop Henry, and especially Rudy.

I would like to acknowledge OSU Distinguished Provost's Fellowship, P.F. Yerex and Nellie Buck Yerex Fellowship, and NIH R01-GM119227 for funding. I also thank the Department of Biochemistry and Biophysics for all of the amazing support and opportunities: I am so glad I decided to move across the country to Oregon.

Last but not least, I thank my family. Mom and Dad, thank you for always supporting me in every endeavor, including moving halfway across the country for college and all the way across the country for graduate school. Thank you for encouraging me, helping me with Science Olympiad as a middle schooler, and giving me Nalgene bottles to play with as a little kid.

CONTRIBUTION OF AUTHORS

In Chapter 2, Andrew Osborn, Taifo Mahmud, and P. Andrew Karplus wrote the paper.

In Chapter 3, Sara J. Codding planned and performed experiments; Shumpei Asamizu expressed protein; Taifo Mahmud and P. Andrew Karplus planned experiments, analyzed data, and wrote the paper.

In Chapter 4, Andrew Osborn planned and performed experiments, analyzed data, and wrote the paper. Khaled M. Alseud, Khaled M. Almabruk, Shumpei Asamizu, and Janet A. Lee planned and performed experiments and analyzed data. Taifo Mahmud and P. Andrew Karplus planned experiments, analyzed data, and wrote the paper.

In Chapter 5, Callia K. Elkhal planned and performed experiments and analyzed data; Derek Parsonage planned and performed experiments, contributed reagents, and analyzed data; Somchart Maenpuen and Pimchai Chaiyen planned and performed experiments and analyzed data; Al Claiborne and P. Andrew Karplus planned experiments, analyzed data, and wrote the paper.

In Chapter 6, Russell A. Carpenter planned and performed experiments and analyzed data; Vittorio Pandini, Andrea R. Hall, and Rick Faber planned and performed experiments; Giuliana Zanetti planned experiments; Alessandro Aliverti planned and performed experiments, analyzed data, and wrote the paper; P. Andrew Karplus planned experiments, analyzed data, and wrote the paper.

In Chapter 7, Zhan Deng planned and performed experiments; P. Andrew Karplus planned experiments, analyzed data, and wrote the paper.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

Figure Page

LIST OF FIGURES (Continued)

Figure Page

LIST OF FIGURES (Continued)

Figure Page

LIST OF TABLES

LIST OF ABBREVIATIONS

aDHQS	aminodehydroquinate synthase
AHBA	3-amino-5-hydroxybenzoic acid
*Am*EVS	*Actinosynnema mirum* EVS
AML	artificial mother liquor
*An*DHQS	*Aspergillus nidulans* DHQS
Ava	*Anabaena variabilis* DDGS
*Av*DDGS	*Anabaena variabilis* DDGS
*Bc*DOIS	*Bacillus circulans* DOIS
*Bp*3253	protein encoded by *Bordetella pertussis* gene 3253
*Bs*GlyDH	Bacillus stearothermophilus glycerol dehydrogenase
CBP	carbaphosphonate
CDO	cysteine dioxygenase
CG6P	carbaglucose 6-phosphate
cGlpDH	cytosolic glycerol 3-phosphate dehydrogenase
CTC-1	charge-transfer complex 1
CTC-2	charge-transfer complex 2
DAAO	D-amino acid oxidase
DAHP	3-deoxy-D-arabinoheptulosonate 7-phosphate
DDG	desmethyl-4-deoxygadusol
DDGS	desmethyl-4-deoxygadusol synthase
DHAP	dihydroxyacetone phosphate
DHQ	3-dehydroquinate
DHQS	dehydroquinate synthase
DOI	2-deoxy-scyllo-inosose
DOIS	2-deoxy-scyllo-inosose synthase
*Ecass*GlpO	*E. casseliflavus* GlpO
*Ec*GlpD	*E. coli* GlpDH
EEV	2-epi-5-epi-valiolone
EEVS	2-epi-5-epi-valiolone synthase
EV	2-epi-valiolone
EVS	2-epi-valiolone synthase
FCB2	flavocytochrome b2
Fd	ferredoxin
FNR	ferredoxin:NADP+ reductase
G6P	glucose 6-phosphate
GCE	genetic code expansion
Glp	glycerol 3-phosphate; α-glycerophosphate
GlpDH	glycerol 3-phosphate dehydrogenase; α-glycerophosphate dehydrogenase
GlpF	glycerol facilitator
GlpK	glycerol kinase
GlpO	glycerol 3-phosphate oxidase; α-glycerophosphate oxidase
GOX	glycolate oxidase
HCAII	human carbonic anhydrase II

LIST OF ABBREVIATIONS (Continued)

HCAIIts	thermostable variant of human carbonic anhydrase II
HGT	horizontal gene transfer
LMO	lactate monooxygenase
LOX	lactate oxidase
MAA	mycosporine-like amino acid
MDH	mandelate dehydrogenase
mGlpDH	mitochondrial glycerol 3-phosphate dehydrogenase
*Mp*GlpO	*Mycoplasma pneumoniae* GlpO
MT-Ox	methyltransferase/oxidoreductase
NOX	NADPH oxidase
O-MT	O-methyltransferase
OYE	Old Yellow Enzyme
SH7P	sedoheptulose 7-phosphate
SH7PC	sedoheptulose 7-phosphate cyclase
*Sh*EEVS	*Streptomyces hygroscopicus* subsp. *jinggangensis* 5008 EEVS
SPC	sugar phosphate cyclase; C7-sugar phosphate cyclase
*Sp*GlpO	*Streptococcus pneumoniae* GlpO
*Ssp*GlpO	*Streptococcus* sp. GlpO
TLC	thin layer chromatography
ValA	*Streptomyces hygroscopicus* subsp. *jinggangensis* 5008 EEVS
WT	wild-type

Structural Insights into Novel Sugar Phosphate Cyclases and Flavoenzymes

Chapter 1

Thesis Overview

Protein Structure-Function Studies

Proteins are among the most abundant and functionally diverse biomolecules. Among their many functions, proteins provide structure (e.g. collagen), transport small molecules (e.g. hemoglobin) and larger cargo (e.g. dynein), protect against infection (e.g. lysozyme), and provide movement (e.g. myosin). Among these broad functions of proteins, perhaps the most amazing are those that act as enzymes and catalyze specific biochemical reactions (e.g. sedoheptulose 7-phosphate cyclases and flavoenzymes). Enzymes have evolved over time to carry out a remarkable range of chemistries including transferring a chemical group (e.g. protein kinases), joining chemical compounds (e.g. ubiquitin ligase), and breaking chemical bonds (e.g. proteases). Enzymes continue to evolve in nature as well as in the laboratory where research is focused on designing new proteins or modulating existing proteins to carry out different functions (as recently reviewed [1]). The ability to design, modulate, exploit, and really understand these functions and catalytic abilities lies largely within the structure of these proteins which are uniquely and exquisitely set up to carry out their chemistries. Therefore, the goal of structural biology is to utilize structural information to understand how proteins work, both when working correctly and when behaving aberrantly.

The overall goal of the work reported in this dissertation is to combine structural and functional studies to advance our understanding of four enzyme families: sedoheptulose 7-phosphate cyclases (SH7PCs), glycerol 3-phosphate oxidase, ferredoxin:$NADP^+$ reductase, and lactate monooxygenase. Each of these studies are illustrative of the power of utilizing structure to understand enzyme function and mechanism. In the remainder of this introductory chapter, I will provide a brief introduction to the general aspects and history of SH7PCs and the three flavoenzyme families. Then, I will outline the contents of the remaining chapters of this dissertation.

Novel Sugar Phosphate Cyclases and Natural Products

Natural products, commonly produced as part of secondary metabolism, represent molecules with diverse molecular architecture that are useful to both their

native sources and also often to humanity. Natural products have been used medicinally since ancient times, with the earliest records of natural products to treat human ailments dating as far back as 2600 B.C. when oils from cypress, myrrh, and licorice were used to treat coughs, colds, and inflammation in Mesopotamia.[2] A number of very well-known and still widely used drugs originally came from natural sources: for example aspirin, derived from the natural product salicin from the bark of the willow tree;[3] morphine and codeine, originally from the opium poppy;[4] the anti-malarial quinine, isolated from the bark of *Cinchona* trees;[5] and penicillin, produced by the fungus *Penicillium notatum*.[6] Today, natural products continue to be an important source of biodiverse molecules. Between 1981 and 2010, 64% of small molecule approved drugs were somehow related to or derived from natural products.[7] Especially in a time of increasing drug resistance, natural products continue to serve as a source of both material and inspiration for small-molecule drug development.

Understanding the enzymes and pathways that generate these diverse biomolecules allows us to exploit existing, natural pathways to generate new molecules and natural product analogs. This is generally done in two ways: by re-engineering individual proteins for modified functionality and/or re-designing modular pathways involving multiple enzymes to generate new products (reviewed in [8]). Both of these methodologies are largely guided by our understanding of structure and how individual proteins and modular pathways or complexes work. Secondarily, understanding protein structure and function can guide our understanding of the targets of these small molecules, especially drug targets (typically proteins).

The family of enzymes involved in natural product production that are one focus of this dissertation are the sedoheptulose 7-phosphate cyclases (SH7PCs). This family contains three types of enzymes: 2-epi-5-epi-valiolone synthase (EEVS), desmethyl 4-deoxygadusol synthase (DDGS), and 2-epi-valiolone synthase (EVS). Each of these enzymes utilizes sedoheptulose 7-phosphate (SH7P) as substrate to generate its unique product. The first SH7PC was identified in 1999 by Dr. Taifo Mahmud when feeding experiments showed that the valienamine moiety of the natural product α-glucosidase inhibitor, acarbose, derived from 2-epi-5-epi-valiolone (EEV).[9] Subsequent experiments identified the first member of this new class of

enzymes in the SH7PC family: an EEVS that utilizes sedoheptulose 7-phosphate to generate EEV.[10] Subsequently, two additional SH7PCs were identified as part of this family: DDGS[11] and EVS,[12] generating desmethyl 4-deoxygadusol (DDG) and 2-epi-valiolone (EV), respectively. In general, the SH7PCs generate a core C7-cyclitol found in different natural products with diverse properties including UV protection (mycosporine-like amino acids[13] and gadusol[14]), cytotoxicity (cetoniacytone A[11]), and antibacterial and fungicidal (validamycin[15]) (Figure 1.1A).

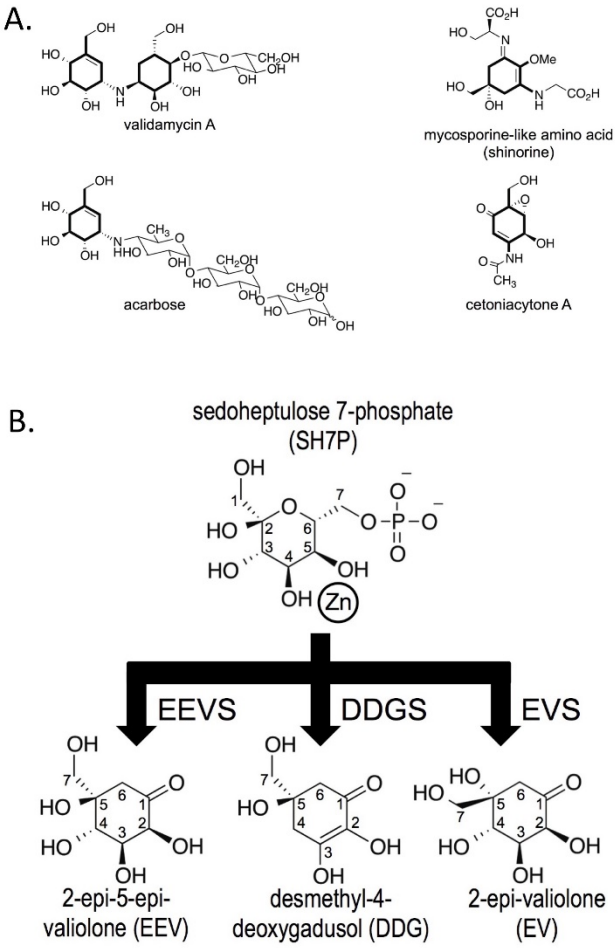


Figure 1.1. Sedoheptulose 7-phosphate cyclases and their products.
A. Four cyclitol-containing natural products are shown and labeled by name, with their C7-cyclitol units made by SH7PCs highlighted in bold. **B.** The substrate and products of sedopheptulose 7-phosphate cyclases are shown. An enzyme-bound divalent metal cation (Zn^{2+}) is drawn next to the two substrate hydroxyls proposed to coordinate it. Abbreviations for the substrates and products are shown, with each enzyme abbreviation being that of its product followed by an additional S for "synthase."

In the main body of this thesis, I have included three chapters regarding work on SH7PCs. In lieu of including a more extensive introduction to SH7PCs here, I lead the body of my thesis off with a retrospective review published in 2017[16] on sugar phosphate cyclases (Chapter 2). This review provides a broad and extensive background SH7PCs and also summarizes more recent work on SH7PCs including the primary publications I was involved in that are presented in the following two chapters.

In a collaborative project with Dr. Taifo Mahmud's research group, we targeted this enzyme family as an interesting and compelling target for structural studies. The primary motivation for undertaking structural studies of SH7PCs was to identify how each of these enzymes utilize the same substrate to generate one of three distinct products (Figure 1.1B), despite highly similar sequences and active site compositions. The first structure of an SH7PC – that of an EEVS from *Streptomyces hygroscopicus* (ValA) at 2.1 Å resolution – was reported in 2014 (Chapter 3).[17] This structure revealed that EEVS is similar to the well-studied related enzyme, dehydroquinate synthase (DHQS), in overall fold and active site composition. Comparisons with DHQS revealed an unexpected active site difference that led us to propose a novel anomer selection hypothesis for how two of the SH7PCs, EEVS and DDGS, are differentiated from the third, EVS, based on the anomer of the substrate each selectively binds. While this paper was in press, I solved the higher resolution structure (1.7 Å) of DDGS from *Anabaena variabilis* (Ava) which allowed me to revise the interpretation of an otherwise difficult to model β-turn in ValA. This revealed a unique feature in the topology of EEVS and DDGS compared with DHQS. A very short correction was published describing this altered topology and is included here as an additional section in Chapter 3.

Beyond providing higher-resolution information, obtaining the structure of a representative DDGS gave us additional insights into these enzymes. The representative structures of EEVS and DDGS guided corrections and additions to the "fingerprint" used to differentiate EEVSs from DDGSs, inferences about substrate binding in EEVS and DDGS, and the design of point mutation studies to demonstrate the importance of characteristic differences between EEVS and DDGS.[18] The insights

from these structures were coupled with bioinformatics studies to more accurately classify and differentiate SH7PCs identified through genome mining and assess their distribution. In this thesis, the text of the chapter (Chapter 4) has been modified from that of the original publication to bring into the text some of the content that was originally published as Supporting Information.

While not a focus of this dissertation, the direction of continuing studies of SH7PCs has shifted away from structure to physiology and is oriented toward a set of unexpected sources of EEVSs and their characterization. Through genome mining, Dr. Taifo Mahmud's research group surprisingly identified EEVS genes in non-mammalian vertebrates (e.g. fish, birds, reptiles, and amphibians). Using both *in vitro* studies using recombinant protein and *in vivo* studies using zebrafish embryos they found that in these vertebrates, EEVS works with a dual-function methyltransferase-oxidoreductase to generate the natural product gadusol.[14] While it is still not known how gadusol functions *in vivo*, one hypothesis is that gadusol functions as a UV protectant. Initially, an important part of this project was verifying in multiple ways that these EEVSs found in vertebrates (traditionally thought to be limited to bacteria) are truly EEVSs. My contribution was to analyze sequence and residue conservation patterns among sugar phosphate cyclase superfamily members, including new vertebrate and algal EEVSs. The Mahmud research group is continuing work in this exciting direction to better understand EEVSs and the role of gadusol in the non-mammalian vertebrates as well as targeting potential pharmaceutical and other commercial applications for gadusol.

Fantastic Flavoenzymes

The first flavin-dependent enzyme was isolated from Brewer's Yeast in 1932 in an attempt to understand the nature of biological oxidations.[19] Originally referred to as "Gelbe Ferment" (yellow ferment), we now know this enzyme by the name "Old Yellow Enzyme" (OYE). This yellow enzyme led to a hallmark discovery in our understanding of enzymes and how they work, and also a Nobel Prize for Hugo Theorell in 1955. Theorell was able to reversibly divide OYE into two distinct components: a yellow coenzyme part (FMN) and a colorless protein part.[20]

Individually these components were inactive, but together they could catalyze the oxygen-dependent oxidation of NADPH to $NADP^+$. As flavins are derived from vitamin B_2 or riboflavin, this identified the first biochemical role for a vitamin as part of a coenzyme required for enzyme catalysis.

Following the discovery of OYE, a "new" yellow enzyme, D-amino acid oxidase, an important representative member of the superfamily discussed in Chapter 5, was discovered.[21-22] Since then, "yellow enzymes" have emerged as a remarkable example of diversity in both function and structure. These enzymes come in many different folds and topologies and all utilize one of two different flavin cofactors (more commonly FAD than FMN) to carry out a wealth of different chemistries, including dehydrogenation, one- and two-electron transfers, light emission, and activation of oxygen for oxidation and hydroxylation reactions.[23-25] Flavoenzymes prove to be important for diverse and interesting physiological processes. For example, the human genome contains 90 genes encoding flavin-dependent enzymes, about 60% of which have been implicated in human disease.[26] As another example, the flavoenzyme cryptochrome regulates circadian rhythm in plants and animals and has been implicated in magnetosensing in birds.[27-28]

Since the first structure of a flavoenzyme was solved in 1978,[29] one of the things that has most developed and guided our understanding of these remarkable enzymes is protein X-ray crystallography – providing atomic level information about both the active sites and the rest of the surrounding protein, which in the case of flavoproteins is especially important for tuning the chemical properties of the flavin present.

Flavoenzymes have been a long-standing area of interest in the research group of Dr. Andy Karplus. Luckily for members of the Karplus group, these amazing enzymes are also quite nice to work with; unlike many systems, flavoenzymes have an automatic "tell" if your crystals are protein or salt as flavoenzymes (and their resulting crystals) are typically yellow. In this dissertation, three different flavoenzymes are studied: glycerol 3-phosphate oxidase (GlpO), a FAD-dependent enzyme which catalyzes the oxidation of glycerol 3-phosphate to dihydroxyacetone phosphate (Chapter 5); ferredoxin:$NADP^+$ reductase (FNR), a FAD-dependent

enzyme which catalyzes the transfer of electrons from photoreduced ferredoxin to $NADP^+$ in photosynthesis (Chapter 6); and lactate monooxygenase (LMO), a FMN-dependent enzyme which catalyzes the conversion of lactate to acetate, carbon dioxide, and water (Chapter 7). As will be illustrated in the body of this dissertation, in all of these cases, understanding the structure of these enzymes allows us to gain valuable insights into their function, catalytic mechanism, and regulation. Flavoenzymes prove to be interesting not only because of their chemistry, but also, to me, because of their histories, which I will detail briefly as part of this introduction.

Glycerol 3-Phosphate Oxidase

There are two different types of flavin-dependent enzymes involved in catalyzing the oxidation of glycerol 3-phosphate to dihydroxyacetone phosphate (DHAP): glycerol 3-phosphate oxidase (GlpO) and glycerol 3-phosphate dehydrogenase (GlpDH). In terms of their chemistry, these enzymes are very similar in that they are FAD-dependent and catalyze a hydride transfer to convert a hydroxyl to a ketone. They differ in their final electron acceptors: GlpO uses oxygen as an electron acceptor, generating hydrogen peroxide, and GlpDH uses ubiquinol as an electron acceptor, generating ubiquinone.

Glycerol 3-phosphate oxidation activity was first identified in muscle and liver tissue in 1919 by Meyerhof.[30] The enzyme responsible for this activity was later identified as a glycerol 3-phosphate dehydrogenase (or α-glycerophosphate dehydrogenase) and systematically characterized as a FAD-dependent, mitochondrial flavoenzyme that utilizes glycerol 3-phosphate as substrate to generate DHAP and does not react with oxygen.[31-32]

In 1945, scientists made the observation that lactic acid bacteria could utilize glycerol under aerobic conditions, generating hydrogen peroxide in the process, and suggested this involved an enzyme that differed from the GlpDH already identified in mammals.[33] In 1969, this enzyme (isolated from *Streptococcus*) was identified as a glycerol 3-phopshate oxidase (GlpO; or α-glycerophosphate oxidase) and characterized as a FAD-dependent enzyme that uses oxygen and glycerol 3-phosphate to generate stoichiometric amounts of H_2O_2 and DHAP.[34]

In mammals, GlpDH is found in mitochondria (mGlpDH) where it works with cytosolic glycerol 3-phosphate dehydrogenase (cGlpDH; confusingly, a different, unrelated enzyme) as part of a glycerol 3-phosphate shuttle that links glycolysis, fatty acid metabolism, and oxidative phosphorylation (Figure 1.2A).[35] Together, mGlpDH and cGlpDH re-oxidize cytosolic NADH to sustain ATP production without making lactate, an important function in some conditions and tissues as indicated by increased expression levels of both mGlpDH and cGlpDH. As an important element of metabolism, it is not surprising that dysregulation of mGlpDH has been implicated in disease states, including cancer.[36] However, as a membrane-bound enzyme, mGlpDH has been difficult to study, especially in terms of its structure.

GlpDHs are found across eukaryotes with a high degree of sequence conservation; for example, human mGlpDH has about 30% sequence identity with *E. coli* GlpDH. GlpOs are apparently limited to bacteria, but also share significant sequence identity with GlpDHs; for example, human mGlpDH also has about 30% sequence identity with *Streptococcus sp.* GlpO. Therefore, bacterial GlpO/DHs can and have been used as a proxy for learning about mGlpDHs, especially in terms of structure.[37-38]

Beyond contributing to our understanding of mGlpDH, GlpOs are important in their own right for their role in human disease. GlpOs and the hydrogen peroxide they produce have been implicated as pathogenicity factors in organisms like *Streptococcus pneumoniae*,[39-40] responsible for pneumococcal meningitis, and *Mycoplasma pneumoniae*,[41-44] responsible for about half of community-acquired pneumonias. This makes GlpO an enticing drug target. In these organisms, GlpO works with glycerol facilitator (GlpF) and glycerol kinase (GlpK) as part of glycerol metabolism and produces hydrogen peroxide as a byproduct (Figure 1.2B). In Chapter 5, I present the structure of GlpO from *Mycoplasma pneumoniae* (*Mp*GlpO), a potential drug target. In this study, structural comparisons to closely related, structurally known proteins proved to be especially informative – leading us to the insight that this flavin-dependent family of GlpOs and GLpDHs includes two subtypes. Also, using a liganded structure of a close structurally known homolog produced as part of the Protein Structure Initiative as a template, we proposed the first

plausible binding mode of glycerol 3-phosphate and a detailed catalytic mechanism for all GlpO/DHs.

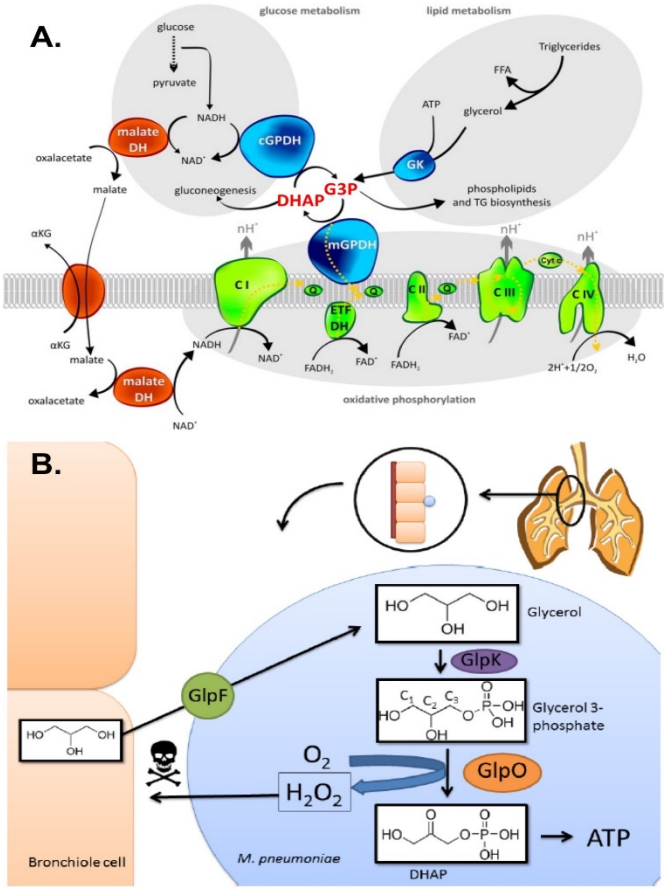


Figure 1.2. Physiological roles of glycerol 3-phosphate oxidases and dehydrogenases in mitochondria and pathogens.

A. Mitochondrial glycerol 3-phosphate dehydrogenase (mGlpDH) and its role as a hub for metabolism. Enzymes involved in the glycerol 3-phosphate shuttle (blue), the electron transport chain (lime), and the malate/aspartate shuttle (orange) are shown. DHAP: dihydroxyacetone phosphate, G3P: glycerol 3-phosphate, CI: complex I, CII: complex II, CIII: complex III, CIV: complex IV, ETF DH: electron-transferring-flavoprotein dehydrogenase, Q: Coenzyme Q, FFA: free fatty acids, cGlpDH: cytosolic glycerol 3-phosphate dehydrogenase, GK: glycerol kinase. Adapted from [35]. **B.** Glycerol 3-phosphate oxidase (GlpO) and its role in pathogenicity of *M. pneumoniae*. *M. pneumoniae* adheres to lung epithelial tissue where it uptakes and metabolizes glycerol from host phospholipids using glycerol facilitator (GlpF), glycerol kinase (GlpK), and GlpO. GlpO utilizes glycerol 3-phosphate to generate dihydroxyacetone phosphate (DHAP) and hydrogen peroxide. DHAP is utilized in glycolysis to generate ATP while hydrogen peroxide is shuttled out of the bacterial cell, where it is destructive to the host tissue. Adapted from [45].

Ferredoxin:$NADP^+$ Reductase

FNR is well-known for its involvement in photosynthesis. FNR exists in both photosynthetic tissues like leaves, where it generates NADPH, and non-photosynthetic tissues like roots, where it takes electrons from NADPH to reduce ferredoxin. While FNR is interesting and important in its own right because of its involvement and necessity in photosynthesis and in fueling a bulk of Earth's organisms, it also serves as a model for an important, more complex and biomedically relevant family of enzymes. These include enzymes like cytochrome P450 reductase (involved in drug metabolism), NADPH oxidase (responsible for generating superoxide involved in immune response), and nitric oxide synthase (catalyzing the production of nitric oxide, an important signaling molecule). While all of these enzymes differ in their overall structure and function, they have one thing in common: they all utilize a FNR-like module to carry out their chemistry involving a remarkable interconversion between single electrons and hydrides.

The interesting origin story of FNR begins in 1951, when over the course of about one month, three separate reports were published in Nature claiming that cell-free photosynthesis resulted in reduced pyridine nucleotides.[46-48] This led to an explosion of inquiry trying to identify how this chemistry was occurring and what enzymes were involved and eventually to the identification of ferredoxin (beautifully recounted by Daniel Arnon, one of the scientists involved [49]) and its partner enzyme, which we now know as FNR. FNR itself was identified four different times by multiple different groups and given four different names: TPNH diaphorase,[50] pyridine nucleotide transhydrogenase,[51] NADPH-cytochrome f reductase,[52] and ferredoxin-TPN reductase.[53] All of these enzymes had some features in common – they were yellow, FAD-dependent, and specific to NADP(H) (or TPN(H) as it was known at the time). It was not until 1966 that is was finally realized that these were all really the same ferredoxin-$NADP^+$ reductase involved in generating NADPH as part of photosynthesis.[54]

Since the middle of the 20th century, FNR has been a subject of scientific inquiry with the first structure solved in 1991 by the Karplus research group.[55] More of the recent history, particularly with regards to structure, is detailed in Chapter 6,

but today FNR still serves as an important model system, especially for guiding our understanding of and experimental design to study these more complex FNR-like superfamily members.

While the Karplus research group and others have contributed significantly to our understanding, especially in terms of structure and mechanism, of FNR and FNR-like superfamily members,[56-66] we wanted to unveil more detailed information about factors promoting hydride transfer in FNR and more broadly the FNR-like superfamily. This work included a set of 11 different high (~1.5 Å) resolution structures and spanned about 10 years and five different graduate students and post-docs in the Karplus Lab.[67] This study (Chapter 6) combines structural and functional characterizations of FNR variants that allow us to capture and describe the productive binding mode of the nicotinamide of NADP(H), something that is otherwise impossible with wild-type members of the FNR-like superfamily, as well as reanalysis of kinetic work calling into question the validity of these variants for studying catalysis. These high resolution structures come with additional information with regards to anisotropic movement of individual atoms as well as deviations from standard geometry and provide evidence for active site compression as a factor contributing to catalysis in FNR and FNR-like superfamily members.

Lactate monooxygenase

Lactate monooxygenase (LMO), originally known as lactate oxidase, was first identified by Edson in 1947.[68] This enzyme was also identified more than once, with lactate oxidase being "discovered" a second time by Yamamura et al. in 1952.[69] Guided by the color of his crude extract, Edson guessed this enzyme responsible for consuming lactate and molecular oxygen to generate stoichiometric amounts of acetate and carbon dioxide was a flavoenzyme and utilized pyruvate as an intermediate, and indeed he proved to be correct. However, it was a number of years before everyone agreed with him. Work from other groups claimed this lactate oxidizing enzyme was not a flavoprotein,[70] that neither FAD nor any other riboflavin-containing coenzyme could be detected in the pure enzyme, and that pyruvate could not be an intermediate.[71] Now, it is understood that LMO as isolated from

mycobacteria is indeed an FMN-containing flavoenzyme which forms an octamer and utilizes pyruvate as an intermediate.

Since those early days, our foundational understanding of LMO as well as many other flavoenzymes has increased primarily due to work from one lab: that of Vincent Massey at the University of Michigan.[72-77] We now know LMO is a member of a broader family of α-hydroxy acid oxidases, including lactate oxidase, glycolate oxidase, mandelate dehydrogenase, and flavocytochrome b_2. While these enzymes are catalytically and structurally very similar, LMO deviates from the other members of this family kinetically and functionally – utilizing a coupled pathway in which lactate is oxidized to pyruvate and pyruvate undergoes oxidative decarboxylation to acetate and water while all other family members utilize an uncoupled pathway whereby the oxidized intermediate (equivalent to pyruvate) is released as product. However, attempts by the Massey group and others to unravel the mystery behind the distinct activities of LMO and other α-hydroxy acid oxidases (particularly through generating simple point mutations to interconvert activities with lactate oxidase[72, 77]) proved unsuccessful. Thus, the question of how LMO catalyzes a coupled reaction while other α-hydroxy acid oxidases catalyze an uncoupled reaction remained. With Vincent Massey's passing in 2002 and the availability of a true lactate oxidase that is better suited for biotechnological applications, lactate monooxygenase seemed to fall out of favor and our direct study and understanding of LMO has not advanced over about the last twenty years.

However, it was long thought, especially by Vincent Massey, that the structure of LMO would help unravel this longstanding mystery. In the mid-1990s, Vincent Massey reached out to the Karplus group to undertake structural studies of LMO. While beautiful, yellow crystals were readily grown and diffraction data were collected, the structure proved to be a difficult one to solve and eluded multiple lab members over the years. These purified proteins and data sets were left to languish in the freezer and in archival records until 2016. In Chapter 7, I present the structure of LMO from *Mycobacterium smegmatis*. This is the first structure of any LMO and the chapter is a still developing draft of a manuscript with analyses still to be completed. Gratifyingly, through solving the structure of LMO and comparing it with other

family members, we have developed an unexpected, but compelling hypothesis for how structural differences distal to the active site are responsible for the kinetic differences among these enzymes.

Graduate work not included in this dissertation

Before closing, I would like to acknowledge three additional projects, including a pedagogical project and two structure-function research projects, to which I have contributed during my graduate career. These projects are reported in two peer reviewed publications for which I am lead author (one published and one in revision) [78] and one on which I am a co-author.[79] I outline these studies briefly below.

Kean, K.M.*, van Zee, K., and Mehl, R.A. (2018) "Unnatural Chemical Biology: Research-Based Laboratory Course Utilizing Genetic Code Expansion" *Journal of Chemical Education*, under revision.

Course-based undergraduate research experiences are becoming an essential part of undergraduate STEM education, and while the prevalence of these courses is growing, they rarely include chemical biology tools to manipulate biomolecules and robustly engineer proteins. This paper describes a course-based undergraduate research experience utilizing the chemical biology tool, genetic code expansion (GCE), in which students characterize proteins containing non-canonical amino acids of their own choosing based on a hypothesis of their own design (offered as BB494/594: Biochemistry Laboratory Molecular Techniques 2 at OSU). This paper presents an intellectual overview of the curriculum and provides example student results and suggestions for implementation and adaptation at other institutions. In hopes of exposing more students to the power of GCE, we present this as a research-based laboratory course that can readily and easily be adapted, scaled, and implemented as at institutions ranging from primary undergraduate institutions to research universities.

For this project, I revised and updated course materials and wrote the manuscript. I also wrote a proposal with my co-authors and obtained funding through

OSU's Open Oregon State initiative to adapt and publish the laboratory manual for this course as an open source textbook.

Kean, K.M.*, Porter, J.J., Mehl, R.A., and Karplus, P.A. (2018) "Structural insights into a thermostable variant of human carbonic anhydrase II" *Protein Science*, 27, 573-577.

This project is near and dear to my heart as it is based upon my first rotation project as a graduate student at OSU, rotating in Dr. Ryan Mehl's research group in the Fall of 2012. Carbonic anhydrase catalyzes the conversion of carbon dioxide and water to bicarbonate and is of interest for many biotechnological applications, in particular carbon sequestration. Carbon sequestration often requires harsh conditions and involves attaching proteins to a surface, for which technology is limited and results in compromised enzymatic activity. Towards this end, the Mehl research group was interested in using genetic code expansion for site-specific surface attachment of a thermostable variant of carbonic anhydrase.

Human carbonic anhydrase II (HCAII) is among the best studied and most active carbonic anhydrases, and during my rotation, I identified a thermostable variant of HCAII in the literature (which we abbreviate HCAIIts) with six different point mutations as being ideally suited for our protein engineering work. Together the six point mutations impart a remarkable increase in thermostability – this sextuple mutant remains >85% active after 2 hours at 65°C while wild-type HCAII is <15% active after 2 hours at 55°C. How do these six mutations impart such a remarkable increase in thermostability? The ultra-high resolution crystal structure (0.9 Å) of HCAIIts revealed that each of the six mutations involved led only to local structural changes consistent with them acting largely independently. Furthermore, the structures showed that each mutation variously resulted in increased hydrogen bonding, improved packing, reduced side chain entropy loss upon folding, and improvements in backbone conformation energetics to yield this dramatic increase in stability.

Driggers, C.M., **Kean, K.M.**, Hirschberger, L.L., Cooley, R.B., Stipanuk, M.H., and Karplus, P.A. (2016) "Structure-Based Insights into the Role of the Cys-Tyr Crosslink and Inhibitor Recognition by Mammalian Cysteine Dioxygenase" *Journal of Molecular Biology*, 20, 3999-4012.

Cysteine Dioxygenase (CDO) helps regulate cysteine levels in mammals by converting cysteine to cysteine sulfinic acid. In order to gain insight into how a Cys-Tyr crosslink promotes activity and how select competitive inhibitors bind, Dr. Andy Karplus' research group undertook a thorough study (ultimately including 20 high-resolution structures of mammalian CDO) involving wild-type and crosslink-deficient C93A and Y157F variants of CDO with various substrates and inhibitors bound. For this project, I conducted experiments, determined the structures of additional soaks of CDO and CDO variants, and provided input on the analysis and text.

Chapter 2

The Sedoheptulose 7-phosphate Cyclases and Their Emerging Roles in Biology and Ecology

Andrew R. Osborn*, Kelsey M. Kean*, P. Andrew Karplus, and Taifo Mahmud

*These authors contributed equally to this work

Published in *Natural Product Reports, 2017, 34, pp 945-956*

Abstract

This highlight covers a family of enzymes of growing importance, the sedoheptulose 7-phosphate cyclases, initially of interest due to their involvement in the biosynthesis of pharmaceutically relevant secondary metabolites. More recently, these enzymes have been found throughout Prokarya and Eukarya, suggesting their broad potential biological roles in nature.

Scope and purpose

This highlight focuses on a subgroup of emerging importance among enzymes that catalyze the cyclization of C_7-sugar phosphates to cyclic compounds as a part of primary and/or secondary metabolism. Related to the well-studied 3-dehydroquinate synthases (DHQS) from the shikimate pathway, this subgroup of cyclases use sedoheptulose 7-phosphate (SH7P) as a substrate. There are now three known sedoheptulose 7-phosphate cyclases (SH7PC), which convert their common substrate to three different cyclic products.[12] They are 2-*epi*-5-*epi*-valiolone synthases (EEVS), converting SH7P to 2-*epi*-5-*epi*-valiolone; desmethyl-4-deoxygadusol synthase (DDGS), converting SH7P to desmethyl-4-deoxygadusol; and 2-*epi*-valiolone synthase (EVS), converting SH7P to 2-*epi*-valiolone. EEVS was originally thought to be present only in certain actinomycetes where it is part of the pathways to C_7N-aminocyclitol natural products; however, analyses of genome sequences revealed that EEVS is broadly distributed in both prokaryotes and eukaryotes, including bacteria, fungi, stramenopiles, and animals.[18] A recent study demonstrated that the role of EEVS in animals and some algae is to produce the sunscreen-like compound gadusol.[14] DDGS was originally discovered based on its role in the biosynthesis of mycosporines and mycosporine-like amino acids (MAAs),[13] another class of sunscreen compounds believed to play a critical role in ecology, whereas EVS was found to be involved in the biosynthesis of bioactive C_7N-aminocyclitol natural products.[80] In this highlight, we describe the discovery, biochemistry, and structure-function relationships of SH7PCs and discuss the distribution and the emerging roles of these less characterized sugar phosphate cyclases (SPCs) in biology and ecology.

Sugar phosphate cyclases and their roles in primary and secondary metabolism

As noted above, SPCs catalyze the cyclization of six- and seven-carbon sugar phosphates to a variety of cyclitol products, which are part of primary and secondary metabolism. The one SPC superfamily member directly involved in primary metabolism is DHQS, which catalyzes one of the rate limiting steps of the shikimate pathway – the cyclization of 3-deoxy-D-arabinoheptulosonate 7-phosphate (DAHP) to 3-dehydroquinate (DHQ) – and is necessary for the synthesis of aromatic amino acids.[81] Seminal work by Floss and others have shown that a close homolog of DHQS, known as aminoDHQS, is involved in secondary metabolism (e.g., in the biosynthesis of rifamycin and mitomycin), but it follows the template of DHQS in the cyclization of an amino sugar phosphate.[82] AminoDHQS uses aminoDAHP to form aminoDHQ, which is further modified to 3-amino-5-hydroxybenzoic acid (AHBA).[83] The latter compound is then usually used as a starter unit for polyketide synthases. In its own right, DHQS can also be linked to secondary metabolism through aromatic amino acid-derived compounds, such as esmeraldic acid and scytonemin (Figure 2.1).

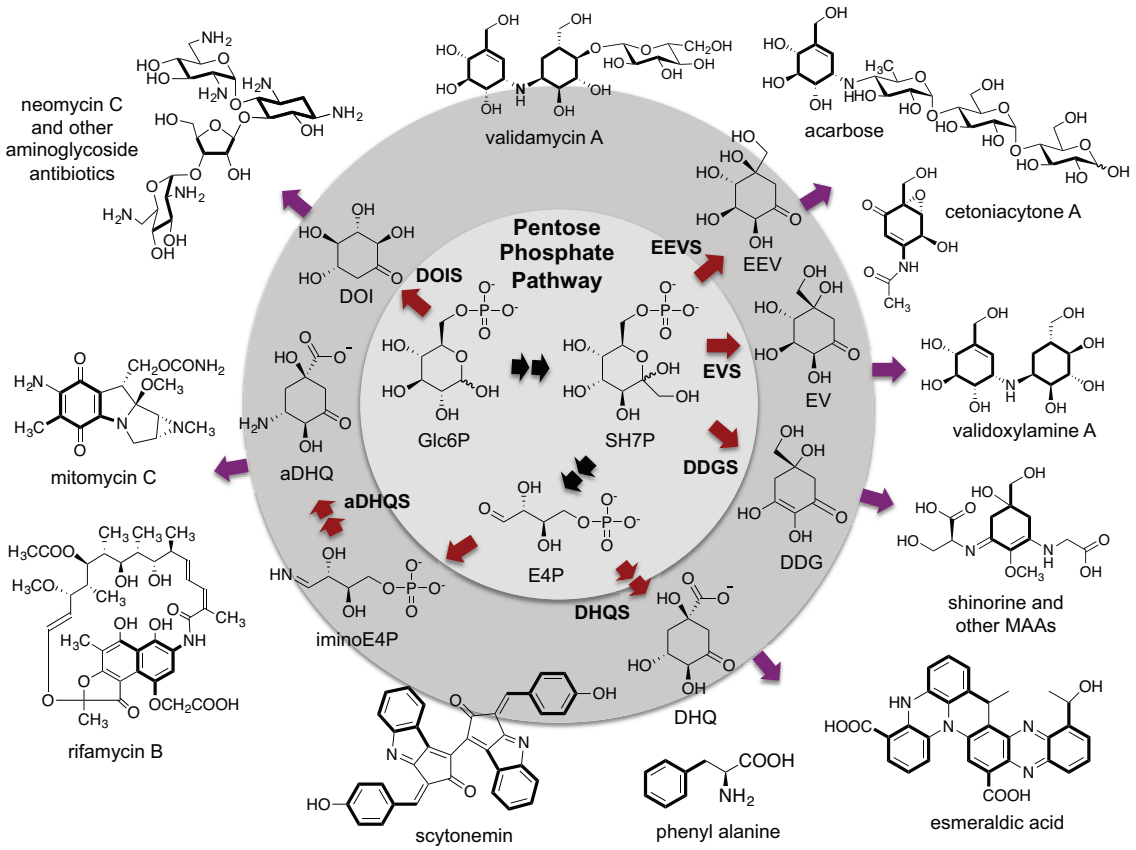


Figure 2.1 Sugar phosphate cyclases (SPCs), their products, and representative natural products resulting from their pathways.
Glc6P, glucose 6-phosphate; SH7P, sedoheptulose 7-phosphate; E4P, erythrose 4-phosphate; iminoE4P, imino-erythrose 4-phosphate; EEV, 2-*epi*-5-*epi*-valiolone, EV, 2-*epi*-valiolone, DDG, desmethyl-4-deoxygadusol; DHQ, 3-dehydroquinate; aDHQ, 3-aminodehydroquinate; DOI, 2-deoxy-*scyllo*-inosose.

Other known SPCs use intermediates of the pentose phosphate pathway, which is a key primary metabolic pathway that serves as a source of NADPH, ribose 5-phosphate, and erythrose 4-phosphate. 2-Deoxy-*scyllo*-inosose synthase (DOIS) uses glucose 6-phosphate as substrate and generates 2-deoxy-*scyllo*-inosose (DOI), a key scaffold in aminoglycoside antibiotics (e.g., neomycin and kanamycin). The sedoheptulose 7-phosphate cyclases (SH7PCs) catalyze the cyclization of SH7P to give 2-*epi*-5-*epi*-valiolone (EEV), desmethyl-4-deoxygadusol (DDG), or 2-*epi*-valiolone (EV). These cyclitol molecules can be modified and functionalized further to generate molecules that are both diverse in terms of structure and function, such as the fungicide validamycin, the cytotoxin cetoniacytone A, and the UV-absorbant

MAAs. In a sense, SPCs can be considered to be a gateway between primary and secondary metabolism, as they convert intermediates from the pentose phosphate pathway to secondary metabolites. Due to the diversity of these secondary metabolites and the organisms that synthesize them, it is also possible that aside from their functions as secondary metabolites they may have a larger role in primary metabolism than what is currently known.

Sedoheptulose 7-phosphate cyclases (SH7PCs)

2-Epi-5-epi-valiolone synthases (EEVSs)

The identification of this new class of enzymes was first reported in 1999 when feeding experiments revealed that the valienamine moiety of acarbose, an α-glucosidase inhibitor used for treating type II diabetes, was derived from EEV.[9] Furthermore, biochemical characterization of a DHQS-like protein (AcbC) within the acarbose pathway in *Actinoplanes* sp. 50/110 demonstrated that the protein could convert SH7P to EEV.[10] Multiple sequence alignments and the crystal structure of DHQS, showed that AcbC and DHQS matched for 7 of 13 key residues associated with catalysis and substrate binding, metal binding, and NAD^+ binding, and with notable differences postulated to be localized around the varying substituents of their SH7P and DAHP substrates. Based on this, the catalytic mechanism for EEVS was proposed to be similar to DHQS involving 5 step reactions: dehydrogenation by NAD^+, elimination of the phosphate group, reduction of the C-5 ketone, ring opening and rotation along the C5 – C6 bond, and intramolecular aldol cyclization (Figure 2.1A).[10]

After the initial identification of the AcbC protein as an enzyme that uses SH7P to make EEV, subsequent work showed 2-*epi*-5-*epi*-valiolone is incorporated into the secondary metabolites validamycin,[84] cetoniacytone A,[11] and pyralomicin,[85] suggesting these biosynthetic pathways also contain an AcbC-ortholog. The corresponding AcbC-ortholog genes found within the biosynthetic gene clusters of validamycin (ValA),[15] cetoniacytone A (CetA),[11] BE-40644 (BE-Orf9),[11] salbostatin (SalQ),[86] acarbose in *Streptomyces glaucescens* (GacC),[87] and pyralomicin (prlA),[88] were identified and shown to be EEVSs.

While originally thought to be limited to bacteria, EEVS genes have recently been identified in algae, marine invertebrates, non-mammalian vertebrates, and a fungus and are implicated in the *de novo* production of gadusol in some of these organisms.[14] Based on heterologous expression and characterization of this gene from zebrafish, this too is an authentic EEVS – using SH7P to make EEV. And still more recently, a group of phylogenetically distinct, bacterial EEVS (referred to as EEVS*), have been identified and shown to have EEVS activity.[18]

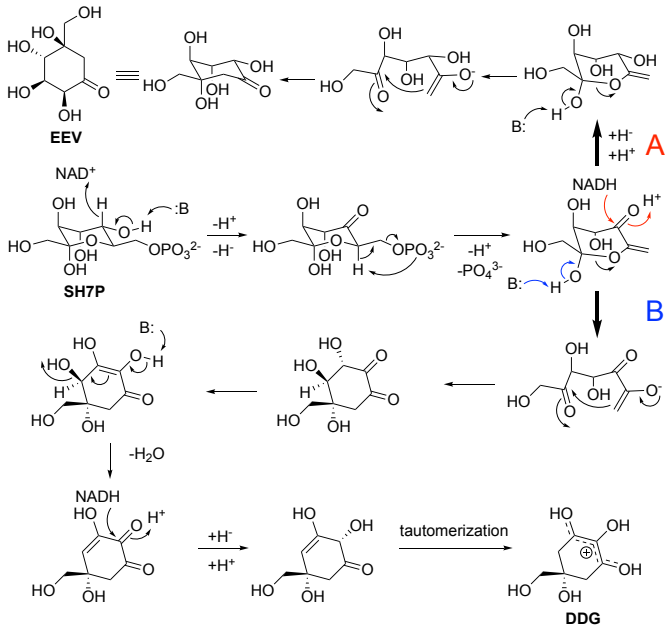


Figure 2.2. Proposed catalytic mechanisms for EEVS and DDGS.
(**A**) Proposed mechanism for EEVS. (**B**) Proposed mechanism for DDGS.

Desmethyl-4-deoxygadusol synthases (DDGSs)

The identification of DDGSs in a number of cyanobacteria as a distinct group of SPCs was first reported in 2007.[11] In 2010, Balskus and Walsh experimentally demonstrated that a DDGS, as well as an *O*-methyltransferase (O-MT), a ATP-grasp protein, and a NRPS-like protein, are responsible for the synthesis of MAAs, contradicting the longstanding assumption that MAAs are synthesized via the shikimate pathway.[13] DDGS, e.g., Ava_3858 of *Anabaena variabilis* and Npun_5600 of *Nostoc punctiforme*, converts SH7P to DDG. A proposed mechanism for DDGS is similar to those of other SPCs, but with additional enolization, dehydration, and tautomerization steps (Figure 2.2B).[12]

In addition to these DDGSs, other putative and divergent DDGSs[18] have been identified and characterized to varying degrees.[89-93] Likewise, these genes are found with an O-MT or even fused with an O-MT and have been shown to be responsible for MAA production via heterologous expression and characterization.[93] Many of these proteins contain N-terminal extensions or plastid targeting sequences and in some cases evidence supports horizontal gene transfer as part of the evolutionary history of these enzymes[89-92, 94] (see Evolution of SH7PC by horizontal gene transfer).

2-*Epi*-valiolone synthases (EVS)

EVSs appear to be the least common among the SH7PC family of enzymes and were only identified and differentiated from other SPCs in 2012.[12] Only two EVSs have been heterologously expressed and characterized for their activity, showing they use SH7P as substrate to make EV, but more EVS were identified through bioinformatics studies.[12, 14, 18] Prior to the characterization of EVS, its product (EV) had not been identified in nature. Despite its similarity in function to EEVS and DDGS, EVSs appear to be more similar to DHQS in their active site.[12] The catalytic mechanism for EVS was proposed to be very similar to those for DHQS and EEVS, but accounted for the difference in configuration at C-5 between EEV and EV by requiring a 180° rotation prior to ring closing. In an alternative hypothesis, EEVS and EVS differ by the anomer of SH7P they bind (see section 4.2).[17]

EVSs continue to remain to be the most elusive and least well-understood members of the SH7PC family, in part due to their more recent identification, the greater challenge in differentiating EVS from DHQS, and the apparent limited presence of EVS in nature. However, EVS has been identified as replacing EEVS in some C_7N-aminocyclitol gene clusters.[80] For instance, gene knockout and comparative metabolomics with the wild-type *Actinosynnema mirum* identified Amir_2000 as being involved in the biosynthesis of validoxylamine A, a precursor to validamycin A. This was both the first C_7N-aminocyclitol identified in *A. mirum*, and represented a new route for the biosynthesis of validoxylamine A.

Structural aspects of SH7PCs

Until recently, most inferences about structure-function relations for SH7PCs have been based on the well-studied DHQS.[10-13] In 2014, Kean et al.[17] published the first structure of an SH7PC, that of ValA, the EEVS from *S. hygroscopicus* in the validamycin pathway. In 2017, the structure of a DDGS, Ava_3858 from *A. variabilis* involved in the production of MAAs, was published.[18] These structures are highly similar to each other and reveal features shared among the SPC superfamily as well as features unique to the SH7PCs and the individual enzymes EEVS and DDGS. Unfortunately, no structural information is yet available for the third SH7PC, EVS, which surprisingly has more similarity to DHQS than to the other SH7PCs. Also no complexes of ligand-bound structures have been solved.

Overall structures and active site

In terms of the fold and broad aspects of active site organization, EEVS and DDGS are quite similar to DHQS. Both form functionally obligatory homodimers, with both chains of the dimer contributing to each active site. Each chain contains bound NAD and Zn prosthetic groups and consists of an N-terminal NAD-binding domain with a parallel β-sheet core and a C-terminal, mostly α-helical metal-binding domain (Figure 2.3). The active site is in a cleft formed where N and C-terminal domains come together. Among the SPC superfamily, the identities of the metal-binding residues are conserved as 1 Glu and 2 His residues. A structural feature observed in EEVS and DDGS but not in DHQS is a domain swapping interaction in which the N-terminal residues make an extended β-strand (labeled β1/β2) that reaches across the back of the dimer and contributes a β-strand to the opposite monomer of the dimer (Figure 2.3), rather than forming a β-hairpin as observed in DHQS.

In the ValA active site, it was noted that many features are well conserved with DHQS. One conserved feature that had apparently not been highlighted in the DHQS literature was a conserved pair of Asp residues (Asp138 and Asp165 in ValA) which have their carboxylate oxygens positioned to electrostatically stabilize the oxidized form of the nicotinamide ring during catalysis (See Fig 2 of [17]). Also, key active site differences have been identified in EEVS, DDGS, and EVS, which

differentiate them from each other and from the broader SPC superfamily. Based on sequence comparisons, EEVS and DDGS are the more similar members of the SH7PC family, and the structures of ValA and Ava_3858 confirmed two previously identified pairs of residues predicted to differ in their active sites (Asp281 and His360 in ValA vs. Ala268 and Thr347 in Ava_3858), and identified a third difference (Leu267 in ValA vs. Glu254 in Ava_3858).[17-18] To assess how these differences contribute to the different catalytic activities of EEVS and DDGS, a series of mutants were generated and activity assays showed that Asp281 and Leu267 in ValA and the equivalent residues Ala268 and Glu254 in Ava_3858 are essential for activity.[18] However, the specific roles these residues play in catalysis and how they lead to differential activity are not yet understood.

The comparison of the ValA structure (an EEVS) with DHQS also revealed an important surprise related to an active site Arg located at the beginning of helix 4 in the C-terminal domain.[17] This Arg is also present in DDGS, but does not correspond to that found in the DHQS active site; instead it is the next residue. This alternate Arg is associated with a robust difference in a loop conformation and the sequence of EVS implies that it will conserve the Arg found in DHQS (See Fig. 7 of [17]). Interestingly, these Arg residues are positioned near where we expect the substituents of the anomeric carbon (C2) of SH7P to bind, making it plausible that this arginine is involved in differential anomer recognition between EVS and EEVS/DDGS.

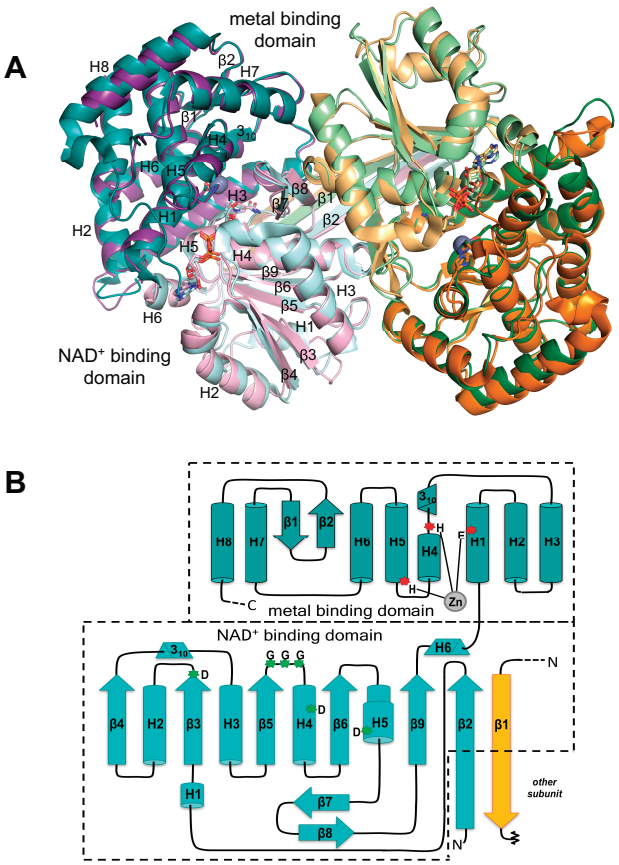


Figure 2.3 Overall structure and topology of known SH7PCs, EEVS and DDGS.
(**A**) Overlay of the dimers of the Ava_3858 (teal and orange; PDB code 5TPR) and ValA (purple and green; PDB code 4P53) are shown with the N-terminal NAD^+-binding domains in light hues and C-terminal metal-binding domains in dark hues. The NAD^+ and Zn^{2+} with its coordinating ligands are shown for each. Secondary structural elements in each domain of one monomer are labeled. (**B**) Topology diagram showing α-helices (cylinders), β-strands (arrows), 3_{10} helices (triangular prisms) and π-helices (wider cylinder) common to EEVS and DDGS. The domains are colored light and dark teal as indicated, and helices (H) and strands (β) common to the SPCs are named within each domain. The domain swapped β-strand β1 from the other subunit of the dimer but contributing to the NAD^+-binding domain of the teal subunit is colored in light orange. The three Zn^{2+}-binding residues (red asterisks) and glycine-rich turn and conserved aspartic acids (green asterisks) important for NAD^+ binding are indicated.

An anomeric selection hypothesis

In light of these structures and conservation patterns among the SPC superfamily, we proposed a new hypothesis to explain how the SH7PCs can generate such different products from the same substrate.[17] The SH7PCs had previously been viewed as binding a single substrate, the α-pyranose form of SH7P, and this meant that the mechanism proposed for EVS required a portion of the ring-open enzyme

bound intermediate to undergo a 180° rotation after ring opening so that the C-C bond formation ring-closure step would yield the product with the correct stereochemistry.[12] However, in reality, SH7P can readily interconvert in solution between an open form and the α and β anomers of pyranose and furanose ring forms. Based on the EEVS structure, and specifically the fact that EEVS and DDGS conserved a different Arg residue than did EVS, we proposed that a more structurally compelling hypothesis would be that these enzymes bind different anomers of SH7P that correspond to the stereochemistry of their final product (Figure 2.4). This would allow each enzyme to produce the correct product without requiring a dramatic rotation within the spatial constraints of the active site pocket.

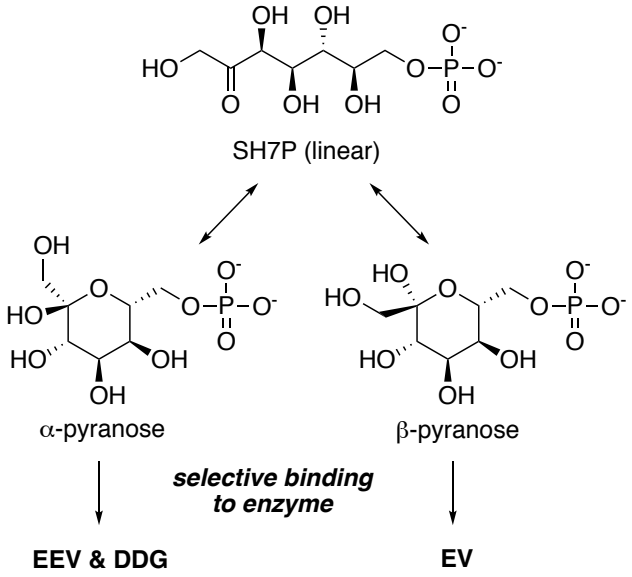


Figure 2.4 Proposed enzyme-specific selection of forms of sedoheptulose 7-phosphate (SH7P).
EEV, 2-*epi*-5-*epi*-valiolone; DDG, desmethyl-4-deoxygadusol; EV, 2-*epi*-valiolone.

A speculative proposal for SH7P binding to DDGS/EEVS

General features of substrate binding by the SH7PCs have been proposed based on what was seen for DHQS, but it has been acknowledged that the DHQS complex (PDB code 1DQS) cannot match in detail the substrate binding by the SH7PCs, in part because of the stereochemical differences in key metal-binding hydroxyls of their substrates (Figure 2.5A): in DAHP and its substrate analogue,

CBP, both metal-coordinating hydroxyls are equatorial while in SH7P one metal-coordinating hydroxyl is equatorial while the other is axial. As part of preparing this highlight, we thought it would be valuable to better define what can be reasonably predicted about how the α-pyranose form of SH7P might bind to the DDGS and EEVS active sites. We created an idealized chair conformation of the substrate and manually docked it into the DDGS active site using the positions of ordered solvent molecules (water and a sulfate) to approximate the locations of the SH7P hydroxyl and phosphoryl substituents. Gratifyingly, this led to a unique and plausible binding mode shown in Figure 2.5B.

In this docked model, both the axial C4-hydroxyl and the equatorial C5-hydroxyl of SH7P are positioned well to coordinate the Zn. Also, the phosphoryl group matches well with the bound sulfate and many of the substituent hydroxyls are positioned to form hydrogen bonds with the protein. Especially notable is that the hydroxymethyl of the anomeric carbon (C2) is positioned to hydrogen bond with both Lys156 and Arg265 (the above mentioned Arg distinguishing EEVS and DDGS from EVS), consistent with this Arg being responsible for anomer recognition. The arginine conserved in EVS would be in a different position, consistent with EVS binding the opposite anomer. Important to note is that sidechain positions, especially of Lys156 and Arg265, should be taken as very rough because upon substrate binding, we expect that the DDGS structure used for modeling will undergo some domain closure similar to that observed in DHQS.[95]

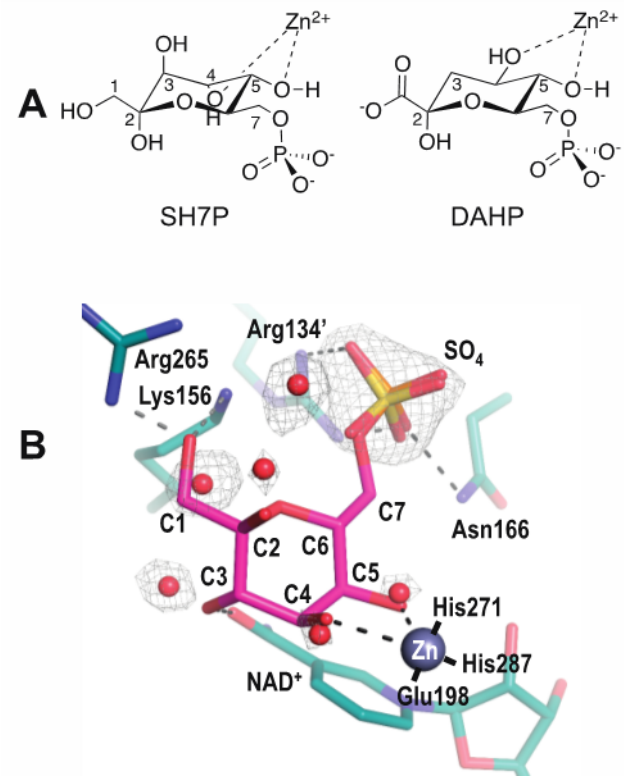


Figure 2.5 Substrate recognition by SH7PCs.
(A) The boat conformation of DAHP as it coordinates Zn2+ in the DHQS active site has both Zn-ligating hydroxyls equatorial, yet in SH7P one putative Zn-ligating hydroxyl is equatorial while the other Zn-ligating hydroxyl is axial. (B) Speculative binding mode of SH7P in EEVS and DDGS. The α-pyranose anomer of SH7P (magenta) manually docked into the active site of Ava_3858 (teal) with bound zinc (silver sphere). The electron density (grey, contoured at 1.3ρrms) evidence for the bound waters (red spheres) and sulfate (yellow) is shown along with proposed interactions of the docked SH7P with the active site (dashed lines). While Lys156 and Arg265 have short approach distances in this model, we expect these sidechains will move due to an anticipated conformation change upon substrate binding. The coordination of Zn with its three ligating residues are also shown (solid black lines). The carbons of SH7P the protein side chains and NAD+ are labeled, with a prime symbol meaning the side chain comes from the other subunit of the dimer.

Natural distribution of SH7PCs

SH7PCs in bacteria

Bacterial SH7PCs have been found in Gram-(+) and Gram-(–) bacteria isolated from soils, aquatic systems (fresh and salt), symbionts, pathogens, and extreme environments.[12, 14, 18] This diversity suggests SH7P-derived natural products have many biological roles, which will be discussed later in this Highlight. The majority of identified bacterial SH7PCs are from actinobacteria (EEVS) and cyanobacteria (DDGS).[18] Bacterial EEVS have been reported in a number of C_7N-

aminocyclitol gene clusters (e.g., acarbose and validamycin), whereas DDGS has only been found in MAA gene clusters.[18] EVS appears with much fewer frequency and is limited to soil actinobacteria and myxobacteria.[12] Divergent DDGS and EEVS* sequences have also been identified through phylogenetics studies, although only EEVS* was enzymatically characterized.[12, 14, 18]

The bulk of known EEVS sequences are from the genus *Streptomyces,* in which the C_7N-aminocyclitol gene clusters, usually resembling the acarbose and validamycin gene clusters, are widespread.[18] However, EEVS is also found in diverse Gram-(+) bacteria, including marine bacteria (e.g., *Salinispora* spp.), pathogenic bacteria (e.g., *Rhodococcus fascians*, *Mycobacterium marinum,* and *Clostridium botulinum*), other soil-dwelling bacteria (e.g., *Amycolatopsis, Nonomuraea*, *Nocardia*, *Kitasatospora*) and plant symbionts (e.g., *Frankia alni*).[12, 14, 18] Of these, only *C. botulinum* is not an actinobacteria. Likewise, few EEVS* genes are found outside of actinobacteria and myxobacteria, most notably in *Bacillus cereus*. EEVS is also found in Gram-(–) bacteria, which include the genera *Burkholderia, Paraburkholderia, Pseudomonas,* and *Rhodanobacter*. The EEVS of *P. kirkii* is involved in the biosynthesis of kirkamide (Figure 2.6), an aminocyclitol with cytotoxicity towards aquatic anthropods and insects, while the gene clusters including EEVS in other Gram-(–) bacteria remain uncharacterized.[18, 96]

Cyanobacteria have the most species with DDGS encoded in their genome and are found in MAA gene clusters. MAAs have been found in cyanobacteria from fresh and salt water, high salt conditions, and terrestrial environments.[18, 93, 97] Compared to products deriving from EEVS-containing gene clusters, the products of MAA gene clusters are easier to detect due to the structural similarity that essentially only diverges by which amino acids are attached to the 4-deoxygadusol core. However, this similarity likely requires further structural elucidation to identify individual MAAs. Interestingly, to date, only one putative EEVS has been identified in cyanobacteria, *Gloeocapsa* sp. PCC 7428, which lacks a DDGS.[18] This EEVS is located in an uncharacterized gene cluster.

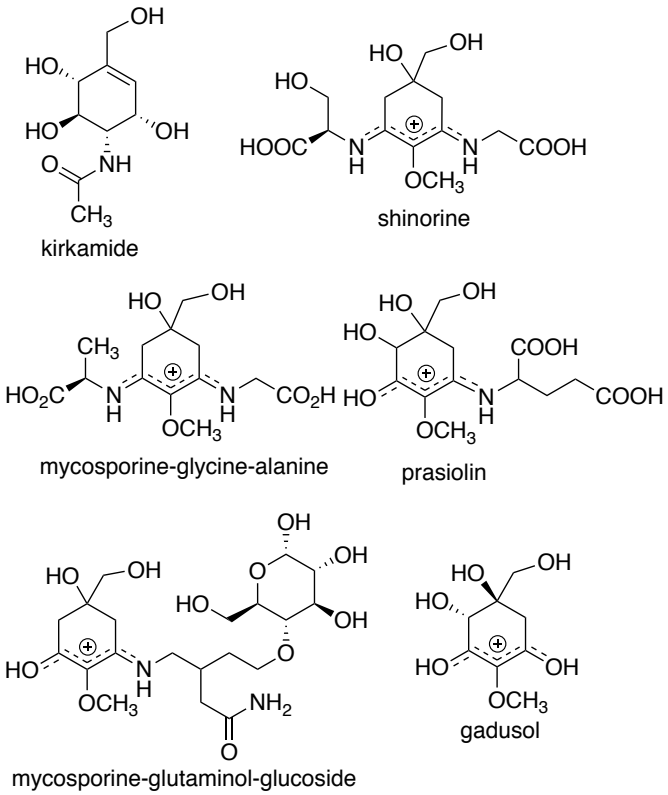


Figure 2.6 Chemical structures of kirkamide, gadusol, and representatives of mycosporines and MAAs.
Some stereoconfigurations were not assigned.

In actinobacteria, DDGS and other MAA biosynthetic genes were first identified through genome mining of *Actinosynnema mirum* DSM 43827 and *Pseudonocardia* sp. P1.[98] Heterologous expression of both gene clusters in *S. avermitilis* SUKA22 revealed that shinorine was the primary MAA made by each gene cluster, but the *A. mirum* cluster also produced a new MAA mycosporine-glycine-alanine (Figure 2.6).[98] This opened the door to exploring actinobacterial MAA gene clusters as sources of novel MAAs, especially now that more DDGS have been identified in other actinobacteria.[93] Although DDGS and EEVS have been identified in actinobacteria, it is rare for both to exist in the same organism. The lone exception in actinobacteria is *R. fascians*, a plant pathogen.

SH7PCs in fungi

Other than one putative EEVS in a yeast (*Saitoella complicata*), all SH7PCs in fungi are predicted to be DDGS and are present in Ascomycota and Basidiomycota.[18] While mycosporines, MAA analogs with only one amino acid attached to a 4-deoxygadusol core, were first found in fungi, the corresponding gene clusters in fungi have not yet been characterized experimentally. Since mycosporines may be involved in inhibiting sporulation, it may not be surprising to see mycosporine genes broadly distributed in fungi.[99] Fungi that have mycosporine genes includes many important plant fungal pathogens, such as *Botrytis cinerea*, *Colletotrichum* spp., *Fusarium graminearum*, *F. oxysporum*, *Magnaporthe oryzae*, and *Ustilago maydis*.

More recently, mycosporines and MAAs have been identified in yeast and lichen. In basidiomycetous yeast, the ability to make mycosporines appears to be taxonomically split, where five of the seven Pucciniomycotina classes produced mycosporine-glutaminol-glucoside.[100] Cyanobacterial lichens and a tripartite lichen containing a green alga can make a suite of MAAs, suggesting they contain DDGS.[101]

SH7PCs in marine organisms

MAAs and their corresponding gene clusters have been identified in marine invertebrates, dinoflagellates, and algae.[18, 89-90, 102] These organisms have a fused divergent DDGS and O-MT, unlike traditional MAA gene clusters where the DDGS and O-MT are discrete. Phylogenetic analysis suggests that multiple horizontal gene transfer (HGT) events spread this fusion protein, with origins in cyanobacteria and dinoflagellates.[89] Organisms that have this fused divergent DDGS include cyanobacteria from the genus *Synechoccocus*, the dinoflagellates *Oxyrrhis marina* and *Heterocapsa triquetra*, and the sea anemones *Nematostella vectensis* and *Aiptasia pallida*, and the coral *Acropora digitifera.* These fusion proteins also had an N-terminal addition that was proposed to target these proteins to the plastid in dinoflagellates, though all homologs seem to have this addition.[91]

Marine red algae are a major source of MAAs, however, few genome sequences are publically available; thus, the identification of algal SH7PCs is

hindered. Nevertheless, divergent DDGS has been identified in the marine red algae *Pyropia haitanensis* and *Chondrus crispus*.[18] Currently, more EEVS sequences (12) have been identified in algae than DDGS/divergent DDGS (4).[18] Most organisms have EEVS or DDGS, and not both, yet two algae do have both: *C. crispus* also has an EEVS, while the pelagophyte *Aureococcus anophagefferens*, which causes harmful brown tides, has an EEVS and DDGS. The EEVS from *A. anophagefferens* is more closely related to the vertebrate EEVS, which is involved in gadusol biosynthesis (see below), than other algal EEVS sequences.[14, 18] While there has been no direct evidence that algae can produce gadusol, the terrestrial green alga *Prasiola calophylla*, does produce an unusual MAA, prasiolin, which consists of glutamic acid attached to a gadusol core.[103] No publically available genome exists for *P. calophylla*, so EEVS is not known to exist in its genome. In addition, it remains possible that prasiolin is formed from a 4-deoxygadusol core, which is then hydroxylated, rather than being derived from gadusol.

SH7PCs in vertebrates

MAAs have been found in the skin mucus and eyes of marine fishes, and since they are produced by microbes and found in marine invertebrates, are proposed to be acquired by fishes through their diet.[104] A structurally similar compound named gadusol was isolated from the roes of *Gadus morhua* (Atlantic cod).[105] Due to this similarity, it was also suggested to be acquired through the diet. More recently, however, a surprising discovery was made on the presence of EEVS-like genes in the genomes of fish, amphibians, reptiles, and birds.[14] Since vertebrates lack DHQS and the shikimate pathway, it was thought they would not have any DHQS-related SPCs. Clustered with the EEVS-like gene was a methyltransferase/oxidoreductase (MT-Ox), and together they work to make gadusol, a possible sunscreen compound.[14] The majority of EEVS-like sequences known so far come from fish and birds.[18] To date, EEVS has been identified in 99 vertebrates.[18] However, the presence of the gadusol genes does not mean they are functional in all organisms. A recent bioinformatics analysis identified mutations in the EEVS and MT-Ox genes in four crocodylians.[106] These mutations in both genes were predicted to occur ~190 million years ago, and

their concurrent loss is consistent with their function. The crocodilian EEVS has multiple deletions and insertions, including a chompy repeat element, two premature stop codons, and a splice-donor mutation.[106] The MT-Ox also has multiple deletions. This raises the possibility that not all non-mammalian vertebrates have active gadusol genes, thus the ability to make gadusol needs be confirmed in more vertebrate species.

Evolution and distribution of SH7PCs by horizontal gene transfer

While HGT events are often difficult to identify and characterize, multiple independent studies looking at SH7PCs and MAA biosynthesis in cyanobacteria, dinoflagellates, and cnidarians have provided phylogenetic evidence for the transfer of what we now call divergent DDGSs across diverse genera by HGT. These enzymes have been particularly useful to study because the divergent DDGS always exist with a downstream O-MT or fused O-MT. In 2006, Waller et al.[89] presented a rare and surprisingly well-defined case of HGT of a divergent DDGS and O-MT in which the donor (cyanobacteria) and recipient (dinoflagellates) are clear. Interestingly, both a fusion and subsequent fission of these two genes can be tracked through evolution of the dinoflagellates. Phylogenetic analysis also revealed a second, separate HGT event of these genes between terrestrial cyanobacteria and fungi. The physical association of terrestrial cyanobacteria with fungi and marine cyanobacteria with dinoflagellates are consistent with HGT events.[89]

Additional support for HGT involving a SH7PC has also been observed in some marine holobionts.[90] For instance, the sea anemone *Anthopleura elegantissima* contain the same MAAs regardless of the absence of its dinoflagellate symbiont or diet,[107-108] suggesting that *A. elegantissima* is capable of synthesizing MAAs independently. Furthermore, the sea anemone *Nematostella vectensis*, the first basal metazoan with its genome sequenced, had a divergent DDGS:O-MT, suggesting that it could synthesize MAAs.[90] Interestingly, the closest relatives of the *N. vectensis* divergent DDGS:O-MT were the fusion proteins from the dinoflagellates *Oxyrrhis marina* (~65% identical DDGS part of fusion protein) and *Heterocapsa triquetra*, which strongly suggests that the cnidarian divergent DDGS was derived through

HGT from a symbiotic dinoflagellate to an ancestral cnidarian. This type of HGT event has also been proposed to occur between cyanobacteria, dinoflagellates, and metazoans.[91-92]

Likewise, work by Osborn et al.[14] suggests that the EEVS and genetically linked MT-Ox found in vertebrates has arisen by HGT. The only non-vertebrate organisms found to have orthologous adjacent EEVS and MT-Ox genes are two algae, the stramenophile *A. anophagefferens* and the microalgae *Coccomyxa subellipsodisea*. As in a phylogenetic tree these genes group closely with those of vertebrates, it is proposed that they are the source for these genes in non-mammalian vertebrates. Recently, the only known fungal EEVS (from *S. complicata*) was also found to clade with the algae in the vertebrate EEVS clade.[18] Thus, it is possible that algae also transferred this EEVS to *S. complicata*. The exact time point of when vertebrates gained the genes encoding gadusol production remains unknown, but they appear to have gained the gadusol genes during the evolution of bony fishes, while certain lineages (e.g., mammals and coelacanths) lost them again. Even early diverging mammals, like the duck-billed platypus, lack the gadusol genes. If gadusol does biologically function as a UV-protectant, mammals may have lost these genes during the nocturnal bottleneck.[18]

The biology and ecology of SH7PCs and their products

While SH7PCs are widely distributed in different environments and organisms, the biological function of the final secondary metabolites is generally unknown or unclear. In soil bacteria, EEVS- and EVS-derived C_7N-aminocyclitols are the most common SH7P-derived natural products. The amino-pseudooligosaccharides (e.g., acarbose, oligostatins, and trestatins), which have potent α-glucosidase inhibitory activity, could play a role in preventing the utilization of sugars by surrounding organisms, and may also function as 'carbophores' that tag and import extracellular glucose.[109] Validoxylamine A, which strongly inhibits the fungal enzyme trehalase, can prevent the fungi from accessing its energy storage.[110] Thus, some C_7N-aminocyclitols seem to be used for self-defense, competition for food, or niche protection.

The most common SH7PC in marine and freshwater organisms is DDGS, and the known biological functions of MAAs are constantly expanding. MAAs have features of keystone molecules, due to their presence throughout many trophic levels, yet their function differs not only by organism but also within single organisms.[102] In bacteria, production of MAAs has been associated with providing protection against UV damage, salt stress, and high ammonium due to increased MAA levels under these conditions.[93, 98, 111-112] In the cyanobacteria *Microcystis aeruginosa*, shinorine was found in the slime layer, where it appears to have a primary structural role in the sheath and a limited role in UV protection.[113] Stronger evidence for MAAs primarily acting as sunscreen exist in other cyanobacteria, e.g., *A. variabilis*, where shinorine production is directly linked to UV exposure.[97, 112] However, MAAs may still have multiple functions in those cyanobacteria, possibly observed only under stressful conditions (e.g., sulfate deficiency, salt shock, and UV exposure). [93, 98, 111-112] Additionally, the location of MAAs could also signify their different roles, possibly by organism, as MAAs can be found in the cytoplasm and/or in the outer sheath.[97, 113-114]

The major marine ecological source of MAAs are macro- and microalgae, which are capable of producing a wide range of MAAs.[97, 115-116] Algae can serve as the dietary source of MAAs for marine invertebrate herbivores, which is one route for MAA bioaccumulation up the food chain (Figure 2.7).[102, 117] UV light exposure increases production of MAAs in algae, and has been reviewed previously.[97, 104] This matches the tendency of elevated MAA concentrations in red algae that live closer to the ocean's surface than those at increased depth.[116] Microalgae (e.g., dinoflagellates) can be the source of MAAs during symbiosis.[118]

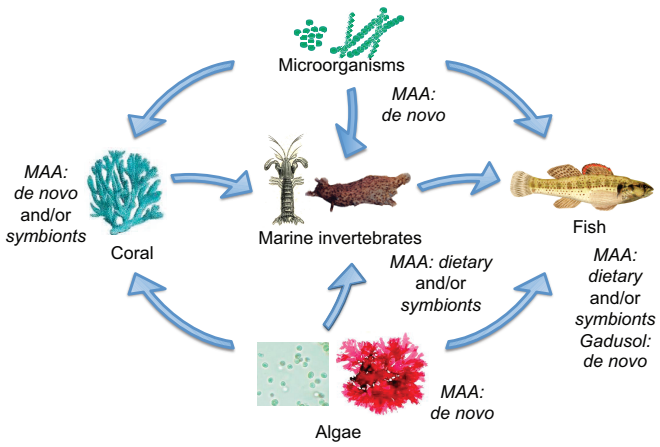


Figure 2.7 Distribution of gadusol and MAAs through different tropic levels of the marine environment.
Some microorganisms, algae, and corals can produce MAA *de novo*, whereas gadusol is mostly produced by non-mammalian vertebrates, such as fish.

The functions of MAAs in marine invertebrates are even more diverse than in bacteria and algae; in addition to UV protection, MAAs have been implicated in alarm cues and vision. MAAs can be acquired by marine invertebrates through their diet or symbionts, and can be produced *de novo*.[102, 117-121] In corals, MAAs accumulate with exposure to UV light, suggesting MAAs contribute to UV protection.[118-119] In the sea hare *Aplysia californica*, MAAs (which are acquired through diet) have been identified as a component of the sea hare's defensive secretion, opaline, which elicited a behavior response in juvenile sea hares.[102] Moreover, MAAs are found in the eyes of the mantis shrimp *Neogonodactylus oerstedii*, where they function as part of the shrimp's complex color vision that includes a UV visual system.[122] Beyond diet, symbiosis, and *de novo* synthesis, MAAs can also be inherited, as evidenced by the slipper limpets *Crepipatella peruviana* and *C. dilatata*.[121] Females of these two species have different MAA levels in general and pass MAAs to their embryos through two different mechanisms, i.e., through the embryonic yolk and nurse eggs, respectively. MAAs have also been quantified in intertidal egg masses of mollusks and polychaetes, where the adult diet, phylogeny, and egg viability impact MAA composition.[123] Eggs from mollusk herbivores had higher MAA concentrations than carnivores, likely because herbivores are directly consuming MAAs in their diet.[123]

Fish are incapable of making MAAs, yet they have been found in the skin mucus of fish and in fish eyes, where they may act as sunscreens, UV vision, and/or UV communication.[104, 124-125] Some reef fish and cleaner fish increase UV absorption in their mucus when UV-B exposure is increased. Depth and sampling location also effect MAA accumulation in the skin mucus.[126] Since mollusk herbivores had more MAAs than carnivores, it is tempting to speculate that a similar phenomenon will occur with marine fishes.

Although fish cannot make MAAs, evidence implies that most can synthesize gadusol.[14] Gadusol was thought to be derived from 4-deoxygadusol in MAA producing organisms and obtained only through the diet, but a recent study has shown that fish can synthesize gadusol *de novo* using an EEVS.[14] This represents a transition from DDGS derived products to EEVS derived products in the marine environment. Like MAAs, gadusol has UV-protective and antioxidant activities and likely provides protection from UV light and/or oxidative stress, but gadusol may also play a physiological role during embryonic development. In zebra fish (*Danio rerio*), the EEVS and MT-Ox genes were expressed during development with maximum expression at 72 hours post fertilization. Thus, gadusol may provide some function for hatching embryos or newly hatched larvae. Other than the various live-birthing guppy species, the non-mammalian vertebrates with gadusol genes are oviparous; however, gadusol may not be linked to oviparity. The platypus, an egg laying mammal that diverged early in mammalian evolution, does not have the gadusol genes,[14] and crocodylians appear to have non-functional gadusol genes.[106] While MAAs have been identified in fish, it is possible that MAA-like compounds derived from gadusol exist. In fact, prasiolin, a mycosporine from the terrestrial green algae *Prasiola calophylla*, contains a gadusol core.[103] However, homologs of the amino acid-attaching MAA genes, the ATP-grasp-like and D-Ala D-Ala ligase-like or NRPS-like genes, have not been identified in vertebrate genomes.

SH7PCs and Their Products in Modern Symbiotic Relationships

As described above, perhaps the most studied symbiotic relationship involving MAAs is that of corals and dinoflagellates. In this mutualistic symbiotic relationship,

corals provide shelter and inorganic nutrients for dinoflagellates and dinoflagellates provide their hosts with photosynthetic and secondary metabolites, such as MAAs. Coral susceptibility to bleaching is partially dependent on the presence as well as the type of its dinoflagellate symbionts,[127] the latter which may be related to the repertoire of MAAs different dinoflagellates are capable of synthesizing. Due to the threat to coral reefs posed by global warming, understanding these relationships is especially important and could aid us in developing strategies to promote coral reef health. MAAs are essential for protecting these coral from environmental factors, especially UV radiation. There is evidence that cnidarians can synthesize MAAs *de novo*,[90] but their symbiotic dinoflagellates are also potential source of these protective compounds.[119] Coral dinoflagellates also have symbiotic relationships with other marine invertebrates that involve MAAs.[127] Regardless of the source, UV protection by MAAs (and therefore DDGS) is essential for maintaining this delicate but ecologically important marine habitat.

SH7PCs and their products have been implicated in other, diverse symbiotic relationships, including lichen communities, fish cleaning, leaf nodules, and kleptoplasty, where they are intimately involved in the dynamics and maintenance of these delicate relationships. The studies of these more diverse relationships are limited but we provide brief vignettes detailing the potential roles of SH7PCs and their products in symbiosis below.

Lichens are composite organisms arising from algae and/or cyanobacteria living among fungi in a symbiotic relationship. Lichens are tolerant of many different environmental conditions, such as high altitudes with high UV exposure. Cyanobacterial lichens are known to produce MAAs and four different cyanobacteria-containing lichen species belonging to the genus *Peltigera*, *Stereocaulon*, or *Lobaria* were analyzed for MAA production.[101] Despite being collected from a similar high altitude in the Himalayas, each species produced a unique suite of MAAs. Since free-living algae, cyanobacteria, and fungi make different MAAs and mycosporines, each symbiont could be contributing to the suite of MAAs produced. The production of MAAs may protect this delicate symbiotic community and contributes to the adaptability and survival of lichens under extreme environmental conditions.

Cleaning behaviors in fish serve as a model system for understanding cooperative relationships. In these relationships, cleaner fish benefit the client fish by removing harmful gnathiid isopods but also remove epidermal mucus, perhaps at the detriment of the client fish. The dynamics surrounding these relationships are complex and one aspect of interest is what determines food preference of mucus over gnathiids by cleaner fish.[125] The epidermal mucus of representative client fish contains varying amounts of MAAs, which may be sequestered to the mucus of cleaner fish. While the mucus load varies among client fish species, overall epidermal mucus has greater caloric value than gnathiid isopods and also contains more MAAs, providing the added bonus of a UV protective compound to cleaner fish which may contribute to the feeding preference of cleaner fish for mucus and the dynamics of this mutualistic relationship.

Bacteria from the genus *Burkholderia* and species from the *Rubiaceae* and *Primulaceae* plant families share a unique symbiotic relationship in which neither species can thrive on its own; the bacteria cannot survive independently but are critical for plant development and are vertically inherited. The secondary metabolite kirkamide, a C_7N-aminocyclitol with cytotoxic effects on aquatic antropods and insects, was isolated from the leaf nodules of the plant *Psychotria kirkii.*[96] Genomic evidence revealed that kirkamide is synthesized by *P. kirkii*'s bacterial symbiont, *Candidatus Burkholderia kirkii*, utilizing a biosynthetic gene cluster including an EEVS.[94] All but one of the eight leaf nodule symbionts considered contain a functional EEVS and more than half are capable of synthesizing kirkamide, indicating kirkamide has a beneficial defensive role, but is not solely responsible for the obligate nature of the symbiosis. However, the acquisition of the kirkamide biosynthetic cluster may have enabled the shift in lifestyle in bacterial leaf nodule symbionts from commensal to mutualistic.

Kleptoplasty is the process in which non-photosynthetic host cells engulf algae and utilize their chloroplasts, referred to as kleptoplasts, for energy production. The active lifespans of these kleptoplasts vary dramatically but the reasons why are not understood. On one end of the spectrum are Ross Sea dinoflagellates, which are capable of engulfing and maintaining active kleptoplasts for a remarkably long time.

A comparative study of Ross Sea dinoflagellates and free algae identified both differential profiles and localization of MAAs.[128] Based on the excitation spectra of chlorophyll a, it was suggested that MAAs in kleptoplasts may serve as a non-chlorophyll sensitizer, reducing the dependence of the organism on chlorophylls for light harvesting. In fact, Ross Sea dinoflagellates cells were found to have diminished Photosystem II activity. Therefore, the reduction in photosynthetic activity and reactive oxygen species production, along with supplementary light-harvesting by MAAs, was thought to be the source of the long active lifespan of kleptoplasts involved in this symbiotic relationship.

Conclusions and Perspectives

The first enzyme known to cyclize the pentose phosphate pathway intermediate sedoheptulose 7-phosphate to a cyclitol product was discovered less than two decades ago, and research on these enzymes continues to thrive. Today, three highly homologous proteins have been found to catalyze the cyclization of sedoheptulose 7-phosphate, but interestingly, they produce distinct catalytic products. Recent efforts in protein crystallography and structural studies are beginning to reveal how these enzymes use the same substrate but form different products. Genome sequencing and bioinformatics studies showed that SH7PCs are widely distributed throughout prokarya and eukarya, including pathogenic bacteria, plant symbionts, nitrogen-fixing bacteria, myxobacteria, cyanobacteria, fungi, stramenopiles, and animals. Acting as a bridge between a primary metabolic pathway (the Pentose Phosphate Pathway) and secondary metabolism, SH7PCs lead to the production of various bioactive natural products. DDGS appears to be involved only in the biosynthesis of a class of natural products, 4-deoxygadusol and its related mycosporines/MAAs, and the role(s) of these compounds in symbiosis and ecology has been established. EEVSs appear to be involved in the biosynthesis of more diverse bioactive natural products. While their wide distribution in various organisms was not realized until recently, preliminary evidence has emerged to suggest their function(s) for the producing organisms, including but not limited to defense mechanism, competition for food, mutualistic symbiosis, and UV protection. While

interest began with their involvement in the biosynthesis of pharmaceutically relevant natural products, their biochemistry, evolution, and expanding biological roles have continued to make SH7PCs an alluring subject to study.

Acknowledgements

This work was supported in part by grant GM112068 (to TM) from the National Institute of General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health (NIH).

Chapter 3

Structure of sedoheptulose 7-phosphate cyclase: ValA from *Streptomyces hygroscopicus*

Kelsey M. Kean, Sara J. Codding, Shumpei Asamizu, Taifo Mahmud and P. Andrew Karplus

Published in *Biochemistry, 2014, 53, pp 4250–4260*
Minor correction published in *Biochemistry, 2014, 53, pp 4316*

Abstract

Sedoheptulose 7-phosphate cyclases (SH7PCs) encompass three enzymes involved in producing the core cyclitol structures of pseudoglycosides and similar bioactive natural products. One such enzyme is ValA from *Streptomyces hygroscopicus* subsp. *jinggangensis* 5008 which makes 2-epi-5-epi-valiolone as part of the biosynthesis of the agricultural antifungal agent validamycin A. We present, as the first SH7PC structure, the 2.1 Å resolution crystal structure of ValA in complex with NAD^+ and Zn^{2+} cofactors. ValA has a fold and active site organization resembling the sugar phosphate cyclase dehydroquinate synthase (DHQS) and contains two notable, previously unrecognized interactions between NAD^+ and Asp side chains conserved in all sugar phosphate cyclases that may influence catalysis. Because the domains of ValA adopt a nearly closed conformation even though no sugar substrate is present, comparisons with a ligand-bound DHQS provide a model for aspects of substrate binding. One striking active site difference is a loop that adopts a distinct conformation as a result of an Asp → Asn change with respect to DHQS and alters the identity and orientation of a key Arg residue. This and other active site differences in ValA are mostly localized to areas where the ValA substrate differs from that of DHQS. Sequence comparisons with a second SH7PC making a product with distinct stereochemistry lead us to postulate that the product stereochemistry of a given SH7PC is not the result of events taking place during catalysis, but is accomplished by selective binding of either the α or β pyranose anomer of the substrate.

Introduction

Natural products have served as a major source of pharmaceuticals and bioactive molecules for centuries and continue to play key roles in guiding the development of new therapeutics today. Among these are pseudooligosaccharides[129], such as the antidiabetic drug acarbose, the crop protectant validamycin A, the antitumor agent cetoniacytone A, and the sunscreen mycosporin-like amino acids which have similar core cyclitol structures (Figure 3.1A). The core cyclitols of these molecules are generated from the pentose phosphate pathway intermediate

sedoheptulose 7-phosphate (SH7P) by one of three enzymes present in some bacteria and fungi that allow SH7P to be used in secondary metabolism. The enzymes – 2-epi-5-epi-valiolone synthase (EEVS), 2-epi-valiolone synthase (EVS), and desmethyl-4-deoxygadosol synthase (DDGS) – each catalyze the cyclization of SH7P to a distinct C7-cyclitol product (Figure 3.1B). These enzymes, the first of which was identified about 15 years ago,[15, 130] are known as SH7P cyclases (SH7PCs) and are a part of the sugar phosphate cyclase (SPC) family of enzymes, all of which require NAD^+ and a metal ion, either cobalt or zinc, as prosthetic groups.[11, 131-133]

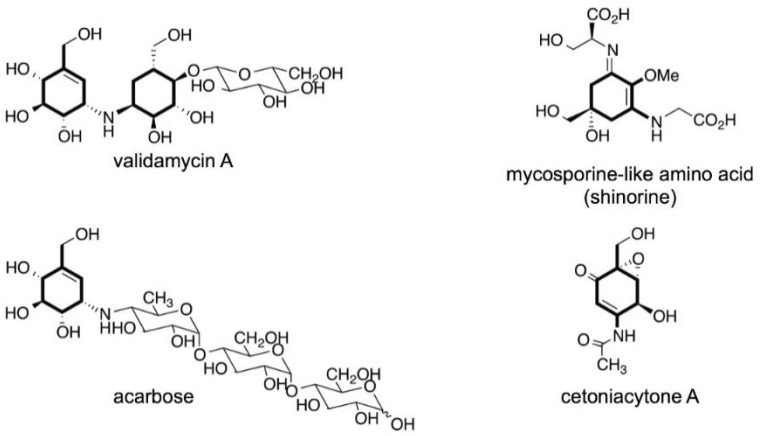
The SH7PCs are structurally uncharacterized and current understanding of their enzyme mechanisms are based mostly on studies of two other sugar phosphate cyclases: dehydroquinate synthase (DHQS) and 2-deoxy-*scyllo*-inosose synthase (DOIS). DHQS acts on the substrate 3-deoxy-D-arabinoheptulosonate 7-phosphate (DAHP; Figure 3.1B) as part of aromatic amino acid biosynthesis and has been well-studied because it is a target for antimicrobial drug development[131]. Numerous structures of DHQS have been solved from bacteria, fungi, and plants under a variety of conditions, including in the presence and absence of carbaphosphonate (CBP), a substrate analogue inhibitor, and in the presence or absence of NAD^+ [95, 131, 134-137]. DOIS acts on the substrate glucose 6-phosphate (G6P; Figure 3.1B) and is not as well studied, but structures of it have been solved both in the presence and absence of carbaglucose 6-phosphate (CG6P), a mechanism-based inhibitor[132].

Both DHQS and DOIS form homodimers. Each individual subunit is composed of two domains connected by a hinge, and in DHQS, the domains move closer together upon substrate binding. The two domains are a N-terminal NAD^+-binding α/β-sandwich and a C-terminal metal-binding α-helical domain. These same two domains are found in class III metal-dependent polyol dehydrogenases (polyol-DHs), that use $NAD(P)^+$ and a bivalent metal to carry out distinct but related chemistry to SPCs[138]. All together, these enzymes form what is known as the DHQS-like superfamily[139]. In SPC structure descriptions, the NAD^+-binding domain has been identified as a Rossmann fold[131-132, 140], but it has been noted that this assignment is not certain[138].

Extensive studies[131, 141] have established that in converting DAHP to dehydroquinate (DHQ), the DHQS active site coordinates the substrate at its active site metal via two hydroxyls (Figure 3.1B), and then catalyzes a remarkable five reactions: alcohol oxidation by NAD^+, phosphate β-elimination, carbonyl reduction by the earlier formed NADH, ring opening, and intramolecular aldol condensation. As is common among homologous enzyme pairs with distinct but related chemistry, the first step of the reaction in DHQS and the polyol-DH enzymes, NAD^+ promoted oxidation of an alcohol, is conserved[142]. Based on its crystal structures with and without substrate analogs, DOIS is proposed to have a similar mechanism to DHQS[132], and analogous mechanisms involving the same five steps have been proposed for EEVS, EVS, and DDGS[133]. Based on comparisons of SH7PC sequences with those of DHQS and DOIS, 14 putative active site residues were identified in EEVS, EVS, and DDGS sequences that were mostly identical, but showed characteristic variations in each of the three SH7PC types[133]. These patterns of variation, however, provided no insight into how the active sites produce distinct products from the same substrate, especially regarding the differing stereochemistry at the C2-position of the EEVS and EVS products (Figure 3.1B); this remains a major unanswered question. Developing a better understanding of the structure and function of SH7PCs will provide a foundation for their use in generating new bioactive compounds through synthetic biology and semisynthetic production.[143]

Here, we present a crystal structure of ValA, the EEVS from *Streptomyces hygroscopicus* 5008 that is involved in the biosynthetic pathway of the agricultural antifungal agent validamycin A.[15] This first structure of a SH7PC fortuitously includes tightly-bound Zn^{2+} and NAD^+ cofactors and provides an informative view of the residues lining the active site. We combine sequence comparisons with the various SH7PCs sequences and structural comparisons with DHQS and DOIS substrate analog complexes and develop an unexpected hypothesis for how these different SH7PCs can use the same substrate to generate different products.

(A)



(B)

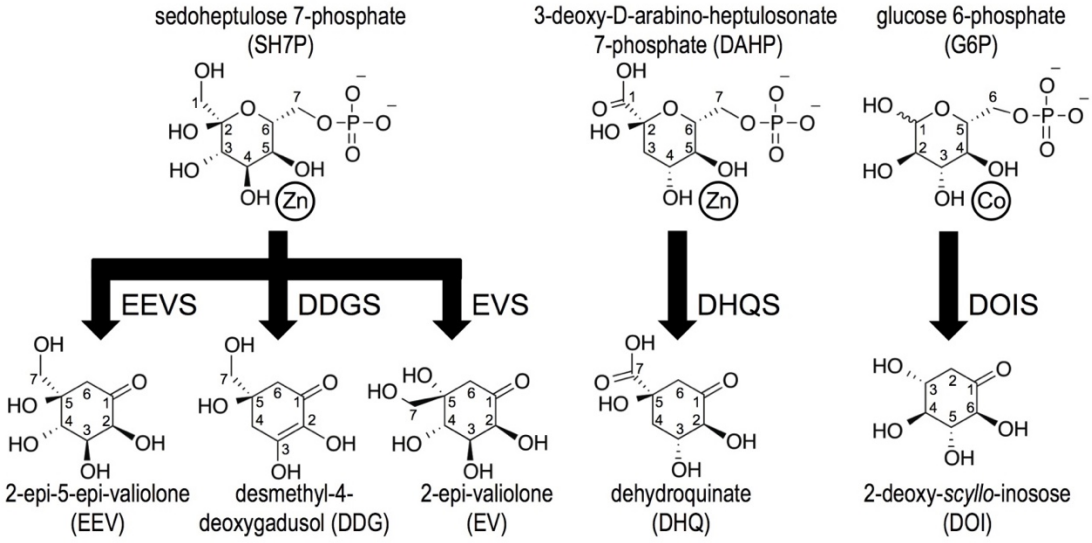


Figure 3.1. Reactions catalyzed by known sugar phosphate cyclases.
(A) Four cyclitol-containing natural products are shown and labeled by name, with their C7-cyclitol units made by SH7PCs highlighted in bold. **(B)** The substrates (above) and products (below) of five sugar phosphate cyclases are shown. A divalent metal cation is drawn next to the two substrate hydroxyls seen (for DHQS and DOIS) to coordinate it. SH7PCs and DHQS may utilize Zn^{2+} or Co^{2+} as the divalent metal cation[133, 144] but are shown here with Zn^{2+}, the metal present in this structure of ValA and in structures of *An*DHQS. DOIS exclusively uses Co^{2+} [145] and is depicted as such. Abbreviations for the substrates and products are introduced, with each enzyme abbreviation being that of its product followed by an additional S for "synthase." Due to resonance, DDG has an internal symmetry so the stereoconfiguration at C5 of the product after it is released into solution is not uniquely defined. We draw it here with the same stereochemistry as EEV anticipating the proposal we make in this work that the cyclization products of EEVS and DDGS have the same C5 configuration.

Materials and Methods

Expression, Purification, and Crystallization

The expression of recombinant ValA was done as previously described[133]. For purification, all done at 4 °C, cell pellets from 100 ml cultures were each resuspended in ~5 ml of 40 mM HEPES, 300 mM NaCl, pH 8.0 (buffer A) with 10 mM imidazole, sonicated (13 watts, 4 x 1 min), and centrifuged (14,500 rpm, 30 min). The supernatant was loaded onto a Ni-NTA resin column (5 ml resin; 0.8 ml/min). After washing with 100 ml buffer A with 20 mM imidazole, the protein was eluted using a 200 ml gradient from 20 to 500 mM imidazole in buffer A. Fractions (~6 ml each) containing protein were combined and dialyzed overnight against 2 L 10 mM Tris-HCl, 300 mM NaCl, 5 mM imidazole, pH 8.0. A second phase of purification was done similarly using a TALON column (~40 ml run at 0.3 ml/min) in buffer B (20 mM Tris-HCl, 300 mM NaCl, pH 8.0) with 5 mM imidazole for column equilibration, 10 mM imidazole for washing, and a 200 ml gradient from 10 mM to 200 mM imidazole for elution. Fractions (~4 ml each) containing pure ValA as judged by SDS gel electrophoresis were combined and dialyzed against 2 L 10 mM Tris-HCl, pH 7.5 (3 x 3 h). The protein solution was concentrated by ultrafiltration (10 K cutoff membrane) to 10 mg/mL, flash frozen in liquid nitrogen, and stored at -80 °C.

The enzyme was crystallized at 4 °C in hanging drops formed from 4 μL protein stock and 1 μL of a 0.6 M succinic acid, pH 6.5 reservoir solution. The resulting crystals were rod-like with dimensions of about 50 x 50 x 200 μm^3.

X-ray diffraction data collection

For diffraction data collection (at –170 °C), crystals were briefly passed through a solution containing 20% PEG 400 and then flash frozen in loops by plunging into liquid nitrogen. Data were collected from two crystals using $\lambda = 1.0$ Å X-rays and $\Delta\varphi = 1°$ steps at beamline 5.0.2 at the Advanced Light Source (Berkeley, CA). From both crystals, 120 2.0 s images were collected at detector distance d = 250 mm, and from the second, an additional 200 2.0 s images were collected at d = 350 mm. All these images were integrated using Mosflm[146] and merged using the CCP4

suite of programs[147-148] to obtain the data set used for structure solution and refinement. The merged data set was usable out to 2.1 Å using $CC_{1/2}$ ~0.2 as the cutoff criteria (Table 3.1), and a random 5% of reflections were marked for cross-validation. In addition, a third crystal was used for a fluorescence scan and to collect a data set at beamline 5.0.2 using $\lambda = 1.282$ Å X-rays to maximize the anomalous signal from the bound zinc. This data set included two sets of 60 $\Delta\varphi = 1°$, 4.0 s images offset by $\Delta\varphi = 90^{\circ}$ to collect the bijovet pairs, and yielded data useful to 3.5 Å resolution (data not shown).

Structure determination

The phase problem was initially solved by molecular replacement using MR-Rosetta with default settings[149]. As search models, we tried both Chain A of *V. cholerae* DHQS (PDB code 3OKF) and Chain A of *H. pylori* DHQS (PDB code 3CLH) which were the known structures that a BLAST search of the Protein Databank (PDB) showed as having the highest sequence similarity to ValA (33% and 29% identity, respectively). *V. cholerae* DHQS did not yield a solution, but *H. pylori* DHQS gave a result with R/R_{free} of 0.25/0.31 at 2.1 Å resolution and 327 residues built. The electron density map from this solution allowed us to build almost all the side chains, the active-site Zn^{2+}, and the NAD^{+} prosthetic group. In contrast, conventional molecular replacement approaches yielded models with R_{free} near 50% and maps that were very difficult to interpret (data not shown). All manual model building was done in Coot[150]. Refinements at various stages were carried out using Phenix[151] or Buster[152] with TLS refinement, with the final rounds being carried out using Phenix. Water molecules were manually placed based on typical criteria: electron density of $\geq 3\ \rho_{rms}$ in F_o-F_c maps and $\geq 0.8\ \rho_{rms}$ in the $2F_o$-F_c maps and a reasonable potential H-bond partner. Three regions at or near crystallographic two-fold axes were challenging to interpret. The first was a five-residue stretch, residues 46 to 50, that was near and crossing over a crystallographic two-fold axis that was a non-physiological crystal packing interaction. Into this weak helix-like main chain electron density we eventually modeled a portion, residues 48-50 and the side chain of Gln41, at 50% occupancy sharing the space with the same segments from its

symmetry mate. As the program would not ignore contact of the side chain of Gln41 with itself, this side chain position was not allowed to move in the final refinement calculations. We conclude that the segment of only one of the monomers is ordered at a time, and that in solution this part of the protein would be fairly dynamic. The second challenging region centered on a few residues N-terminal to residue 26. These residues had some positive density, but the electron density extended across the same crystallographic two-fold axis noted above and then weakened and we left this small section of density uninterpreted. The third challenging region was a β-hairpin turn (residues 32 to 33) located at the two-fold axis generating the expected physiological dimer interaction. This turn showed weak electron density while the associated β-strands showed strong and clear density. In order to follow the path of the electron density with a single conformation, we modeled the turn with an unfavorable *cis*-peptide bond before Lys32. This model also has a very short non-bonded collision (<2.5 Å) with its own symmetry mate, and so we suspect that it does not represent a true conformation, but only approximates the average chain path associated with a set of multiple conformations that allow reasonable packing at the interface. The final R/R_{free} were 0.179/0.262 with reasonable geometry (Table 3.1).

Structural comparisons and analyses

Secondary structure assignments were made using DSSP[153-154] and structure-based sequence alignments were generated using the Dali server[155].

Results and Discussion

Overall structure

The structure of recombinant ValA from *S. hygroscopicus* 5008 presented a challenging molecular replacement problem, as the most similar known structures were only ~30% identical in sequence. The structure determination was greatly facilitated by the MR-Rosetta algorithm[149], which yielded solutions of much higher quality than did conventional molecular replacement (see Materials and Methods). In addition to being aided by MR-Rosetta, the quality of the solution was also enhanced by the inclusion of weak high-resolution data that would have been discarded based

on conventional high-resolution cutoff criteria. For the data set used here, the conventional high-resolution cutoff criteria of R_{meas} ~ 60% or $<I/\sigma>$ of ~ 2 would lead to limits of 2.85 and 2.3 Å, respectively, whereas the more generous criteria of $CC_{1/2}$ ~ 0.2, shown in recent work to produce better refined models[156-158], leads to a limit of 2.1 Å (Table 3.1). To test how the inclusion of weak high-resolution data impacted the molecular replacement calculations, we carried out the MR-Rosetta runs using these three justifiable resolution cutoffs. Based on R_{free} values, using the 2.1 Å resolution cutoff yielded the best solution, with the 2.3 Å and 2.85 Å cutoffs being slightly worse and much worse, respectively (Table 3.2). This example thus shows that weak high-resolution data (out to $CC_{1/2}$ ~0.2 and $<I/\sigma>$ ~0.9 in this case) can help with challenging molecular replacement solutions as well as produce better refined models.

Further refinement of the molecular replacement solution yielded a model for the one chain in the asymmetric unit with a final R/R_{free} of 17.9%/26.2% to 2.1 Å resolution (Table 3.1). The large majority of the main chain as well as an active site NAD^+ and Zn^{2+} are well ordered with strong and clear density, and an absorption scan and anomalous difference map clearly confirm the presence and placement of the active site Zn^{2+} (Figure 3.2). The final structure includes 360 of the 414 expected residues, 188 waters, 1 PEG, 1 Zn^{2+}, and 1 NAD^+. The missing residues (1-25, 46-47, 58-62, 244-249, and 399-414) are not modeled due to weak or unclear electron density. Additionally, three sections, including the residues just N-terminal to residue 26, a β-hairpin turn at residues 32-33, and a weakly ordered helix at residues 46-50, laid on or near crystallographic two-fold axes and had weak, ill-formed density, making them challenging to model (see Materials and Methods). A crystallographic two-fold axis brings two ValA chains together to form a dimer that, according to the PISA server[159] buries 4220 $Å^2$ surface area (i.e. 2110 $Å^2$ per monomer). This dimer (Figure 3.3A) is equivalent to those observed for the homologous enzymes DHQS and DOIS[131-132, 136-137, 141, 160] and the dimer interface is well conserved, implying that it is the physiological form of ValA.

Table 3.1. Data collection and refinement statistics[a]

A. Data	
Resolution limits (Å)	66.9 - 2.10 (2.21-2.10)
Unique observations	20232 (2875)
Multiplicity	26.8 (19.3)
Completeness (%)	99.4 (98.9)
Average I/σ	11.2 (0.9)
R_{meas} (%)	38 (676)
$CC_{1/2}$ (%)	0.99 (0.22)
B. Refinement	
No. residues	360
No. solvent atoms	188
No. atoms total	3004
<B> protein ($Å^2$)	48
<B> solvent ($Å^2$)	56
R_{cryst} (%)	17.9 (28.1)
R_{free} (%)	26.2 (34.7)
rmsd bonds (Å)	0.010
rmsd angles (°)	1.28

[a] Numbers in parentheses correspond to values in the highest resolution bin.

Table 3.2. Resolution dependence of MR-Rosetta results

Resolution (Å)	**R**	**R_{free}**	**Residues Built**[a]
2.85	0.30	0.40	306 (221)
2.30	0.26	0.33	334 (318)
2.10	0.25	0.31	327 (298)

[a] Gives total number of residues built in backbone segments and, in parentheses, the number of these modeled as specific residues in the sequence of the target structure

Each chain of ValA encompasses the expected N-terminal NAD^+-binding domain and C-terminal metal-binding domain common to the DHQS-like superfamily. We describe here the domain topologies (Figure 3.3B) using a secondary structure nomenclature that takes into account which elements are conserved among the SPCs (Figure 3.4). The NAD^+-binding domain has a core 7-stranded β-sheet (with strand order 1-2-9-6-5-3-4) surrounded by five α helices, one β-hairpin (β7 and β8), and two short 3_{10} helices. The metal-binding domain is mainly α helical and includes eight α-helices, one 3_{10}-helix, and one β-hairpin. This domain not only contains the Zn^{2+} binding residues, but based on what has been seen in DHQS and DOIS, it also contains the majority of the residues involved in substrate recognition and so has also been called the substrate-binding domain[132]. However, the sugar-phosphate substrate actually binds in a cleft between the two domains and its recognition involves residues from both domains.

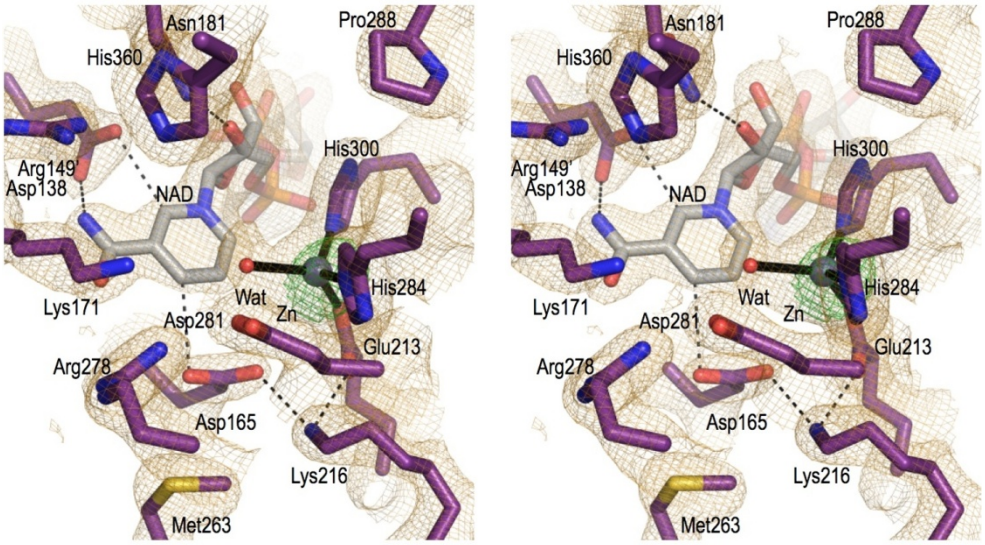
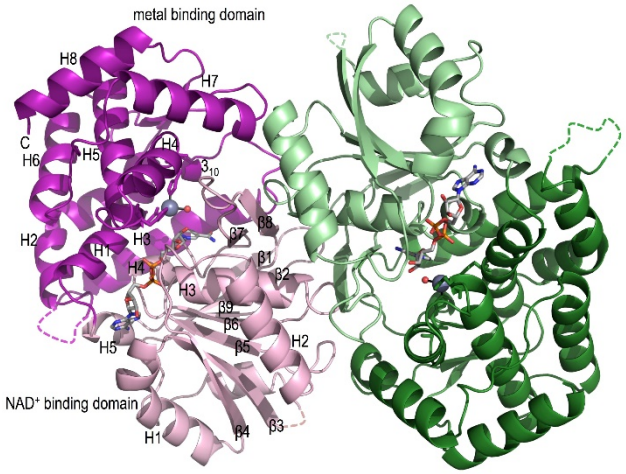


Figure 3.2. Electron density map quality and active site structure.
Stereoview of the ValA active site residues (purple carbons) and a water (red sphere) that are near the NAD^+ (grey carbons) and the Zn^{2+} (silver sphere) cofactors. Coordination bonds (black lines) and select H-bonds (black dashes) are shown along with the $2F_o$- F_c electron density (orange; contoured at 1 ρ_{rms}) and an anomalous difference map (green; contoured at 12 ρ_{rms}).

(A)



(B)

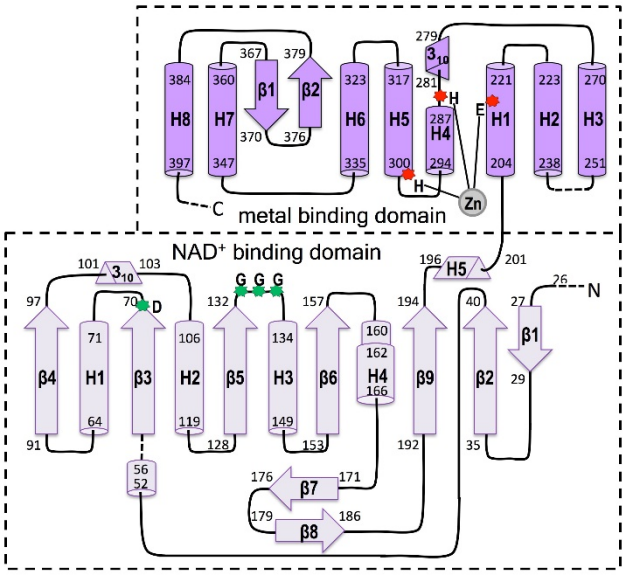


Figure 3.3. Overall structure and topology of ValA.
(A) Ribbon diagrams of the two chains of the ValA dimer are shown in purple and green tones, respectively, with the N-terminal NAD^+-binding domains in light hues and the C-terminal metal-binding domains in dark hues. Dashed lines indicate internal unmodeled backbone segments. The NAD^+ and the Zn^{2+} with its coordinating ligands are shown (colored as in Figure 3.2). Secondary structural elements in each domain of one monomer are labeled. **(B)** Topology diagram showing α-helices (cylinders), β-stands (arrows), 3_{10} helices (triangular prisms), and π-helices (wider cylinder) with their respective first and last residues given. The minimal length α- and 3_{10} helices (5 and 3 residues, respectively) are left out of the family secondary structure nomenclature. The domains are colored light and dark purple as indicated, and helices (H) and strands (β) common to the SPCs are named sequentially within each domain. Dashed lines denote unmodeled backbone segments. The three Zn^{2+} binding residues (red asterisks) and the glycine-rich turn and acidic residue (green asterisks) important for NAD^+-binding are indicated.

Relationships to other structurally known proteins

A structural similarity search performed using the DALI server[155] showed that ValA is most similar to assorted DHQSs (rmsd ~2.2 – 2.6 Å, Z-scores ~35-45), followed by DOIS (rmsd ~2.2 Å, Z-score 34) and then various polyol-DHs (rmsd ~2.7 – 3.6 Å, Z-scores ~23-28). Interestingly, although the structures used as search models in molecular replacement were the two PDB structures (PDB code 3OKF and PDB code 3CLH) most sequence similar to ValA, in terms of structural similarity these rank only eighth (Z-score = 41) and twenty-first (Z-score =36), respectively. This underscores why it can be useful to try all potential homologs in molecular replacement rather than just the most sequence similar ones[161]. The DALI search further showed that there are no known protein structures outside of the DHQS-like superfamily that share noteworthy structural similarity to ValA or to either of its individual domains.

Based on these results, representative enzymes were chosen for a structure-based sequence alignment (Figure 3.4): *Aspergillus nidulans* DHQS (*An*DHQS; PDB code 1DQS), the most well-studied DHQS[131]; *Bacillus circulans* DOIS (*Bc*DOIS; PDB code 2D2X), the only structurally known DOIS[132]; and *Bacillus stearothermophilus* glycerol dehydrogenase (*Bs*GlyDH; PDB code 1JQ5), the most structurally similar member of the polyol-DH family[138]. Representatives from the two structurally unknown types of the SH7PCs were also included in this sequence alignment: a DDGS from *Anabaena variabilis* (*Av*DDGS) and an EVS from *Actinosynnema mirum* (*Am*EVS). The structure-based alignment between ValA and DHQS is largely consistent with alignments that led to the previously proposed putative active site residues in ValA [11, 133]. The only change is that Lys356 in *An*DHQS had been previously aligned with Pro370 in ValA, but the structure-based alignment identifies the equivalent residue as His360.

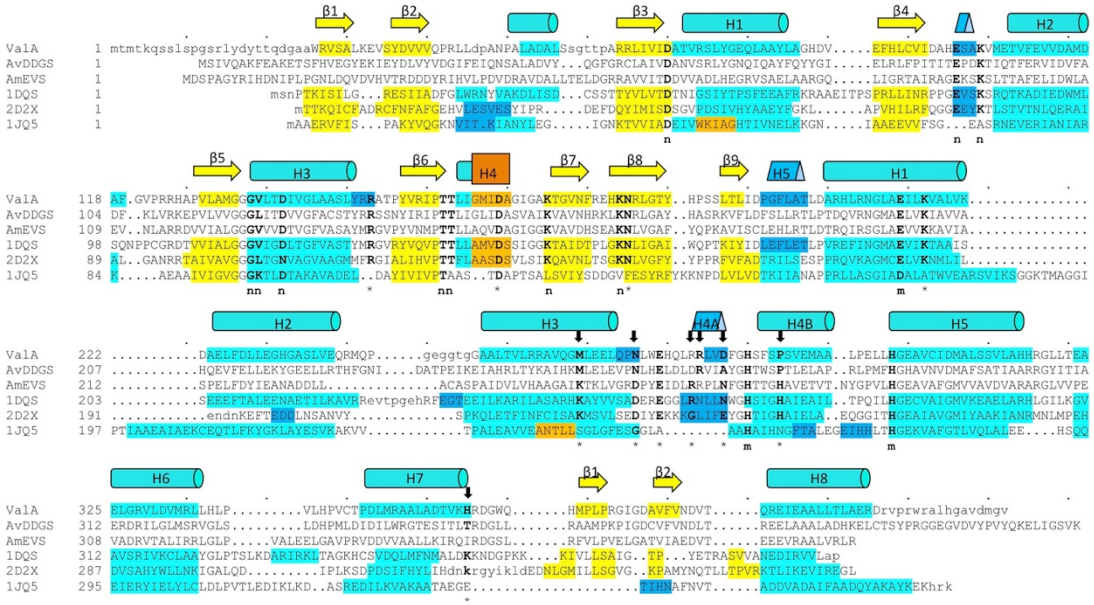


Figure 3.4. Sequence alignment of ValA with representative related enzymes.
The sequence of ValA is listed first and its secondary structure elements are schematically shown above the sequence. Other sequences in order are *Av*DDGS (*A. variabilis* DDGS; Ava_3858), *Am*EVS (*A. mirum* EVS; Amir_2000), 1DQS (*A. nidulans* DHQS), 2D2X (*B. circulans* DOIS), and 1JQ5 (*B. stearothermophilus* glycerol dehydrogenase). For the structurally known proteins, the residues in β-strands (yellow), α-helices (teal), 3_{10} helices (blue), and π helices (orange) are highlighted. Residues involved in metal-binding (m), NAD^+-binding (n), and substrate binding and/or catalysis (*) are indicated below the sequences, and active site residues with notable variation (↓) are indicated above the sequences.

Zinc and NAD^+ binding

Although zinc and NAD^+ were not added during sample preparation or crystallization, the electron density maps showed their unambiguous presence in the crystal structure (Figure 3.2), presumably meaning that they were bound by ValA already in the *E. coli* cytosol and carried along during the purification. The zinc and NAD^+ are both bound quite similarly to what has been described for DHQS[131] and DOIS[132] and so aside from some novel observations, we will here only briefly summarize the features of the binding. All residues in direct contact with zinc and NAD^+ are designated in Figure 3.4, and most are conserved among the SPCs. The Zn^{2+} ion is coordinated by Glu213, His284, and His300, all from the C-terminal metal-binding domain, and a water. As inferred from the liganded structures of DHQS and DOIS, this active site water will be displaced upon substrate binding.

The binding of NAD^+ includes characteristic residues conserved among the SPCs such as Asp70 at the end of strand β3 that H-bonds with the adenosine ribose O2' hydroxyl, the glycine-rich turn connecting β5 and H3 providing backbone amides that H-bond to the pyrophosphate oxygens, and Glu101, Lys104, Lys180, and Asn181 that H-bond with the nicotinamide ribose hydroxyls. The nicotinamide amide nitrogen donates H-bonds to the Asp138 side chain and the Lys171 backbone oxygen, and the oxygen forms H-bonds with surrounding waters in this structure. A fascinating pair of interactions that is conserved in known DHQS-like superfamily structures but has not been described before places carboxylate oxygens from Asp138 and Asp165 each roughly in the plane of the nicotinamide ring where they are respectively in position to accept weakly polar H-bonds from the nicotinamide C2 and C4 carbons (see Figure 3.2). We expect that these interactions will preferentially stabilize the positive charge on the oxidized form of the nicotinamide ring, which is distributed between the ring carbon atoms 2, 4, and 6 via resonance forms. These interactions may help explain the high affinity of these enzymes for NAD^+, and especially the interaction with the reactive nicotinamide C4 position may play a role in modulating the nicotinamide redox properties during the catalytic cycle.

The ValA sugar phosphate binding site

Although no substrate or substrate analog is bound in this crystal form of ValA, we can still gain insight into its substrate binding by comparisons with the ligand-bound structures of *An*DHQS and *Bc*DOIS structures (henceforth just referred to as DHQS and DOIS, respectively). For convenience, we refer to sequence differences between ValA and DHQS as mutations or changes with respect to DHQS, even though ValA did not evolve from a modern DHQS. As noted in the introduction, DHQS undergoes a conformational change from 'open' in the absence of a sugar-phosphate ligand to 'closed' upon binding the substrate analog CBP via a domain rotation of ~12° that brings the N and C-terminal domains closer together.[95] A recent structure of DHQS from *Actinidia chinensis* reinforces the relevance of the closed conformation seen, as the same closed conformation appears to be stabilized by the binding of inorganic phosphate and glycine in a way that mimics the substrate.[162]

DOIS, in contrast, was reported to not undergo such a domain closure based on comparisons of its structures with or without a substrate analog.[132]

A set of overlays of ValA with representative liganded and unliganded forms of DHQS and DOIS show that our ValA structure has a conformation between the open and closed DHQS forms but much closer to the closed form, varying by only ~2° (Figure 3.5). They further show that both the unliganded and liganded DOIS structures do indeed have minimal differences in their domain orientations, but that the DOIS conformation is ~10° more open in chain A and ~7.5° more open in chain B than the DHQS closed structure (data not shown), suggesting that it might not accurately represent the ligand bound structure. Also supporting this possibility is that the DOIS liganded structure was obtained by soaking crystals of the unliganded enzyme with inhibitor, during which some crystal cracking was observed.[132] This implies that the enzyme could not undergo a complete domain closure without compromising the integrity of the crystal. For this reason, we focus in the following comparisons solely on the DHQS·CBP complex, which by all evidence accurately represents a true inhibitor-bound conformation.

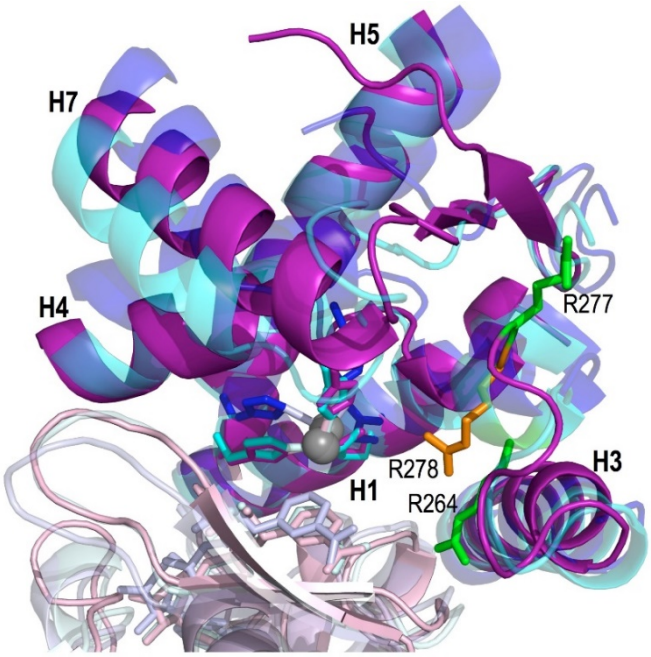


Figure 3.5. ValA overlay with closed and open DHQS structures.
Shown are ribbon diagrams of an unliganded, open DHQS (blue; PDB code 1NRX), a CBP-bound, closed DHQS (cyan; PDB code 1DQS) and ValA (purple), all overlaid based on their NAD^+-binding domains. Pale hues are used for the NAD^+-binding domains and NAD^+, and dark hues are used for the metal-binding domains. For clarity, only select secondary structure elements of the metal binding domain are shown along with the three zinc ligating residues (sticks) and the zinc (grey sphere). The active site side chain that does not align well between DHQS (Arg264) and ValA (Arg277) is shown for both structures in green, and the alternate equivalent ValA residue (Arg278) is shown in orange. That other 'open' DHQS structures, such as those of *Tt*DHQS (PDB code 1UJN) and *Hp*DHQS (PDB code 3CLH), are up to 5° different in domain orientation compared with *An*DHQS (PDB code 1NRX) does not alter the conclusions of this analysis.

With only this unliganded structure of ValA available, we cannot make any claims about what ligand-induced conformational changes may occur. However, the similarity of ValA to the closed conformation of DHQS is fortunate as it means that the NAD^+, the zinc, and nearly all of the ValA residues equivalent to DHQS active site residues align rather well (Figure 3.6), giving us confidence that this comparison provides an informative picture of which ValA residues will play a role in substrate binding. Of the ligand-binding residues in the DHQS complex, only one, Arg264, is not in close proximity to its corresponding residue in ValA. The equivalent residue in ValA is Arg277 (Figure 3.4), and it points in the opposite direction (see green side chains in Figure 3.5). Interestingly, Arg277 is not conserved among EEVSs (data not shown), suggesting it is not a key residue for this enzyme. Even more interestingly,

due to a different nearby loop conformation, the following residue in ValA, Arg278, has its side chain close to that of DHQS Arg264 (Figure 3.6) and is conserved among EEVSs, suggesting that it may be the functionally equivalent residue. An important question then becomes if the different loop conformation is a robust difference between ValA and DHQS, or if it may be simply due to the ValA structure not having a ligand bound.

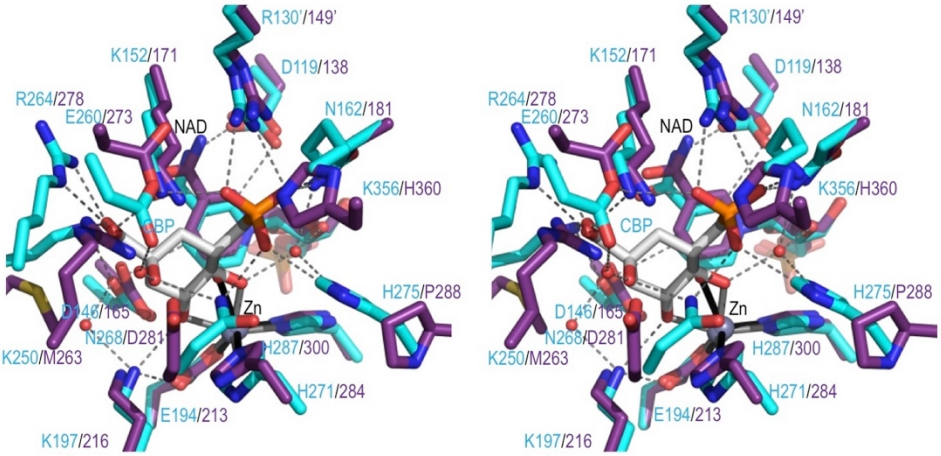


Figure 3.6. Comparing the ValA active site region with the DHQS·CBP complex. Stereoview of select active site residues in ValA (purple) overlaid on the DHQS (cyan) in complex with CBP (white) shown in roughly the same orientation as DAHP is drawn in Figure 3.1. H-bonding interactions in the DHQS active site (dashed lines) and coordination bonds with Zn^{2+} (solid lines) are shown. A prime on a residue number means it is from the other subunit of the dimer.

A closer look at the loop (residues 257-264 in DHQS and 270-278 in ValA) identifies another key active site position and confirms that the difference in loops is robust (Figure 3.7). In both unliganded and liganded DHQS structures, the loop wraps around the side chain of Asp257 which accepts multiple backbone amide H-bonds to stabilize the conformation. The equivalent residue in ValA, Asn270, is not compatible with the DHQS loop conformation, but plays an equally central role in stabilizing the alternate less-compact loop path (Figure 3.7).

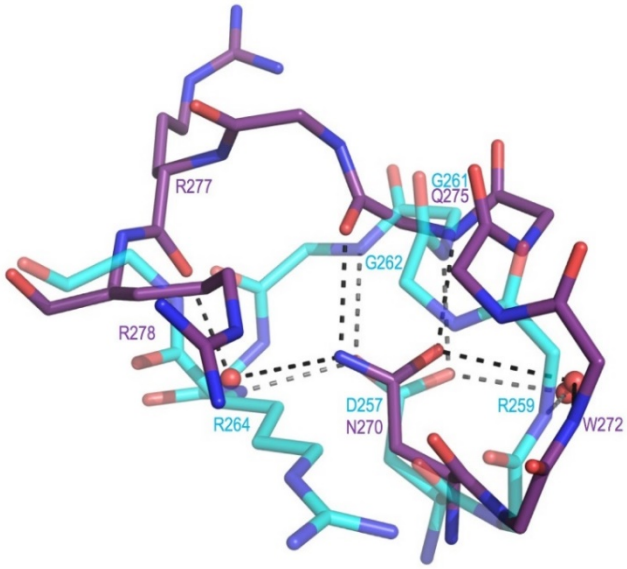


Figure 3.7. Active site loop difference relates to presence of Asp257 in DHQS *vs.* Asn270 in ValA.

Shown are residues 269-278 of ValA (purple) and 256-264 of DHQS (cyan; PDB code 1DQS) after the proteins have been overlaid as in Figure 3.5. H-bonding interactions (dashed lines) involving the loop residues and associated waters are shown. In DHQS, the Asp257 carboxylate receives H-bonds directly or indirectly (via water) from four backbone nitrogens (from Arg259, Gly261, Gly262, and Arg264). In ValA, the Asn270 side chain amide directly or indirectly makes H-bonds with two backbone nitrogens (from Trp272 and Glu275) and two backbone oxygens (from Gln275 and Arg277).

Taking ValA Arg278 as the equivalent of DHQS Arg264, Figure 3.6 compares the DHQS residues surrounding the substrate analog CBP with their ValA equivalents. Among these, just four ValA residues are different types: Met263 replaces a Lys, Asp281 replaces an Asn, Pro288 replaces a His, and His360 replaces a Lys. Using atom numbering for the substrate (see Figure 3.1) rather than the CBP inhibitor, the essential features of CBP binding to DHQS (clockwise from the top of Figure 3.6) are a phosphate binding pocket (at two o'clock), bidentate metal coordination by the C5 and C4 hydroxyls (at three to six o'clock) which also serves to point the C5-hydrogen at the nicotinamide C4 atom in good geometry for hydride transfer, and then a pocket for the C2 hydroxyl and carboxylate groups (around ten o'clock). In ValA, the metal and nicotinamide are nearly identically positioned as are key residues interacting with the phosphate (Lys171, Arg149', and Asn181) and the metal-coordinating hydroxyls (Asp165 and Lys216). We take this to mean that the analogous parts of the ValA substrate will be bound similarly to CBP in DHQS.

Having a high spatial conservation of these parts of the substrate makes sense, as they are where most of the chemistry takes place.

In contrast, the significantly shifted or mutated residues (Arg278, Met263 and Asp281) are present at the pocket around the C2 hydroxyl and carboxylate groups where the ValA substrate has different substituents. In particular, the Lys→Met change makes sense with the absence of the substrate carboxylate. The two remaining changes, involving ValA residues Pro288 and His360, create more space around the phosphate group, but we do not understand why that might be. The DHQS residue replaced by Pro288 is His275, which has been proposed to serve as an acid/base during catalysis[131].

Based on the above, we conclude that ValA will bind its substrate, SH7P, with the phosphate group and the C5 and C4 metal-coordinating hydroxyls in similar positions to those in the DHQS·CBP complex. However, other aspects of the binding mode such as ring conformation and/or orientation must differ from CBP, especially because the configuration of the C4 hydroxyl in SH7P differs (see Figure 3.1B) such that it and the C5 hydroxyl cannot simultaneously be equatorial. This difference makes predicting details of the binding mode of SH7P more challenging.

Variations among the SH7PCs and a proposal for how they catalyze different reactions

As noted in the introduction, a major open question about SH7PC enzymes is how they bind the same substrate and produce different products (Figure 3.1B), with the most conceptually confusing aspect being how EEVS produces one stereochemistry at the C5 position of the product (derived from the substrate C2 atom; as shown in Figure 3.1B) while EVS produces the other.[133] Because of an internal symmetry in the DDGS product, it could be produced with either C5 stereochemistry (Figure 3.1B). Interestingly, EEVS and DDGS are more sequence similar to each other than they are to EVS, with both being reported to vary from DHQS in the identities of four putative active site residues while EVS varied in only one[11, 133] (Figure 3.8). Our structural results on ValA (an EEVS) strengthen this pattern, in that the Asn residue associated with its distinct loop conformation and the

alternate active site Arg residue (Figure 3.7) are also both present in DDGSs (Figure 3.4). In contrast, EVS, like DHQS, has an Asp in the loop and conserves the first Arg (Figure 3.4). This implies that these two subsets of the SH7PCs (EEVS and DDGS vs. EVS) have distinct binding environments for the substituents of the substrate C2 atom, and leads us to hypothesize that the discrimination related to the stereochemistry at C5 of the product actually occurs upon substrate binding rather than during catalysis.

Specifically, each enzyme would selectively bind either the α- or β-pyranose form of SH7P and the rapid interconversion between the various pyranose and furanose forms would allow these normally less-populated SH7P forms (~16% and <1%, respectively[163]) to accumulate in the enzyme active sites (Figure 3.9A). In this way, rather than the aldol acceptor geometry variation between EEVS and EVS requiring a 180° rotation of the polar C5 substituents in the middle of the catalytic cycle, each enzyme would start with a substrate preorganized to generate the correct configuration at that position (Figure 3.9B). As EEV has the configuration at C5 that would be derived from α-pyranose SH7P, this is the form of substrate that EEVS and DDGS would bind, and as EV has the same configuration at C5 as would be derived from β-pyranose SH7P, the EVS active site would select for that form of the substrate. This hypothesis is completely consistent with the behavior of DHQS and DOIS, which both bind their substrate analogs with a fixed configuration corresponding to that of their product. In the case of DOIS, the enzyme's selectivity is directly observed in its preferential binding from a racemic mixture of the inhibitor CG6P only the form that mimics the β-anomer of substrate.[132, 164]

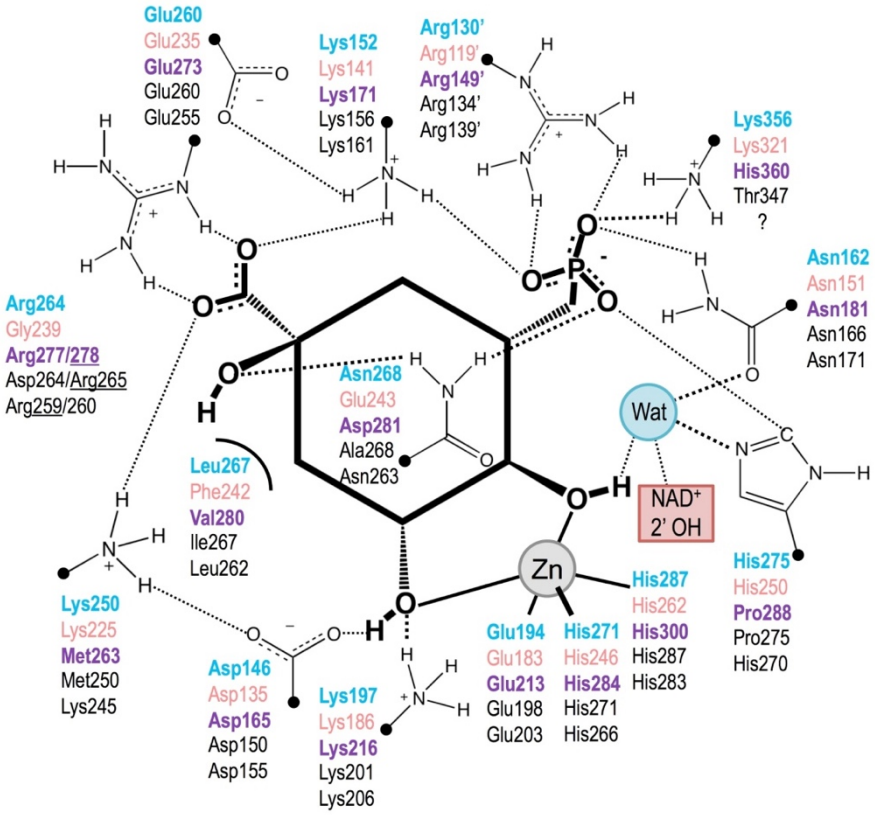


Figure 3.8. Variation in active site residues among sugar phosphate cyclases.
Schematic drawing of residues lining the substrate binding pocket in the DHQS·CBP complex shown in roughly the same orientation as in Figure 3.6. Each DHQS residue shown is labeled (cyan) and under that label are listed the corresponding residues found in the structures of DOIS (pink) and EEVS (purple), or based on the alignment in Figure 3.4, residues expected to be equivalent in DDGS and EVS (black). The CBP ligand is bold. H-bonding interactions (dashed lines) and coordination bonds with Zn^{2+} (solid lines) are shown. Residue numbering corresponds to the representative proteins used in Figure 3.4 (*An*DHQS, *Bc*DOIS, ValA, *Av*DDGS, *Am*EVS).

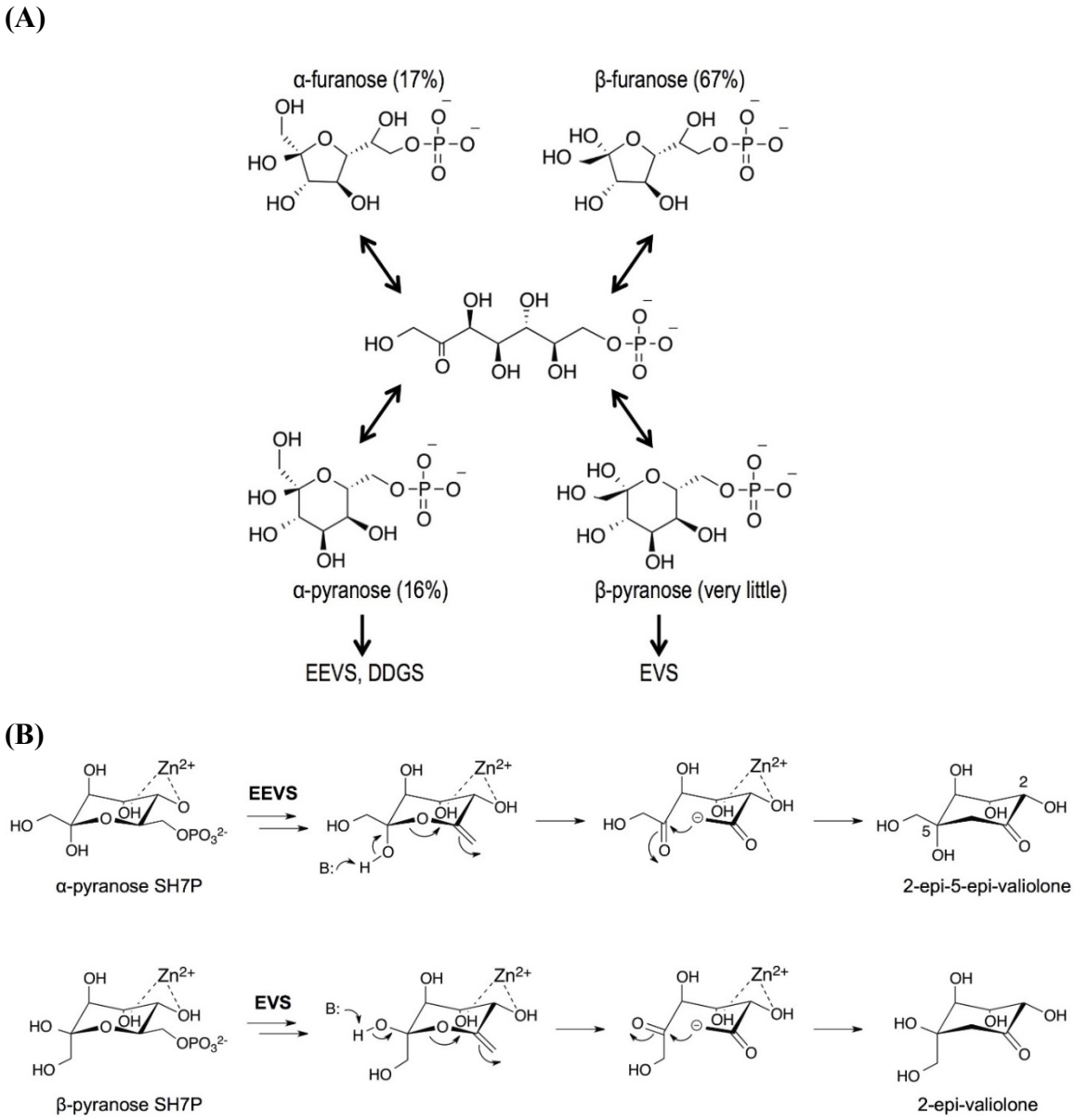


Figure 3.9. Proposed enzyme-specific selection of forms of sedoheptulose 7-phosphate. **(A**) The α and β pyranose and furanose anomers of SH7P with their relative abundance as determined by NMR[163] are shown along with the linear form via which they interconvert. Also shown is our proposal that EEVS and DDGS bind the α-pyranose anomer while EVS binds the β-pyranose anomer. **(B)** The ring opening and intramolecular aldol condensation steps of the proposed reaction mechanisms of EEVS and EVS are shown, emphasizing how the α- and β-pyranose anomers of SH7P are preorganized for the generation of the respective stereochemistries at C5 in the products. "B:" represents an active site base that may aid ring opening.

In this way, rather than viewing the SH7PCs as a family of enzymes which use one substrate to specifically give two stereochemically distinct types of products, one can then view them as enzymes that bind distinct substrates, either α-pyranose or β-pyranose SH7P, to give products with stereochemistries matching those substrates. In terms of the evolutionary origin of the SH7PCs, since EVS sequences are more similar to those of DHQSs,[133] we suggest that EVS evolved from DHQS with only slight changes to the active site being needed to allow binding of the β-pyranose anomer of SH7P. In contrast, the key differentiating step in EEVS evolution would have been the Asp→Asn mutation that changed which Arg pointed into the active site pocket, as this allowed binding of the α-pyranose anomer of SH7P leading to the production of EEV. Then further mutations, including two active site mutations of Asp →Ala and His→Thr (Figures 3.4 and 3.8), gave rise to the mechanistic difference that characterizes the DDGS enzymes. Although we do not understand this transition yet, the more minor nature of differences between the DDGSs and the EEVSs is emphasized by phylogenetic trees showing that EEVS and DDGS are more closely related to each other than to the other SPCs.[133]

Although much remains to be learned about ValA and the SH7PCs in general, the ValA structure presented here sheds much light on this enzyme family. Especially seeing the spatial orientation of active site residues (Figure 3.8) has provided insight into how variation in the active site pocket allows for the different specificities found in the SPC superfamily. Nevertheless, the detailed roles of catalytic residues in the EEVS, DDGS and EVS mechanisms still remain open questions. For example, the residue which takes the catalytic role of an essential histidine in DHQS (His275) and DOIS (His250) has not been identified in these enzymes. We expect that answering such questions for the SH7PCs will best be approached through kinetics and structural studies carried out with stereospecific carbacyclic phosphonate analogs of SH7P that do not yet exist, but that are in concept similar to the DAHP analogs used in informative studies of DHQS. [95, 131, 165-166] We are now initiating studies in this direction.

Coordinates

The coordinates and the structure factors have been deposited in the Protein Data Bank (www.rcsb.org/pdb/) as entry 4P53.

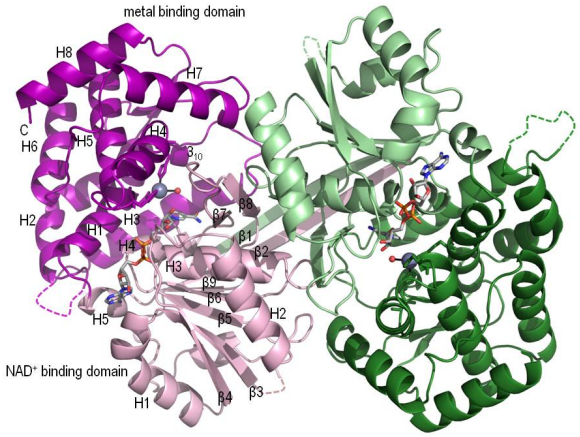
Acknowledgements

We would like to thank Dale Tronrud and Camden Driggers for help with the crystallographic work, and Murugesh Padmanarayan for growing and analyzing the first crystals.

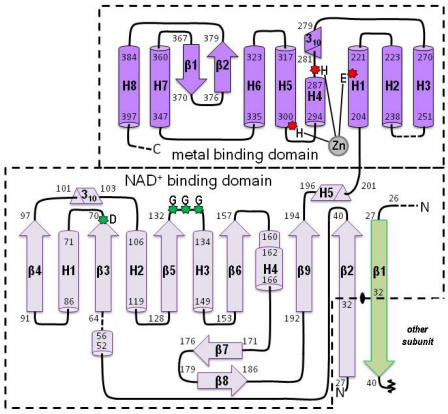
Addition/Correction to Structure of a Sedoheptulose 7-Phosphate Cyclase: ValA from *Streptomyces hygroscopicus*

Shortly after this article describing the crystal structure of ValA from *Streptomyces hygroscopicus* was accepted for publication, we were able to solve the structure of the related enzyme desmethyl-4-deoxygadusol synthase from *Anabaena variabilis* (*Av*DDGS) at a higher (~1.7 Å) resolution (work still in progress). In light of the new structure, we are now able to satisfactorily interpret what we had described as a difficult to fit β-hairpin turn at residues 32 and 33 of ValA which collided with its symmetry mate across the crystallographic 2-fold axis. Now we are able to say that rather than these residues forming a β-hairpin to match the chain topology seen in dehydroquionate synthase structures, a domain swapped arrangement exists in which the residues N-terminal to 33 continue in a linear direction, making an extended β-strand that participates in the core β-sheet in the other subunit of the dimer. We have updated the PDB deposition to reflect this altered topology and provide here an updated version of our original Figure 3.3 (Revised Figure 3.3) that provides a corrected overview of the topology of ValA. Further details of this new topology will be reported in a future publication of the *Av*DDGS structure.

(A)



(B)



Revised Figure 3.3. Overall structure and topology of ValA.
(A) Ribbon diagrams of the two chains of the ValA dimer are shown in purple and green tones, respectively, with the N-terminal NAD^+-binding domains in light hues and the C-terminal metal-binding domains in dark hues. The extended β-strands of each subunit involved the domain swapped arrangement are visible in the back of the dimer. Dashed lines indicate internal unmodeled backbone segments. The NAD^+ and the Zn^{2+} with its coordinating ligands are shown (colored as in Figure 3.2). Secondary structural elements in each domain of one monomer are labeled. **(B)** Topology diagram showing α-helices (cylinders), β-stands (arrows), 3_{10} helices (triangular prisms), and π-helices (wider cylinder) with their respective first and last residues given. The minimal length α- and 3_{10} helices (5 and 3 residues, respectively) are left out of the family secondary structure nomenclature. The domains are colored light and dark purple as indicated, and helices (H) and strands (β) common to the SPCs are named sequentially within each domain. The domain-swapped β-strand containing residues 27-32 from the other subunit of the dimer, but contributing to the purple domain, is light green. The crystallographic two-fold rotation axis (indicated as ⬮) relates this β-strand to the residues extending from residue 32 of the purple domain to be part of the β-sheet of the other subunit. We retain the β1 and β2 names for the two parts of the

long N-teminal β-strand both because they participate in different β-sheets and so as to maintain in this report a consensus secondary structure nomenclature relevant to the sugar phosphate cyclase superfamily. Dashed lines denote unmodeled backbone segments. The three Zn^{2+} binding residues (red asterisks) and the glycine-rich turn and acidic residue (green asterisks) important for NAD^{+}-binding are indicated.

Chapter 4

Evolution and Distribution of C7-Cyclitol Synthases in Prokaryotes and Eukaryotes

Andrew R. Osborn, Kelsey M. Kean, Khaled M. Alseud, Khaled M. Almabruk, Shumpei Asamizu, Janet A. Lee, P. Andrew Karplus, and Taifo Mahmud

Published in *ACS Chemical Biology,* 2017, 12, pp 979–988

Abstract

2-epi-5-epi-Valiolone synthase (EEVS), a C_7-sugar phosphate cyclase (SPC) homologous to 3-dehydroquinate synthase (DHQS), was discovered during studies of the biosynthesis of the C_7N-aminocyclitol family of natural products. EEVS was originally thought to be present only in certain actinomycetes, but analyses of genome sequences showed that it is broadly distributed in both prokaryotes and eukaryotes, including vertebrates. Another SPC, desmethyl-4-deoxygadusol synthase (DDGS), was later discovered as involved in the biosynthesis of mycosporine-like amino acid sunscreen compounds. Current database annotations are quite unreliable, with many EEVSs reported as DHQS, and most DDGSs reported as EEVS, DHQS, or simply hypothetical proteins. Here, we identify sequence features useful for distinguishing these enzymes, report a crystal structure of a representative DDGS showing the high similarity of the EEVS and DDGS enzymes, identify notable active site differences, and demonstrate the importance of two of these active site residues for catalysis by point mutations. Further, we functionally characterized two representatives of a distinct clade equidistant from known EEVS and known DDGS groups, and show them to be authentic EEVSs. Moreover, we document and discuss the distribution of genes that encode EEVS and DDGS in various prokaryotes and eukaryotes, including pathogenic bacteria, plant symbionts, nitrogen-fixing bacteria, myxobacteria, cyanobacteria, fungi, stramenopiles, and animals, suggesting their broad potential biological roles in nature.

Introduction

2-epi-5-epi-Valiolone synthase (EEVS) is a member of the sugar phosphate cyclases (SPCs), a group of homologous enzymes that catalyze the cyclization of sugar phosphates to cyclic compounds in primary and secondary metabolism.[11, 167] Members of the SPC family (Figure 4.1) share significant sequence and structural similarity with 3-dehydroquinate synthase (DHQS).[17, 168] DHQS converts 3-deoxy-D-arabinoheptulosonate 7-phosphate (DAHP) to 3-dehydroquinate (DHQ), the first committed step in the shikimate pathway, which leads to aromatic amino acids, folates, ubiquinones, and many secondary metabolites. Other members of the SPC

family include aminodehydroquinate synthase (aDHQS, a variant of DHQS),[169] 2-deoxy-*scyllo*-inosose synthase (DOIS),[170] desmethyl-4-deoxygadusol synthase (DDGS)[11, 13] and 2-*epi*-valiolone synthase (EVS).[12, 80] aDHQS converts amino-DAHP to amino-DHQ, the precursor of 3-amino-5-hydroxybenzoic acid, which is involved in the biosynthesis of important polyketide antibiotics, such as rifamycin, geldanamycin, mitomycin, and ansamitocin.[171-174] DOIS catalyzes the conversion of glucose 6-phosphate to 2-deoxy-*scyllo*-inosose, which is the precursor of deoxystreptamine-containing aminoglycoside antibiotics, *e.g.,* butirosin, neomycin, kanamycin, and tobramycin.[110] Finally, EEVS, DDGS, and EVS use sedoheptulose 7-phosphate (SH7P; a pentose phosphate pathway intermediate) as substrate to give 2-epi-5-epi-valiolone, 2-desmethyl-4-deoxygadusol, and 2-epi-valiolone, respectively. 2-epi-5-epi-Valiolone and 2-epi-valiolone are precursors of aminocyclitol natural products, such as the antidiabetic drug acarbose and the antifungal agent validamycin A,[80, 175-176] whereas desmethyl-4-deoxygadusol is the precursor of the mycosporine-like amino acid sunscreen compounds.[13] Of particular interest is that EEVS is also involved in the formation of gadusol, another sunscreen-like compound found in fish, and possibly in other vertebrates, e.g., amphibians, reptiles, and birds, but not mammals.[177]

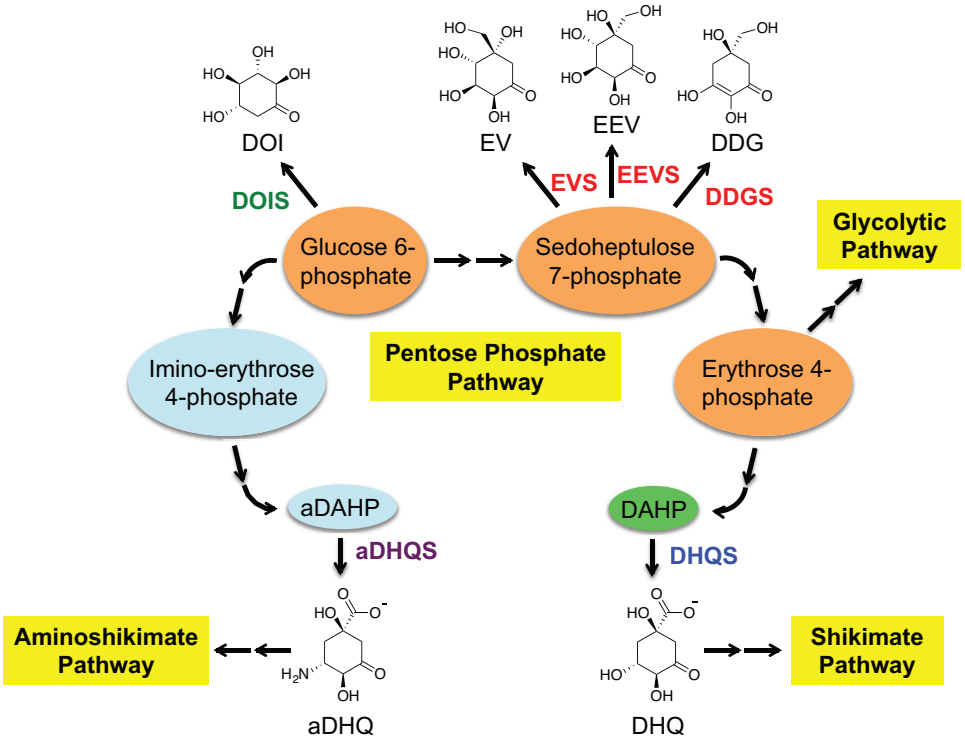


Figure 4.1. Six sugar phosphate cyclase family members, their substrates and products, and their connections to the Pentose Phosphate Pathway.
DOIS, 2-deoxy-*scyllo*-inosose synthase; EVS, 2-epi-valiolone synthase; EEVS, 2-epi-5-epi-valiolone synthase; DDGS, desmethyl-4-deoxygadusol synthase; aDHQS, aminodehydroquinate synthase; DHQS, 3-dehydroquinate synthase.

For catalysis, all SPC superfamily members require NAD^+ and a metal ion, either Zn^{2+} or Co^{2+}, as prosthetic groups. Among them, DHQS has been particularly well-studied because of its involvement in primary metabolism; thus, it is a potential target for antibacterial drug development. The primary amino acid sequences of EEVSs and DDGSs are highly similar to each other, and to some extent to those of DHQSs,[11-12] and we have noticed that these enzymes are often misannotated in genome databases, particularly involving the misassignment of DDGS as EEVS or DHQS. This inconsistent/inaccurate functional assignment of these enzymes has hampered correct prediction of their roles in nature.

This has prompted us to evaluate the genes and their encoded protein sequences with the goal of establishing parameters for more accurately annotating EEVS, DDGS, and other SPCs. Here, we describe a bioinformatics study and an

approach to more accurately assign putative functions to EEVS and DDGS encoding genes in databases. We also report a crystal structure of a representative DDGS, Ava_3858 from *Anabaena variabilis* (*Av*DDGS),[12] compare it with those of EEVS and DHQS, identify notable active site differences, and demonstrate the importance of two of these active site residues for catalysis.[17, 131] In addition, we confirm through biochemical experiments the function of two putative EEVS proteins that form a separate clade from the previously characterized EEVSs, and demonstrate the broad distribution of EEVS and DDGS in prokaryotes and eukaryotes.

Results and Discussion

Bioinformatics analysis and reassignment of EEVS and DDGS

During the past decade, there has been a significant upsurge in deposition of genes annotated as EEVS or DHQS-like proteins in public databases, and as noted in the introduction, many of them have been incorrectly annotated. To address this in a comprehensive manner, we have gathered for evaluation all genes that had e-values < 7e–68 in a BLAST search of the NCBI database using the known EEVS from the validamycin A pathway (ValA or *Sh*EEVS) and the DDGS from the shinorine pathway (Npun_R5600) as queries. The 630 amino acid sequences were aligned using MUSCLE and the phylogenetic tree (Figure 4.2) was constructed using FastTree, employing the JTT model of protein evolution and the CAT approximation. From the 630 proteins studied, we preliminarily concluded based on the phylogenetic tree (Figure 4.2) that 335 of them are EEVS and 295 are DDGS. As expected, ValA groups together with other known EEVS proteins: AcbC, CetA, PyrA, BE-Orf9, and SalB.[10-11, 88, 178-179] About 30% of the EEVS proteins had been annotated as DHQS or hypothetical proteins (Table S1). All (i.e. 100%) of the DDGS proteins had been misannotated, most commonly as DHQS, EEVS, or hypothetical proteins (Table S2; the new annotations are also available on our website at http://people.oregonstate.edu/~mahmudt/?page_id=396).

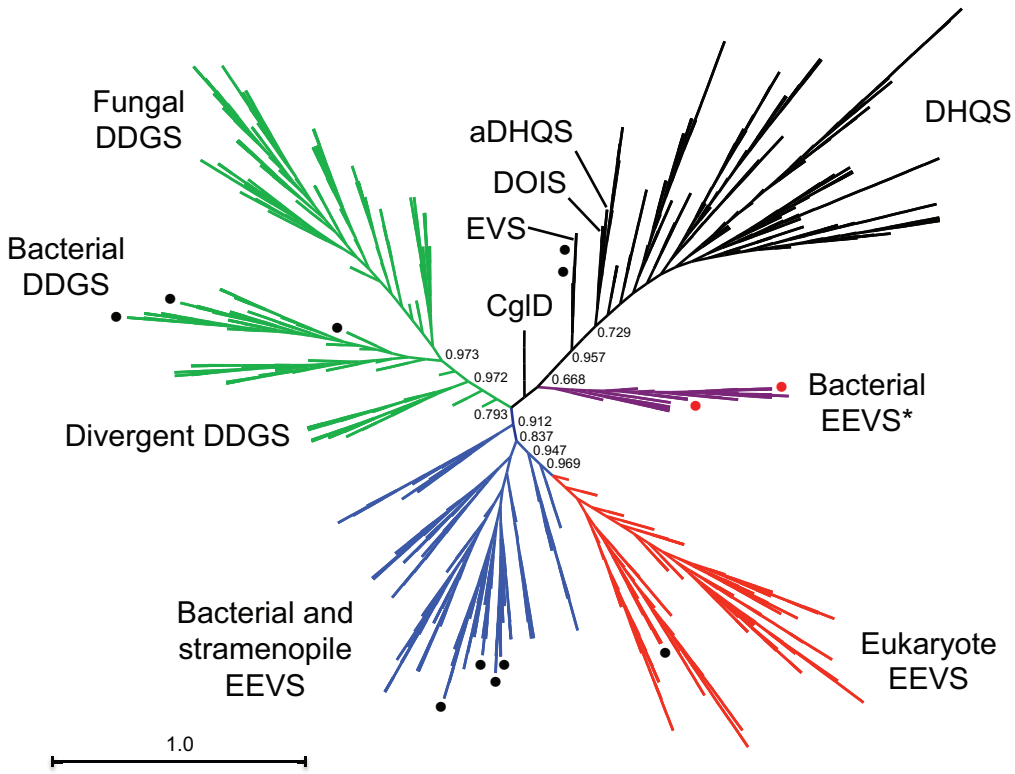


Figure 4.2. Radial cladogram of the sugar phosphate cyclase superfamily.
A group of putative bacterial EEVS (drawn in purple and designated "EEVS*") occurs as a separate clade from the known EEVS clades (drawn in blue and red). Numbers show local support values. Black circles show sedoheptulose 7-phosphate cyclases (EEVS, EVS, and DDGS) that have been biochemically characterized. Red circles identify the two bacterial EEVS* (GacC and Staur_1386) that are characterized in this study.

Conserved motifs for EEVS and DDGS

Through detailed comparative bioinformatics analysis of these preliminary groupings, we identified two stretches of residues that appeared most useful for distinguishing the enzymes from each other. In these segments, the conserved amino acid sequences for EEVS were MLEELxPNLxE and xxRxxDxGH, which were recognizably different from the DHQS sequences, and also adequately distinguishable from those of DDGS [MLELExPNLHE and LDRVIAxGH] (Figure 4.3). Specifically, in the first conserved regions, the EEVS proteins contain an MLEEL motif, whereas the DDGS proteins contain a MLELE motif. In the second conserved regions, the EEVS proteins contain an Asp, in place of an Ala in the DDGS proteins. Some proteins have slight variations from these conserved motifs, for example KL instead of EL for EEVS. However, the overall conserved motif can still sufficiently

distinguish EEVS and DDGS enzymes despite the sequence variations.

Independent analysis and comparison of crystal structures of *Sh*EEVS and *Av*DDGS also identified these characteristic residue differences as a means to differentiate EEVS and DDGS. Leu267 in *Sh*EEVS *versus* Glu254 in *Av*DDGS (part of the first conserved region MLEE**L**/MLEL**E**) and Asp281 in *Sh*EEVS *versus* Ala268 in *Av*DDGS (part of the second conserved region xxRxx**D**xGH/xxRxx**A**xGH) contribute to the active site pocket where they may also be responsible for the different activities of these enzymes (see below).

Additionally, a new putative DDGS gene whose product shows low sequence similarity to the known DDGS was recently reported in the halotolerant cyanobacterium *Aphanothece halophytica*.[93] Inactivation of this gene in this organism resulted in mutants that no longer produce MAAs. In our phylogenetic analysis, forty-one proteins formed a new clade with the *A. halophytica* DDGS (Figure 4.2) with their predicted protein sequences having some conserved motifs (e.g. YxxxEY(G)xNxxET and QC(D)RPHA(G)YGHTWSP) distinct from the mainstream DDGS sequences (Figure 4.3). This "divergent DDGS" clade includes proteins from cyanobacteria, algae, and marine invertebrates (Table S3). Some of the divergent DDGS genes (e.g. that of *A. halophytica*[93]) are not clustered with the other MAA biosynthetic genes, but others are (e.g. that of *Anabaena* sp. 90).

As can be seen in Figure 4.3, there is more variation in the conserved regions of bacterial EEVSs than DDGSs. This could be due to the differences in the reaction performed by these enzymes. EEVSs perform a cyclization reaction through a mechanism similar to those of DHQS and DOIS, involving five step reactions (Figure 4.4).[10, 131, 180] On the other hand, DDGSs perform a cyclization and a dehydration reaction, involving a more complex mechanism.[13] Due to this, more residues in the conserved region of DDGS may play key roles, and/or the spatial requirements for DDGS catalysis may be more stringent. Thus mutations in DDGS enzymes could negatively impact its catalysis more so than would changes in the EEVS enzymes. On the other hand, the relatively low sequence divergence observed among the vertebrate EEVSs is consistent with a recent report describing decelerated amino acid substitution in modern vertebrates.[181] Also, the transfer of EEVS gene from

microorganisms to vertebrates is predicted to occur later during evolution.[177]

Figure 4.3. Comparisons of partial amino acid sequences of the sugar phosphate cyclases.

Conserved amino acid residues in two fingerprint segments are shown from bacterial DHQS, aminoDHQS, DOIS, and EVS and various EEVS and DDGS groups as labeled. Black circles above the sequences indicate residues found in the catalytic pocket of the enzymes. Stars above the EEVS and DDGS sequence groups indicates residues that differentiate DDGS from EEVS.

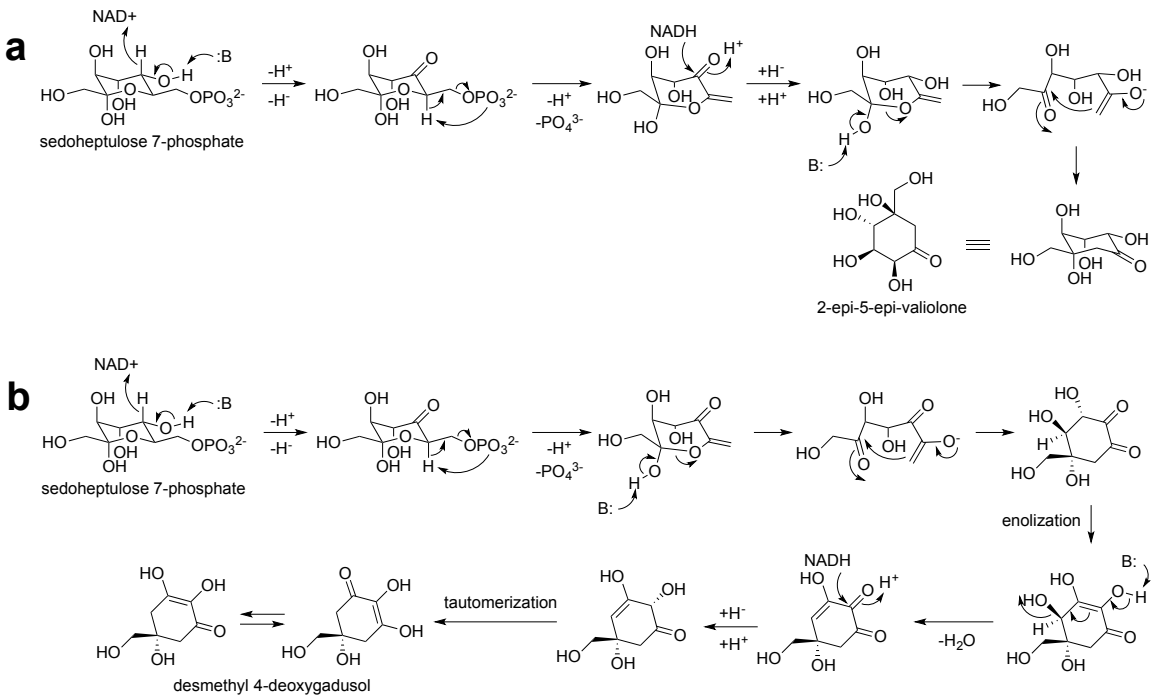


Figure 4.4. Proposed catalytic mechanisms for EEVS (a) and DDGS (b).

Crystal structure of *Av*DDGS

To investigate the catalytic pocket and unique features of a representative DDGS, we solved the X-ray crystal structure of *Av*DDGS. Crystals of recombinant His-tagged *Av*DDGS yielded diffraction data to 1.7 Å resolution and the structure was easily solved by molecular replacement using the *Sh*EEVS structure. The final refined model contains two chains, making up one dimer, in the asymmetric unit (Figure 4.5a); each chain included 400 of the 444 expected residues, one Zn^{2+}, one NAD^+, and a sulfate in the active site with a final R/R_{free} of 15.6/18.3% (Table 4.S5). The N-terminal tag and residues 1–2 and 403–410 are not modeled in either chain, but otherwise the main chain is well-ordered with clear density in both chains. Zn^{2+} and NAD^+ were not added during sample preparation or crystallization, but fortuitously are both present and have clear, unambiguous electron density in the electron density maps (Figure 4.5b). In both chains, the estimated occupancies of Zn^{2+} and NAD^+ are 0.5 and 0.75, respectively, and a sulfate (at occupancy 0.25) binds at the NAD^+ pyrophosphate position when NAD^+ is not there.

Table 4.S5. Data Collection and Refinement Statistics[a]

Data quality statistics

Wavelength (Å)	0.976
Space group	P $2_12_12_1$
Unit cell	
a, b, c (Å)	66.5, 120.3, 133.3
Resolution range (Å)	66.63–1.70 (1.73–1.70)
Reflections	3250204 (110159)
Unique reflections	118029 (5652)
Completeness (%)	99.9 (98.1)
Multiplicity	27.5 (19.5)
R_{meas} (%)	47.6 (627)
$\langle I/\sigma \rangle$	9.5 (0.8)
$CC_{1/2}$ (%)[b]	99.5 (21.5)

Refinement statistics

R_{work} (%)	15.6
R_{free} (%)	18.3
Amino acid residues	800
Solvent atoms	1294
Non-hydrogen atoms	7962
RMS bonds (Å)	0.950
RMS angles (°)	0.008
φ, ψ outliers (%)[c]	0.24
<B-factor> ($Å^2$)	24.5
PDB code	5TPR

a. Numbers in parentheses correspond to values in the highest resolution bin
b. Defined by Karplus & Diederichs[156]
c. As defined in MolProbity[182]

The dimer seen in the asymmetric unit (Figure 4.5a) is consistent with those observed in other SPC family enzymes and thus is thought to be biologically relevant. The overall structure is highly similar to *Sh*EEVS (rmsd = 0.9 A for 353 C_α atoms). All core secondary structural elements are conserved (Figure S1), and this includes a domain-swapped interaction observed in *Sh*EEVS (PDB 4P53). In what appears to be a common feature of this subset of sedoheptulose 7-phosphate cyclases (SH7PCs), instead of having a β-hairpin near position 31 in *Av*DDGS, the N-terminal residues continue in a linear direction, making an extending β-strand (here called β1/β2 as it

combines what are two β-strands in other SPC enzymes like DHQS) that reaches across the back of the dimer, and effectively contributes a β-strand to each monomer of the dimer (Figure 4.5a). In *Av*DDGS, a β-hairpin near residue 12 means that one short β-strand (β0) adds an 8th strand to the typically 7-stranded β-sheet observed in other SPC family members.

Each chain of *Av*DDGS consists of the expected N-terminal NAD^+-binding domain and C-terminal metal-binding domain seen in the SPC family (Figure S1). The N-terminal NAD^+-binding domain has a core 8-stranded β-sheet surrounded by 5 α-helices, 2 short 3_{10} helices, and 1 β-hairpin (strands 7 and 8). The C-terminal metal-binding domain is primarily α-helical and consists of 8 α-helixes, 1 β-hairpin and 1 3_{10} helix. The Zn-coordinating residues as well as most of the residues making up the active site cavity come from the metal-binding domain although the active site itself is located in the cleft between the two domains.

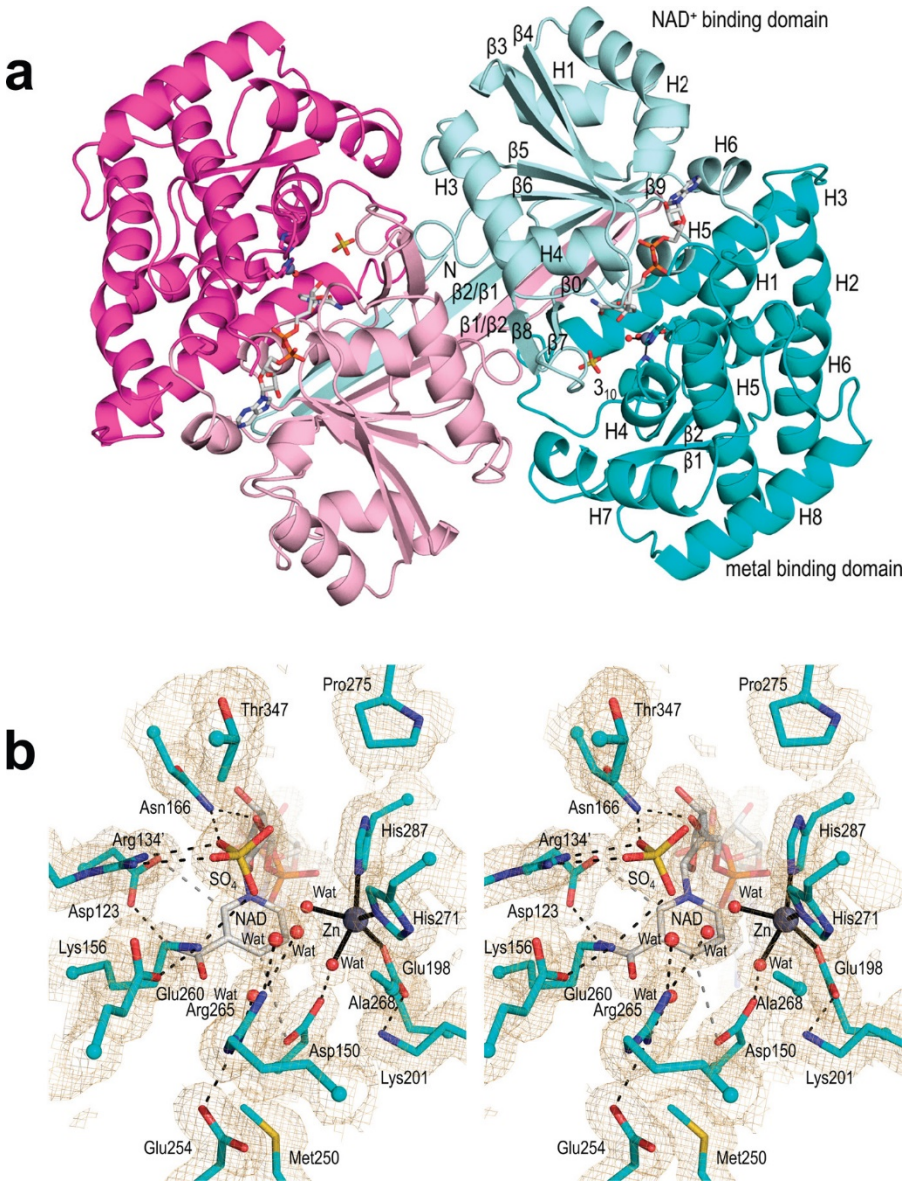


Figure 4.5. Overall structure and active site of *Av*DDGS.

a. Ribbon diagram of the two chains of the *Av*DDGS dimer are show in cyan and magenta tones, respectively, with the N-terminal NAD^+-binding domains in light hues and the C-terminal metal-binding domains in dark hues. The NAD^+, Zn^{2+} with its coordinating residues, and sulfate bound in the active site are shown. Secondary structural elements and domains of one monomer are labeled. The extended β-strands involved in the domain-swapped interaction (labeled β1/β2 and β2/β1) represent β1 and β2, respectively, in the labeled monomer. **b**. Stereoview of the *Av*DDGS active site residues (cyan carbons), waters (red spheres), sulfate bound in the active site (yellow), NAD^+ (grey carbons), and Zn^{2+} (silver sphere). Coordination bonds (black lines), select hydrogen bonds (black dashes), or approaches of interest (grey dashes) and $2F_o$-F_c electron density (orange, contoured at $1\rho_{rms}$) are also shown.

Active Site of *Av*DDGS

The active site is well defined (Figure 4.5b) with the Zn^{2+} and NAD^+ bound as in DHQS, DOIS, and *Sh*EEVS, with some key residues briefly noted here (and highlighted in Figure S1). The Zn^{2+} ion is coordinated by Glu198, His271, His287, and two waters. For NAD^+ binding, the adenosine ribose O2' hydroxyl hydrogen bonds with Asp56 and Asn58, located at the end of β3, and the adenine forms hydrogen bonds with the Thr186 and Thr143 side chains and backbone carbonyls of Thr143 and Leu183. The pyrophosphate oxygens (and the sulfate oxygens found in their place) hydrogen bond with the backbone amides of Gly119 and Leu120 in the glycine-rich turn connecting β5 and H3, as well as with the Thr144 side chain. The nicotinamide ribose hydroxyls hydrogen bond with the side chains of Glu87, Lys90, Lys165, and Asn166, and the nicotinamide amide group hydrogen bonds with the Asp123 carboxylate and the backbone carbonyl of Lys156. Also, as observed in *Sh*EEVS,[17] the carboxylates of Asp150 and Asp123 (Figure 4.5b) are roughly in the plane of the nicotinamide ring where they can make interactions with atoms C2 and C4 of the positively charged NAD^+. In addition, the nicotinamide amide oxygen and pyrophosphate oxygens hydrogen bond with ordered waters.

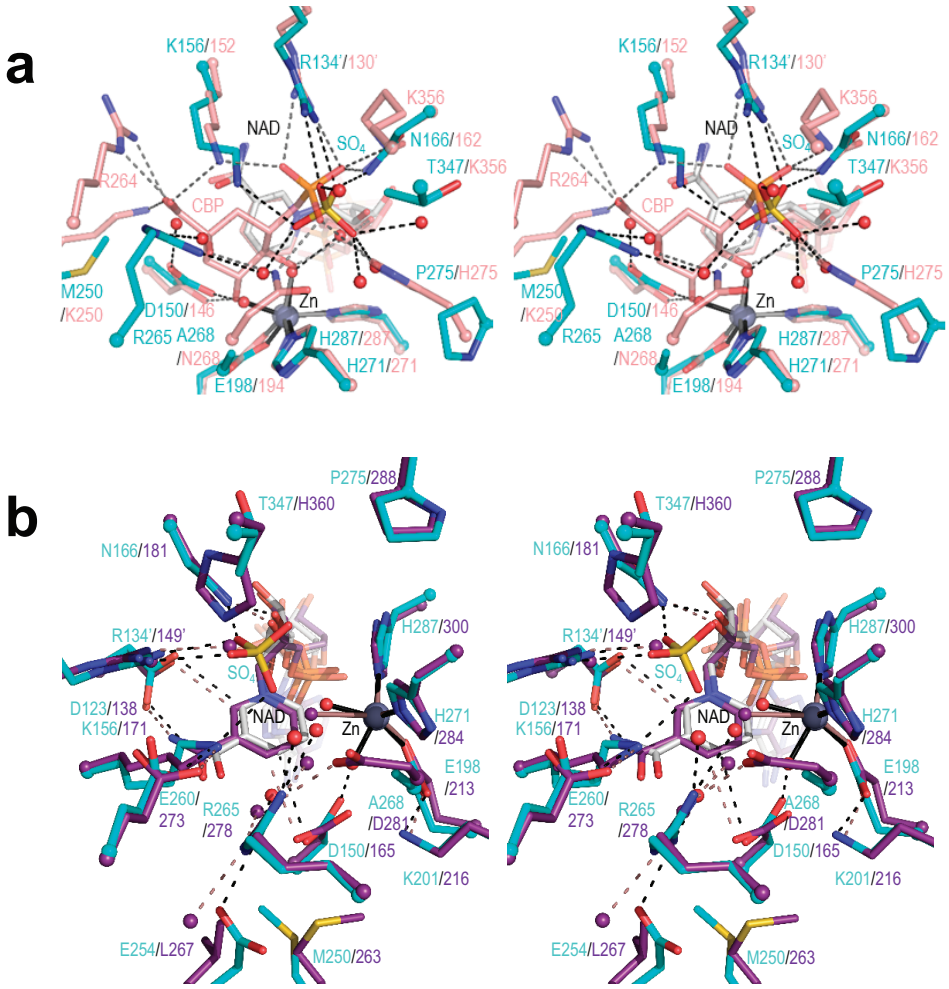


Figure 4.6. Comparisons of the *Av*DDGS active site region with the DHQS·CBP complex and with *Sh*EEVS.

a. Stereoview of select active site residues in *Av*DDGS (cyan carbons) with NAD^+ (grey carbons) overlaid on DHQS in complex with CBP (salmon carbons; PDB 1DQS). Hydrogen bonding interactions (dashed lines) and coordination bonds with Zn^{2+} (solid lines) in the *Av*DDGS (black) and DHQS (grey) active sites are shown. The sulfate (yellow) and waters (red spheres) in the active site of *Av*DDGS in agreement with components of CBP or water (salmon sphere) in the active site of DHQS as well as the zinc ions (silver spheres) are shown. Labels for Arg265 (*Av*DDGS) and Arg264 (DHQS) are separated due to their disparate positions in the active site and in the sequence alignment. A prime on a residue number means it is from the other subunit of the dimer. View of active site is rotated roughly 90° counter-clockwise with respect to the view in Figure 4.5b and 4.6b. **b**. Stereoview of active site residues in *Av*DDGS (cyan carbons) with NAD+ (grey carbons), sulfate bound in the active site (yellow), and select waters (red spheres) overlaid on *Sh*EEVS with NAD+ (both purple carbons) and select waters (purple spheres). Hydrogen bonds (dashed) and coordination bonds (solid) are shown in the *Av*DDGS active site (black lines) and *Sh*EEVS active site (light pink lines). A prime on a residue number means it is from the other subunit of the dimer.

The SH7P Binding Site in *Av*DDGS

Although this structure of *Av*DDGS is unliganded, we are able to make inferences about substrate binding via a comparison to DHQS with its bound substrate analog, carbaphosphonate analog (CBP) (Figure 4.6a). Supporting the value of this comparison, the bound sulfate in *Av*DDGS overlays well with the CBP phosphate and *Av*DDGS ordered water sites overlay well with the CBP C2, C4, and C5 hydroxyls. As these substituents are also present in *Av*DDGS's substrate, SH7P, we hypothesize that it will bind in a similar way.

One difference between SH7P and CBP occurs at the second substituent of C2, where in place of the CHP carboxylate, SH7P has a hydroxymethyl group. In the *Av*DDGS structure, one ordered water overlays roughly with one of the CBP carboxylate oxygens, perhaps indicating the position of the hydroxymethyl in SH7P. Notably, this is also the site of major differences between the DHQS and EEVS/DDGS active sites. In DHQS, Arg264 and Lys250 hydrogen bond with the C2-carboxylate (Figure 4.6a) and in DDGS (and EEVS) neither residue is conserved. The Lys is replaced by Met250 and the residue equivalent to Arg is not positionally conserved (even when it remains an Arg), greatly decreasing the local positive charge. Also the Lys→Met change allows Asp150 to be closer to the pocket and it binds to the water that may mimic the C2-hydroxymethyl substituent (Figure 4.6a). As we noted in describing the EEVS structure,[17] a different Arg that is one position later in the sequence (Arg265 in *Av*DDGS) is conserved in *Sh*EEVS and *Av*DDGS. This alternate Arg hydrogen bonds with the water noted above that mimics the C2-hydroxyl in CBP (Figure 4.6a).

Comparison of *Av*DDGS with *Sh*EEVS

The active sites of *Sh*EEVS and *Av*DDGS are quite similar but have key "fingerprint" differences[11-12, 17] that are putatively responsible for their differences in activity. Two binding pocket differences that had been previously identified are Asp281 in *Sh*EEVS *vs.* Ala268 in *Av*DDGS and His360 in *Sh*EEVS *vs.* Thr347 in *Av*DDGS. As noted above, here we have added a third active site fingerprint residue: Leu267 in *Sh*EEVS *vs.* Glu254 in *Av*DDGS. An active site overlay of these two

enzymes (Figure 4.6b) reveals just a few notable differences that essentially involve these three residues. In *Av*DDGS, the Glu254 side chain *vs.* Leu267 in *Sh*EEVS displaces an ordered water and hydrogen bonds directly to Arg265. Interestingly, the presence of Ala 268 *vs.* Asp281 in *Sh*EEVS opens room for a unique ordered water in *Av*DDGS that also hydrogen bonds to Arg265. The presence of Thr347 *vs.* His360 in *Sh*EEVS opens room near the putative binding site. For each of these changes it is not immediately obvious how it may contribute to the different catalytic activity of *Av*DDGS.

Looking beyond the active site, two of these positions, Glu254 and Ala268, are part of the sequence motifs we have used for distinguishing DDGS from EEVS (Figure 4.3). As seen in Figure 4.S2, despite their variation in amino acid sequence, in the folded protein each segment has the same secondary structure in DDGS and EEVS and orients the side chains in similar directions.

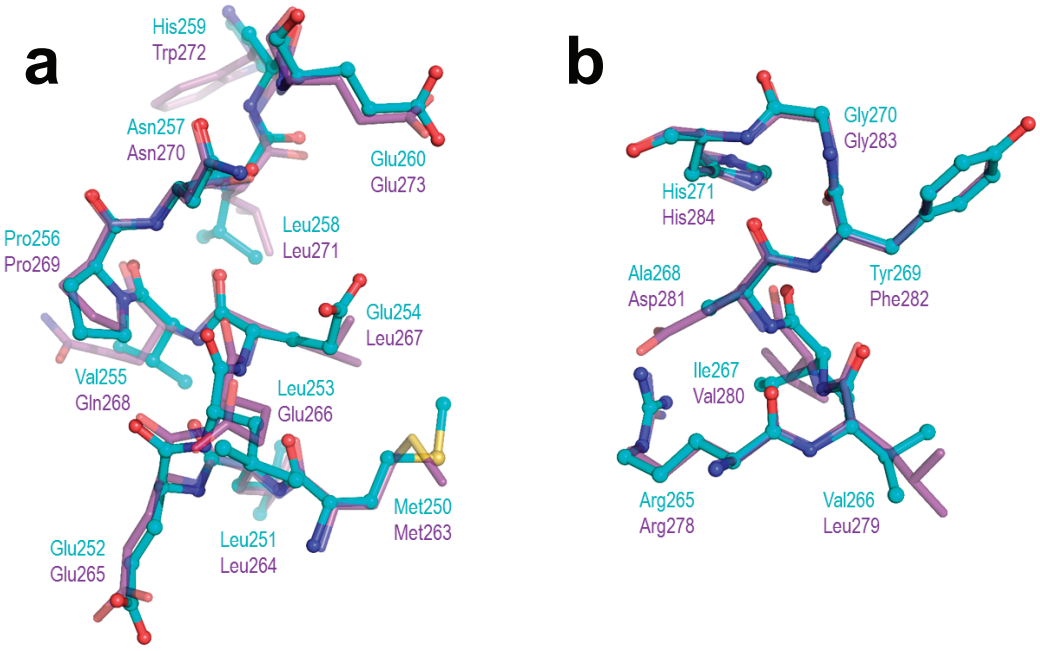


Figure 4.S2**.** Conserved motifs in *Av*DDGS and *Sh*EEVS.
a. First conserved motif discussed in the text maintains backbone interactions in *Av*DDGS and *Sh*EEVS. Residues 250–260 in *Av*DDGS (cyan ball and stick) and 263–273 in *Sh*EEVS (purple semitransparent) are shown. Met250, Glu254, and Glu260 in *Av*DDGS and Met263, Leu267, and Glu273 in *Sh*EEVS are active site residues in the respective proteins. **b**. Second conserved motif discussed in the text maintains backbone interactions in *Av*DDGS and *Sh*EEVS. Residues 265–271 in *Av*DDGS (cyan ball and stick) and 278–284 in *Sh*EEVS (purple semitransparent) are shown. Arg265, Ala168, and His271 in *Av*DDGS and Arg278, Asp281, and His284 in *Sh*EEVS are active site residues in the respective proteins.

Point mutations in *Sh*EEVS and *Av*DDGS

To test the importance of the three active site residues noted in the previous section as being characteristic to *Sh*EEVS or *Av*DDGS, we generated a total of 14 mutants of *Sh*EEVS (ValA) and *Av*DDGS (Ava_3858), consisting of six single point mutants (L267E, D281A, and H360T for *Sh*EEVS; A268D, E254L, and T347H for *Av*DDGS), six double point mutants (D281A/H360T, L267E/D281A, and L267E/H360T for *Sh*EEVS; E254L/A268D, E254L/T347H, and A268D/T347H for *Av*DDGS), and two triple point mutants (L267E/D281A/H360T for *Sh*EEVS; E254L/A268D/T347H for *Av*DDGS). All proteins were recombinantly produced in *Escherichia coli* and characterized for their activity under the conditions previously described.[12] For consistency between the proteins, they were tested fresh upon cell disruption and centrifugation without additional purification steps (Figure S3). Also, we developed a thin layer chromatography (TLC) protocol that separates the EEV and DDG products, and a staining reagent, *p*-anisaldehyde, that differentiates EEV and DDG as yellow and purple spots, respectively (Figure S4). The results revealed that *Sh*EEVS Leu267 and Asp281 are critical for its activity, and the equivalent *Av*DDGS residues Ala268 and Glu254 are as well (Figure S4). This is consistent with their high conservation in the two sequence motifs we used for distinguishing DDGS from EEVS. Interestingly, we found that *Sh*EEVS His360 and the equivalent *Av*DDGS Thr347 do not directly contribute to their respective catalytic activity. Yet, our bioinformatics studies revealed that these residues are highly conserved among proteins from their respective classes (94% of EEVS have His360 and 96% of DDGS have Thr347).

The lack of activity of A268D and E254L mutants of *Av*DDGS may suggest that one or both of these residues play a role in the early ring opening and aldol cyclization steps or in the unique dehydratase reaction by *Av*DDGS (Figure 4.4b). However, as there are no detectable intermediates produced by either of these mutants, their actual role(s) in *Av*DDGS catalytic activity are still unclear. Further elucidation of the DDGS catalytic mechanism, including the residues responsible for its proposed dehydratase activity, will be a subject of future investigations.

Phylogenetically distinct putative EEVS genes in some Gram-(+) and Gram-(–) bacteria

Phylogenetic studies also revealed a group of putative EEVSs (indicated as EEVS* here) that are arranged in a separate clade from the known EEVS clade and more similar to DHQSs (Figure 4.2). Those include GacC, a putative EEVS in the acarbose pathway from *Streptomyces glaucescens* GLA.O and Staur_1386, from the myxobacterium *Stigmatella aurantiaca* DW 4/3–1. A previous report has noted that although the *gac* cluster in *S. glaucescens* GLA.O has similarity to the *acarbose* cluster in *Actinoplanes* sp. SE50/110, but with enough differences to be uncertain of their equivalence.[183] Staur_1386 is part of a cryptic biosynthetic gene cluster in *S. aurantiaca* DW 4/3–1 with notable similarity to a cluster in a phylogenetically distant bacterium, *Cellvibrio japonicus* Ueda107 (Figure 4.7). Our discovery of putative pseudoglycosyltransferase genes within these clusters[184-186] has allowed us to predict their involvement in pseudo-oligosaccharide biosynthesis, but the exact end products remain unknown.

To confirm the catalytic function of GacC and Staur_1386, we cloned the corresponding genes from *S. glaucescens* GLA.O and *S. aurantiaca* DW 4/3–1 and heterologously expressed them in *E. coli*. There are two possible start codons in the *gacC* gene that would result in proteins with 388 and 410 amino acids. As the smaller protein is similar in size to AcbC from the acarbose pathway[10] and the larger protein is similar to ValA from the validamycin pathway,[187] we produced both versions of GacC. The recombinant GacC-388, GacC-410, and Staur_1386 proteins (unpurified, Figure S5) were characterized using SH7P as substrate in the presence of NAD^+ and Zn^{2+} or Co^{2+}. Analysis of the products by TLC and GC-MS revealed the production of 2-epi-5-epi-valiolone by all of these proteins (Figure S6), and this EEVS activity was further confirmed by assays of the purified recombinant GacC-388 and Staur_1386.

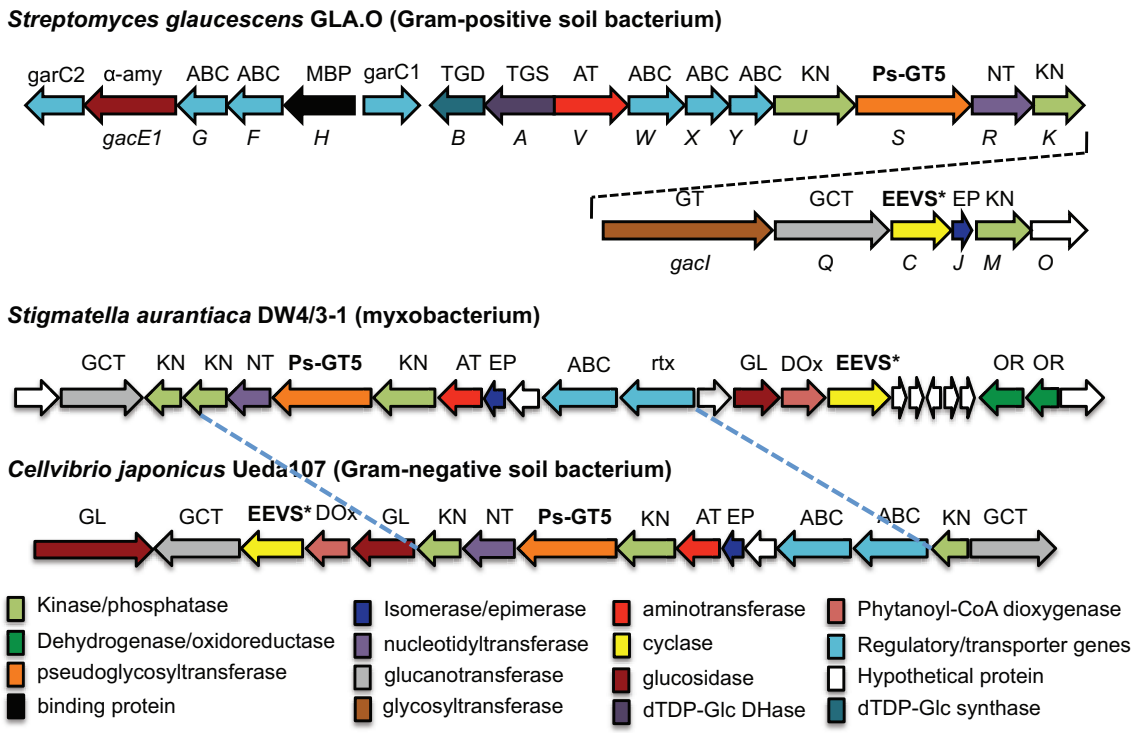


Figure 4.7. Biosynthetic gene clusters in *Streptomyces glaucescens* GLA.O, *Stigmatella aurantiaca* DW4/3–1, and *Cellvibrio japonicus* Ueda107 (previously known as *Pseudomonas fluorescens* subsp. *cellulosa*).

Unexpectedly, our phylogenetic analysis revealed that *S. hygroscopicus* subsp. *hygroscopicus* strain NRRL B-1477 had an EEVS* (WP_030827434.1) (Table S4), and only shared 41% identity with ValA from *S. hygroscopicus* subsp. *jinggangensis*. A recent analysis of *Streptomyces* lineages showed that *Streptomyces* from this clade descended from multiple lineages.[188] So we propose that the EEVS* gene or the gene cassette has spread through horizontal gene transfer (HGT)[94] and, since homologous recombination is not uncommon in *Streptomyces*, possibly formed through homologous recombination events. Interestingly, 14 of the total 51 known EEVS* proteins have a common insertion (Figure S7), and all were from Gram-(–) bacteria (i.e. 14 of the 19 Gram-(–) bacterial EEVS*).

In our phylogenetic analysis and in previous analyses,[177] this EEVS* clade branches off before the main EEVS clade and DDGS branches split. This implies a gene duplication event followed by divergence of paralogs, which may explain why DDGS appears to be more closely related to one EEVS clade than the other. However, alternative routes of evolution and/or convergent evolution cannot be

entirely ruled out. For example, another plausible scenario is that the ancestor of EEVS and DDGS underwent a gene duplication event, followed by those paralogs diverging into EEVS and DDGS enzymes. Whereas the evolutionary history of EEVS and DDGS remain uncertain, based on their unequal distributions across kingdoms of life (see Figures 4.2 and 4.8), it appears that both enzymes may have spread through multiple speciation and HGT events.

Distribution of EEVS and DDGS genes in bacteria and eukaryotes

EEVS catalyzes the first committed step in the biosynthesis of bacterial-derived C_7N-aminocyclitol natural products,[96, 167, 189-190] and was long thought to be present only in certain secondary metabolite-producing bacteria. However, the present study reveals a broad distribution of EEVS in various Gram-(+) and Gram-(–) bacteria, including gliding bacteria (myxobacteria) and cyanobacteria (Table S1). Some of them are clustered with other major biosynthetic enzymes such as terpene synthases, polyketide synthases, and non-ribosomal peptide synthetases, suggesting involvement in diverse natural products biosynthesis (data not shown). Strikingly, EEVS genes are also present in vertebrates, e.g., fish, amphibians, reptiles, and birds.[177] These genes are always paired with a methyltransferase-oxidase gene and together they are responsible for the biosynthesis of the sunscreen compound gadusol.[177] Interestingly, despite that many reports of putative EEVS genes in fungal genomes, based on our bioinformatics analysis EEVS is conspicuously absent in fungi. Except a putative EEVS in the yeast *Saitoella complicate*, we would reclassify all of the annotated fungal EEVS proteins as DDGSs (Table S2).

Microbial DDGS are mostly distributed in cyanobacteria, fungi, and Gram-(+) bacteria. However, we identified Gram-(–) bacteria of the genus *Lewinella* and *Halomonas* as containing the divergent DDGS protein (Table S3). Recent reports also suggest the presence of DDGS in some marine invertebrates and stramenopiles (Figure 4.8).[120] In our analysis, the putative marine invertebrate DDGS groups with the divergent DDGS clade. However, the stramenopiles have representatives in both the main DDGS clade and the divergent DDGS clade.

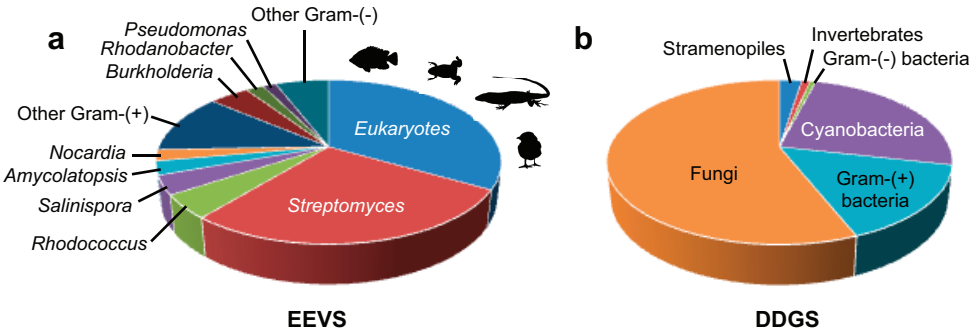


Figure 4.8. Distribution of EEVS and DDGS in the sequenced microorganisms. With the exception of a fungus and seven stramenopiles, all eukaryotic EEVS are from vertebrates (fish, amphibians, reptiles, and birds).

Figure 4.8 also shows that EEVS and DDGS enzymes are distributed quite differently. For example, while DDGS is widely distributed among fungi, there is only a single fungus, *S. complicate*, with an EEVS (Table S1 and S2). The EEVS encoded by this fungus, clades with the vertebrate EEVS branch. Given that most fungi lack EEVS genes and the vertebrate EEVS appears to be gained through HGT, it seems reasonable to suggest that *S. complicate* also gained its EEVS gene through HGT.

Interestingly, the available genomes usually either contain an EEVS or a DDGS instead of having both. The few exceptions include *Rhodococcus fascians* (Gram-(+) bacteria), *Chondrus crispus* (alga), and *Aureococcus anophagefferens* (alga). It is not uncommon for organisms to have multiple SPCs, as many organisms have some combination of DHQS, DOIS, EVS, or aDHQS in addition to an EEVS or DDGS. This strong anti-correlation between the presence of EEVS and DDGS genes within the same organism is an oddity that warrants further investigation.

Methods

Molecular Phylogenetic Analysis

Publically available amino acid sequences were obtained from the NCBI. Sequences were aligned using MUSCLE.[191] Approximate maximum likelihood phylogenetic analysis was performed using FastTree 2.1.3 with a JTT+CAT model.[192] MUSCLE and FastTree were performed on the Center for Genome and Biocomputing (Oregon State University) server. Sources of proteins for the analyses are listed on

Tables S1–S4. Archaeopteryx was used to view and edit the phylogenetic tree.[193] Amino acid sequences were analyzed and viewed using the software Geneious (Biomatters).

Expression, Purification, and Crystallization of *Av*DDGS

The protein was expressed and purified as previously described.[12] After purification, the proteins were dialyzed against 2 L of 10 mM Tris-HCl pH 7.5. The protein solution was concentrated by ultrafiltration to 16 mg mL^{-1}, flash-frozen in liquid nitrogen, and stored at –80 °C. The enzyme was crystallized at 4 °C in hanging drops made of 1 µL protein stock, 2 µL of a reservoir solution (5% (v/v) PEG 8000, 0.2 M ammonium sulfate, 0.1 M sodium cacodylate pH 7.0) and 0.5 µL seed solution. The resulting crystals grew in clusters of plates which needed to be separated to isolate single crystals.

X-ray Diffraction Data Collection

For diffraction data collection at –170 °C, crystals were briefly passed through a solution containing 30% (v/v) glycerol and then cryo-cooled by being plunged into liquid nitrogen. Data were collected from three crystals using $\lambda = 0.976$ Å and $\Delta\varphi=1°$ steps at beamline 5.0.3 at the Advanced Light Source (Berkeley, CA). From the crystals, 360 images were collected at a detector distance of 250 mm. For two crystals, 5 s images were collected. For the third crystal, 3 s images were collected. All images were integrated using Mosflm[194] and merged using the CCP4 suite of programs.[195-196] In data processing, images which showed too much decay based on visual examination were excluded. For the first, second, and third crystals, the first 210 images, 360 images, and 310 images, respectively, were included. The merged data set was usable to 1.7 Å resolution using a $CC_{1/2}$ of ~0.2 as the cutoff criteria. A random 5% of reflections were marked for cross-validation. The atomic coordinates have been deposited in the Protein Data Bank (PDB entry 5PTR).

Structure Determination

The phase problem was initially solved by molecular replacement using MR Rosetta with default settings.[197] The search model was *Sh*EEVS (PDB entry 4P53), the closest structurally known homolog with 39% sequence identity, and resulted in a preliminary solution with R and R_{free} values of 0.20 and 0.22 with 772 residues built (two chains in the asymmetric unit). All manual model building was done in Coot.[198] Refinements were done using Phenix[199] with TLS refinement and riding hydrogens. Water molecules were manually placed on the criteria of having electron density $\geq 4\rho_{rms}$ in $F_o–F_c$ maps and $\geq 0.9\rho_{rms}$ in $2F_o–F_c$ maps and a reasonable potential H-bond partner. The electron density map from this solution allowed us to build 2 chains in the asymmetric unit including residues 3-402 with all but one side chain (Arg 400 is stubbed at C_β) modeled in both chains, the active site Zn^{2+}s, NAD prosthetic groups, 5 sulfates, 2 glycerol, and 2 polyethylene glycols. Alternate side chain and/or backbone conformations are included in chain A for Gln5, Glu12, Thr13, Ser130, Gln191, Leu327, Ser343, and the Gly401-Gly402 backbone and in chain B for Glu12, Thr13, Arg61, Asp104, Ser130, Gln191, Ser343, and Ser387. This construct of *Av*DDGS contained a 34 amino acid N-terminal 6x His tag extension, and this was ordered in the crystal structure. Final statistics are reported in Table S5. The secondary structure assignments were made using DSSP,[153, 200] and structure-based sequence alignments were generated using the Dali database.[201] For figures and RMSD calculations, crystal structures were overlaid using SSM Superposition.[202]

*Av*DDGS and *Sh*EEVS Mutagenesis and Characterization

For details, see the Supporting Information.

Cloning and Expression of *gacC* and *staur_1386*, Protein Purification, Enzyme Assay, and GC-MS Analysis

For details, see the Supporting Information.

Acknowledgements

This work was supported in part by grant GM112068 (to TM) from the National Institute of General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health (NIH).

Chapter 5

Structure and proposed mechanism of α-glycerophosphate oxidase from *Mycoplasma pneumoniae*

Callia K. Elkhal*, Kelsey M. Kean*, Derek Parsonage, Somchart Maenpuen, Pimchai Chaiyen, Al Claiborne, and P. Andrew Karplus

*These authors contributed equally to this work

Published in *The FEBS Journal,* 2014, 16, pp 3030–3042

Abstract

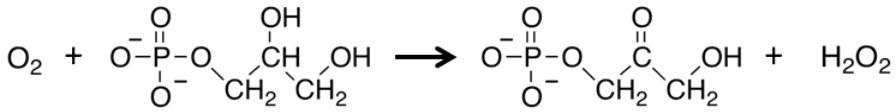
The formation of hydrogen peroxide (H_2O_2) by the FAD-dependent α-glycerophosphate oxidase (GlpO), is important for the pathogenesis of *Streptococcus pneumoniae* and *Mycoplasma pneumoniae*. The structurally known GlpO from *Streptococcus* sp. (*Ssp*GlpO) is similar to the pneumococcal protein (*Sp*GlpO) and provides a guide for drug design against that target. However, *M. pneumoniae* GlpO (*Mp*GlpO), having <20% sequence identity with structurally known GlpOs, appears to represent a second type of GlpO we designate as Type II GlpOs. Here, the recombinant His-tagged *Mp*GlpO structure is described at ~2.5 Å resolution, solved by molecular replacement using as a search model the *Bordetella pertussis* protein 3253 (*Bp*3253) a protein of unknown function solved by structural genomics efforts. Recombinant *Mp*GlpO is an active oxidase with a turnover number of ~580 min^{-1} while *Bp*3253 showed no GlpO activity. No substantial differences exist between the oxidized and dithionite-reduced *Mp*GlpO structures. Although, no liganded structures were determined, a comparison with the tartrate-bound *Bp*3253 structure and consideration of residue conservation patterns guided the construction of a model for α-glycerophosphate (Glp) recognition and turnover by *Mp*GlpO. The predicted binding mode also appears relevant for the type I GlpOs (such as *Ssp*GlpO) despite differences in substrate recognition residues, and it implicates a histidine conserved in type I and II Glp oxidases and dehydrogenases as the catalytic acid/base. This work provides a solid foundation for guiding further studies of the mitochondrial Glp dehydrogenases as well as for continued studies of *M. pneumoniae* and *S. pneumoniae* glycerol metabolism and the development of novel therapeutics targeting *Mp*GlpO and *Sp*GlpO.

Introduction

Mycoplasma pneumoniae is a human respiratory tract pathogen that causes 40% or more of community-acquired pneumonias,[203] with the typical syndrome being tracheobronchitis in children. This pathogen initiates colonization of the host airway mucosal epithelium *via* a special attachment organelle,[204] which also provides an important gliding function. Among other distinctive features of *M. pneumoniae* are a

small genome, the lack of a rigid cell wall, and limited metabolic capabilities. As in the Gram-positive streptococci from which the mycoplasmas diverged ca. 600 million years ago,[203] the tricarboxylic acid cycle, electron transport chain, and respiratory cytochromes are absent. However, glycerol metabolism appears to be an important pathogenicity factor for *M. pneumoniae* [41-44] and thus the enzymes involved provide potential drug targets for combating respiratory infectious diseases.

Of particular interest in this context is the *M. pneumoniae glpD* gene (MPN051) that is annotated as encoding a glycerol-3-phosphate dehydrogenase,[44] and based on its sequence can be identified as a member of the D-amino acid oxidase (DAAO) superfamily[205] of FAD-dependent enzymes combining a 'glutathione-reductase-2' type FAD-binding domain and an antiparallel β-sheet based substrate-binding domain. Despite the annotation as a dehydrogenase, the encoded enzyme has been shown to be a constitutively expressed cytosolic FAD-dependent α-glycerophosphate oxidase,[41] using O_2 as the final electron acceptor and producing dihydroxyacetone phosphate (DHAP) and H_2O_2:



We therefore refer here to the encoded protein as *M. pneumoniae* GlpO (*Mp*GlpO), rather than GlpD. The gene is tightly linked with that for glycerol kinase (*glpK*), and together these enzymes can catalyze the ATP-dependent conversion of glycerol to the glycolytic intermediate dihydroxyacetone phosphate.[41] The peroxide produced by *Mp*GlpO has been shown to be crucial for pathogenicity, and the ortholog from the animal pathogen, *Mycoplasma mycoides* subsp. *mycoides* SC, has also been implicated as a primary virulence factor.[43] Similarly, Mahdi *et al.*[39] have shown that *Streptococcus pneumoniae* GlpO (*Sp*GlpO) is responsible for H_2O_2-mediated cytotoxicity against human brain microvascular endothelial cells and for promoting pneumococcal meningitis.

Given the clinical significance of *M. pneumoniae* and *S. pneumoniae* and the importance of *Mp*GlpO and *Sp*GlpO in cytotoxicity and virulence, structural studies of these enzymes would provide a valuable foundation for drug design. A good understanding of the *Sp*GlpO structure is in hand because it is quite similar (~62%

sequence identity) to the structurally known *Streptococcus* sp. GlpO (*Ssp*GlpO).[38] However, insufficient information exists for modeling the 43 kD *Mp*GlpO enzyme, as it is surprisingly divergent from the higher molecular weight (~65 kD) streptococcal enzymes: it is not only missing a C-terminal α-helical domain, but it has only ~20% sequence identity with *Ssp*GlpO and is actually more sequence similar to other DAAO superfamily enzymes, such as glycine oxidase, than it is to *Ssp*GlpO. To provide a foundation for understanding catalysis and guiding drug design against *Mp*GlpO, we report here its crystal structure. As part of this work, we also carried out some functional characterizations of the protein encoded by *Bordetella pertussis* gene 3253 (*Bp*3253), which is a related protein of unknown function solved by the NorthEast Structural Genomics group (PDB entry 3DME, deposited 2008). It became of interest because it is the structurally known protein most similar to *Mp*GlpO (27% sequence identity).

Results and Discussion

Expression and biochemical properties of *Mp*GlpO and the *B. pertussis* protein *Bp*3253.

Previously published studies of *Mp*GlpO were carried out with an N-terminal His-tagged protein derived from *M. pneumoniae* M129.[41] In the present study, the pET28a vector introduces a 34-residue His-tag in-frame with the codon-optimized synthetic gene, and this tag was present in all studies. The characterization of *B. pertussis Bp*3253 was carried out using the same N-terminally His-tagged construct used in the crystal structure determination of PDB entry 3DME. Both recombinant *Mp*GlpO and *Bp*3253 were expressed and purified (yielding ~5 and ~60 mg/L culture respectively), and the visible absorption spectra for the purified proteins are characteristic of properly folded flavoenzymes (Figure 5.1). The native molecular weight for His-tagged *Mp*GlpO as determined by gel filtration, at 41.7 kD, is reasonably close to the value of 46.3 kD calculated for His-tagged *Mp*GlpO indicating that the recombinant protein is a monomer in solution.

Although the oxidase activity of *Mp*GlpO was documented previously,[41] we assayed the recombinant His-tagged *Mp*GlpO purified here and found a specific

activity of 12.8 U/mg at 25 ˚C with 200 mM D/L-Glp and ca. 260 μM O_2 (the saturating concentration in water at 25 ºC). This is equivalent to a k_{cat} of 9.7 s^{-1} (i.e. 580 min^{-1}). This level of activity is reasonable for an oxidase even though it is somewhat lower than was seen for native and recombinant GlpOs from *E. casseliflavus* (*Ecass*GlpO)[206] and *Streptococcus* sp. (*Ssp*GlpO)[207] which gave k_{cat} values of 70-90 s^{-1} at 25 ºC. When *Bp*3253 was tested using the same assay, it showed no ability to catalyze the Glp-dependent formation of H_2O_2.

Many flavoprotein oxidases show a pronounced reactivity with sulfite to form a reversible adduct at the flavin N5-position.[76] The *Ecass*GlpO and *Ssp*GlpO giving similar values of K_d = 0.82 mM and 1.5 mM,[206-207] and we here performed similar sulfite titrations with freshly prepared *Mp*GlpO and *Bp*3253 (Figure 5.1A). For *Mp*GlpO, the measured K_d was 3.0 mM, but flavin absorbance was only partially bleached with about half of the *Mp*GlpO flavin being refractory to sulfite adduct formation, even after an incubation of 30 min at 48 mM sulfite. A parallel experiment with the *Bp*3253 protein (Figure 5.1B) shows a monophasic sulfite titration giving essentially complete bleaching of the flavin absorbance spectrum with a K_d of 0.42 mM, about seven-fold more favorable than that for *Mp*GlpO. We do not have any good proposals at this time for why only roughly half of the *Mp*GlpO reacts with sulfite.

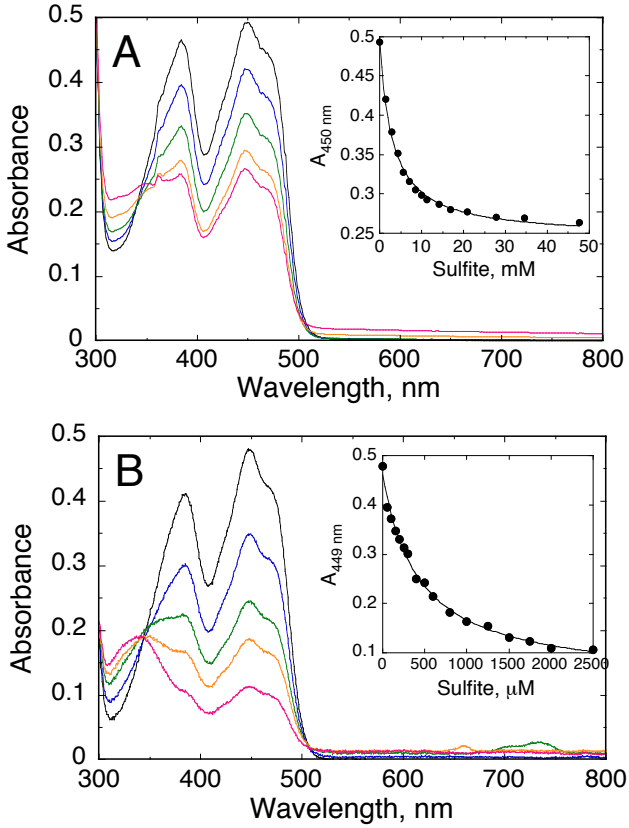


Figure 5.1. Sulfite titrations of *Mp*GlpO and *Bp*3253.
A. MpGlpO [35.4 μM in 0.8 mL of 50 mM potassium phosphate, 0.5 mM EDTA, pH 7.0] was titrated aerobically with a 1 M sodium sulfite solution. The spectra shown correspond to the addition of 0 (black), 1.43 (blue), 4.27 (green), 11.3 (orange) and 47.6 mM (red) total sulfite. The inset shows the absorbance at 450 nm as a function of added sulfite. **B.** The titration of *Bp*3253 (36.2 μM in 1 mL) was carried out as described in (A), with 25 and 100 mM solutions of sodium sulfite. The spectra shown are with 0 (black), 0.15 (blue), 0.50 (green), 0.80 (orange) and 2.0 mM (red) total sulfite. The inset shows the absorbance changes at 449 nm as sulfite is added.

Structure determination of oxidized and reduced *Mp*GlpO

Crystals obtained of recombinant His-tagged *Mp*GlpO yielded diffraction data to ~2.5 Å resolution. Attempts to solve the structure by molecular replacement with a structure of *Ssp*GlpOΔ (PDB entry 2RGH) as the search model were unsuccessful, which was not surprising given the low ~20% sequence identity between *Mp*GlpO and *Ssp*GlpO. A PDB database query for better search models, led to the identification of the *B. pertussis* protein *Bp*3253 (PDB entry 3DME; 27% identity) as the known structure with the highest sequence similarity. Using this as a search model, molecular replacement was successful and led to ~2.5 Å resolution

structures with acceptable statistics (Table 5.1) for both oxidized and dithionite-reduced *Mp*GlpO.

Table 5.1. Data Collection and Refinement Statistics for *Mp*GlpO structures. Numbers in parentheses represent data for the high-resolution shell.

	oxidized	reduced
PDB entry code	**4X9M**	**4X9N**
Data quality statistics		
Wavelength (Å)	0.9765	0.9765
Space group	*P*23	*P*23
Cell dimensions, $a = b = c$ (Å)	111.59	111.61
Resolution range (Å)	50-2.4 (2.53-2.4)	60-2.5 (2.64-2.5)
Reflections	430,576	836,296
Unique reflections	18,401	16,368
Completeness (%)	100 (99)	100 (100)
Multiplicity	23.4 (23.6)[a]	51.1 (28.6)
R_{pim}	0.022 (0.19)	0.072 (0.28)
R_{meas}	0.109 (0.90)	0.533 (1.51)
I/σ	22.7 (4.4)	19.9 (3.0)
Refinement statistics		
R_{work} (%)	14.8 (21.9)	14.7 (23.5)
R_{free} (%)	20.4 (25.3)	21.1 (34.3)
No. amino acid residues	384	384
No. solvent atoms	200	200
No. non-hydrogen atoms	3301	3313
$<B>$ protein (Å^2)	35	39

The refined models contain a single *Mp*GlpO chain in the asymmetric unit including residues 1-384, the FAD cofactor, ordered water sites, and a Ni^{2+} atom. The $2F_o$-F_c map for the FAD cofactor and some of its environment illustrates the quality

of electron density in well-ordered regions of the protein (Figure 5.2). Missing from the models and presumed mobile are all 34 residues of the N-terminal His-tag and two side-chains; also, a few residues are modeled with alternate conformations (see Experimental Procedures). The Ni^{2+} ion, identified using an X-ray fluorescence scan and presumably introduced during Ni-affinity chromatography, is situated at a crystallographic three-fold packing interface where it is coordinated by three His59-side chains. As the gel filtration result described above shows that the recombinant His-tagged *Mp*GlpO is a monomer, this three-fold interface and a two-fold crystal packing interaction involving strand β15 and burying ca. 1600 $Å^2$ of surface area are not physiologically relevant. So the nickel, possibly picked up during purification, can be considered a fortuitous crystallization aid.

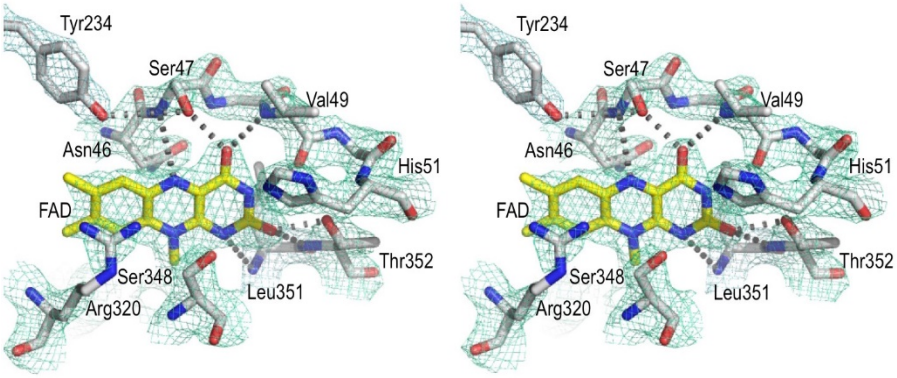


Figure 5.2. Electron density quality for the flavin and nearby side chains.
Stereoview of the final $2F_o$-F_c electron density map (cyan; contoured at $2.0\sigma_{rms}$) for the *Mp*GlpO flavin (yellow carbons) and the adjacent protein atoms (off-white carbons) and one water (red sphere). Hydrogen bonds (dashed lines) to the flavin and the water are indicated. Residues are labeled.

Overall structure

As expected for a DAAO superfamily member, the *Mp*GlpO chain is organized into an FAD-binding domain with a predominantly parallel, six-stranded β-sheet and a substrate-binding domain with a core antiparallel, eight-stranded β-sheet (Figure 5.3). The two domains are discontinuous, with the FAD-binding domain including residues 1-87, 149-219, and 330-364, and the substrate-binding domain formed by residues 86-148 and 227-323.

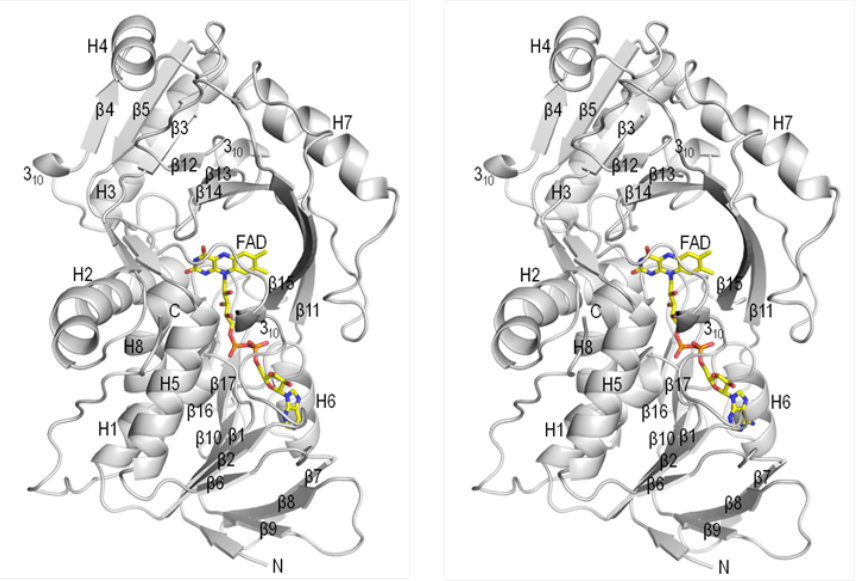


Figure 5.3. Tertiary structure of the *Mp*GlpO monomer.
Stereo ribbon diagram of the *Mp*GlpO monomer showing the FAD (sticks with yellow carbons) and labeling the secondary structural elements. The FAD-binding domain is at the bottom and the substrate-binding domain is at the top.

A search using the DALI server[201] to identify structural homologs of *Mp*GlpO, confirms that there are no highly similar known structures. The most similar structure is that of the molecular replacement search model *Bp*3253 (PDB code 3DME; Z-score=44, 26% sequence identity and 2.0 Å rmsd over 352 residues), and a bit less similar is a glycine oxidase from *Bacillus subtilis* (PDB code 1NG4; Z-score=42; 20% sequence identity and 1.9 Å rmsd over 340 residues). An overlay with these proteins shows the high similarity extends throughout the protein chains (Figure 5.4A). Remarkably, the enzymes with more similarity in function, *Ssp*GlpO (PDB codes 2RGO) and *E. coli* GlpD (PDB code 2QCU), are structurally less similar, giving *Z*-scores of only 28 and 36, respectively (with higher rmsd values of 2.9 and 3.7 Å), and showing only ~17% sequence identity over the ~350 residues that are in the two domains common to the proteins. Visually, the overlay of *Mp*GlpO with *Ssp*GlpO and *E. coli* GlpD shows their greater divergence (Figure 5.4B). A structure-based sequence alignment of these proteins (Figure 5.4C) provides further details of the comparisons, and also is useful for tracking the similarities and differences in active site residues that will be discussed in the active site section.

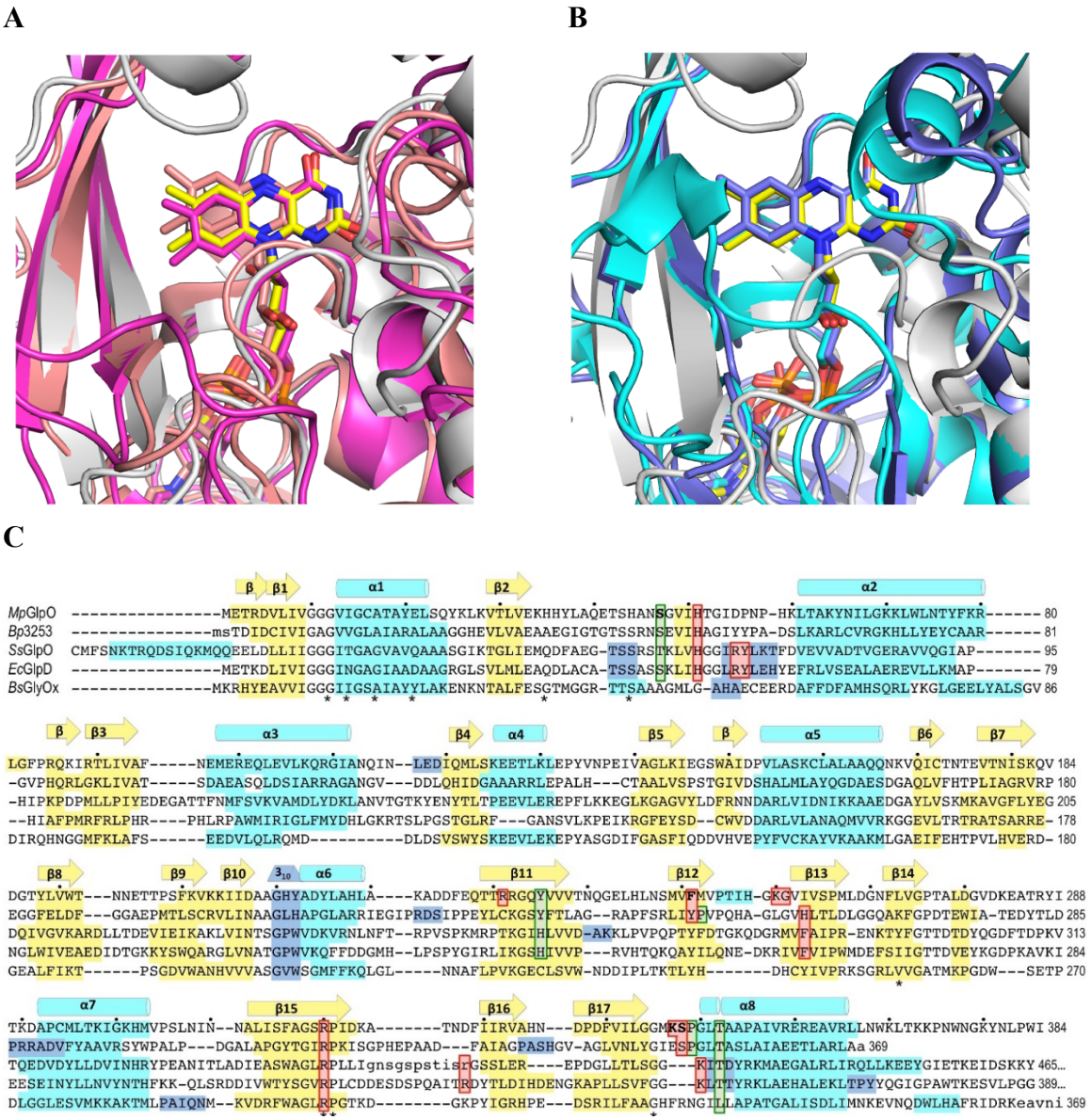


Figure 5.4. Comparisons of the *Mp*GlpO structure and sequence with select homologs. **A.** The flavin region in an overlay of *Mp*GlpO (off-white protein and yellow FAD) on *Bp*3253 (violet; PDB entry 3DME) and glycine oxidase (salmon; PDB entry 1RYI;[208]). For this view the Figure 5.3 molecule was rotated 180˚ around a vertical axis (i.e. this view is from the back of that image). Note the similar paths of the loop in front of the flavin, in *Mp*GlpO containing Ser 348. **B.** Same as (A) but overlaying *Mp*GlpO (colored as in A) with a form of *Ssp*GlpO missing a 50 residue segment (blue; PDB entry 2RGH;[38]) and *Ec*GlpD (cyan; PDB entry 2QCU; [37]). Note the different paths of the loop in front of the flavin of the type I enzymes compared with *Mp*GlpO. **C.** Structure-based sequence alignment of the five enzymes shown in panels (A) and (B). Conserved residues (*) and residues involved in β-strands (yellow), α-helices (cyan) and 3_{10}-helices (blue) are indicated. Also highlighted are residues discussed as important for in flavin binding (green boxes; and including the *cis*-Pro in *Mp*GlpO and *Bp*3253) and substrate binding (red boxes). Dots above the *Mp*GlpO sequence mark its every 10th residue, and at the end of each line is a numbered residue for each of the sequences.

That *Mp*GlpO is more similar to other DAAO superfamily enzymes than it is to *Ssp*GlpO led us to consider that whether GlpO activity may have independently evolved twice in the DAAO superfamily. To explore this possibility further we generated a relatedness tree of structurally known proteins similar to *Mp*GlpO (Figure 5.5). In this tree, despite the DALI scores, the *Mp*GlpO and *Ssp*GlpO and *E. coli* GlpD actually do come out as being more closely-related to each other than to other functionally characterized DAAO superfamily members. The *Ssp*GlpO-like enzymes that cluster together include both GlpOs and the mitochondrial/bacterial GlpD dehydrogenases.[37] They represent a very widely distributed group that we are designating as 'Type I GlpO/DH' enzymes. In contrast, the *Mp*GlpO-like enzymes, which we are designating 'Type II GlpO/DH' enzymes, are more narrowly distributed, being found only bacteria of the class mollicutes (including *M. pneumoniae*) and the closely related low G+C Gram-positive anaerobic bacteria of the class Erysipelotrichia.[209]

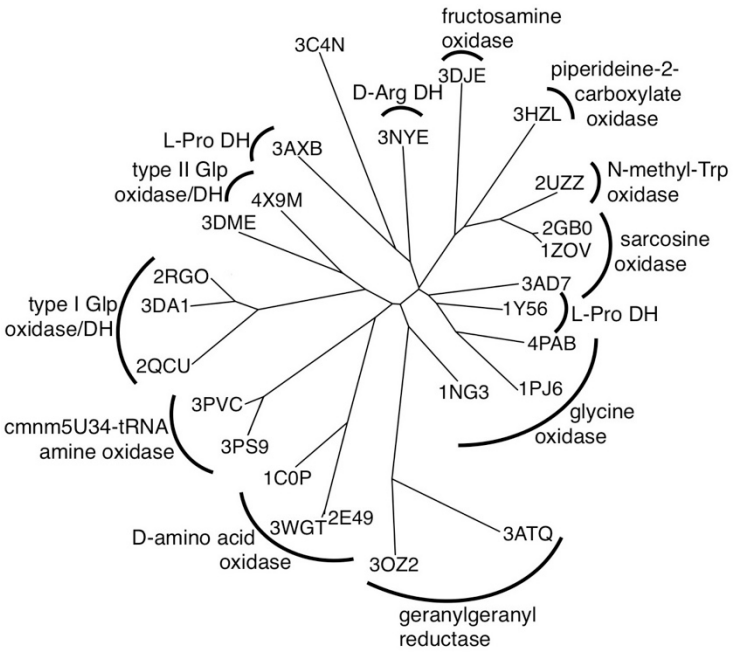


Figure 5.5. Relatedness tree of structurally known DAAO superfamily members most similar to *Mp*GlpO.

A DALI[201] search in November 2014 using the PDB90 database option and oxidized *Mp*GlpO as the search model, provided a gap-removed alignment for hits with Z-scores higher than 20. These were used to generate a tree with PhyML.[210] Branches are labeled with individual PDB entry names and known enzyme types are indicated.

Interestingly, the Type II GlpO/DHs from anaerobic bacteria appear to all have an additional ca. 85 residue C-terminal domain that is not closely related to any known structure but includes two segments with conserved pairs of Cys residues (CxCE and CQxGFC) and has some similarity (e-value $4x10^{-10}$) with the Pfam family of bacterioferritin-associated ferredoxin-like [2Fe-2S] domains (Pfam04324). The one biochemically studied Type II GlpO/DH containing such a C-terminal domain is the *E. coli glpA* gene product that is expressed under anaerobic conditions (replacing GlpD); and consistent with the assignment of the additional domain having an iron-sulfur center is the report that *E. coli* GlpA binds both FAD and non-heme iron.[211] Our proposal based on these observations is that these Type II GlpO/DHs with the additional C-terminal domain are all dehydrogenases for which the Type II GlpO/DH module converts Glp to DHAP, and the C-terminal domain serves as a conduit to receive electrons from the flavin and pass them on to a further (anaerobic) acceptor.

Flavin binding and active site

General features of FAD binding to DAAO superfamily enzymes have been well-described,[205] so we will not detail those here but will focus in on the flavin and the substrate binding site. In *Mp*GlpO, the bound flavin is slightly twisted with a 3_{10}-helix involving residues Thr42 - Asn46 closely covering its *si* face, and Ser47 – Val49 interacting with the N5, O4, and N3 atoms of the flavin (Figure 5.2). Optimizing the hydrogen-bond between flavin-N3 and Val49-O appears to be a cause of the flavin twist. Above the flavin (on the *si* side), there is open space above and in front of atom N5, creating a cavity lined by His51, Arg320, and Ser348 (Figure 5.2). Interestingly, we do not see any substantive differences between the active sites of oxidized and reduced *Mp*GlpO structures at this resolution, even though the pale color of the reduced *Mp*GlpO crystals provided visual evidence that they truly were reduced. Given the lack of differences, in the remaining discussion of the active site features, we will focus solely on the somewhat higher resolution oxidized structure.

Unfortunately, attempts to obtain an *Mp*GlpO structure with a substrate or substrate analog bound were unsuccessful (see methods), as was also true for our earlier structural work on *Ssp*GlpO.[38] However, the protein structure that is the most

similar to *Mp*GlpO, that of *Bp*3253, fortuitously has a ligand – L-tartrate – bound in its active site. Although the function of *Bp*3253 is not yet known, it binds tartrate similarly to how glycine oxidase [208, 212]and DAAO[213] bind their substrate analogs (overlays not shown), implying that the *Bp*3253:tartrate complex is an informative one. Specifically, one tartrate α-carbon (i.e. alpha to one of the carboxylates) is placed just 3.7 Å from flavin N5 and with excellent geometry for hydride transfer (Figure 5.6).

An overlay of *Mp*GlpO with *Bp*3253 (Figure 5.6) shows that the flavins and most nearby peptide backbone segments align well. In the *Bp*3253 complex there are just six side chains (and no backbone atoms) making van der Waals or hydrogen-bond contacts with the tartrate. Starting with the key Arg sitting above the flavin C7 and C8 methyl groups, the *Bp*3253 residues, with their *Mp*GlpO equivalents in parentheses, are Arg316(Arg320), Pro272(Pro274), His259(Ile261), Tyr248(Phe250), His52(His51), and Ser350(Ser348). Hydrogen bonds connect Arg316 with the C1-carboxylate, His259 via a water molecule with the C3 hydroxyl, and Tyr248, His52, and Ser350 with the C4-carboxylate (Figure 5.6). Four of these positions are perfectly conserved in *Mp*GlpO, both in identity and potential placement, recognizing that Arg320 in *Mp*GlpO could easily shift to match the position seen for the equivalent *Bp*3253 residue. The conserved placement of Ser350(Ser348) projecting over the flavin depends on a conserved *cis*-peptide bond between the Ser and Pro351(Pro349). Notable differences in the active sites are the Tyr248->Phe replacement in *Mp*GlpO and the His259->Ile261 replacement that is compounded by a large difference in the path of the β12-β13 loop, with the backbone of *Mp*GlpO Gly259 occupying the space filled by the His259 side chain in *Bp*3253.

Regarding this β12-β13 loop difference, a key question is why the path of *Mp*GlpO residues 256-261 is different. Are they adopting an arbitrary conformation in this crystal form or a conformation that reliably represents the chain path of the Type II GlpO/DHs? Three evidences suggest it is representative rather than arbitrary: first, the *Mp*GlpO path is similar to the paths seen in broader superfamily members such as glycine oxidase meaning that it is the *Bp*3253 path that is unusual; second, the distinct *Bp*3253 path appears to be directly related to it having a proline (Pro249) at

the end of strand β12 that disrupts the normal β-sheet hydrogen bonding (Figure 5.6) and is not present in *Mp*GlpO (Figure 5.4C); and third, the loop is well ordered and not involved in crystal contacts.

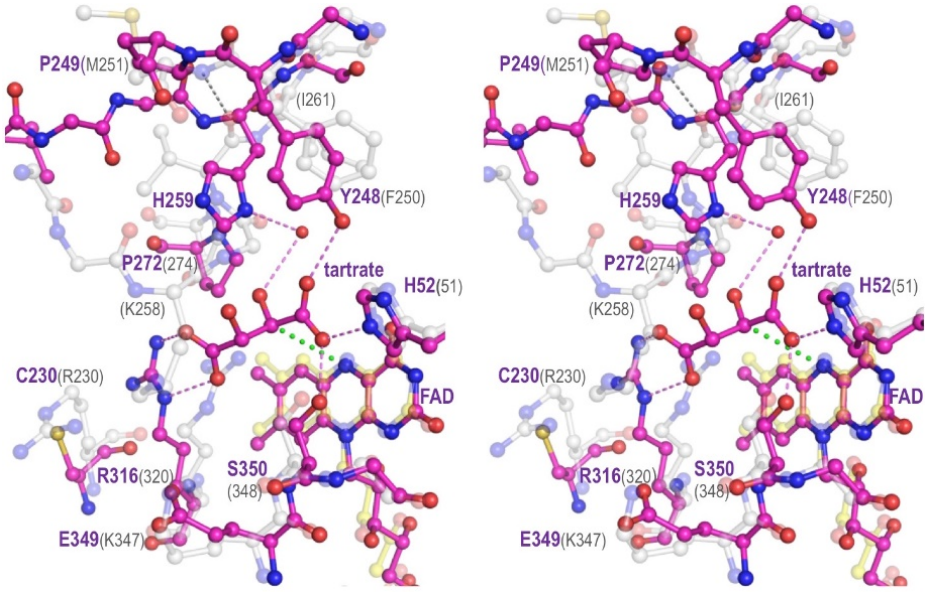


Figure 5.6. Comparison of the *Bp*3253 tartrate binding pocket and *Mp*GlpO.
Stereoview shown of the *Bp*3253-tartrate complex (violet carbons) overlaid on the equivalent region of *Mp*GlpO (semi-transparent off-white carbons for protein and yellow for flavin). Also indicated are H-bonds (grey dashed lines for *Mp*GlpO, violet dashed lines for *Bp*3253-tartrate complex) and the close approach of the tartrate C3 atom to the flavin N5 (green thick dotted line). Select residues in *Bp*3253-tartrate complex (violet) and *Mp*GlpO (grey) are identified. H259 in the *Bp*3253-tartrate complex and I261 in *Mp*GlpO are equivalent residues, but their labels are placed near their respective side chains.

Modeling Glp binding to type II GlpO/DHs

Given the reliability of the β12-β13 loop conformation, we can now consider how *Mp*GlpO binds substrate (Figure 5.7A). Compared with *Bp*3253, the presence of Phe250 and the Gly259 methylene make the pocket above the pyrimidyl portion of the flavin less polar, leading us to speculate that this region would recognize the C1-end of Glp and the region near Arg320 would accommodate the negatively charged phosphoryl group. As to what could provide additional recognition of the phosphoryl, the side chains of Arg230, Lys258 and Lys347 are nearby and appear adjustable (B-factors at the side-chain tips being ~20 $Å^2$ higher than at the backbone). These side chains are also well conserved among the 70 sequences having >40% sequence

identity with *Mp*GlpO (i.e. BLAST e-value $<10^{-80}$): Lys258 is fully conserved; Lys347 is conserved as either a Lys or a Gln; and Arg230 is Arg or Lys in all but four sequences, but in these it is a Leu implying a lesser importance. Given these insights, we created a model for docked Glp guided by the constraints that the C2-hydrogen be oriented for transfer to the flavin-N5, that either the C1- or C2-hydroxyl replaces the water binding to His51 and Ser348 (see Figure 5.2), and the phosphate interacts with Arg320. The binding mode obtained (Figure 5.7A) fortuitously places His51 ideally for serving as a catalytic base that could deprotonate the C2-hydroxyl in the forward reaction.

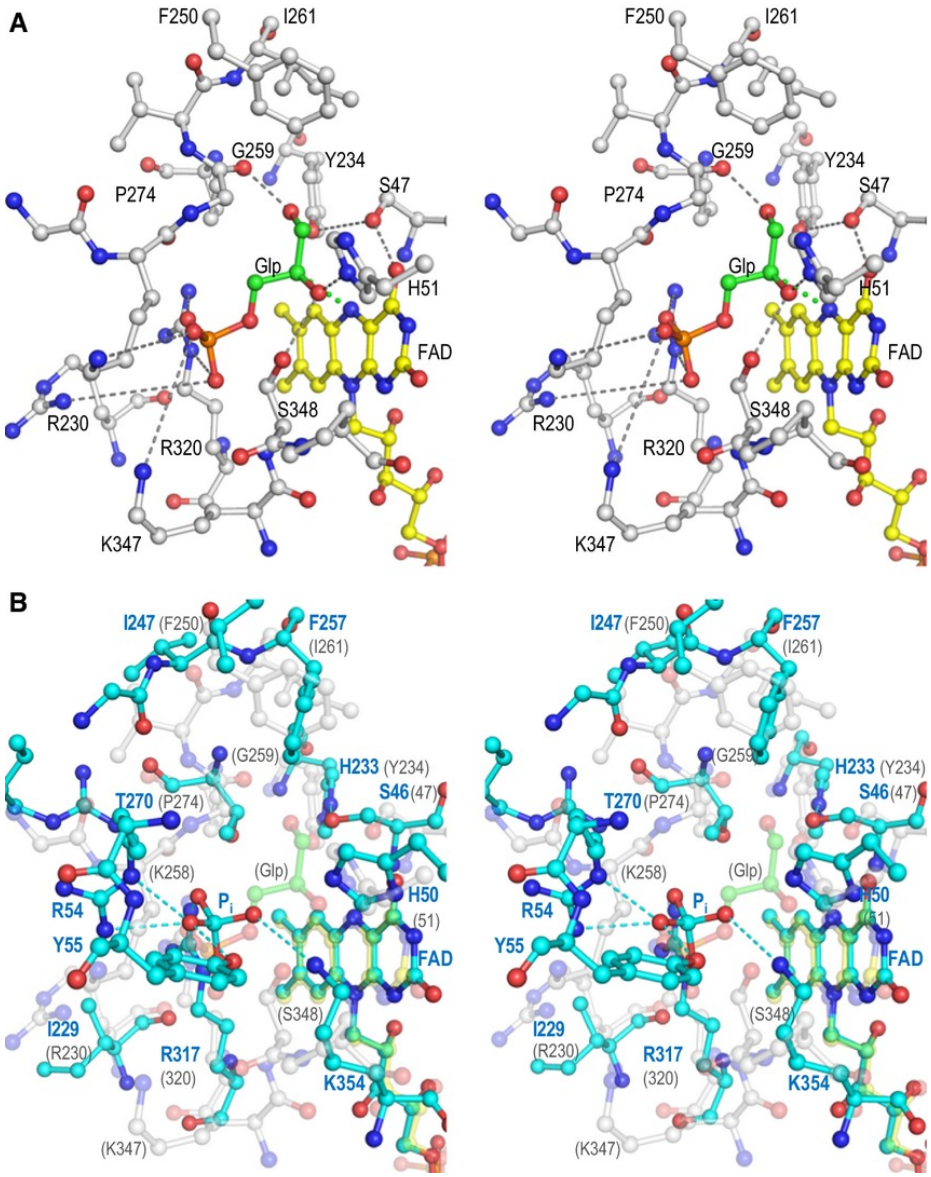


Figure 5.7. A predicted *Mp*GlpO-Glp complex and its comparison with *Ec*GlpD.
A. Stereoview of the *Mp*GlpO active site with a roughly positioned Glp-bound (off-white carbons for protein, yellow for flavin, and green for Glp). H-bonds (grey dashed lines) and the close approach of the modeled Glp C2 atom to the flavin N5 (green thick dotted line) are also shown. Criteria used for placing the Glp are described in the text. **B.** Stereoview of the *Mp*GlpO (colored as in A, but semitransparent and without H-bonds shown) overlaid on the equivalent region of *Ec*GlpD (cyan carbons) with its bound inorganic phosphate (cyan phosphorus) and H-bonds (cyan dashed lines). Select residues in *Ec*GlpD (cyan) and *Mp*GlpO (grey) are identified.

Extrapolation of Glp binding to type I GlpO/DHs

Interestingly, this proposed mode of substrate binding differs from those proposed previously for the type I GlpO/DHs *Ssp*GlpO[38] and *Ec*GlpD,[37] neither of which were very satisfactory. For *Ssp*GlpO, we had proposed that the substrate would be oriented the other way, with the 3'-phosphoryl moiety near the equivalent of *Mp*GlpO His51 and the C1-hydroxyl interacting with the equivalent of *Mp*GlpO Arg320.[38] We had also noted that the equivalent of His51 seemed the only potential base, but how this could be achieved was not clear. For *Ec*GlpD,[37] a phosphate bound in the native crystal structure interacted with Arg317 (equivalent of *Mp*GlpO Arg320), Arg54 and Tyr55, but ligand soaks were taken to indicate binding modes that placed the substrate C2-atom far from flavin-N5.

Given the unsatisfactory nature of those proposals, we explored whether the mode of binding we predict for *Mp*GlpO could also work for the Type I enzymes. An overlay of *Ec*GlpD onto *Mp*GlpO with its docked Glp shows a remarkable compatibility, with the predicted position of the phosphoryl of Glp matching closely with the experimentally observed phosphate of *Ec*GlpD (Figure 5.7B). The ligand fits reasonably into the *Ec*GlpD active site, and it can be seen that the functionality of key *oxid*Lys258 and Lys347 in interacting with the phosphoryl, and Phe257 replacing Phe250 in providing a non-polar environment for the C1-methylene (Figure 5.7B). These functional substitutions are most easily seen in schematic drawings of the interactions (Figure 5.8). We emphasize that whereas the general features associated with this rough placement of Glp are plausible and have explanatory power, the details are not reliably defined because the side chain and backbone positions of protein groups are expected to shift during ligand binding from the positions they adopt in the unliganded *Mp*GlpO structure used to guide the modeling.

Catalytic mechanism and outlook

Given the predicted mode of Glp binding to the *Mp*GlpO (Figure 5.8A) and the *SSp*GlpO (Figure 5.8B) active sites, how the electrons flow during catalysis and how the forward reaction is enhanced by base catalysis become readily apparent. As indicated by the arrows in the Figure 5.8A schematic, in *Mp*GlpO the substrate is

well-aligned for His51 to deprotonate the C2-hydroxyl, promoting its electrons to fold in to form a carbonyl, and facilitating loss of the C2-hydrogen as a hydride that can attack the flavin N5; the ensuing shifting of flavin electrons would lead to the formation of the reduced flavin N1-anionic form, with the negative charge at the N1/O2 locus stabilized by three hydrogen bonds donated by the Leu351 and Thr352 backbone amides and the Thr352 hydroxyl. The geometries of the interactions are also stereoelectronically reasonable, in that the His51 interaction with the C2-hydroxyl has a roughly "anti" orientation with respect to the hydride leaving group. The Ser47 hydroxyl – supported by its interactions with Tyr234 – is well-placed to help stabilize the protonated N5 of the reduced flavin through minor shifts in its position in association with flavin reduction. A set of equivalent interactions exist in the *SSp*GlpO active site (Figure 5.8B).

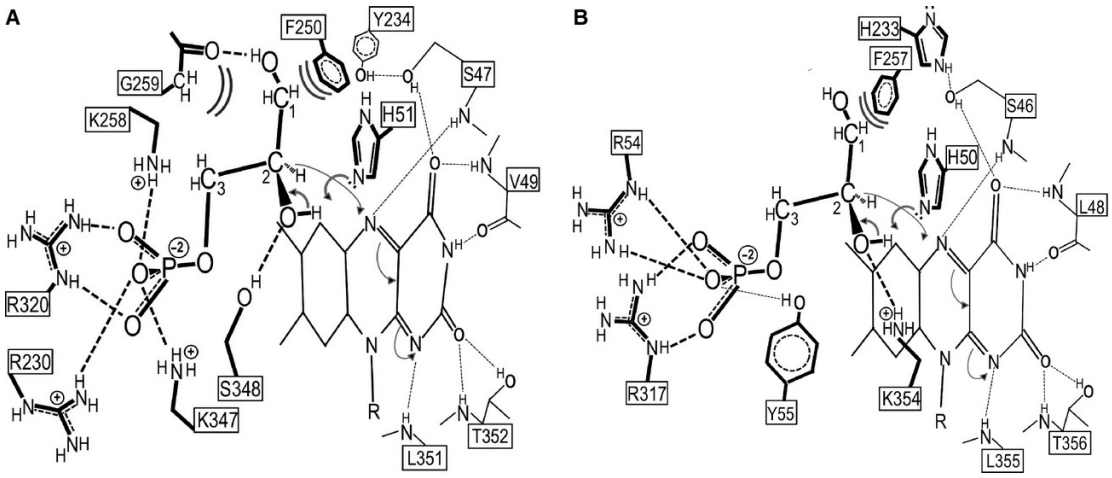


Figure 5.8. Schematic drawings of residues involved in substrate binding and catalysis in *Mp*GlpO and *Ec*GlpD.

A. The *Mp*GlpO active site atoms roughly in the plane of the flavin are shown smaller and with thinner bonds and hydrogen bonds (dashed), and atoms in front of the flavin are shown with thicker bonds and hydrogen bonds (dashed). Residues shown interacting with the flavin and the Glp are in Figure 5.4C highlighted in red and green boxes, respectively. Curved arrows indicate the proposed flow of electrons during the reductive half-reaction. **B.** The same except for showing the *Ec*GlpD active site as a representative of a Type I GlpO.

A full kinetic analysis of *Mp*GlpO has recently been completed,[214] and our structural results allow us to propose explanations for some of the observations made

in that study. One such area is the redox potential difference between the *Mp*GlpO flavin, which at -167 mV is much lower than those seen for type I GlpOs (e.g. *Enterococcus casseliflavus* GlpO at -118 mV; [206]). While many subtle factors can influence redox potential, we note that one very clear difference in the flavin electrostatic environment consistent with this shift is that the Type I GlpO/DH enzymes conserve an active site lysine (equivalent to Lys354 in *Ec*GlpD; Figure 5.8B, 5.7B, 5.4C) having its amino group just at van der Waals distance above the flavin N1/O2 locus, whereas the type II enzymes have a conserved neutral serine side chain in that place (equivalent to Ser348 in *Mp*GlpO). The additional local positive charge in the type I enzymes would make them more easily reduced. As was pointed out by Maenpuen et al, the difference in redox potential may be the reason that *Mp*GlpO can catalyze the reverse reaction (i.e. DHAP oxidation of the reduced flavin) whereas those type I enzymes tested cannot.

Another very interesting result of the kinetics study was the observation of two enzyme populations, partitioning as 70:30 with 70% of the oxidized enzyme reacting rapidly with Glp, and 70% of the reduced enzyme reacting slowly with DHAP in the reverse reaction, but reacting rapidly with O_2. Based on the structure, a plausible single explanation for all of these observations is that the two populations are defined by the protonation state of the catalytic acid/base His-51, with 70% being deprotonated and 30% protonated at the pH of 7 at which the studies were done. For the reaction with Glp, the deprotonated form of His51 is required for abstracting the C2-hydroxyl proton, so 70% of the enzyme would react rapidly. The slower population would reasonably be limited by the rate of deprotonation of His51 in the ligand bound form. For the reaction of the reduced flavin with DHAP, the deprotonated 70% of the enzyme reacts poorly (in bimolecular limited fashion at 56 M^{-1} s^{-1}), while the protonated 30% reacts rapidly enough that the k_{off} of the Glp produced (at 6 s^{-1}) is rate limiting. In terms of the O_2 reactivity of the reduced enzyme, it is reasonable that the His51-protonated/charged and the His51-deprotonated/neutral active sites would react differently and our proposal implies that the deprotonated 70% reacts rapidly with oxygen (at ~600 s^{-1}) while the protonated 30% reacts more slowly (at ~100 s^{-1}). This explanation predicts that the enzyme will

be more active above pH=7, and a set of assays looking at the enzymes pH dependence at a single high substrate concentration shows the pH optimum is near 8. More extensive studies of k_{cat} and K_M as a function of pH are now planned to test these ideas and better define the enzyme mechanism.

The proposed binding mode of Glp also implies the importance of the negatively charged phosphoryl group for substrate recognition, as both the Type I and Type II enzymes have 3 to 4 positively charged side chains involved in its recognition (Figure 5.8). The importance of the phosphoryl is consistent with the binding studies reported by Maenpuen et al[214] showing that *Mp*GlpO effectively binds the negatively charged glyceraldehyde-3-phosphate, lactate, and malate, but not the neutral glycerol. While much remains to be done, we see the most influential contributions of these analyses are two-fold: first, is the recognition of the Type I and Type II GlpO/DH enzymes as distinct variants with striking active site differences that were not predicted from sequence alignments; and second, is the plausible concrete predicted mode for substrate recognition and catalysis that is relevant not only for bacterial GlpO enzymes, but also for the widespread mitochondrial GlpD dehydrogenases, and that will guide mutational studies that test the proposed mechanism and dissect the roles of specific active site residues. Also, the distinction between the Type I and Type II GlpO/DH active sites raises the encouraging possibility that it will be possible to design inhibitors that may selectively block activity of bacterial Type II enzymes such as *Mp*GlpO while not inhibiting the mitochondrial GlpD dehydrogenase of the host.

Experimental Procedures

Expression and Purification of *Mp*GlpO and *Bp*3253

The codon-optimized MPN051 gene encoding *Mp*GlpO was synthesized by GenScript (Piscataway, NJ, USA) and subcloned into the expression plasmid pET28a (Novagen, Darmstadt, Germany). *Mp*GlpO was expressed with an N-terminal His-tag in *E. coli* B834(DE3) cells using autoinduction medium at 28ºC. All steps of purification were conducted at 4ºC. Harvested cells were resuspended in 50 mM potassium phosphate, pH 7.0, containing 200 mM NaCl, 20 mM imidazole, 0.5 mM

4-(2-aminoethyl)benzenesulfonyl fluoride (AEBSF) protease inhibitor and 10 % glycerol. Cells were disrupted using an Avestin EmulsiFlex-C5 homogenizer, and centrifuged (27,000 g for 60 min). The clarified extract was loaded onto a 25 mL Co Sepharose High Performance column (GE Healthcare, Piscataway, NJ, USA), and *Mp*GlpO protein was eluted with 0.5 M imidazole after washing with 20 mm imidazole. The pooled yellow fractions were dialyzed overnight against 50 mM potassium phosphate, pH 7.0, containing 0.5 mM EDTA. FAD (0.25 mM) was added to the enzyme and incubated for 45 min on ice before loading on a 75-mL SP-Sepharose HP column (GE Healthcare). The column was washed with 50 mM potassium phosphate, pH 7.0, 0.5 mM EDTA, 100 mM NaCl before *Mp*GlpO was eluted with a 100 mM to 1 M NaCl gradient. Fractions were analyzed by SDS-PAGE, pooled, buffer-exchanged into 50 mM potassium phosphate, pH 7.0, 0.5 mM EDTA and concentrated to 10 mg/mL before freezing in aliquots at -80ºC.

The pET21_NESG clone for expression of His-tagged *Bp*3253 (corresponding to PDB entry 3DME) was purchased from the DNASU Plasmid Repository, and its expression and purification followed the protocol described above for *Mp*GlpO, with the following modifications. Recombinant *E. coli* B834(DE3) cells were grown in TYP medium, prior to induction with 0.5 mM IPTG and overnight protein expression at 16 °C. The cells were broken in a solution containing 50 mM Tris-Cl, pH 8.0 (4 °C). Nucleic acids were removed by adding 2% (w/v) streptomycin sulfate and centrifugation. Crude extract was loaded to the Co column in 50 mM sodium phosphate pH 8.0, 300 mM NaCl, 20 mM imidazole, the column was washed with more buffer, and pure *Bp*3253 was eluted with 500 mM imidazole. *Bp*3253 protein was concentrated to 10 mg/ml in an Amicon ultrafiltration cell with a YM30 membrane and for freezing was buffer-exchanged into 50 mM potassium phosphate, pH 7.0, containing 0.5 mM EDTA.

Biochemical characterizations

Extinction coefficients for His-tagged *Mp*GlpO and *Bp*3253 were determined by standard methods, using an Agilent model 8453 diode-array spectrophotometer. The specific activities of the two proteins were measured as described earlier for

*Ssp*GlpO,[207] using the standard spectrophotometric assay with a Cary 50 spectrophotometer (Varian) thermostatted at 25ºC. The native M_r for recombinant His-tagged *Mp*GlpO was determined by gel filtration at 25ºC, in a pH 7 phosphate buffer with 150 mM NaCl, using a Sephadex G-200 medium HR 10/30 column calibrated with six standard proteins covering the range 15,600 to 440,000. Titrations of both *Mp*GlpO and *Bp*3253 proteins with sulfite followed established protocols.[206-207]

Crystallization and Data Collection

Thawed aliquots of wild-type His-tagged *Mp*GlpO were subjected to a variety of crystal screens at 4 ˚C and the most promising crystal leads grew using a reservoir of 2.68 M NaCl, 3.35% v/v isopropanol, and 0.1 M HEPES pH 7.5 (Wizard Precipitant Synergy Screen Block 1 condition B7). Optimization at 4 ºC using hanging-drops led to yellow, trigonal-pyramidal crystals measuring approximately 0.3 x 0.3 x 0.3 mm^3 growing within one week using drops made from 1 µL protein at 5 mg/ml plus 2 µL of a reservoir solution as above but containing 2% (v/v) isopropanol. For data collection, crystals in an artificial mother liquor (AML) of 3 M NaCl in 0.1 M Bis-Tris, pH 7.0 were placed for 3 minutes in AML with 15% glycerol as a cryoprotectant before being flash-frozen by plunging into liquid nitrogen. Data sets to first 2.50 Å and then 2.40 Å resolution were collected at beamline 5.0.3 at the Advanced Light Source synchrotron. For analyses of additional forms of *Mp*GlpO, crystals were soaked in either 10 mM dithionite, 10 mM L-Glp, 10 mM L-tartrate, 10 mM 2-phosphoglycerate, or 10 mM phosphoenolpyruvate (PEP) in AML for 1 h prior to freezing and data collection. Two experiments with dithionite-soaked crystals gave a complete merged data set to 2.50 Å. Data sets at between 2.5 and 3.0 Å resolution were collected for the ligand soaks.

*Mp*GlpO Phasing and Structure Refinement

Diffraction data for oxidized *Mp*GlpO were processed with iMOSFLM[194] and with SCALA in the CCP4 suite.[195, 215] *Mp*GlpO crystallizes in space group *P*23, and there is one molecule per asymmetric unit. Attempts to solve the structure by molecular replacement using AMoRe[216] with the *Ssp*GlpO coordinates were not successful.

However, using chain A of PDB entry 3DME – annotated as the "conserved exported protein (BP3253) from *Bordetella pertussis*" – as a search model gave a solution using the 2.50 Å data set and AutoMR.[217] Manual rebuilding of the initial model was carried out in COOT,[198] and this model was refined with the PHENIX software suite.[199] Six rounds of simulated annealing and minimization refinement gave a partially refined model with R/R_{free} of 32%/45%, respectively. At this point, the improved 2.40 Å resolution *Mp*GlpO data were used, and six rounds of refinement with BUSTER[152] led to an R_{free} of 28%. Water molecules were added, as indicated by both electron density peak height and hydrogen-bonding interactions, and refinement continued with REFMAC.[218] A nickel ion (see "Results") was introduced late in the process and was confirmed by X-ray fluorescence, and further manual modeling and refinement led to the final oxidized *Mp*GlpO structure with R/R_{free} 15.8%/16.4%. The model does not include the Lys79 and Lys298 side chains beyond -CB, as these surface residues have little-or-no side chain density. Alternate side chain and/or backbone conformations are included for Gln40, His244, and the Trp375-Asn376-Gly377 backbone.

The 2.50 Å refinement of the dithionite-reduced GlpO structure began from the oxidized structure. Rounds of manual modeling and REFMAC refinement gave a model with only minor changes, and with R/R_{free} values of 16.4%/22.6%. Of the soaks with Glp, 2-phosphoglycerate, PEP, and tartrate, no substantive interpretable density differences were observed in the vicinity of the isoalloxazine and so these structural analyses were not pursued further.

Acknowledgements

We thank Dale Tronrud for helpful guidance with the crystallographic work. This work was supported by North Carolina Biotechnology Center Grant 2011-MRG-1116 (to A.C.), by Thailand Research Fund MRG5580066 (to S.M.), by Thailand Research Fund RTA5680001 (to P.C.), and Howard Hughes Medical Institute grant 52005883 (to C.K.E.). Synchrotron data were collected at the Advanced Light Source, supported by contract DE-AC02-98CH10886 from the Office of Basic Energy Sciences of the U.S. Department of Energy.

Chapter 6

High resolution studies of hydride transfer in the ferredoxin:$NADP^+$ reductase superfamily

Kelsey M. Kean*, Russell A. Carpenter*, Vittorio Pandini, Giuliana Zanetti, Andrea R. Hall, Rick Faber, Alessandro Aliverti, and P. Andrew Karplus

*These authors contributed equally to this work

Published in *The FEBS Journal, 2017, 19 (21), pp 3302–3319*

Abstract

Ferredoxin-$NADP^+$ reductase (FNR) is an FAD-containing enzyme best known for catalyzing the transfer of electrons from ferredoxin (Fd) to $NADP^+$ to make NADPH during photosynthesis. It is also the prototype for a broad enzyme superfamily, including the NADPH oxidases (NOXs) that all catalyze similar FAD-enabled electron transfers between NAD(P)H and one-electron carriers. Here we define further mechanistic details of the NAD(P)H ⇌ FAD hydride-transfer step of the reaction based on spectroscopic studies and high resolution (~1.5 Å) crystallographic views of the nicotinamide-flavin interaction in crystals of corn root FNR Tyr316Ser and Tyr316Ala variants soaked with either nicotinamide, $NADP^+$, or NADPH. The spectra obtained from FNR crystal complexes match those seen in solution and the complexes reveal active site packing interactions and patterns of covalent distortion of the FAD that imply significant active site compression that would favor catalysis. Furthermore, anisotropic B-factors show that the mobility of the C4 atom of the nicotinamide in the FNR:$NADP^+$ complex has a directionality matching that expected for boat-like excursions of the nicotinamide ring thought to enhance hydride transfer. Arguments are made for the relevance of this binding mode to catalysis, and specific consideration is given to how the results extrapolate to provide insight to structure-function relations for the membrane-bound NOX enzymes for which little structural information has been available.

Introduction

Plant ferredoxin $NADP^+$ oxidoreductases (FNR) are flavoenzymes that catalyze the reversible electron transfer between NADP(H) and the iron-sulfur protein ferredoxin (Fd). In photosynthesis, plastidic FNRs serve to catalyze electron transfer from Fd to $NADP^+$ via the FAD prosthetic group, and in non-photosynthetic tissues distinct FNRs use NADPH as an electron source for the reverse reaction.[219-221] Plant-type FNRs – as seen in three-dimensional structures of FNRs from cyanobacteria,[222] spinach (leaf [223]), pea (leaf [59]), paprika (leaf [224]), and corn (leaf [225] and root [56]) – are highly similar and contain two distinct domains: one for binding FAD and one for binding $NADP^+$.[55] Importantly, this two-domain FNR-fold is also the prototype for an

enzyme superfamily that uses a bound FAD or FMN to transfer redox equivalents between the hydride carrying NAD(P)$^+$ cofactors and diverse one-electron carriers [55, 226-227]. Included among the FNR superfamily are the NADPH oxidases (NOXs), a biomedically important group of membrane bound enzymes that produce superoxide or hydrogen peroxide as part of many signaling pathways as well as during the oxidative burst of macrophages that is a key part of our immune defenses (recently reviewed in [228-231]).

Structures of plant type FNRs and other superfamily members, such as nitric oxide synthase,[232] cytochrome P450 reductase,[64, 66] and phthalate dioxygenase reductase [233] showed that NADP(H) appears to bind non-productively to the wild-type enzymes because an aromatic side chain (a C-terminal Tyr residue in the case of plant type FNRs) sits in the site the nicotinamide ring must occupy for hydride transfer (Figure 6.1A). Deng *et al.* [59] resolved this mystery by showing for pea leaf FNR at 1.8 Å resolution that a mutant missing the aromatic placeholder residue (Tyr308Ser or Y308S) could bind NADP(H) tightly and in an apparently productive manner with the nicotinamide ring C4 atom adjacent to the FAD N5 atom and with a geometry reasonable for hydride transfer. The Y308S mutant bound its cofactor so tightly that NADP$^+$ co-purified with the enzyme. This mutant also showed an unexpected 500-fold change in cofactor specificity as the stronger binding of NAD(H) *increased* its steady-state turnover,[58] but the stronger binding of NADP(H), *decreased* its steady state turnover, not because of impaired hydride transfer but because the NADP binding was so tight that k_{off} became rate-limiting.[58, 222]

With the support of spectroscopic studies, it was concluded that the wild-type enzyme binds NADP(H) in a bipartite fashion in which the 2'-P-AMP binds strongly to anchor the cofactor to the enzyme and the thermodynamics of nicotinamide displacing the aromatic placeholder is such that the nicotinamide ring remains largely in a solvent exposed conformation ("nicotinamide-out") and is in rapid equilibrium with a smaller population of molecules in which the aromatic placeholder swings out and the nicotinamide ring takes its place ("nicotinamide-in" conformation) to allow for hydride transfer. The Tyr to Ser mutant changes the binding thermodynamics so that the nicotinamide is ~100% "in". Since that report, studies of aromatic blocker

variants of other superfamily members have shown a consistent geometry for the "nicotinamide-in" nicotinamide-flavin interaction [64, 222] (Figure 6.1B) as well as changes from 50 to 1000-fold in $NADP^+$ *vs* NAD^+ specificity.[60, 65, 222, 234-235]

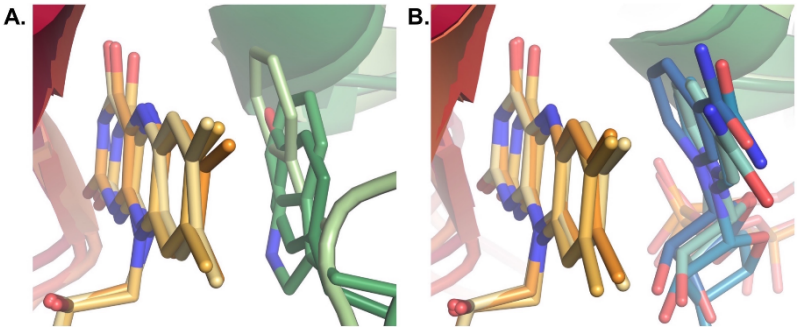


Figure 6.1 Aromatic placeholder and nicotinamide binding in three FNR superfamily members**.**
(A) Stacking of aromatic placeholder side chain onto the flavin in wild-type forms of spinach FNR (PDB 1FNC; Tyr314 in forest green, FAD in orange), cytochrome P450 reductase (PDB 1AMO; Trp677 in green, FAD in dandelion), and NO synthase (PDB 1F20; Phe1395 in pale green, FAD in sand). **(B)** Mode of nicotinamide binding in the $NADP^+$ complexes of Y303S *Anabaena* FNR (PDB 2BSA; $NADP^+$ in blue), W677X cytochrome P450 reductase (PDB 1JA0; $NADP^+$ in cerulean), and Y303S pea FNR (1QFY; $NADP^+$ in seafoam); FAD colors as in panel A.

Although there is considerable structural information about FNRs and FNR-like modules in related enzymes, the mechanism of hydride transfer is still not fully understood. For instance, the previous NADP(H) complexes of the pea FNR mutants gave indications that the C4 in the nicotinamide ring of $NADP^+$ was mobile in a way that could favor hydride transfer,[59] but higher resolution data are needed to better define the details of this mobility. Also, it has been proposed that the geometry of the complex seen in the structures that have the aromatic placeholder mutated are artifacts that are not relevant to catalysis.[236-238] Here, we obtain further insight into the hydride-transfer step in FNR-like enzymes by combining spectroscopic studies and a series of higher resolution structures of FNR crystals soaked with nicotinamide, $NADP^+$, and NADPH. To accomplish this, we chose corn root FNR, for which crystals of the wild-type enzyme diffract to near 1 Å resolution.[57] We also argue for the general relevance of these insights for broader members of the FNR superfamily, including the NOX enzymes, which is further supported by the very recent structures

of the core catalytic subunit domains of NOX5, published while this work was under review.[239]

Results and Discussion

Strategy

An inspection of the molecular packing in the well-diffracting crystals of corn root FNR that we reported earlier [56-57] revealed that the NADP(H) binding site was not involved in crystal packing interactions. We hypothesized that we could obtain high resolution structures of NADP(H) complexes of this enzyme by making mutants of Tyr316, the aromatic blocking residue. Not knowing which mutant would be more informative, we mutated Tyr316 to Ala (Y316A), Ser (Y316S), and Phe (Y316F). Crystal structures of $NADP^+$ soaks of Y316F FNR showed no nicotinamide binding, so this variant was not subjected to further study. For Y316S and Y316A, however, we present extensive characterization of the in-solution spectroscopic and catalytic properties, as well as crystal structures and in some cases the in-crystal spectroscopic properties of their complexes with NADP(H).

Solution Properties of Mutants

Wild-type and mutant enzyme forms were produced in *Escherichia coli* and purified essentially as reported elsewhere.[56] The two protein variants were expressed at slightly lower levels than that the wild-type enzyme, and their purification required additional hydrophobic interaction chromatography on butyl Sepharose. This resulted in a substantial drop in overall yields, which in the case of Y316A was low enough to preclude its full functional characterization. As shown in Figure 6.2A, the two replacements induced very similar perturbations in the visible absorption spectrum of the flavoprotein, strongly reminiscent of those observed in the equivalent Tyr to Ser replacement in both pea and *Anabaena* FNRs.[56, 240] Phenol – a mimic of the Tyr side-chain – was found to induce a spectral change in Y316S that qualitatively matched the difference between the absorption spectra of the wild-type and the variant protein (Figure 6.2B), implying that it can stack against the isoalloxazine ring in the active-

site pocket. However, the affinity of Y316S FNR for phenol was too low to allow the accurate determination of the K_d of the complex.

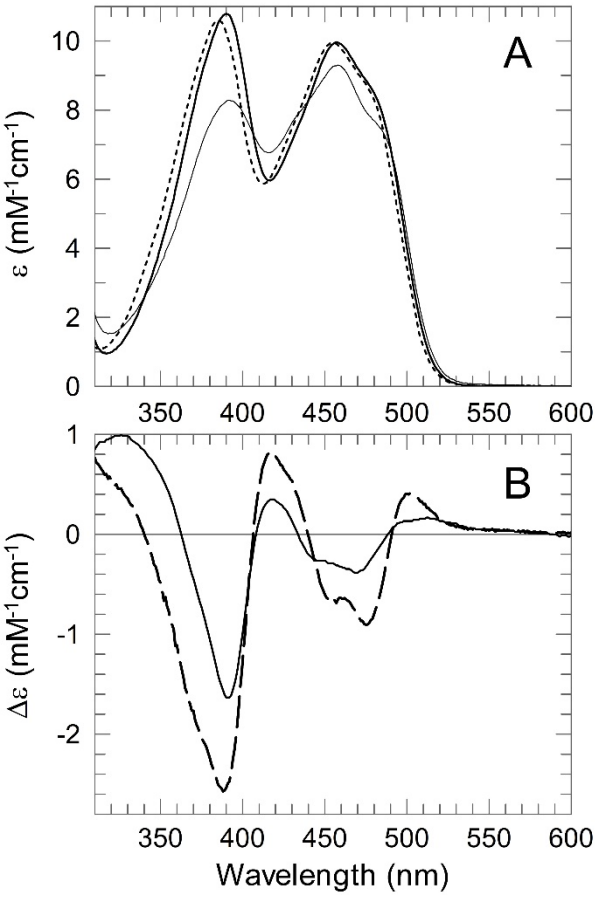


Figure 6.2 Effect of the replacements of the Tyr316 residue on the visible absorption spectrum of FNR and rescue of the wild-type spectral features by phenol binding to the Y316S variant.

All spectra were determined in 50 mM Tris-HCl, pH 7.4 at 25 °C. (**A)** Extinction coefficients in the visible region of Y316S (solid line) and Y316A (dotted line) FNRs in comparison to that of wild-type enzyme (thin line). (**B)** Perturbation of the visible absorption spectrum of the Y316S variant (ca. 9 μM) induced by the presence of 150 mM phenol. The difference spectrum was rescaled (solid line) to match that expected for a theoretical enzyme concentration of 1 mM in order to be compared to the difference between the extinction coefficients of wild-type and Y316S FNR forms (dashed lines).

Titrations with NADP(H). Anaerobic titration of oxidized Y316S FNR with $NADP^+$ induced perturbations in the visible absorption spectrum of the FAD prosthetic group (Figure 6.3) very similar to those observed for the Y308S mutant of pea FNR.[59] The K_d of the complex was estimated to be ≤0.02 μM, well below that of the wild-type

complex (0.3 μM).[56] Whereas no interaction between NAD^+ and wild-type FNR was detectable spectrophotometrically, NAD^+ induced in Y316S FNR a spectral shift virtually identical to that produced by $NADP^+$, displaying a K_d of 145 ± 5 μM (Figure 6.3).

Next, using an EDTA-deazariboflavin system, stepwise anaerobic photoreductions of the FAD group of wild-type and mutant FNR forms were done in the absence and in the presence of roughly equimolar amounts $NADP^+$ or NAD^+ (Figure 6.4). These showed that the Y316S replacement decreased the amount of FAD semiquinone accumulated during the process and greatly increased the intensity of the broad long wavelength (~800 nm) absorption band attributable to charge-transfer electronic transitions of the interacting oxidized nicotinamide and 2-electron reduced flavin.[56] Upon complete reduction of the system, the charge-transfer band disappeared implying that the limited amount of $NADP^+$ present became reduced and no charge transfer interaction occurs between the reduced flavin and NADPH. Furthermore, the presence of $NADP^+$ favored the protonation of the dihydroquinone form of the prosthetic group of Y316S FNR, as indicated by the shift from 360 to 420 nm of the local maximum of the spectra of the fully reduced species (Figure 6.4). The results with NAD^+ (not shown) are qualitatively similar to those obtained with $NADP^+$, although a significantly less intense charge-transfer band was observed, consistent with the lower affinity of the enzyme for NAD^+ as compared to $NADP^+$.

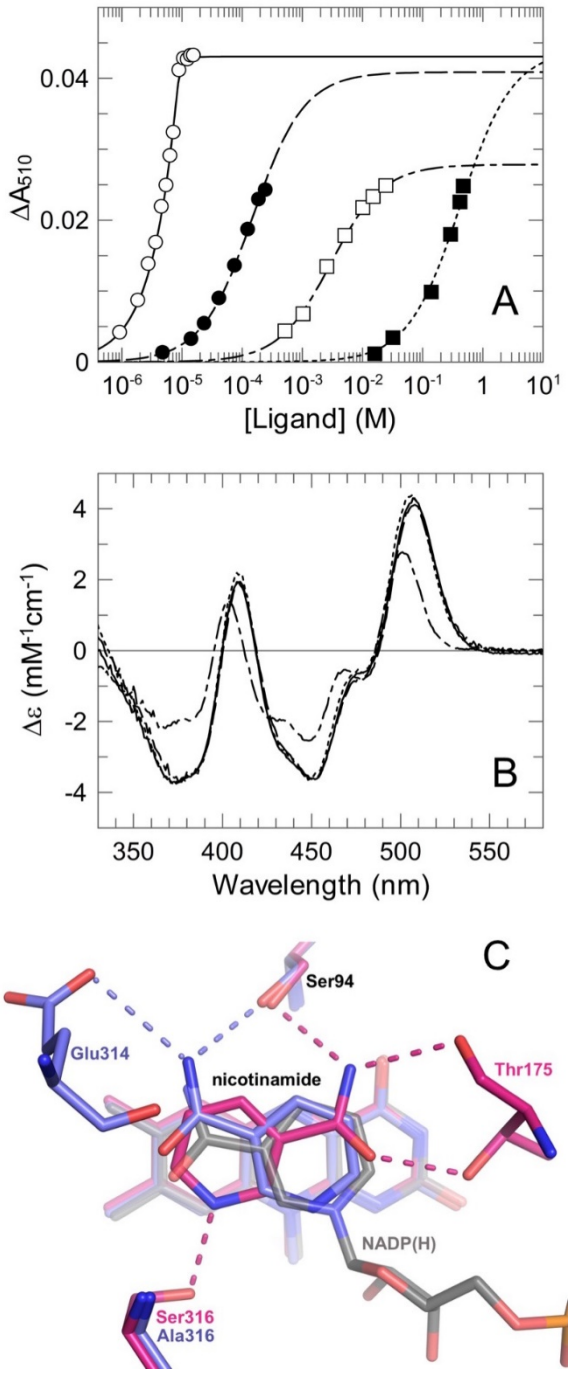


Figure 6.3 Interaction of Y316S FNR with nicotinamide-containing ligands.
(A) Progress of the titrations of 10 μM Y316S FNR with $NADP^+$ (hollow circles), NAD^+ (filled circles), nicotinamide (hollow squares), and methyl-nicotinamide (filled squared). All titrations were performed in Tris-HCl, pH 7.7, at 15 °C. Data points and fitting curves were corrected to account for protein dilution. The logarithmic concentration scale was chosen to allow comparison of ligands displaying huge differences in affinity. (**B)** Computed difference spectra in the visible region of protein-ligand complexes. The difference between extinction coefficients of complexed and uncomplexed protein forms is shown for $NADP^+$ (solid line), NAD^+ (dashed line), nicotinamide (dash-dotted line), and methyl-nicotinamide (dotted line)

are shown. (**C**) Overlay of $Y316A_{nic}$ (slate) and $Y316S_{nic}$ (hot pink) with nicotinamide bound, and as a reference, bound $NADP^+$ (ghosted grey). Hydrogen bonds to bound nicotinamide are shown (slate and hot pink dashed lines in the respective structures).

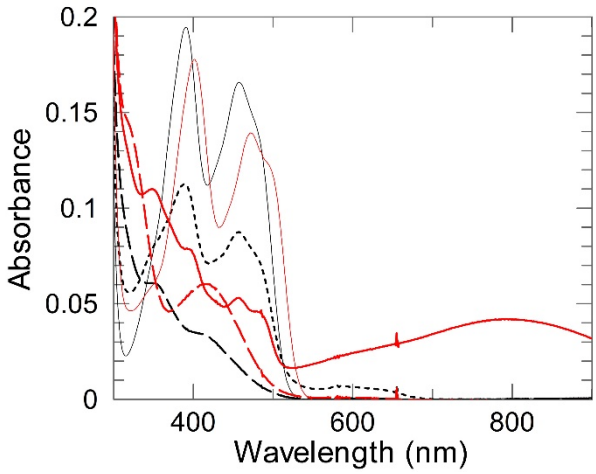


Figure 6.4 Effect of the presence of $NADP^+$ on photoreduction of Y316S FNR. Significant spectra recorded during the stepwise photoreduction of the FAD prosthetic group in the absence (black lines) and in the presence of a slight excess of $NADP^+$ (red lines). About 16 μM protein solutions in 10 mM HEPES, pH 7.0, were photoreduced at 15 °C in the presence of an EDTA-deazariboflavin system. Spectra corresponding to the maximal accumulation of semiquinone (dotted line) and charge-transfer complex (solid line) species are shown, and compared to those of the oxidized (thin lines) and fully reduced (dashed lines) respective enzyme mixtures.

Anaerobic titrations of wild-type FNR and its variants with NADPH were performed to further analyze the interactions between FAD and NADP(H). In such titrations, at one equivalent of NADPH added, for the "nicotinamide-in" species an equilibrium should exist between two charge transfer complexes depending on whether the nicotinamide or the flavin is in the reduced state. Charge-transfer complex 1 (CTC-1) has the nicotinamide ring of NADPH interacting with oxidized flavin (*i.e.* FNR_{ox}:NADPH) and an absorption band with a maximum near 600 nm, and charge transfer complex-2 (CTC-2) has the nicotinamide ring of $NADP^+$ interacting with reduced flavin (*i.e.* FNR_{red}:$NADP^+$) and a broad absorption band with a maximum at ca. 800 nm.[241-242] The spectra at one equivalent of NADPH show that for wild-type the CTC-1 absorption band dominates (Figure 6.5A), whereas for both mutants the CTC-2 band dominates (Figure 6.5B,C), apparently accounting for ~90% of the population, based on the level of reduced flavin reduction as indicated by A_{460}.

For the two mutants, the CTC-2 band is maximal at roughly one equivalent NADPH and decreases to about half of that height as the concentration of NADPH is increased to an ~135-fold excess (Figure 6.5B,C). The further additions of NADPH presumably result in the progressive decrease in the amount of CTC-2 due to the partial displacement of $NADP^+$ by NADPH. For the wild-type enzyme, the amount of CTC-1 similarly decreases as NADPH concentration increases, but it is progressively replaced by a species attributable to a complex between NADPH and FNR carrying FAD semiquinone. We do not understand how the semiquinone is formed under these conditions, but wonder if it may reflect some flavin disproportionation between the CTC-2 and CTC-1 complexes or is the result of a small amount of residual oxygen or other one-electron acceptor in the titrating NADPH solution. Prolonged incubation of NO synthase with NADPH was also seen to promote semiquinone buildup.[243]

Spectral properties during turnover as an NADPH oxidase. To monitor the spectral changes that occur during steady-state NADPH oxidase activity of FNR, air was admitted into the anaerobic cell after the titration was complete and NADPH was >100-fold in excess. In each case, the enzyme underwent turnover until the NADPH was all converted to $NADP^+$. For wild-type FNR (Figure 6.5D), as has been seen before [56], there was substantial stabilization of the blue semiquinone form of FAD in the early parts of the reaction and CTC bands were hardly detectable. For Y316S FNR (Figure 6.5E), less FAD semiquinione built up and the reaction had two distinct phases. During the first phase, the CTC-2 concentration progressively increased until at ~15 min it reached ~65% of the maximal value observed during the anaerobic titration (thin red line), then it decreased to near zero. The final spectrum (at ~1.7 h) had the 456 nm band shifted as is characteristic of an oxidized enzyme in complex with $NADP^+$ (Figure 6.5E). The behavior of Y316A FNR (Figure 6.5F) was qualitatively similar to that of the Y316S form, but with the reaction slowed by about 10-fold as the steady-state CTC-2 formation was reached after about 2.5 h (*vs* ~15 min) and the reaction was complete in 17 rather than 1.7 h. The Y316A variant also had still less stabilization of the FAD semiquinione, and its maximal CTC-2 accumulation during turnover was ~85% of the highest amount observed in the anaerobic titration (thin red line). The slower reaction and higher CTC-2 buildup are

consistent with the k_{off} for $NADP^+$ being slower for the Y316A mutant. Also, the destabilization of the semiquinone seen in these variants mirrors what was seen for the corresponding mutants of *Anabaena* FNR [240] and NO synthase.[60, 243]

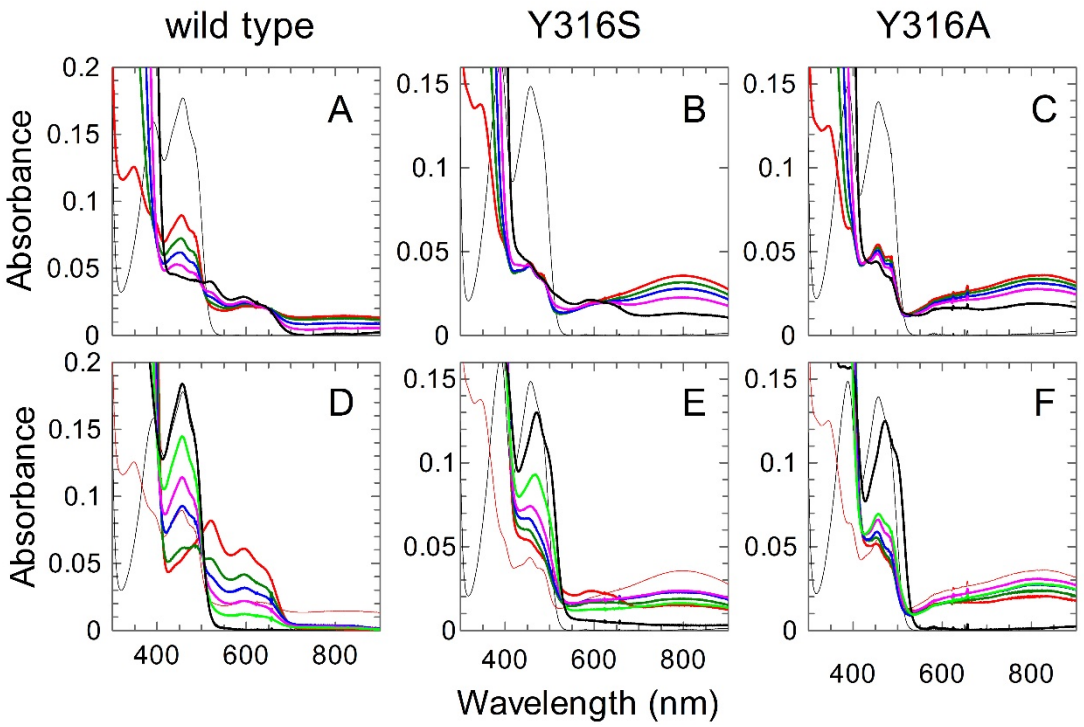


Figure 6.5 Spectra of wild-type FNR and its variants recorded during anaerobic titrations with NADPH and during NADPH-O_2 turnover.
(Upper panels) titrations of wild-type **(A)**, Y316S **(B)** and Y316A **(C)** FNRs showing the spectra of the free oxidized enzymes (thin black traces) and those recorded after the addition of ca. 1.2, 2.5, 5, 12, and 120-fold excess of NADPH (red, green, blue, magenta, and black bold traces, respectively). **(Lower panels)** Spectra recorded at different reaction times after air was admitted into the cuvettes containing the previously anaerobic solutions of the FNR forms in the presence of 120-fold excess NADPH. The spectra of the free oxidized FNR forms (thin black traces) and those of the respective CT complexes recorded in the presence of 1.2-fold NADPH excess (thin red traces) are reported for comparison. **(D)** Spectra recorded after 15 s, 25 min, 35 min, 40 min, 45 min, and 60 min (red, green, blue, magenta, lime, and black bold traces, respectively) turnover catalyzed by wild-type FNR. **(E)** Spectra recorded after 2.5 min, 6 min, 11 min, 19 min, 46 min, and 100 min (red, green, blue, magenta, lime, and black bold traces, respectively) turnover catalyzed by Y316S FNR. **(F)** Spectra recorded after 1 min, 4 min, 53 min, 2.5 h, 3.5 h, and 17 h (red, green, blue, magenta, lime, and black bold traces, respectively) turnover catalyzed by Y316A FNR.

Ferricyanide reductase activity. For the wild-type enzyme, ferricyanide is the most effective electron acceptor from reduced FNR, with k_{cat} = 520 s^{-1}.[56] The Y316S

and Y316A replacements impaired the steady-state diaphorase activity of FNR to a similar extent, and k_{cat} and K_m^{NADPH} values determined for the Y316S variant were ca. 170 and 9-fold lower than those of the wild-type enzyme (Table 6.1),[56] respectively. Such large catalytic impairment parallels that reported for the Tyr to Ser variants of pea and *Anabaena* FNRs,[58, 240] with the slower turnover shown to be limited by the rate of $NADP^+$ (*i.e.* product) release. Consistent with this, we found that nicotinamide partially rescued the diaphorase activity of Y316S FNR in a concentration-dependent fashion, increasing its k_{cat} to 280 s^{-1} at 800 mM nicotinamide (Table 6.1). As has been noted previously,[59, 220] by competing with the corresponding moiety of $NADP^+$ for stacking onto the flavin, nicotinamide will favor product dissociation and speed enzyme turnover if the k_{off} of $NADP^+$ is rate limiting.

Table 6.1 Effect of the Y316S replacement on the kinetic parameters of the NADPH—$K_3Fe(CN)_6$ reductase reaction catalyzed by FNR.[a]

FNR form	k_{cat} s^{-1}	K_m^{NADPH} μM	k_{cat}/K_m^{NADPH} s^{-1} μM^{-1}
Wild type[b]	520 ± 10	12 ± 1	43 ± 4
Y316S	3 ± 0.3	1.4 ± 0.3	2.1 ± 0.5
Y316S, 0.8 M nicotinamide	280 ± 18	45 ± 10	6 ± 1

[a]Initial rate data were measured as moles of ferricyanide reduced per second.
[b]Data on wild type FNR, taken from,[56] are reported here for comparison.

Interactions with small-molecule ligands. The findings that added nicotinamide increased the catalytic activity of Y316S FNR, prompted us to quantify the affinities of nicotinamide and N-methyl-nicotinamide. As shown in Figure 6.3A, both were found able to interact with Y316S FNR, although only the former had an affinity high enough for its K_d value to be determined (3 ± 0.1 mM). The difference spectrum of the Y316S-nicotinamide complex was blue-shifted by ca. 7 nm with respect to that of the $NAD(P)^+$ complexes, but the spectral change induced by N-methyl-nicotinamide was almost identical to that induced by $NAD(P)^+$ (Figure 6.3B). This indicates that for Y316S FNR, N-methyl-nicotinamide but not nicotinamide is an excellent mimic of the redox-active moiety of the dinucleotides.

Reanalysis of Stopped-Flow Results Reported for *Anabaena* FNR.

As noted in the introduction, in earlier work Lans *et. al.* [236] called into question the relevance of the NADP(H) complexes that form in FNR when the aromatic placeholder is mutated. The origin of this view came from stopped-flow results for the Tyr303Ser (Y303S) mutant of *Anabaena variabilis* FNR. Per their interpretation, these results showed that Y303S promoted the hydride transfer from NADPH to FNR_{ox} "slightly slower than the wild-type enzyme" but that "the reverse process is undetectable" (*e.g.* hydride transfer from FNR_{red} to $NADP^+$ does not occur). This interpretation was based on the observation that when $NADP^+$ was mixed with Y303S FNR_{red} the CTC-2 (~800 nm) charge transfer band that formed upon mixing did not substantially decrease over time (see Figure 3A,B of Lans *et al.* [236]). While the conclusion that no hydride transfer occurs is consistent with the unchanging CTC-2 band, there could be other explanations. We were skeptical of the inference both based on the thermodynamic principle that if the Y303S mutant catalyzed hydride transfer from NADPH to FNR_{ox} it must equally well catalyze the reverse hydride transfer from FNR_{red} to $NADP^+$, and based on our photoreduction studies (Figure 6.4 above) implying that Y316S-bound $NADP^+$ is converted to NADPH. To clarify this question, we reanalyzed the Lans *et al.* [236] stopped-flow data for NADPH mixed with Y303S FNR_{ox} and for $NADP^+$ mixed with Y303S FNR_{red}.

Importantly, in the original work, when $NADP^+$ is mixed with Y303S FNR_{red}, even though the 800 nm band changes little over time, the ~460 nm peak systematically increases indicating that oxidation of FAD is occurring. Lans *et al.* attributed this to a "side effect" rather than enzymatic activity; however, when comparing these spectra with those for NADPH mixed with Y303S FNR_{ox}, it is striking that both reactions appear to reach the same endpoint, as would be expected if Y303S FNR were reaching the same equilibrium state independent of the direction of approach. Given this observation, we reanalyzed the reported stopped-flow results for both reactions using the changes in A_{460} as an indicator of FAD oxidation/reduction. This reanalysis yielded excellent fits of k_{obs} of 164 ± 16 s^{-1} for $NADP^+$ mixed with Y303S FNR_{red} and 146 ± 21 s^{-1} for NADPH mixed with Y303S FNR_{ox} (Figure 6.6). These two k_{obs} values should both equal the sum of the

elementary forward and back reaction rate constants ($k_{obs} = k_1+k_{-1}$) and are equal within the error of the analysis, leading us to conclude that the forward and back reactions both function as would be expected. We suggest the best number to use for k_{obs} for the *Anabaena* Y303S FNR is 190 ± 15 s^{-1} which is the value originally reported by Lans *et al.* [236] for the forward reaction using the much more extensive original dataset.

This revised interpretation is consistent with all the data and with basic thermodynamic principles, and implies that the nicotinamide-flavin interaction consistently formed in the aromatic placeholder mutants is in fact productive and relevant for understanding catalysis.

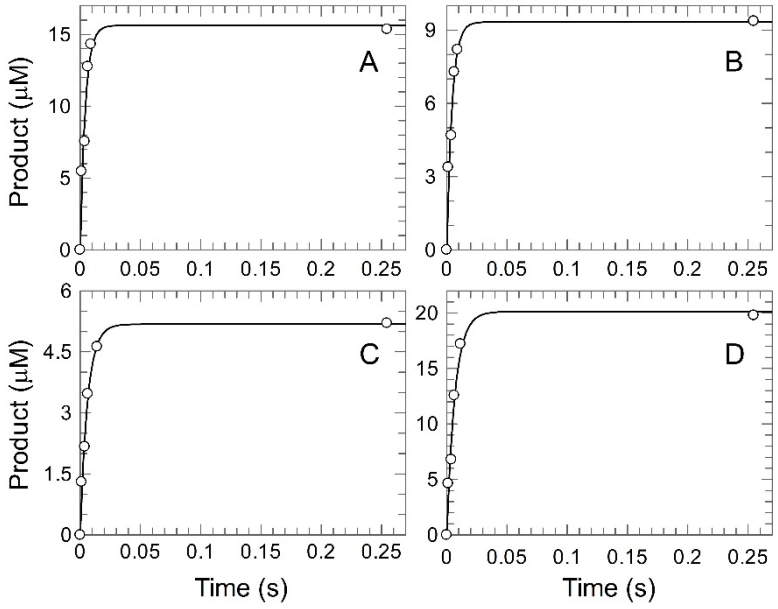


Figure 6.6 Reanalysis of stopped-flow results reported for *Anabaena* FNR.
Data extrapolated from [236] were fit to a single-exponential decay. **(Upper panels)** Time-course of the approach to equilibrium upon mixing **(A)** wild-type FNR_{red} and $NADP^+$, and **(B)** wild-type FNR_{ox} and NADPH. Curve fitting yielded the k_{obs} values of 235 ± 47 s^{-1} and 230 ± 40 s^{-1}, respectively, equivalent to each other within the error of the analysis. Lans *et al.* [236] reported the rate constants for these reactions to be 285 s^{-1} and 270 s^{-1}, respectively. **(Lower panels)** Time-course of the approach to equilibrium upon mixing **(C)** Y303S FNR_{red} and $NADP^+$, and **(D)** Y303S FNR_{ox} and NADPH. Curve fitting yielded the k_{obs} values of 165 ± 16 s^{-1} and 145 ± 20 s^{-1}, respectively, equivalent to each other within the error of the analysis. The raw extrapolated data are available at figshare.

Crystal structures of Y316S and Y316A NADP(H) complexes

Co-crystallization of both FNR variants with nicotinamide yielded crystals that grew readily and could be soaked with $NADP^+$ or NADPH to obtain those complexes. Interestingly, the variants crystallized in two space groups, both the $P3_221$ form seen for wild-type protein[56] and a new $P3_121$ crystal form that had remarkably similar unit cell dimensions and related crystal packing. Here, we report a set of eight refined crystal structures of complexes that represent the Y316S and Y316A variants in complex with three ligands ($NADP^+$, NADPH, and nicotinamide) with most structures refined at between 1.35 Å and 1.6 Å resolution (Table 6.2) and having very well defined active site electron density (Figure 6.7). As the corresponding complexes of Y316S and Y316A largely show equivalent features, we focus mainly on the highest resolution set of structures that provide unique information.

To check whether the complexes in the crystal represent those that form in solution, we obtained single-crystal absorption spectra[244] during data collection for a pair of Y316A crystals soaked with $NADP^+$ and NADPH. These spectra qualitatively match those obtained in solution and indicate that even though the X-ray beam does cause reduction of the FAD, this does not undermine the relevance of the structures (Figure 6.7). Also, since Y316S and Y316A both predominantly form CTC-2 after an NADPH soak (*e.g.* Figure 6.5B,C), we have modeled the NADPH soak structures as an $NADP^+$:$FADH_2$ complex rather than NADPH:FAD.

In the following sections, we first provide a basic description of the complexes and then focus on three detailed aspects of complexes that provide evidence of compression in the active site for enhancing catalysis. In our structural comparisons, the unliganded wild-type structure we use is a 1.05 Å resolution structure refined much earlier (PDB entry 3LO8; released in 2010) and used in tests of new refinement strategies.[245] As the statistics for this structure have not yet been described in the literature, they are included in Table 6.2.

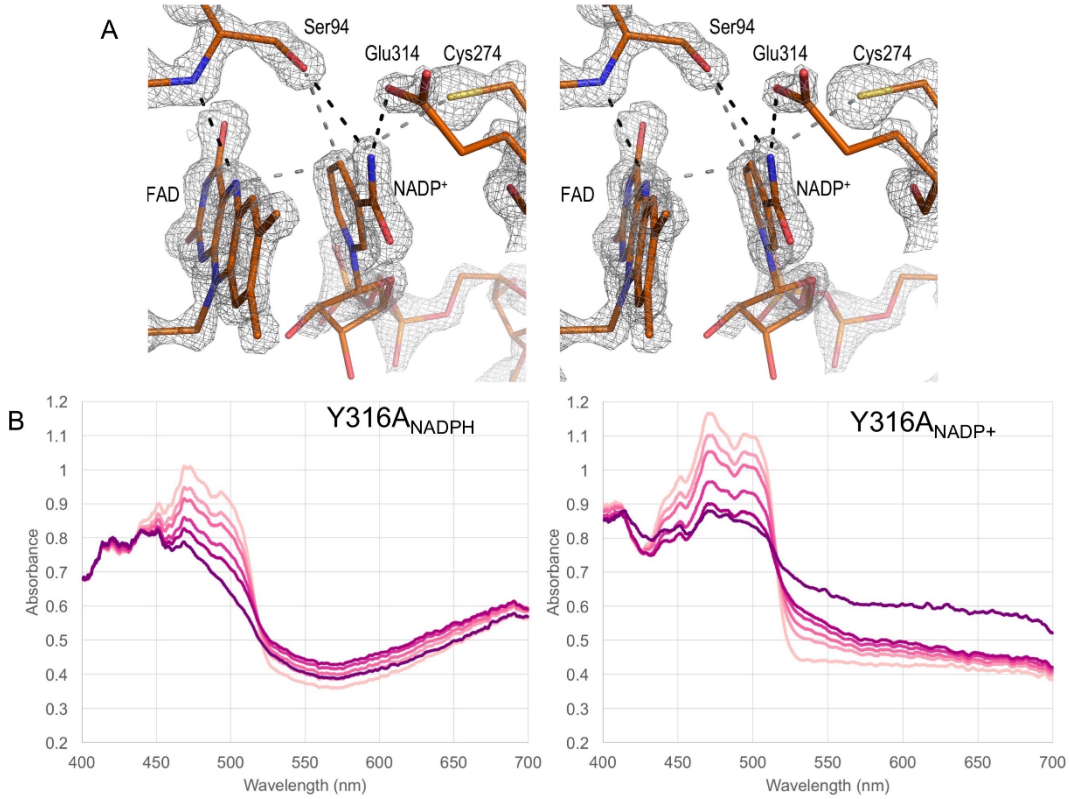


Figure 6.7 Active site environment of FNR.
(A) A stereo view of the active site environment of FNR represented by the structure of $Y316S_{NADP+}$ with the $2F_o$-F_c map shown at 3.0*ρ_{rms}. Key residues within the active site, Ser94, Cys274, and Glu314, which primarily interact with NADP(H), NADP(H), and FAD are shown along with hydrogen bonds (black dashed lines) and interaction distances (gray dashed lines). **(B)** Single crystal visible absorption spectra of Y316A FNR soaks with NADPH (right-hand panel) and $NADP^+$ (left-hand panel) and their change during data collection. Spectra collected corresponding to oscillation images 0, 10, 20, 50, 100 and 360 go from light to dark shades of pink. These spectra qualitatively match those seen in solution in that the CTC-2 absorption band is seen for the crystals soaked in NADPH but not in those soaked in $NADP^+$. The decrease in A_{460} over time indicates reduction of the flavin in the X-ray beam during data collection. Over time the CTC-2 band does not disappear in the NADPH soaked crystal and does not appear in the $NADP^+$ soaked crystal so, we infer that at the cryo-temperatures of data collection, the flavin reduction is not leading to the same structural changes it would lead to in solution.

Overview of the NADP(H) complexes. Globally, all Y316S and Y316A complexes, independent of space group, are very similar to the wild-type structure with rms Cα deviations between 0.1 Å and 0.4 Å. The broad features of NADP(H) binding match those that have been elaborated elsewhere,[55, 59, 222, 227] interacting

mostly with the NADP-binding domain. Briefly, the 2'-phosphate anchor hydrogen bonds with the side chains of Ser237, Arg238, Lys247, and Tyr249, and the adenine is sandwiched between Tyr249 and Leu276. The 5'-phoshoryl group is largely exposed and the nicotinamide-side phosphoryl group hydrogen bonds with the Thr175 N and the Arg114 guanidinium group. The nicotinamide is surrounded by a triad of key side-chains, with Ser94 and Glu314 hydrogen bonding with the nicotinamide carboxamide nitrogen and Cys274 and Ser94 close to the C4 atom (Figure 6.7A). In the higher resolution structures of Y316S, an alternate conformation for the nicotinamide ribose is also observed. In both $Y316S_{NADP+}$ and $Y316S_{NADPH}$ soaks in both space groups, the nicotinamide ribose is modeled as 60% 2'-*endo* and 40% 3'-*endo*. In each Y316A structure, the nicotinamide ribose adopts a single conformation, but in $Y316A_{NADP+}$ it is 2'-*endo* and in $Y316A_{NADPH}$ it is 3'-*endo*. The reason for these conformational differences is not readily apparent but their presence suggests that they are the result of real but subtle differences in nicotinamide binding.

The complexes with free nicotinamide. Surprisingly, the binding mode seen in nicotinamide soaks differ for Y316A and Y316S (Figure 6.3C). In $Y316A_{nic}$, the nicotinamide orientation and hydrogen bonding roughly match how the NADP(H) nicotinamide binds. However, in $Y316S_{nic}$, the nicotinamide binds in a flipped orientation that is stabilized by a hydrogen bond between Ser316-Oγ and nicotinamide-N1 as well as hydrogen bonds of the nicotinamide hydroxyamide with Ser94-Oγ and Thr 175-Oγ and O atoms (Figure 6.3C). Since this binding mode involves Ser316 as a hydrogen-bond partner, we conclude it is an artifact of the Y316S mutation. Also, the existence of the nicotinamide N1 to Ser316-Oγ hydrogen bond in this complex means that neither NADP(H) nor even N1-methyl-nicotinamide could adopt it. This provides a satisfying explanation for why nicotinamide binding to Y316S yields different spectral changes than $NAD(P)^+$, whereas methyl-nicotinamide binds less tightly, but with spectral changes that match those of $NAD(P)^+$ (Figure 6.3B). It also serves as a reminder of the need to be cautious in inferring the binding mode of a molecule of interest based on that of an analog.

The nicotinamide-flavin interactions in the $NADP^+$/NADPH complexes. While the overall structures of the FNR variants are virtually superimposable, subtle

differences exist in the details of nicotinamide binding seen in the $NADP^+$ and NADPH soaks, and having the structures for the two mutants allows us to gain confidence about which variations are reliably due to the difference between the active site redox state rather than the specific mutation or to experimental uncertainty. In terms of the active site, the difference between the structures is a single H-atom: the $NADP^+$ soaks with oxidized flavin and nicotinamide have one H-atom on the nicotinamide C4 and no H-atom on the flavin N5, and the NADPH soaks to produce CTC-2 have one H-atom on the nicotinamide C4 as well as one H-atom on the flavin N5. (The CTC-1 complex would have the same number of H-atoms but distributed with two on the nicotinamide C4-atom and none on the flavin N5-atom.) As described in the following paragraphs, the high resolution structures provide evidence that the single additional H-atom increases active site crowding and a directional compression that we propose is a key promotor of catalysis.

A first line of evidence is provided by the temperature factors of the C4-atom of $NADP^+$ compared to NADPH. In our earlier work on pea FNR,[59] a higher conformational freedom of the $NADP^+$ C4-atom was indicated by weaker electron density (and higher B-factors) for this atom in $NADP^+$ compared to NADPH; but at 1.8 Å resolution, this was not conclusive. Our higher resolution analyses here similarly show the additional mobility of the C4 atom of $NADP^+$ but also provide refined anisotropic B-factors that give a more complete description of dynamics than do isotropic B-factors. The anisotropic B-factors very clearly show that the increased movement for the C4 atom of $NADP^+$ is perpendicular to the plane of the ring and towards N5 of the FAD (Figure 6.8A). This type of motion exactly corresponds to motions that would allow the nicotinamide ring to forms a boat-like conformation that calculations show will promote efficient hydride transfer.[246]

A second observation of crowding derives from a visible distortion of the FAD isoalloxazine. In wild-type FNR, the isoalloxazine group appears to be pushed slightly by the Tyr316 side chain so that it is not coplanar with the N10-C1 ribose bond, but is about 3.5° non-planar (Figure 6.8B). In all FNR variant structures soaked with $NADP^+$, this non-planarity increases to ~10°, and in the structures soaked with NADPH it increases further to ~14° (Figure 6.8B). These deviations from planarity

imply that the nicotinamide displacing the Tyr316 side chain leads to a more tightly packed active site and that the additional H-atom in the active site present in the NADPH soaks increases crowding even more, with the isoalloxazine bending to relieve apparent pressure. It should be noted that the highly crowded complex seen in the NADPH soak is the one that is present during normal catalysis.

A third observation that completes the view of the tightly packed active site of FNR are the interaction distances of the nicotinamide C4 and H4 atoms that are involved in hydride transfer (Figure 6.8C). In all structures, these atoms are sandwiched between the flavin N5 atom in front and the Cys274-Sγ sulfhydryl behind, where it can act as a backstop. In comparing the $Y316S_{NADP+}$ complex to the $Y316S_{NADPH}$ complex which has an additional H-atom on the flavin N5-atom, the C4…N5 distance increases by ~0.3 Å while the C4H… Sγ distance decreases by ~0.1 Å (Figure 6.8C). Taking into account the Y316A complexes (Figure 6.8C), the consistent increase of the C4…N5 distance in the NADPH complex leads us to conclude that Cys274 acts as a firm backstop while the isoalloxazine bends away from the nicotinamide by an additional ~4˚ in order to increase the distance between the flavin and nicotinamide and relieve the pressure due to the presence of the extra H-atom.

These observations lead us to conclude that the active site is experiencing compression, distorting the FAD and pushing the substrates together to enhance catalysis by aligning the substrates in an optimal position and, as has been noted by others, potentially promoting a quantum tunneling mechanism of hydride transfer.[236]

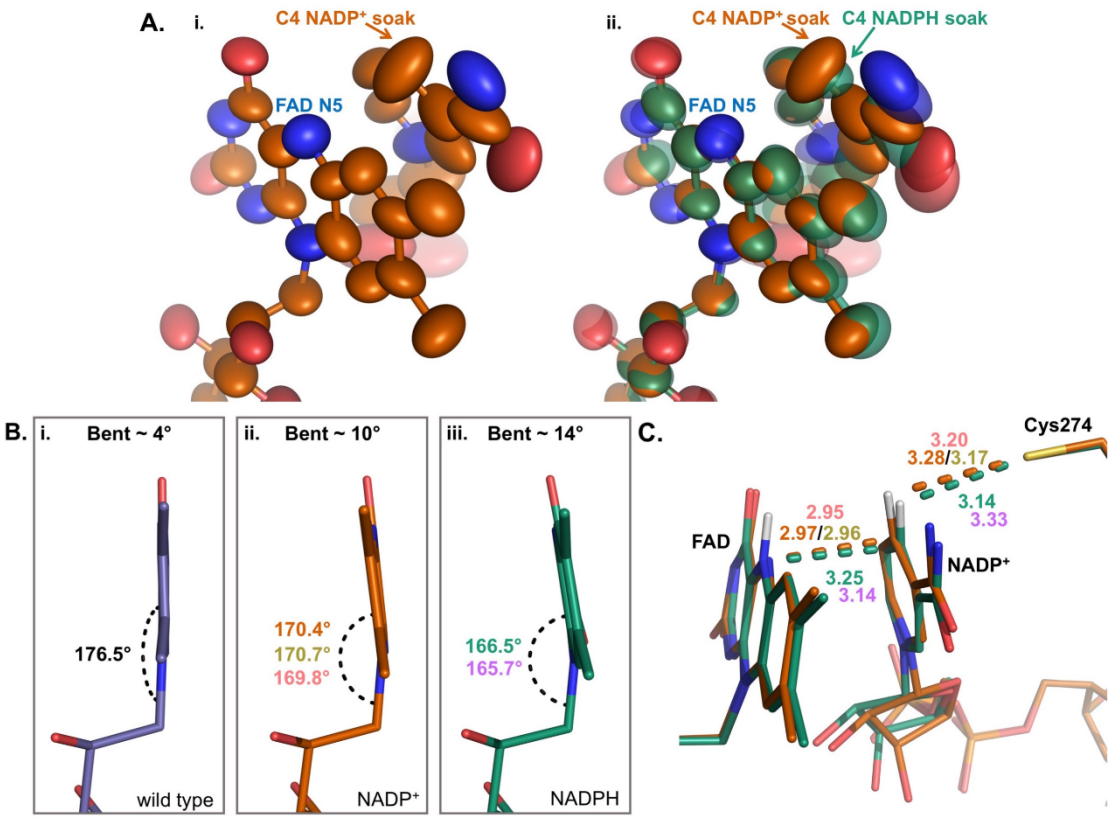


Figure 6.8 Active site compression in corn root FNR as seen by mobility, covalent distortion, and interaction distances.

(A) Mobility of atoms in the active site of Y316S FNR. Atoms in the active sites of $Y316S_{NADP+}$ (orange) (i) and overlaid with $Y316S_{NADPH}$ (semi-transparent green) (ii) shown with ellipsoids, denoting the use of anisotropic B-factors. The nitrogen and oxygen atoms are shown in blue and red, respectively, in the complexes. The mobility of the C4 of $NADP^+$ is in the direction of the N5 of the FAD, forming a boat-like conformation which is favored for hydride transfer. **(B)** Active site compression in corn root FNR as seen by covalent distortion. The angle of distortion of the FAD moiety is shown in: (i) FAD in wild-type corn root FNR (mauve), (ii) FAD in $Y316S_{NADP+}$ (orange) and the corresponding angles from $Y316S_{NADP+}$ ($P3_121$, olive) and $Y316A_{NADP+}$ (salmon), (iii) FAD in $Y316S_{NADPH}$ (green) and the corresponding angle from $Y316A_{NADPH}$ (violet). **(C)** Active site interactions in FNR. Structural overlay of $Y316S_{NADP+}$ (orange) and $Y316S_{NADPH}$ (green) with relevant average distances shown in corresponding colors. The average distances from $Y316S_{NADP+}$ ($P3_121$, olive), $Y316A_{NADP+}$ (salmon), and $Y316A_{NADPH}$ (violet) are also shown. The distances are averages from ten independent refinements of each complex (see Materials and Methods), and all standard deviations were < 0.02 Å. The C4H and N5H distance in $Y316S_{NADPH}$ is 3.0 Å (not labeled). If modeled as $FADH_2$ instead of FAD, the theoretical distance between C4H and N5H in $Y316S_{NADP+}$ would be 2.5 Å. Nitrogen, oxygen, and sulfur atoms colored blue, red, and yellow, respectively.

Insights into factors promoting hydride transfer in FNRs

Taken along with our reanalysis of previous stopped-flow kinetics work that called the relevance of aromatic placeholder FNR mutants into question, these in-

solution spectroscopy, single-crystal spectroscopy, and high-resolution structures provide insights into hydride transfer in FNRs using variants which have near wild-type hydride-transfer kinetics, but importantly, allow us to capture the productive binding mode of NADP(H) in the active site. Such productive complexes have not been possible to capture in any wild-type member of the FNR superfamily with an aromatic placeholder residue (*e.g.* [64, 66, 223, 232-233]) or that instead have a C-terminal peptide extension as an alternate way to block nicotinamide binding.[221, 247] They have also been challenging to obtain for superfamily enzymes that have a shifted domain-domain interaction[248-249] instead of an aromatic blocking residue, but for one of these enzymes, cytochrome b5 reductase, anaerobic co-crystallization has yielded a productive, wild type complex with NADH in which the nicotinamide-flavin interaction geometry is much like that seen here.[61]

The high-resolution structures reported here illustrate the role of specific anisotropic motions as well as active site compression as catalytic strategies that promote hydride transfer. Key evidences are the anisotropic motion of the C4 atom of $NADP^+$, that indicates boat-like perturbations in conformation are a preferred mode of vibrational freedom, as well as a strong decrease in the amplitude of that motion in the NADPH complex. Additionally, there is an incrementally increasing deviation from planarity at the flavin N10 atom in wild type, $Y316S_{NADP+}$, and $Y316S_{NADPH}$ indicating increased crowding and pressure associated with oxidized nicotinamide binding and even more so with reduced nicotinamide binding that is being relieved to some extent by bending of the flavin. Interestingly, a very similar flavin non-planarity was seen in the cytochrome b5 reductase-NADH complex (see Figure 3a of [61]). Finally, comparison of $Y316S_{NADP+}$ and $Y316S_{NADPH}$ reveal close sub van der Waals contact distances consistent with a tightly packed active site. The central role of Cys274 in these interactions provides a rationale for its conservation across the whole superfamily. While steric compression has occasionally been noted as a key factor promoting catalysis for other enzymes,[250-253] there are a limited number of studies which provide direct structural evidence for this,[254-258] as this typically requires atomic or near-atomic resolution crystal structures. Also, that the Y316S and Y316A variants of corn root FNR, like similar mutants of other FNR superfamily members,[60,

[240, 243] show much less stabilization of the FAD seminquinone, is consistent with an earlier proposal[240] that the aromatic side chain is not solely a passive placeholder for the nicotinamide group, but, through its stacking interaction with the flavin, is an active agent that stabilizes the semiquinone form of the flavin to enhance the one-electron transfers required for these dehydrogenase-electron transferases.

Extrapolation of the results to the FNR superfamily members such as NOX enzymes

Clarifying the relevance of the $NADP^+$ binding mode seen in the Tyr316 mutants is not just of interest for understanding catalysis for the whole superfamily, but can guide further studies of these enzymes as well as the generation of superfamily member variants that can be valuable tools. For instance, for the NOX enzymes that play crucial roles in the production of superoxide and hydrogen peroxide for diverse biological processes, the FNR-like module is at the C-terminus of a membrane-bound flavocytochrome catalytic subunit that has been relatively difficult to study.[229] Although a NOX structure has long eluded structural biologists, a crystal structure of the FNR-like module of NOX5 from *Cylindrospermum stagnale* (referred to as the NADPH-dehydrogenase domain) was solved while this work was under review.[239] Sequence alignments of FNR with human NOXs aided by the *C. stagnale* NOX5 structure (Figure 6.9A) show that the NOX isozymes conserve many $NADP^+$ binding residues including the nicotinamide-interacting residues equivalent to corn root FNR Ser94, Cys274, Glu314 and Tyr316 (Phe693 in *C. stagnale* NOX5).

The NOX5 FNR-like module itself was unstable and crystals were only obtained for a construct with a "hyperstabilizing" C-terminal extension that included a Trp two residues after Phe693. Notably, in this structure, the expected natural C-terminal aromatic placeholder (Phe693) does not stack against the isoalloxazine, and, because the anticipated fifth β-strand of the $NADP^+$-binding domain is perturbed and moved away from the protein core, the Trp residue is able to stack against the isoalloxazine ring instead (Figure 6.9B). However, if the NOX5 chain followed the path of the β-strand seen in other FNR superfamily members, Phe693 could stack against the isoalloxazine and we propose that this is what occurs in the native enzyme. Magnani et al [239] noted that the isolated dehydrogrenase domain is

deregulated and predicted that the normal role of the strictly conserved C-terminal aromatic residue would "emerge only in the context of a full-length protein." We agree and further suggest that the instability of the isolated dehydrogenase domain and the loose association of the C-terminal segment (including Phe693) is related to the missing cytochrome transmembrane domain and associated lipid bilayer that interact with this surface of the dehydrogenase domain in a full-length complex (see Fig 5A of [239]).

Consistent with the weak association observed for Phe693 with the protein core is in the isolated dehydrogenase domain, a Phe693Ser mutant had higher than wild-type activity and the stabilizing C-terminal extension mutant (adding the Trp695) dropped activity by 5-fold.[239] Nevertheless, we predict that in the context of a full-length membrane-bound NOX, Phe693 will act as an aromatic placeholder residue and that Ser or Ala mutants will bind both NADPH and NADH more tightly. Such mutants could provide a useful handle for purification (through the tight binding of $NADP^+$) or be a useful tool for probing/controlling the physiological roles of NOXs. A solely NADH responsive version could potentially be made through additional mutations such as the equivalent of FNR Ser237 to Asp (Figure 6.9), as this position has been shown to discriminate against binding the 2'-phosphoryl group in multiple superfamily members.[62-63, 259-260] Interestingly, a full-length human NOX2 C-terminal Phe mutant (F570A) was characterized in 1998 and found to retain ~50% of the wild-type activity[261] rather than dropping 300-800 fold in activity as had been seen for the equivalent pea FNR mutant, and so it was not studied further. However, this study predated the knowledge that the equivalent FNR mutant lost activity due to the slow dissociation of $NADP^+$ and also gained activity with NADH,[56, 58] so those qualities of the variant were never characterized. In retrospect, given that the NOX2 K_d for $NADP^+$ is ~40 μM (over 10-fold higher than is typical for FNRs) and the turnover number is slower, it would not be surprising if the expected enhanced binding of $NADP^+$ in the F570A variant would not be enough to make its dissociation highly rate limiting. Indeed, for NO synthase the equivalent mutation only showed about a 3-fold decrease in steady-state turnover, yet still showed a 50-fold change in specificity.[60, 243]

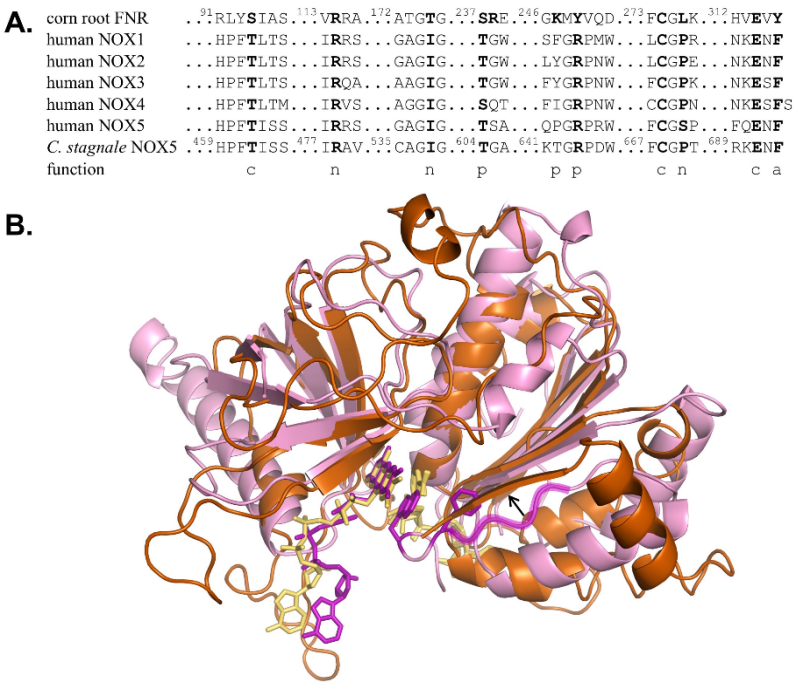


Figure 6.9 Sequence alignment of corn root FNR with NADPH oxidases.

(A) Sequence alignment of segments of corn root FNR with human NOX1, NOX2, NOX3, NOX4, and NOX5 and *C. stagnale* NOX5. Select FNR functional residues and similar aligned residues (in bold) have their function denoted as: aromatic placeholder (a), other catalytic center (c), 2'-phosphate (p) or other NADP interactions (n). Residue numbers are given for the structurally known corn root FNR and *C. stagnale* NOX5. **(B)** Overlay of the *C. stagnale* NOX5 dehydrogenase domain (PDB 5O0X; pink with bound FAD and the natural C-terminal Phe693 and C-terminal extension Trp side chains in purple) and the $Y316S_{NADP+}$ structure (orange with bound FAD and NADPH in yellow). The anticipated fifth β-strand of the $NADP^+$-binding domain of NOX5 which is perturbed and allows the Trp residue to stack against the isoalloxazine ring is highlighted (purple trace) and an arrow indicates the direction we expect it to move in the full-length NOX5 context so that it will align with the rest of the β-sheet as does the fifth β-strand in FNR. The deposited structure of the NADPH binding domain of NOX2 gp91(phox), determined through structural genomics efforts (PDB 3A1F) but not yet described in the literature, also aligns well in a structural overlay.

Table 6.2 Data and refinement statistics for FNR variant structures[a]

	wt			Y316S				Y316A	
Ligand soaked	None	Nicotinamide	Nicotinamide	NADP+	NADP+	NADPH	Nicotinamide	NADP+	NADPH
Space group	$P3_221$	$P3_221$	$P3_121$	$P3_221$	$P3_121$	$P3_221$	$P3_221$	$P3_221$	$P3_221$
Data Statistics									
Wavelength (Å)	1.00	1.00	1.54	1.00	1.54	1.00	1.54	1.00	1.00
Unit cell *a, c* axes (Å)	59.1, 186.7	58.9, 184.8	58.9, 184.4	58.8, 187.5	58.9, 185.0	58.9, 187.3	59.4, 187.5	58.7, 186.7	59.2, 188.0
Resolution (Å)	50-1.05 (1.07-1.05)	51-1.35 (1.42-1.35)	12 – 1.90 (1.97-1.90)	51-1.45 (1.53-1.45)	40-1.80 (1.90-1.80)	51-1.45 (1.53-1.45)	26-1.95 (2.06-1.95)	49-1.50 (1.55-1.50)	49-1.60 (1.66-1.60)
Unique reflections	177114 (8768)	75008 (9295)	29752 (2603)	65831 (8739)	35032(5039)	66408 (8907)	24769 (1325)	44421 (1371)	50821 (4336)
Multiplicity	5.2 (2.9)	5.8(2.6)	8.2 (3.8)	9.6(2.5)	7.3(3.9)	8.4(2.5)	9.1(2.3)	1.7 (1.1)	4.0 (3.2)
Average I/σ	57.0(3.4)	6.2 (1.7)	19.4 (6.9)	14.7(3.0)	6.5(2.2)	12.7(2.7)	34.3(9.6)	10.4 (0.82)	24.5 (6.0)
R_{meas} (%)	9.3 (50.6) [b]	30.2 (46.9)	4.0 (15.3) [b]	9.5(29.0)	33.5(53.2)	13.0(39.5)	4.4(6.4)	5.4 (71.1) [b]	2.8 (16.5) [b]
Completeness (%)	99.8 (99.5)	91.0 (78.4)	97.6 (86.6)	97.9 (91.7)	99.9 (99.9)	98.1 (92.1)	85.8 (32.5)	72.9 (23.0)	98.6 (86.4)
Refinement Statistics									
Refinement method	Aniso	Aniso	TLS	Aniso	TLS	Aniso	TLS	TLS	TLS
Amino acid residues	302	309	309	309	309	309	309	309	309
Solvent atoms	646	484	413	491	297	500	556	451	530
Non-H atoms	3252	3117	3035	3160	2923	3217	3165	2999	3155
RMS bonds (Å)	0.017	0.009	0.009	0.008	0.010	0.008	0.010	0.009	0.009
RMS angles (˚)	2.3	1.0	1.1	1.2	1.3	1.2	1.1	1.2	1.2
$\langle B_{protein} \rangle$ (Å^2)	21.7	14.6	28.9	20.9	34.5	21.3	17.8	26.7	24.0
$\langle B_{FAD} \rangle$ (Å^2)	15.1	10.4	18.3	16.6	22.0	16.7	13.1	21.1	18.5
$\langle B_{NADP(H)} \rangle$ (Å^2)	–	13.8	22.7	24.1	28.3	25.5	23.7	27.0	24.7
R_{work} (%)	12.5	13.5	13.1	10.8	17.2	11.0	13.0	14.9	13.1
R_{free}(%)	15.5	17.3	17.3	14.4	21.1	14.7	18.0	18.5	15.6
PDB code	3LO8	5VW4	5VW9	5VW3	5VW8	5VW2	5VW5	5VW6	5VW7

[a] Numbers in parentheses are in the highest resolution shell

[b] R_{merg} reported in place of R_{meas}

Materials and Methods

Production of Recombinant Corn Root FNR and its Variants

Plasmids for the bacterial expression of the Y316A and Y316S proteins were generated from the pETrFNR2 [9] using the QuikChange II Site-Directed Mutagenesis kit (Agilent) and two appropriate oligonucleotide couples following the manufacturer directions. The wild-type and variant proteins were produced in *E. coli* HMS174(DE3) by induction with 0.1 mM IPTG for 4 h at 30 °C, and purified through a procedure similar to that previously reported for the wild-type protein[9] using an ÄKTA FPLC (GE Healthcare) apparatus. Briefly, the crude cell lysate was brought to 40% NH_4SO_4 and centrifuged. The supernatant was applied to a Sepharose 4B column (GE Healthcare), eluted with appropriate buffer and precipitated with 75% $(NH_4)_2SO_4$. In the case of the protein variants, the pellet was resuspended in 40% $(NH_4)_2SO_4$ and chromatographed on a butyl Sepharose column (GE Healthcare) with a descending salt concentration gradient. The omission of this step prevented the adsorption of the FNR forms on the next cation exchange resin. After desalting, the sample was loaded on an SP-Sepharose HP column (GE Healthcare) and eluted through a NaCl concentration gradient. Notably, the visible spectra of both mutant forms, but not the wild-type enzyme, underwent a blue shift during the ion exchange step, suggesting release of some unidentified bound ligand at that stage. The resulting proteins were homogenous as judged by SDS-PAGE. The FNR forms were concentrated to ~25 mg/mL and stored at -20 ˚C in 10 mM HEPES, pH 7.0.

Spectral Analyses and Ligand Binding

All spectrophotometric measurements and steady-state enzyme kinetics were performed using an 8453 diode-array spectrophotometer (Agilent). Titration with $NADP^+$, NAD^+, phenol, nicotinamide and N-methyl-nicotinamide of the enzyme forms in their oxidized state were performed in 10 mM Tris-HCl, pH 7.7, at 15 °C.

Photoreductions, Anaerobic Titrations and Activity Assays

Flavin photoreduction experiments were carried out both in the absence and in the presence of $NADP^+$ or NAD^+ in anaerobic cuvettes in 10 mM HEPES-NaOH, pH 7.0, at 15 °C, according to the procedure described elsewhere.[9]

Anaerobic titrations with NADPH of 14-18 μM wild-type, Y316A or Y316S FNR variants were performed on 1.2 mL samples in 50 mM HEPES-NaOH buffer, pH 7.0, at 15 °C, in a sealed cuvette. After recording the spectrum of the oxidized enzyme, five successive additions of anaerobic NADPH were made leading to ligand concentrations of 29 μM, 39 μM, 76 μM, 182 μM, and 1.9 mM for wild-type titrations, 23 μM, 45 μM, 90 μM, 210 μM, and 2.2 mM for FNR Y316A titrations, and 24 μM, 48 μM, 95 μM, 225 μM, and 2.3 mM for FNR Y316S titrations, respectively. In each case, the volume after the final addition of NADPH was 1.4 mL. For each FNR form, the cuvette was then opened to air, and the spectrum of the mixture (kept at 15 °C) was monitored at over time until all the NADPH was oxidized. For technical reasons, the spectra for wild-type FNR during O_2 turnover was produced from a separate sample also prepared with 2.4 mM NADPH.

NADPH—$K_3Fe(CN)_6$ reductase activity assays were performed in 100 mM Tris-HCl, pH 8, at 25 °C, at different concentrations of both substrates, as reported elsewhere.[56] As needed, nicotinamide was included in the assays at concentrations ranging from 50 to 800 mM.

Reanalysis of Stopped-Flow Results Reported for *Anabaena* FNR

Spectra (Figure 3 in [26]) were enlarged by ~345% and printed. The height of each peak at ~460 nm corresponding to the flavin oxidation state was measured in cm from the baseline of the spectra (-0.025 AU). The peak height at time 0 for FNR_{red} with $NADP^+$ (corresponding to fully reduced flavin) was taken as a baseline and used to normalize the measurements by subtracting this height from the other peak heights of the corresponding variant. Whereas the final spectra (0.2547 s) for wild-type FNR for both FNR_{red} with $NADP^+$ and FNR_{ox} with NADPH matched in both peak height and shape, the final spectra (0.2547 s) for Y303S FNR varied slightly in peak height but both correspond to achieved equilibrium so were normalized to each other by

subtracting the difference between the final peak heights from the measured peak heights for Y303S FNR_{red} with $NADP^+$.

Taking the peak height at time 0 for FNR_{ox} with NADPH to correspond to fully oxidized flavin, the percentage of oxidized and reduced flavin and corresponding concentration of product (using the provided concentrations of 25 μM FNR and 125 μM NADP(H)) were calculated. A plot of concentration of product (μM) versus time (s) were fitted using GraFit 5 (Erithacus Software Limited) with a single-exponential decay equation to estimate the apparent rate constants of the hydride-transfer reactions.

Crystallization and Structure Determinations

Initial co-crystallization trials of the corn root FNR variants with $NADP^+$ used similar conditions to those used for the wild-type enzyme[9] but yielded only ill-formed small crystals that grew very slowly. Solving their structure revealed that they contained a bound nicotinamide moiety rather than $NADP^+$, so co-crystallization with nicotinamide was tried, and crystals grew readily. The best crystallizations were at room temperature in hanging drops formed by mixing equal volumes of the protein stock (Y316F: 20.2 mg/mL; Y316A: 7.3 mg/mL; Y316S: 10.8 mg/mL; all in 50 mM Tris-HCl, pH 7.4) with a reservoir solution containing 22-24% PEG 8000, 0.1 M sodium cacodylate (pH 6-7), 0.18-0.22 M magnesium acetate, and 100 mM nicotinamide. Typical crystals were ~0.3 mm on each side and grew within 1 week. Both FNR variants formed many prism-shaped crystals belonging to the space group $P3_221$ (same as wild type;[9]) and fewer hexagon-shaped crystals belonging to space group $P3_121$. Soaks of the nicotinamide-bound crystals were done aerobically at room temperature with either 10 mM $NADP^+$ or 10 mM NADPH for 1 h to obtain desired complexes before being flash frozen.

Data collections. For data collection, crystals were pulled through oil and flash-frozen in liquid nitrogen. In-house data were collected at 140 K on a Rigaku RU300 Cu-Kα rotating anode X-ray source running at 50 kV and 100 mA equipped with a Raxis IV image plate detector. Synchrotron data were collected at beamline 5.0.1 or 5.0.3 at the Advanced Light Source (Lawrence Berkeley National

Laboratory, Berkeley, CA) at 100 K. Oscillation images were collected with $\Delta\varphi = 1°$ and were processed using Denzo and Scalepack[57] or iMosflm[2628] and SCALA.[59] Unit cell parameters and data reduction statistics for each FNR variant are given in Table 6.2. In addition, data sets with $\Delta\varphi = 0.5°$ and simultaneous single crystal visible absorption spectra were collected by Dr. Allen Orville at a National Synchotron Light Source beamline as described in.[35]

Structure solutions and refinements. All $P3_221$ FNR variant crystals were isomorphous with the published wild-type corn root FNR structure (PDB code 3LO8 [36]), and this was used as the starting model for the first refinement, with additional refinements built on partially refined models of the most similar structure. The $P3_121$ crystal form was trivial to solve by molecular replacement and these structures were then similarly refined. The same 10% of data were set aside for cross-validation[60] for each structure as had been set aside for the ~1 Å resolution refinement of the wild-type structure. Refinements were carried out by a number of researchers over many years using various versions of REFMAC[61] with Coot[62] used for manual fitting and Molprobity[63] used to help identify model problems. During iterative manual rebuilding, water molecules were added in Coot using standard criteria (>1 ρ_{rms} intensity in the $2F_o-F_c$ map, >2.4 Å distance from nearest contact, no *B*-factors >80 $Å^2$). For making a consistent set of structures for publication, a set of final refinements of each structure were done by the lead author using Phenix.[64] During this stage, we decided to model all NADPH soaks as having an oxidized nicotinamide (*i.e.* $NADP^+$) and a reduced flavin since the spectra show that CTC-2 is present at higher amounts than is CTC-1. For structures refined with anisotropic B-factors (the $P3_221$ Y316S:NADPH, Y316S:$NADP^+$, and Y316S:nicotinamide structures), no anisotropic restraints were used for the ligand in the final round of refinement to ensure the anisotropic B-factors of the nicotinamide atoms were based as much as possible on the diffraction data. Refinement statistics for all models are shown in Tables 6.2. Structural overlays were performed using the protein structure visualization software PyMOL.[65]

Also, in order to obtain a reliable representation of active site distances for $Y316S_{NADPH}$, $Y316S_{NADP+}$ ($P3_221$ and $P3_121$), $Y316A_{NADPH}$, and $Y316A_{NADP+}$

structures, ten different starting models for each were generated using the "shake" algorithm of Phenix with the setting "modify.sites.shake = 0.5". This level of coordination disruption resulted in starting R/R_{free} values of ~40%. Each of these models was re-refined and the distances between Cys274 Sγ and $NADP^+$ C4H and between $NADP^+$ C4 and FAD N5 were measured. The average distances and standard deviations for each set were calculated and reported.

Accession Numbers. Coordinates and structure factors for the wild-type corn root FNR, Y316S:nicotinamide ($P3_221$), Y316S:nicotinamide ($P3_121$), Y316S:$NADP^+$ ($P3_221$), Y316S:$NADP^+$ ($P3_121$), Y316S:NADPH ($P3_221$), Y316A:nicotinamide ($P3_221$), Y316A:$NADP^+$ ($P3_221$), Y316A:NADPH ($P3_221$), Y316F ($P3_221$), and Y316F:$NADP^+$ ($P3_121$) models have been deposited in the Protein Data Bank with accession numbers 3LO8, 5VW4, 5VW9, 5VW3, 5VW8, 5VW2, 5VW5, 5VW6, 5VW7, 5VWA, and 5VWB, respectively.

Acknowledgements

This work was supported in part by National Science Foundation grant MCB-9982727 and National Institutes of Health grant R01-GM119227. The authors would also like to thank Allen Orville for collecting absorption spectra from single crystals of FNR variants and Peter Zwart for collecting and processing some of the data sets reported here. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Supporting Information

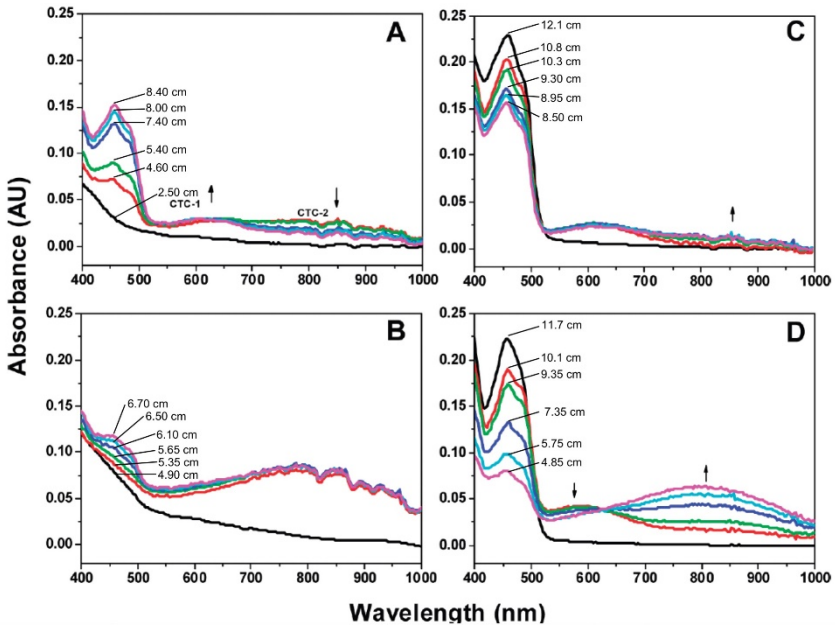


Figure 6.S1 Measurements used for the reanalysis of *Anabaena* FNR Y303S and wild type kinetics.

Reproduction of Figure 3 from [26] showing time courses for the reaction of **(A)** WT FNR_{red} with $NADP^+$, **(B)** Y303S FNR_{red} with $NADP^+$, **(C)** WT FNR_{ox} with NADPH, and **(D)** Y303S FNR_{ox} with NADPH. Spectra were enlarged by ~345% and the height of each peak at ~460 nm corresponding to the major absorption band of oxidized FAD was measured in cm from the x-axis (-0.025 AU). The measured height is listed next to each spectrum. Before using the peak heights to derive concentrations of FNR_{red} and FNR_{ox} in the Y303S experiments, we corrected for the offset in the Y303S $FNR_{red} + NADP^+$ spectra by subtracting from all Y303S $FNR_{red} + NADP^+$ (panel B) measurements the difference of 1.85 cm that was seen between the final Y303S FNR spectra peak heights (purple traces in panels B and D).

Chapter 7

Structure and proposed mechanism of lactate monooxygenase from *Mycobacterium smegmatis*

Kelsey M. Kean, Zhan Deng, and P. Andrew Karplus

In preparation for *Protein Science*

Abstract

Dedicated to Vincent Massey

Lactate monooxygenase (LMO) is a FMN-dependent enzyme that catalyzes the oxidation of lactate to acetate, carbon dioxide, and water and involves pyruvate and hydrogen peroxide as intermediates. While LMO proceeds along a "coupled pathway," all of its α-hydroxy acid oxidase family members, including lactate oxidase, glycolate oxidase, mandelate dehydrogenase, and flavocytochrome b_2, utilize an "uncoupled pathway," wherein the α-hydroxy acid substrate is oxidized and quickly released from the active site as final product before the reduced flavin reacts with oxygen to generate hydrogen peroxide. The how and why LMO deviates in terms of its unique kinetics and resulting function as well as the structure of LMO eluded scientists for decades.

Here, we present the first structures of LMO from *Mycobacterium smegmatis*, including a 2.1 Å wild-type structure and a 1.7 Å structure of C203A variant of LMO. As expected, LMO has an overall fold and active site organization resembling those of other α-hydroxy acid oxidases. Based on structural similarity and sequence identity, LMO is equally distinct from lactate oxidase, glycolate oxidase, mandelate dehydrogenase, and flavocytochrome b_2 and is the first representative enzyme of its type. Structural comparisons with other α-hydroxy acid oxidases highlight the longer length and more compact fold of a variable "loop 4" in LMO, previously shown to have a role in controlling substrate release in these related flavoenzymes. Differences among "loop 4" composition and conformation led us to propose a compelling role for "loop 4" in slowing product release and shifting the kinetic properties of LMO to favor the observed coupled pathway.

Introduction

Lactate monooxygenase, referred to as lactate oxidase at the time, was first discovered in *Mycobacterium phlei* in 1947.[68] At the time of this initial identification, it was proposed that this yellow protein was a flavoenzyme that was able to consume lactate and molecular oxygen to generate stoichiometric amounts of acetate and carbon dioxide and utilized pyruvate as an intermediate. Subsequent characterization

showed this early hypothesis was correct: lactate monooxygenase (LMO) is an FMN-dependent enzyme responsible for catalyzing the oxidation of lactate to acetate, carbon dioxide, and water with pyruvate and hydrogen peroxide as intermediates (Figure 7.1A). However, the nomenclature for this enzyme has changed over time; while the name "lactate oxidase" persisted into the early 1990s,[263] it is now referred to as "lactate monooxygenase" and a different, related flavoenzyme has taken its place as "lactate oxidase."

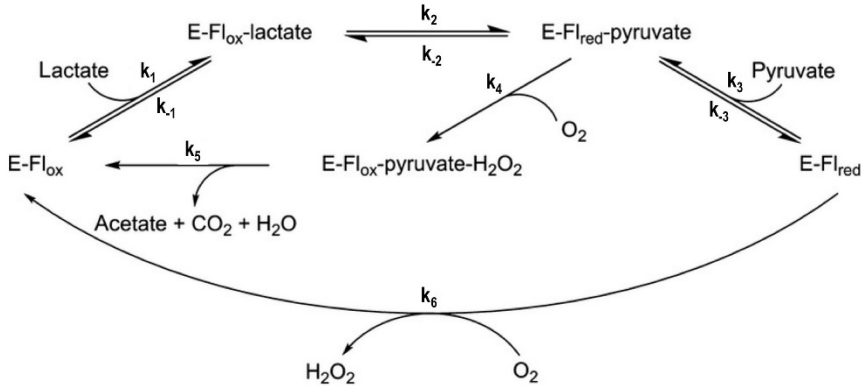
LMO belongs to a larger family of α-hydroxy acid oxidizing flavoenzymes including lactate oxidase (LOX), glycolate oxidase (GOX), flavocytochrome b_2 (FCB2), and mandelate dehydrogenase (MDH). Among these α-hydroxy acid oxidases, LMO is unique in that it proceeds along a "coupled pathway" (Figure 7.1A), wherein lactate is first oxidized to pyruvate and the resulting reduced flavin reacts with oxygen to generate hydrogen peroxide. Rather than releasing the intermediate pyruvate from the active site, pyruvate reacts with hydrogen peroxide, undergoing an oxidative decarboxylation to produce acetate, carbon dioxide, and water. In contrast, other members of the family, including LOX, GOX, FCB2, and MDH, proceed through an "uncoupled pathway" wherein the α-hydroxy acid substrate is oxidized and quickly released as product from the active site. For example, in LOX, lactate is oxidized to pyruvate (product). Separately, the reduced flavin reacts with oxygen to generate hydrogen peroxide as a product of the reductive half-reaction (Figure 7.1A).

LMO has considerable sequence similarity with these other family members and working models for the active site of LMO have been developed based on the structures of LOX ,[264-265] GOX ,[266-267] FCB2,[268] and MDH.[269] Each of these enzymes have the same fold with each monomeric unit forming a core TIM-barrel of 8 α-helices and 8 β-strands with one FMN binding at the C-terminal end of the β-strands. Sequence and structural comparisons along with chemical modification[270-273] and site-directed mutagenesis studies[73-75] indicate conservation of seven key active site residues in the family members (Figure 7.1B), suggesting similar ligand binding and catalytic mechanism across these enzymes.

With such high active site similarity, a central question prompting studies of LMO is what the origin could be for its unique kinetic properties among the α-hydroxy acid oxidases. It was suggested that they must somehow stem from subtle changes associated with the mutation of residues near the LMO active site that could influence the flavin environment and substrate/product binding affinities. This made LMO an intriguing model for understanding and probing the fine-tuning of flavoenzymes through altered protein-flavin-substrate interactions. LMO has been extensively enzymatically characterized[73-77, 263] and attempts to modulate the kinetic properties of LMO to be like those of LOX or vice versa through site-directed mutagenesis have been unsuccessful,[77, 274-276] proving that our understanding of the differentiating features between these lactate-oxidizing enzymes is still lacking. Since the late 1990s, with no structural information to guide further studies, LMO fell out of favor and our knowledge of LMO has been stagnant.

Here, we present the first direct structural information about LMO from *Mycobacterium smegmatis*. This work includes both a 2.3 Å structure of wild-type LMO (WT) and a 1.7 Å structure of an LMO variant in which Cys203 is mutated to Ala (C203A). These structures reveal that LMO bears the expected high level of overall structure similarity to LOX, GOX, FCB2, and MDH with complete conservation of the seven active site residues. Further comparisons lead us to propose that the rationale behind the kinetic deviation of LMO from other members of the family are not due to certain residue changes near the active site but are a result of differences in the composition and folding of a structurally variable and dynamic "loop 4" which occludes the channel leading to the active site in all of these enzymes.

A.



B.

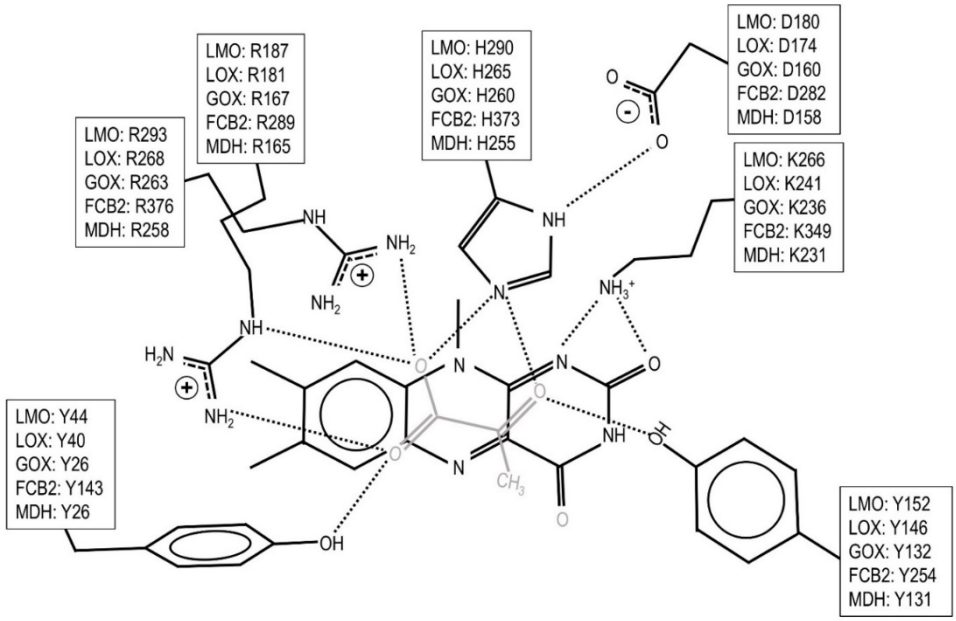


Figure 7.1 Kinetic pathways and active site composition of LMO and related enzymes.
A. Kinetic pathways of LMO and LOX. The inner loop represents the "coupled pathway" observed in LMO where pyruvate is an intermediate and acetate is produced. The outer loop represents the "uncoupled pathway" observed in LOX where pyruvate and H_2O_2 are produced in two uncoupled half-reactions. E, Fl_{ox}, and Fl_{red} represent the enzyme, oxidized flavin, and reduced flavin, respectively. Adapted from [275]. **B.** Schematic of conserved residues in the active site of LMO, LOX, GOX, FCB2, and MDH. Representative hydrogen bonds within the active site and with pyruvate (grey) are shown.

Results and Discussion

Structure Determination and Crystal Packing

Crystals of LMO were first grown and data, including the WT data set used here, were collected in the 1990s.[277] The most similar structurally known protein has only ~35% identical in sequence, and no success was had determining the structure using either molecular replacement or SeMet MAD phasing. The breakthrough allowing successful molecular replacement was with the application of MR-Rosetta.[197] MR-Rosetta was able to place four, well-ordered chains (ABCD – 100% occupancy, average B factor 47 $Å^2$), but this left significant weak but unmodeled density in a large swath of the crystal (Figure 7.2). Two additional, weaker chains (G and H – 80% occupancy, average B factor 112 $Å^2$) were placed with Phaser (see Materials and Methods), resulting in six total chains in the asymmetric unit. While these additional chains have significantly weaker density, indicating increased disorder, both are reliably placed based on clear, unbiased density consistent with the expected positions of an FMN and sulfate in the active site.

Protein crystals normally form a well-ordered and defined crystal lattice, so this unusual and disordered crystal packing made LMO an especially challenging molecular replacement and modeling problem. One similar, unusual crystal lattice has been reported for stefin B crystals.[278] In this rare example, the lattice of the crystal was composed of five tetramer layers, four of which were well-ordered and one of which consisted of a pair of alternatively placed tetramers. In this example, two alternative positions for the pair of tetramers could be modeled. In LMO, a similar layer of difficult to model disorder appeared in the crystal lattice. However, in our structure solution of LMO, only one placement for chains G and H appears needed, and it is not clear why the chains are so highly disordered.

The refined model of LMO C203A at 1.70 Å resolution contains six chains in the asymmetric unit with R and R_{free} values of 17.1% and 19.8%, respectively, (Table 7.1). For five of the chains (A, B, C, D, and G), all 394 residues are sufficiently ordered to be modeled; for the sixth chain (H), the N-terminal methionine is not well-ordered and residues 1-393 are modeled. To be consistent with numbering canonically used for LMO residues, the N-terminal methionine is numbered as

residue 0.{Giegel, 1990 #302} Final refinements for WT LMO are still underway and all subsequent analysis is carried out using the 1.7 Å structure of LMO C203A. Given the stronger electron density of chains A-D, they are the only chains we consider reliable for defining details of the structures, so only they are described from here forward. Also, these chains closely overlay with each other with a Cα rmsd < 0.4 Å, so descriptions based on chain A can be considered as true for all chains unless otherwise noted. Unless noted, all descriptions and analyses are based on the higher resolution structure of C203A.

Table 7.1. Data collection and refinement statistics for LMO structures.[a]

	WT	C203A
Space group	$P42_12$	$P42_12$
Data Statistics		
Unit cell *a, b, c* axes (Å)	148.40, 148.40, 272.60	149.62, 149.62, 274.25
Resolution (Å)	77.49-2.30 (2.38-2.30)	30.90-1.70 (1.74-1.70)
Unique reflections	123105	33380 (24764)
Multiplicity	3.1	26.7 (27.5)
Average I/σ	10.4 (2.6)	15.5 (0.6)
R_{meas} (%)	10.2 (42.2)	18.2 (693.9)
Completeness (%)	91.0 (79.6)	99.9 (100)
$CC_{1/2}$	nd	99.9 (22.7)
Refinement Statistics		
Amino acid residues	2282	2363
Solvent atoms	1252	1464
Non-H atoms	19011	20104
RMS bonds (Å)	0.015	0.014
RMS angles (˚)	1.6	1.4
φ, ψ favored (%) [b]	94.7	96.4
φ, ψ outliers (%) [b]	1.0	0.6
$<B_{protein}>$ ($Å^2$)	71	69
$<B_{solvent}>$ ($Å^2$)	46	52
R_{work} (%)	22.1	17.1
R_{free} (%)	27.3	19.8
PDB code		

[a] Numbers in parentheses represent data for the high-resolution shell.
[b] Ramachandran statistics defined by Molprobity. [279]

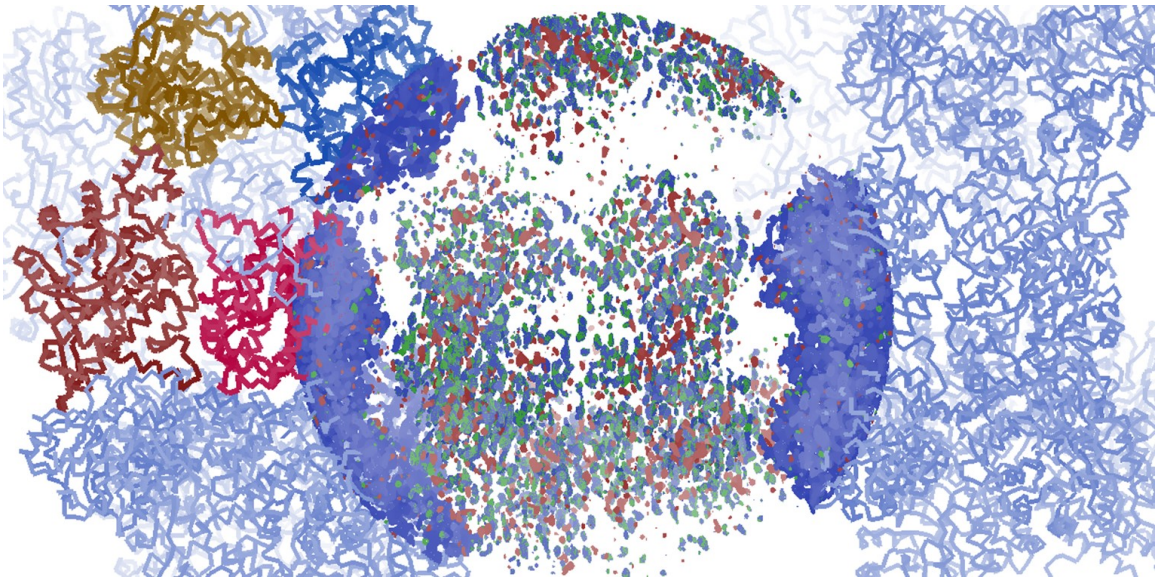


Figure 7.2. Crystal lattice with only four well-ordered chains modeled.
Crystal lattice generated using crystallography symmetry when only four well-ordered chains (Chains A, B, C, and D, displayed as ribbons) are modeled in C203A. A radius of $2F_o$-F_c (blue; contoured at 1.3 ρ_{rms}) and F_o-F_c (red and green; contoured at 3.0 ρ_{rms}) electron density maps are shown centered in the large swath of the crystal in which no model was initially built.

Structure of LMO

As expected for this family, the LMO monomer is organized into a β_8/α_8 TIM-barrel fold (Figure 7.3A). Each chain contains an FMN with sulfate ion nearby, bound above the *si* face of the isoalloxazine ring where substrate would bind (Figure 7.4). LMO is known to assemble into an octamer in solution.[280-281] In these crystals, each of the LMO chains in the asymmetric unit is part of a tetramer created by the crystallographic fourfold symmetry with the C-termini gathered into the center of the tetramer. Pairs of the unique tetramers further assemble into an octamer, with octamers formed from chains A and C, chains B and D, and chains G and H (Figure 7.3B).

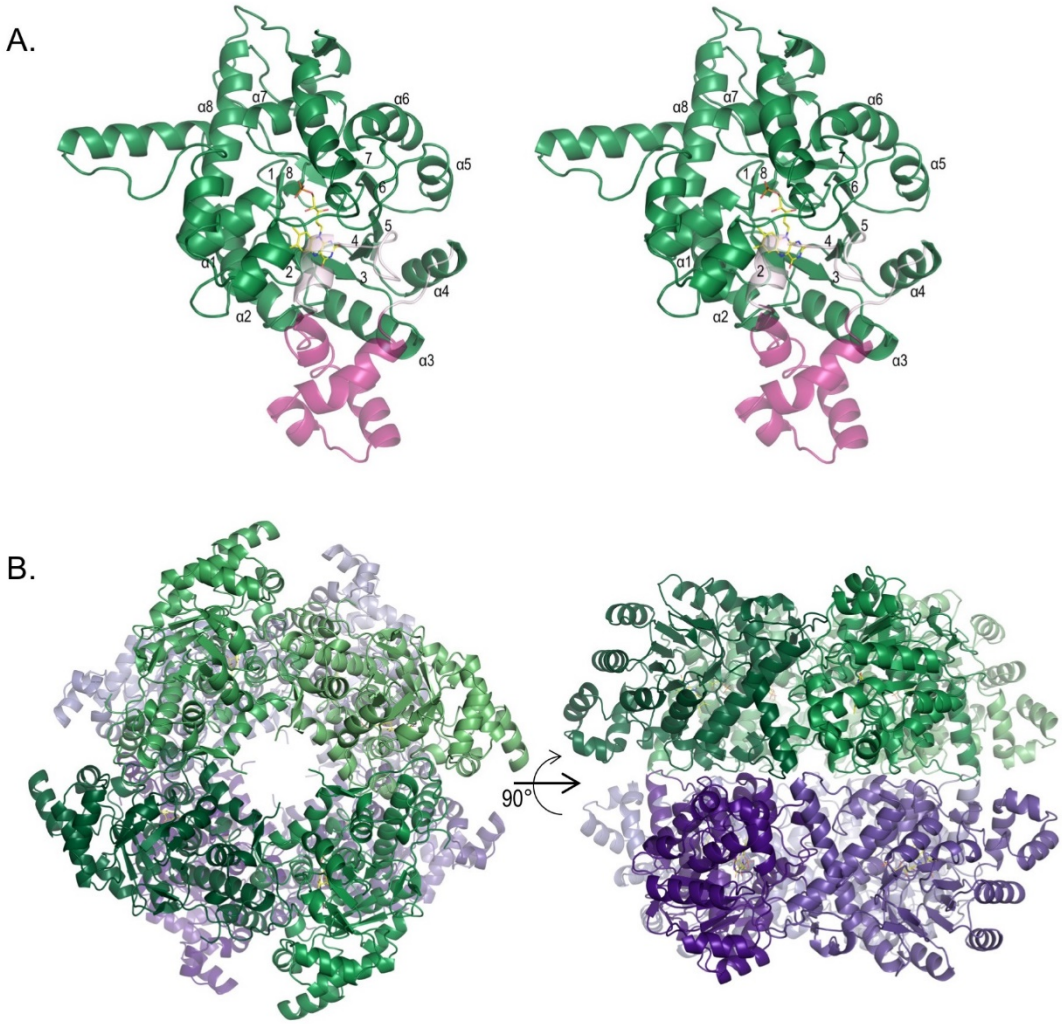


Figure 7.3. Crystal structure of LMO from *Mycobacterium smegmatis* at 1.70 Å resolution.
A. Stereoview of one monomer subunit of LMO (C203A chain A) with FMN (yellow carbons). The α-helices and β-strands (numbers) making up the core TIM-barrel fold are labeled. The section between β4 and α4 (canonical loop 4) is semitransparent with the conserved αD in light pink and variable "loop 4" in dark pink. **B.** Octamer assembly of LMO. Two tetramers formed by crystallographic fourfold symmetry that make up the octamer are shown in green (C203A chain A) and purple (C203A chain C). **C.** Octamer assembly of LMO. View rotated about 90° with respect to B.

The active site region of LMO has very well-defined electron density and defines the positions of all seven residues conserved among the family of α-hydroxy acid oxidases: Tyr44, Tyr152, Asp180, Arg187, Lys266, His290, and Arg293 (Figure 7.4). Tyr44, Arg293, Arg187, and His290 form hydrogen bonds with an ordered sulfate, binding on the *si* side of the flavin where substrate would bind. Both His290

and Lys266 are positioned to carry out the supportive roles in catalysis observed in related enzymes. Asp180 hydrogen bonds with His290 Nδ1 and is well-positioned to stabilize the protonated His290 residue after proton abstraction.[263-264] Lys266 hydrogen bonds with FMN N1 and O2 and is well-positioned to interact with N1 of FMN during catalysis to stabilize an anionic semiquinone.[73, 76, 263-264] Additional ordered waters are observed in the active site and form hydrogen bonds with the sulfate and Tyr152.

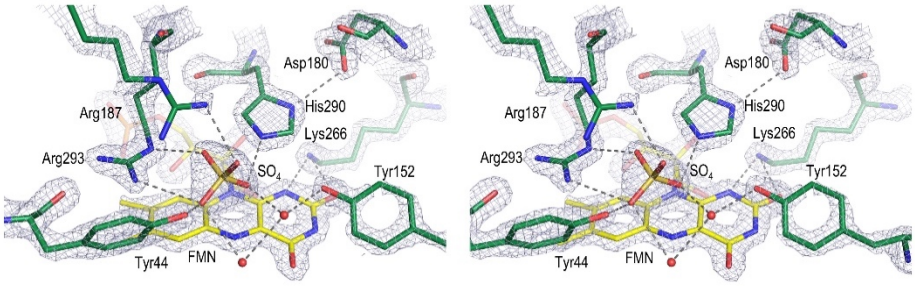


Figure 7.4. Electron density map quality and active site structure of LMO.
Stereoview of the LMO active site residues (green carbons), FMN (yellow carbons), sulfate, and ordered waters (red spheres) with $2F_o$-F_c electron density map (grey; contoured at 3.5 ρ_{rms}). Hydrogen bonds (dashed line) between active site side chains, sulfate, flavin, and waters are shown. Arg187 has an alternate conformation but only the A conformation (50%) interacting with the sulfate is shown.

Relationships to Other Structurally Known Proteins

A structural similarity search performed using the DALI server[201] showed that LMO is most similar to other members of the α-hydroxy acid oxidase family including GOX (rmsd ~1.4-1.9 Å, 33-36% sequence identity), MDH (rmsd 2.0 Å, 32% sequence identity), LOX (rmsd ~1.6-1.8 Å, 31-33% sequence identity) and flavocytochrome b_2 (rmsd 1.7 Å, 31% sequence identity). This indicates that LMO is as distant from each of the other α-hydroxy acid oxidases as they are from each other and is no more similar to LOX than it is to the others. Furthermore, this indicates that even with all of the structures that are known and have been solved from structural genomics projects, this is still the first structure representing the LMO clade.

Surveying the DALI search results for structures with the greatest amounts of their chain ordered, a representative of each enzyme type was chosen for a structure-

based sequence alignment (Figure 7.5): human GOX (2RDU),[267] *Aerococcus viridans* LOX (2E77),[265] *Saccharomyces cerevisiae* FCB2 (1KBI),[268] and *Pseudomonas putida* MDH (1P4C).[269] The structure-based sequence alignment of these family members shows almost all structural elements and all active site residues are conserved. As will be discussed further, a notable feature among the enzymes is that there is substantial variation in both length and structure of "loop 4," a segment between strand β_4 and helix α_4.

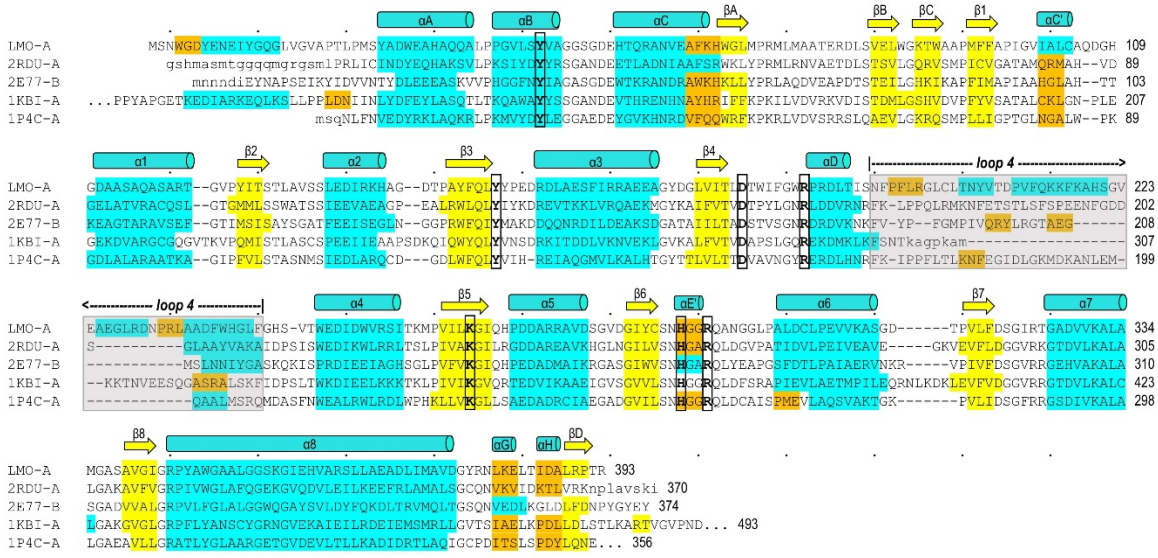


Figure 7.5. Structure-based sequence alignment of LMO with representative α-hydroxy acid oxidase family members.

LMO (WT sequence, structure from LMOC203A chain A) is shown first followed by 2RDU-A (GOX, chain A), 2E77-B (LOX, chain B), 1KBI (FCB2, chain A), and 1P4C-A (MDH-GOX chimera, chain A). Conserved active site residues (bold, black outline) and residues involved in α-helices (cyan), β-strands (yellow), and 3_{10}-helices (orange) are indicated. Only secondary structural elements conserved among all family members are named, in a manner consistent with naming conventions for other family members.[264, 266, 283] "Loop 4" is a dynamic section of these enzymes which covers the active site channel and is not structurally conserved (gray shading) among family members. Dots above the LMO sequence indicate every tenth character and, at the end of each line, a residue number for each of the sequence is shown. Residues in lower case letters are disordered in the structure.

An overlay of LMO with sulfate bound onto LOX and GOX structures having pyruvate and oxalate, respectively, bound in the active sites shows that the active site, FMN, and side chain conformations are quite similar (Figure 7.6). Only one side chain, that of Tyr44 in LMO, has notable variation, with Tyr44-OH of LMO shifted

~2.5 Å compared to LOX Tyr40 in order to hydrogen bond with the bound sulfate. In addition to the seven key, conserved active site residues, in LOX an additional tyrosine (Tyr215) contributes to the active site (Figure 7.6) and forms an additional hydrogen bond with the pyruvate product. This tyrosine is part of the variable "loop 4," and there is no residue in LMO or GOX that is a functional or structural equivalent.

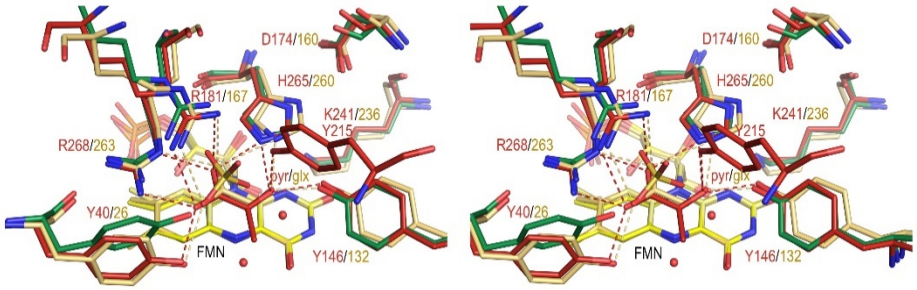


Figure 7.6. Active site overlay of LMO with LOX and GOX.
Stereoview of the LMO active site residues (green carbons), FMN (yellow carbons), sulfate, and ordered waters (red spheres) overlaid with LOX in complex with pyruvate (red carbons; 2E77) and GOX in complex glyoxylate (tan carbons; 2RDU). Hydrogen bonds (dashed line) between active site side chains and ligand in LOX (red) and GOX (beige) are shown.

Each of these active sites are set up in a nearly identical manner to carry out the same chemistry: oxidation of an α-hydroxy acid substrate. For example, in all structures, the catalytic base histidine (His290 in LMO, His265 in LOX, and His260 in GOX) is oriented to abstract a proton from substrate with an aspartic acid (Asp180 in LMO, Asp174 in LOX, and Asp160 in GOX) nearby to stabilize the protonated histidine. Likewise, a lysine (Lys266 in LMO, Lys241 in LOX, and Lys236 in GOX) is positioned to stabilize negative charge at the N1/O2 locus of the flavin during catalysis. The sulfate we see in the LMO active site appears to be a reasonable substrate mimic as the hydrogen bonds with sulfate and ordered waters observed in LMO match well with interactions with ligand observed in LOX and GOX. Specifically, the carboxylate oxygens of the ligands bound in LOX and GOX overlay well with two of the oxygens of sulfate in LMO and form equivalent hydrogen bonds to those sulfate makes with Arg187 and Arg293 in LMO. Similarly, hydrogen bonds

between O2 of pyruvate with residues equivalent to His 290 and Tyr152 mimic the hydrogen bonds made with a sulfate oxygen and ordered water, respectively, in LMO.

Major Differences Occur in "Loop 4" That Covers the Active Site

It has been noted in some structural studies of other α-hydroxy acid oxidases that these enzymes have significant sequence and structural divergence in a segment between β4 and α4, referred to as "loop 4."[267, 283] With the structure of LMO in hand, an obvious feature of LMO compared to other α-hydroxy acid oxidases, in particular LOX and GOX, is that the LMO "loop 4," is both longer and folds differently than the others (Figures 7.5 and 7.7). In LMO, "loop 4" is 49 residues long, and fortuitously, in our LMO structure, "loop 4" is well-ordered, structured similarly, and covers the active site opening in all chains. The structure it adopts includes a portion that folds into a compact bundle of four α-helices.

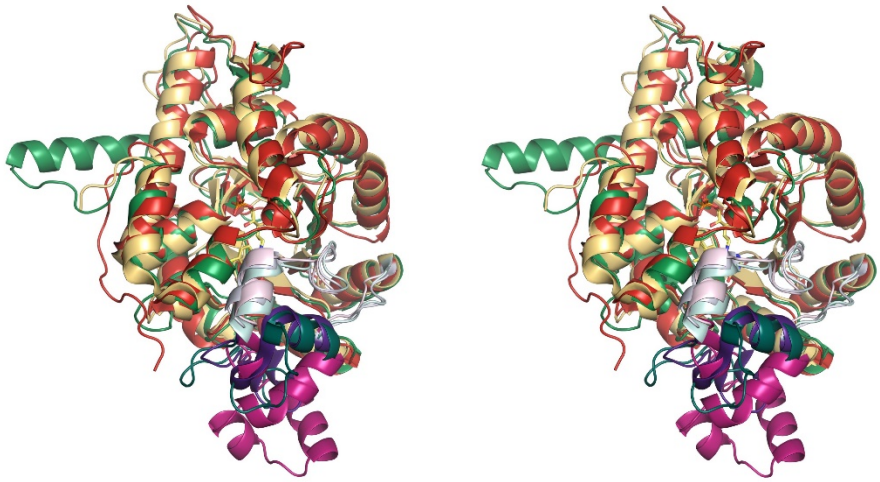


Figure 7.7. Overlays of LMO with LOX and GOX show differences in "loop 4" folding. Stereoview of the LMO monomer (green) overlaid with LOX in complex with pyruvate (red; 2E77) and GOX in complex glyoxylate (tan; 2RDU). The sections between β4 and α4 (canonical loop 4) are highlighted with the conserved elements, including αD, in light colors and variable "loop 4" in dark colors (shades of pink in LMO, shades of purple in LOX, and shades of teal in GOX).

In structures of LOX and GOX, this loop is often disordered,[267, 274-275] indicating its innate dynamic nature, and when they are ordered, they do not occupy the same conformation. However, in all cases, "loop 4" functions as a lid or flap, covering the active site pocket when folded (i.e. closed) and exposing the active site

channel to solvent when unfolded (i.e. open). Thus, movement of this loop is necessary for substrate binding and product release.[265, 267, 274-275, 284-285] As indicated by many different crystal structures, when open, "loop 4" is disordered, rather than occupying a different, open, and ordered conformation that exposes the active site. The differences in "loop 4" length, composition, and structure among these enzymes led us to further investigate the unique properties of the "loop 4" subdomain of LMO with the idea that loop dynamics may dictate the speed of pyruvate release and LMO's kinetics.

There is a large body of literature indicating that buried surface area is correlated to stability in proteins.[286-289] Using LMO, LOX, and GOX as model α-hydroxy acid oxidases, we asked how much surface area is buried when going from an unfolded "loop 4" to the fully folded "loop 4" functioning as an active site lid? To answer this question, the theoretical amount of solvent accessible surface area for an unfolded "loop 4" and the buried surface area upon "loop 4" folding and closure of the active site (taking both buried surface area within "loop 4" as well as at the interface with the rest of the protein into account) were considered (Table 7.2).

First, we asked if differences in loop lengths in the representative structures were characteristic and found that compared to LOXs and GOXs, LMOs consistently have longer "loop 4"s (Table 7.2). Additionally, folding and closure of "loop 4" of LMO over the active site buries the most surface area; in LMO, 4750 $Å^2$ is buried upon "loop 4" folding and closure compared to 3300 $Å^2$ and 3870 $Å^2$ in LOX and GOX, respectively. Purely based on surface area burial, the lid of LMO potentially has the most favorable energetics to remain folded and positioned in a closed orientation as opposed to an unfolded, disordered, open position. We suggest that the unfolding of "loop 4" comes at a lower energetic cost in LOX and GOX than in LMO. This means the "loop 4" lid in LMO will have a greater propensity to be closed, thereby slowing pyruvate release so that pyruvate remains in the active site and reacts further, rather than being released quickly as product.

Other Studies of "Loop 4" Support this Hypothesis

Site-directed mutagenesis studies to probe interactions involving residues in "loop 4" in LOX provide additional support for our hypothesis that "loop 4" dynamics is responsible for the slow dissociation of pyruvate in LMO. In LOX, an additional tyrosine (Tyr215) from "loop 4" contributes to the active site (Figure 7.6) and forms an additional hydrogen bond with the pyruvate product. When Tyr215 was mutated to a phenylalanine (Y215F), pyruvate release was decelerated, with the rate of pyruvate release decreasing by ~7-fold compared to wild-type to become the slowest step overall.[274] Tyr215 is thought to act as a sort of "latch" to hold "loop 4" in place and is involved in the "loop 4" conformational change required for substrate binding and product release. In Y215F, a hydrogen bond is replaced by potentially stronger hydrophobic interactions, so "loop 4" may interact more favorably with the rest of the protein in its closed form, thereby shifting the "loop 4" dynamics towards the closed form and attenuating product release as well as substrate binding. Similarly, mutating Y191, another conserved residue from "loop 4" in LOX, to phenylalanine, leucine, or alanine slowed the pyruvate release step by ~5-fold, 19-fold, and 19-fold, respectively.[275] Y191 does not form any direct interactions with the active site, but mutating this position also appears to shift "loop 4" dynamics to favor a closed conformation and the rate of lid opening. While neither of these studies have generated a mutant with increase flux through a coupled pathway (as observed in LMO), they do change the rate of pyruvate release, a rate constant that is interpreted structurally to represent the "loop 4" lid opening requisite for product dissociation.[274-275]

Table 7.2. Solvent accessible surface area of variable "loop 4" upon folding and positioning over active site channel.

	LMO	LOX	GOX
SASA unfolded [1] ($Å^2$)	6240	3630	4710
SASA folded lid, isolated [2] ($Å^2$)	3670	2490	3050
SASA folded lid, closed [2] ($Å^2$)	2510	1300	1860
Buried SA folded lid, closed ($Å^2$)	3730 (6240-2510)	2330 (3630-1300)	2850 (4710-1860)
Buried SA fold lid, closed, both sides of interface ($Å^2$)	4750 (3730+1020)	3200 (2330+960)	3870 (2850+1020)
Range of lid/loop 4 lengths (residues)*	39-49 [3] (49)	27-35 [4] (29)	17-38 [5] (38)

[1] Average of upper and lower bound proposed by Creamer [290-291]
[2] Calculated with Areaimol in CCP4 [292-293]
[3] BLAST search with LMO (1000 hits, goes to ~50% sequence identity)), CD-Hit clustering with 70% sequence identity (same set used for phylogenetic tree), looking at equivalent loop 4 lengths
[4] BLAST search with 2E77_B (1000 hits, goes to ~50% sequence identity), COBALT alignment, looking at equivalent loop 4 lengths
[5] BLAST search with 2RDU_A (250 hits, goes to ~70% sequence identity), COBALT alignment omitting 2 sequences for low quality (also unusually long so noticeable), looking at equivalent loop 4 lengths
*number in parenthesis residue length in this representative structure
SASA: solvent accessible surface area, SA: surface area

Phylogenetic Studies

As far as we could see in the literature, LMO has exclusively been studied from mycobacteria since it was first identified in 1947.[68, 73, 263, 280] However, the physiological role of LMO remains unknown. With the extensive number of genomes now having been sequenced, it is possible to gain an idea of the range of organisms that have an LMO (Figure 7.9).

LMO is most abundantly represented in Actinomycetales, an order of gram-positive and generally aerobic bacteria including *Mycobacterium*, *Nocardia*, *Streptomyces*, *Rhodococcus,* and *Frankia*.[294] LMO are also identified in bacteria from α- and γ-proteobacteria, plantomycetes, pseudocardiales, thermophilia, geodermatophilales, and bacilli. Surprisingly, LMO sequences are also found in the fungus *Beauveria bassiana* and the archaea *Halopiger salifodinae*. The LMOs from

both of these organisms group with bacteria. *Halopiger salifodinae* is halophilic and groups with bacteria including the halophilic proteobacteria *Salincola salarius* which may thrive and be found in similar environments. Interestingly, *Beauveria bassiana,* an insect pathogen, groups closely with the human pathogens, *Mycobacterium abscessus*, responsible for a multidrug-resistant diseases in humans,[295] and *Tsukamurella pulmonis*, originally isolated from bedbugs and now mostly commonly associated with infections in immunocompromised humans.[296] There are many examples of individual genes and gene clusters arising in fungi and archaea through horizontal gene transfer with bacteria,[297-299] and it is plausible that the presence of LMO in these organisms presents another example of this evolutionary phenomenon. However, these fungal and archaeal sequences could be from contaminations and further analysis needs to be done. In this phylogenetic tree, LOX (2E77), GOX (2RDU), FCB2 (1KBI), and MDH (1P4C) are clearly much more distant from these LMOs than the LMOs are from each other, consistent with all of the proteins included based on our ~50% sequence identity cutoff being truly LMOs rather than one of the other known α-hydroxy acid oxidase enzymes. It will be very interesting to further explore the physiological role and distribution of LMOs in nature.

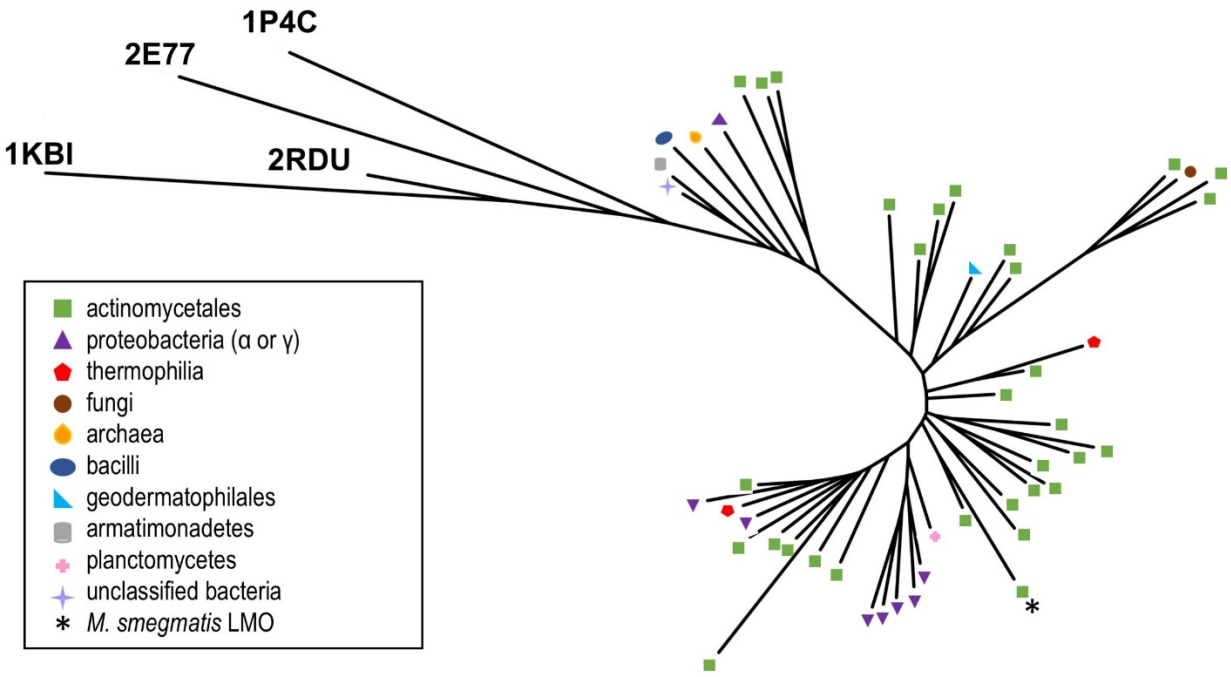


Figure 7.9. Relatedness tree and distribution of lactate monooxygenases and structurally known α-hydroxy acid oxidases.
Branches for putative LMOs sequences are annotated according to the organism where they are found. Branches for individual PDB entries are labeled with PDB deposition codes: 2E77 (LOX), 2RDU (GOX), 1P4C (MDH), and 1KBI (FCB2).

Materials and Methods

Crystallography and structure determination

Recombinant LMO (WT and C203A) from *M. smegmatis* was provided by Dr. Vince Massey from the University of Michigan. The protein sample was originally stored in 1.0 M acetate buffer pH 5.4 at 4 °C. Before crystallization, protein was switched to 5 mM HEPES pH 7.0, 10 μM PMSF and concentrated to 20 mg/mL. Aliquots were frozen and stored at -80 °C.

The proteins were crystallized at 4°C in hanging drops with a reservoir solution of 0.1 M sodium citrate pH 4.6, 0.4 M lithium sulfate, 0.4 M ammonium sulfate. For WT, crystals were transferred into a series of artificial mother liquors containing 0%, 5%, 15%, and 25% glycerol and plunged into liquid nitrogen. Data were collected at CHESS station A1 using $\Delta\varphi = 0.5°$ rotation and processed using DENZO.[300]

For C203A, crystals were transferred to an artificial mother liquor containing 25% glycerol and plunged into liquid nitrogen. Data were collected at beamline 5.0.2 at the Advanced Light Source using λ= 1.0 Å, Δφ = 0.25° rotation, 0.25 second exposure, collecting 360° total. Data were processed using XDS[301] and the CCP4 program suite.[302] A $CC_{1/2}$ cutoff criteria[156] of ~0.2 after merging was used to define the resolution cutoff of 1.70 Å, and a random 5% of reflections in resolution bins were marked for cross-validation.

Largely due to the unusual observed crystal packing, these structures proved challenging for molecular replacement and were tackled in an atypical, multistep manner. The phase problem was initially solved by molecular replacement using MR-Rosetta with default settings[197] using the 2.3 Å WT data set and a homology search model generated by hhpred.[303] The initial solution contained 4 chains with 1687 residues built and R/R_{free} of 0.34/0.38. Subsequent manual modeling building was done in Coot,[198, 304] and this model was refined using phenix.[199]

In an attempt to place and model additional chains in regions of the crystal having substantially weaker electron density (in addition to 4 well-ordered chains A, B, C, and D), MR Rosetta, Autobuild and Phaser were used consecutively. After obtaining a higher resolution C203A data set, the existing WT model was used to seed an MR-Rosetta analysis of the 1.7 Å C203A data set. The resulting solution from MR Rosetta contained 4 chains with 1522 residues built and R/R_{free} of 0.26/0.27. This solution was then used to seed Autobuild,[305] which led to no discernable changes in the models. After additional manual modeling and refinement, this model was defined as an already placed model in Phaser[217] which was run using C203A Chain A as a search model and with the packing criteria turned off. Phaser gave a solution which did not meet standard packing criteria and that contained two additional chains (chains G and H in the final model). These chains corresponded to the weaker electron density region of the crystal and with these additional chains placed, a confirmation of the correctness of the chain placements was provided by a difference map that had strong positive peaks in both chains corresponding to where FAD and sulfate should be found. This model containing 4 well-ordered chains (A, B, C, and D) and 2 less ordered chains (G and H) was used with further manual modeling done

in Coot[198, 304] and all refinements carried out using Phenix.[199] Chains G and H are modeled as 85% occupancy based on the average occupancy for these chains that resulted from allowing the occupancies to refine in Phenix. Water molecules were manually placed based on having electron density $\geq 4.0\ \rho_{rms}$ in Fo-Fc maps and $\geq 0.9\ \rho_{rms}$ in $2F_o$-F_c maps and reasonable hydrogen bonding interactions. Final rounds of refinement on the C203A structure were done with riding hydrogens and TLS using one group per chain (i.e. six total TLS groups). Refinement statistics are provided in Table 7.1.

Final refinements for the WT structure of LMO are still underway. The final LMO WT structure was generated by resolving it by molecular replacement using a nearly final C203A structure as the search model.

Phylogenetic studies

A BLAST search in March 2018 using the *Ms*LMO sequence as the query generated 1000 putative LMO sequences with e-values as low as ~8×10^{-112} and sequence identity to ~50%. To generate a set of diverse representative sequences, the 1000 sequences were clustered using CD-Hit[306] with sequence identity cut off of 70%, resulting in 48 clusters. The representative 48 sequences output by CD-Hit and four sequences for representative PDB structures of GOX, LOX, MDH, and FCB2 were aligned with MUSCLE[307] and this alignment was used to generate a tree with PhyML.[210]

Surface Area Calculations

The solvent accessible surface areas of folded proteins, both with and without loop 4, were calculated using Areaimol.[292-293] The solvent accessible surface areas of unfolded loop 4 were calculated using the upper and lower bound model proposed by Creamer *et al*.[290-291, 308] Values reported in Table 2 are the average between the calculated upper and lower limits of solvent accessible surface area.

The bounds of loop 4 for these calculations were determined based on where representative structures for GOX, LOX, FCB2, and MDH deviated (equivalent to residue 195 in LMO) and became similar again (equivalent to residue 244 in LMO) in

a structural overlay. The range of loop 4 lengths were determined using COBALT alignments generated from a BLAST search using the representative enzyme (2RDU for GOX, 2E77 for LOX) with 1000 target sequences. In the case of GOX, two sequences which were annotated as “poor” and were clear outliers in terms of sequence length were eliminated.

Chapter 8

Concluding Discussion and Outlook

All of the studies presented in this dissertation provide illustrative examples for how protein structure can be used to illuminate details of function and mechanism – both broadly and in a more nuanced manner. Below, I will detail what I consider to be the main highlights and impacts of this work, both in terms of understanding these individual protein families and more broadly protein structure and function. Then, I will suggest some future directions for this work that I feel would be most compelling, particularly in the context of these structural insights.

Impacts and Highlights of Reported Work

Sedoheptulose 7-Phosphate Cyclases

Sedoheptulose 7-phosphate cyclases may differentiate themselves by selectively binding different substrates. In previous work on SH7PCs, SH7P was always depicted in one way: as the α-pyranose anomer. However, as is common for sugars, SH7P is a dynamic molecule that is readily interconverting between its α- and β-furanose, and pyranose, and its linear forms. Understanding the differences in these enzymes required a shift in the way we were thinking about SH7PCs in terms of their chemistry and substrate; rather than viewing SH7P as one, static substrate, we really needed to view it as five interconverting forms of the substrate. This led us to our novel anomer selection hypothesis in which EEVS and DDGS are differentiated from EVS early in catalysis by binding different anomers of SH7P, with EEVS and DDGS, binding the α-pyranose anomer, and EVS binding the β-pyranose anomer. This hypothesis is supported by a characteristic difference near the anomeric position in the active site and is more compelling than the previous proposal that the different stereochemistry in EV is generated by EVS through an intermediate undergoing a 180° rotation before ring closure.

Glycerol 3-Phosphate Oxidase

Flavin-dependent glycerol 3-phosphate oxidases and dehydrogenases come in two types. After solving the structure of *Mp*GlpO, we found that this enzyme is distinct from other, structurally known GlpO/DHs and represents a new type of these enzymes. Mitochondrial and bacterial GlpDHs and *Streptococcus sp*. GlpO group

together as "Type I" GlpO/DHs while *Mp*GlpO represents a "Type II" GlpO/DH. These two types have consistent differences in their active sites and distribution in nature.

We proposed the first plausible substrate binding mode and catalytic mechanism for glycerol 3-phosphate oxidases and dehydrogenases. Guided by a reliable, ligand-bound structure of a closely related protein as well as the positions of ordered solvent in the active site of *Mp*GlpO, we were able to propose a plausible binding mode for glycerol 3-phosphate that further allowed us to propose a specific catalytic mechanism. Despite differences in the active site composition of Type I and Type II GlpO/DHs, both types are functionally set up for the same binding mode and catalytic mechanism, so this proposal applies for all GlpO/DHs, including pathogenic GlpOs and mitochondrial GlpDH that are of particular interest.

Ferredoxin:$NADP^+$ Reductase

Aromatic placeholder variants are a relevant and catalytically active model system. Even though mutating the aromatic placeholder to capture productive NADP(H) binding is a common methodology for studying FNR and FNR-like superfamily members,[59, 64, 222] work from the group of Milagros Medina called the relevance of these variants into question in 2010.[236] Based on their analysis of stopped-flow experiments, they concluded that these variants were not able to generate NADPH (i.e. could not carry out the physiologically relevant hydride transfer), and they further proposed a new catalytic mechanism wherein this aromatic placeholder residue played a more active role in catalysis.[309] Our reanalysis of the data from their stopped-flow experiments showed that these variants are catalytically active in both directions, and thus do represent a productive complex. Most importantly, this mutation strategy allowed us to capture a productive binding for NADP(H) in the active site, otherwise impossible with wild-type enzymes.

Active site compression is an important element promoting catalysis in FNR, and enzymes in general. In FNR, reduced anisotropic mobility of the C4 atom of NADPH compared to $NADP^+$, distortion of FAD geometry from planarity, and a tightly packed active site implicate significant active site compression as a factor

promoting hydride transfer. In this work, we further suggest that compression is an important, but often overlooked and underappreciated element of catalysis that is relevant not just for these enzymes. This underappreciation is at least in part because it is difficult to obtain experimental evidence to indicate the presence of compression; using high or ultra-high resolution structures is one of the ways to obtain this sort of experimental evidence.

This work also provided an illustrative example of the wealth of information contained within high resolution structures. Although 0.3 Å may not seem like much, going from the 1.8 Å structures of FNR solved in 1999[59] to these ~1.5 Å structures provided us with a wealth of additional information and was needed for us to visualize directional movements within the active site and deviations from standard geometry.

Lactate Monooxygenase

LMO is equally evolutionarily distant from other α-hydroxy acid oxidases. While one might expect LMO to be most similar to lactate oxidase due to its similar catalytic activity, LMO is actually equally distinct from lactate oxidase, glycolate oxidase, mandelate dehydrogenase, and flavocytochrome b_2 as indicated by structural similarity as well as sequence identity (both common indicators of phylogenetic distance). This structure is also the first representative structure for the LMO clade of enzymes.

Major differences in "loop 4" in α-hydroxy acid oxidases may drive its kinetic differences. The distinct kinetic and functional properties of LMO compared to other α-hydroxy acid oxidases have been a point of interest for decades, but the underlying mechanism has long eluded scientists. With the first structure of LMO in hand, one thing that stuck out almost immediately in structural comparisons with other α-hydroxy acid oxidases was the unique length and conformation of a variable section, referred to as "loop 4," in LMO. Further analysis of this "loop 4" folded subdomain in LMO and comparison to lactate oxidase and glycolate led us to propose that the folding and stability of "loop 4" in LMO may favor a closed conformation, slowing

product release and shifting the kinetic properties of this enzyme to a coupled pathway.

A General Lesson

Structural comparison is a powerful tool. One common thread that runs through much of this dissertation is the use of structural comparison to gain additional information that any particular structure may be lacking. For example, a liganded structure produced by the Protein Structure Initiative coupled with ordered solvent molecules in the active site of *Mp*GlpO guided us in modeling glycerol 3-phosphate in the active site when we could not obtain a liganded structure ourselves. Similarly, ordered solvent molecules in the active site of Ava (DDGS) guided us in modeling the α-pyranose anomer of SH7P in the active site. While obtaining a true, ligand bound structure is always preferred, the wealth of existing structures, including the thousands produced by the NIH-sponsored Protein Structure Initiative,[310] can be used in a powerful way to guide our understanding of substrate binding and catalysis.

Directions of Future Research

While we have learned much about sedoheptulose 7-phosphate cyclases and the flavoenzymes glycerol 3-phosphate oxidase, ferredoxin:$NADP^+$ reductase, and lactate monooxygenase over the last five years, these projects are not complete. In particular, we have been able to utilize structural information to propose multiple different hypotheses with respect to function and mechanism, but these hypotheses remain to be experimentally substantiated. There are many directions in which these research projects could be taken, and here I provide a few that I think are particularly interesting and derive from our structural insights.

Sedoheptulose 7-Phosphate Cyclases

Does anomer selection differentiate the SH7PCs? Guided by structural information, we proposed a novel anomer selection hypothesis (mentioned above) for how EEVS and DDGS differentiate from EVS before the first step of catalysis by selectively binding different anomers of the substrate, SH7P. While this provides a

more satisfying and compelling explanation for how the products of EEVS and EVS only differ in the stereochemistry at C5 than that originally proposed,[12] this truly is a hypothesis that has yet to be experimentally tested. This hypothesis was guided by structure and is compatible and consistent with the robust loop difference and variation in active site arginine placement in SH7PCs as well as our subsequent docking of α-pyranose SH7P in the active site of Ava.

Obtaining liganded structures of the SH7PCs would be particularly helpful for assessing the proposed anomer selection hypothesis. Attempts thus far to obtain liganded structures of Ava or ValA with SH7P or other sugars have been unsuccessful. A liganded structure of dehydroquinate synthase (DHQS) was obtained using carbaphosphonate (phosphonate mimic of substrate).[131] The equivalent studies with non-catalyzable phosphonate mimics of both α- and β-pyranose SH7P could be undertaken to characterize substrate binding in EEVS, DDGS, and EVS and to determine if these enzymes do selectively bind one anomer over the other.

Additionally, this anomer selection hypothesis is linked to a robust loop difference observed between EEVS and DDGS versus EVS and DHQS. Akin to the point mutation work undertaken in Chapter 4 to test the roles of unique active site residues in EEVS and DDGS, it would be interesting to test the effects of this Asp/Asn change proposed to modulate substrate specificity. Indeed, if the ability to selectively bind α- or β-pyranose SH7P is what differentiates EEVS and EVS, this mutation could interconvert activity between these enzymes.

What does gadusol do in non-mammalian vertebrates? How does the dual-function methyltransferase-oxidoreductase work and what does it look like? Although not a focus of the work in this thesis, the production and role of gadusol is the focus of the continuing work on SH7PCs in the Mahmud Lab. With the identification of EEVS in non-mammalian vertebrates, its partner enzyme MT-Ox, and gadusol (included in Chapter 2),[14] SH7PC work has taken on a new motivation to understand gadusol's function *in vivo* and to utilize gadusol as a natural product with UV-protective properties. Some initial work in the Karplus Lab in this direction (unpublished) has been in trying to solve the structure of each of the MT and OX domains of the fusion protein, and an unpublished structure of the MT domain is in

hand. Additionally, in collaboration with the Mahmud Lab, work is underway to identify and characterize the order in which these individual domains act during gadusol synthesis using MT and Ox constructs we generated.

Glycerol 3-Phosphate Oxidase

Does substrate bind in the proposed manner? Guided by a reliable, ligand-bound structure of a closely related protein, we were able to propose a binding mode for glycerol 3-phosphate and a catalytic mechanism, but this hypothesis needs further testing. Ideally, we would obtain liganded structures of *Mp*GlpO, a Type II GlpO/DH, and a representative Type I GlpO/DH to determine how substrate binds and if this mode is conserved among this enzyme family. Additionally, no site-directed mutagenesis studies to probe the functions of key active site residues, such as the proposed catalytic base histidine (His51 in *Mp*GlpO) or conserved arginine interacting with the phosphate of glycerol 3-phosphate (Arg230 in *Mp*GlpO), have been published. This type of study would further substantiate the roles of these residues in substrate binding and catalysis in both types of GlpO/DHs.

*Can drugs be designed to target only pathogenic glycerol 3-phosphate oxidases? Mp*GlpO was initially selected as a target for structural studies because of its role as a pathogenicity factor. Based on structural analysis, active sites of Type I and Type II GlpO/DHs are set up to carry out the same catalysis but are distinctly different in their amino acid compositions. This presents potential for inhibitor design to target only pathogenic, Type II GlpO/DHs without affecting the Type I, mitochondrial GlpDHs that are necessary for life in a mammalian host.

Ferredoxin:$NADP^+$ Reductase

What can FNR help us understand about FNR-like superfamily members? As FNR has been studied over multiple decades, fewer questions related to our basic understanding of this protein remain. Most of the future work related to FNR lies in using our understanding of FNR to guide our understanding and future studies of other, more complex, and biomedically relevant enzymes like NADPH oxidase (NOX). With one serendipitous mutational strategy in hand to obtain structures of

NOX,[239] hopefully more structures of (at least pieces of) NOX are on the horizon. Our understanding FNR has been a useful guide for studies of cytochrome P450 reductase and nitric oxide synthase – so NOX would the next target.

Lactate Monooxygenase

As mentioned in the Introduction and in Chapter 7, LMO has fallen out of favor over recent decades, replaced by its more biotechnologically relevant relative, lactate oxidase (LOX). I hope that with this structure and the first publication directly related to LMO in about 20 years, interest in LMO will be rekindled and additional questions related to this unique enzyme can be answered.

What is the physiological role of LMO? M. smegmatis is commonly used as a model system for *M. tuberculosis*. But a simple search of the genome of *M. tuberculosis* indicates that it does not have an LMO. The physiological function of LMO is unknown although lactate metabolism has been implicated as a pathogenicity factor in other pathogens (although via other lactate metabolizing enzymes, such as LOX) by enhancing metabolism.[311] It seems feasible that LMO could play a similar role – allowing organisms to utilize lactate, feeding acetate into the citric acid cycle. There is also some evidence that LMO expression levels are regulated in response to nutrient levels[312] and acid stress.[313]

Is the dynamic "loop 4" responsible for the distinct kinetic properties of LMO compared with other α-hydroxyacid oxidase family members? Studies to investigate whether the dynamic "loop 4" really is responsible for shifting the kinetic properties of LMO from other α-hydroxyacid oxidase family members are necessary. Studies of other α-hydroxyacid oxidase family members have acknowledged the dynamic nature of this loop and its plausible role in regulation,[267, 274-275, 314] but this is the first time LMO has been the subject. For example, as mentioned in Chapter 7, studies in LOX targeting a conserved tyrosine which acts as a sort of "latch" to hold "loop 4" in place have shown that mutating this tyrosine to phenylalanine shifts the kinetic behavior of LOX. While this does not fully change the activity from that of LOX to that of LMO, it does slow the release of pyruvate from the active site and is consistent with our hypothesis that "loop 4" dynamics impact product release.[274] Similarly, studies to

perturb the compact folding of "loop 4" in LMO could parse out the roles of "loop 4" stability and folding in dictating the function of LMO.

Concluding Remarks

In this dissertation, I have presented a series of research projects regarding the structure and function of sedoheptulose 7-phosphate cyclases and three families of flavoenzymes and our abilities to use structural understanding to answer a variety of questions and gain insight into a range of details with regards to enzyme function and catalysis. These studies not only contribute to our understanding of these specific enzymes – sedoheptulose 7-phosphate cyclases, GlpO, FNR, and LMO – but also provide illustrative examples of the power of structural studies for understanding enzyme mechanism and function. Additionally, these provide insights into catalysis which are broadly informative for studying enzymes – for example, the role of active site compression in catalysis as demonstrated by FNR, the role of dynamics in modulating enzyme function as demonstrated by FNR and LMO, the utility of using carefully selected, closely related enzymes to guide and improve our understanding of our proteins of interest (all cases). It is my hope that readers can appreciate the enlightening and sometimes unexpected insights and understanding of these enzymes that we would not have without atomic-level structural information obtained from protein crystallography.

Bibliography

1. Regan, L.; Caballero, D.; Hinrichsen, M. R.; Virrueta, A.; Williams, D. M.; O'Hern, C. S., Protein design: Past, present, and future. *Biopolymers* **2015,** *104* (4), 334-50.

2. Cragg Gordon, M.; Newman David, J., Biodiversity: A continuing source of novel drug leads. In *Pure Appl. Chem.*, 2005; Vol. 77, p 7.

3. Mahdi, J. G.; Mahdi, A. J.; Mahdi, A. J.; Bowen, I. D., The historical analysis of aspirin discovery, its relation to the willow tree and antiproliferative and anticancer potential. *Cell Prolif* **2006,** *39* (2), 147-55.

4. Dias, D. A.; Urban, S.; Roessner, U., A historical overview of natural products in drug discovery. *Metabolites* **2012,** *2* (2), 303-36.

5. Achan, J.; Talisuna, A. O.; Erhart, A.; Yeka, A.; Tibenderana, J. K.; Baliraine, F. N.; Rosenthal, P. J.; D'Alessandro, U., Quinine, an old anti-malarial drug in a modern world: role in the treatment of malaria. *Malar J* **2011,** *10*, 144.

6. Fleming, A., On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzæ. *British journal of experimental pathology* **1929,** *10* (3), 226-236.

7. Newman, D. J.; Cragg, G. M., Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* **2012,** *75* (3), 311-35.

8. Bernhardt, P.; O'Connor, S. E., Opportunities for enzyme engineering in natural product biosynthesis. *Curr. Opin. Chem. Biol.* **2009,** *13* (1), 35-42.

9. Mahmud, T.; Tornus, I.; Egelkrout, E.; Wolf, E.; Uy, C.; Floss, H. G.; Lee, S., Biosynthetic studies on the alpha-glucosidase inhibitor acarbose in Actinoplanes sp.: 2-epi-5-epi-valiolone is the direct precursor of the valienamine moiety. *J. Am. Chem. Soc.* **1999,** *121* (30), 6973-6983.

10. Stratmann, A.; Mahmud, T.; Lee, S.; Distler, J.; Floss, H. G.; Piepersberg, W., The AcbC protein from Actinoplanes species is a C7-cyclitol synthase related to 3-dehydroquinate synthases and is involved in the biosynthesis of the alpha-glucosidase inhibitor acarbose. *J. Biol. Chem.* **1999,** *274* (16), 10889-96.

11. Wu, X.; Flatt, P. M.; Schlorke, O.; Zeeck, A.; Dairi, T.; Mahmud, T., A comparative analysis of the sugar phosphate cyclase superfamily involved in primary and secondary metabolism. *ChemBioChem* **2007,** *8* (2), 239-248.

12. Asamizu, S.; Xie, P.; Brumsted, C. J.; Flatt, P. M.; Mahmud, T., Evolutionary divergence of sedoheptulose 7-phosphate cyclases leads to several distinct cyclic products. *J. Am. Chem. Soc.* **2012,** *134* (29), 12219-29.

13. Balskus, E. P.; Walsh, C. T., The genetic and molecular basis for sunscreen biosynthesis in cyanobacteria. *Science (New York, N.Y* **2010,** *329* (5999), 1653-6.

14. Osborn, A. R.; Almabruk, K. H.; Holzwarth, G.; Asamizu, S.; LaDu, J.; Kean, K. M.; Karplus, P. A.; Tanguay, R. L.; Bakalinsky, A. T.; Mahmud, T., De novo synthesis of a sunscreen compound in vertebrates. *Elife* **2015,** *4*.

15. Yu, Y.; Bai, L. Q.; Minagawa, K.; Jian, X. H.; Li, L.; Li, J. L.; Chen, S. Y.; Cao, E. H.; Mahmud, T.; Floss, H. G.; Zhou, X. F.; Deng, Z. X., Gene cluster

responsible for validamycin biosynthesis in Streptomyces hygroscopicus subsp jinggangensis 5008. *Appl. Environ. Microbiol.* **2005,** *71* (9), 5066-5076.

16. Osborn, A. R.; Kean, K. M.; Karplus, P. A.; Mahmud, T., The sedoheptulose 7-phosphate cyclases and their emerging roles in biology and ecology. *Nat. Prod. Rep.* **2017,** *34* (8), 945-956.

17. Kean, K. M.; Codding, S. J.; Asamizu, S.; Mahmud, T.; Karplus, P. A., Structure of a sedoheptulose 7-phosphate cyclase: ValA from Streptomyces hygroscopicus. *Biochemistry* **2014,** *53* (26), 4250-60.

18. Osborn, A. R.; Kean, K. M.; Alseud, K. M.; Almabruk, K. H.; Asamizu, S.; Lee, J. A.; Karplus, P. A.; Mahmud, T., Evolution and Distribution of C7-Cyclitol Synthases in Prokaryotes and Eukaryotes. *ACS Chem. Biol.* **2017**.

19. Warburg, O. C., W., Ein zweites sauerstoffübertragendes Ferment und sein Absorptionspektrum. *Naturwissenschaften* **1932,** *20*, 688.

20. Theorell, H., Purification of the active group of the yellow enzyme. *Biochem Z* **1936,** *275* (344-346).

21. Krebs, H. A., Metabolism of amino-acids: The synthesis of glutamine from glutamic acid and ammonia, and the enzymic hydrolysis of glutamine in animal tissues. *Biochem. J.* **1935,** *29* (8), 1951-69.

22. Warburg, O.; Christian, W., Coenzyme of D-Alanine Oxidase. *Biochem Z* **1938,** *296*, 294.

23. Fraaije, M. W.; Mattevi, A., Flavoenzymes: diverse catalysts with recurrent features. *Trends Biochem. Sci.* **2000,** *25* (3), 126-32.

24. Romero, E.; Gomez Castellanos, J. R.; Gadda, G.; Fraaije, M. W.; Mattevi, A., Same Substrate, Many Reactions: Oxygen Activation in Flavoenzymes. *Chem. Rev.* **2018,** *118* (4), 1742-1769.

25. Massey, V., Introduction: flavoprotein structure and mechanism. *FASEB J.* **1995,** *9* (7), 473-5.

26. Lienhart, W.-D.; Gudipati, V.; Macheroux, P., The human flavoproteome. *Arch. Biochem. Biophys.* **2013,** *535* (2), 150-162.

27. Chaves, I.; Pokorny, R.; Byrdin, M.; Hoang, N.; Ritz, T.; Brettel, K.; Essen, L. O.; van der Horst, G. T.; Batschauer, A.; Ahmad, M., The cryptochromes: blue light photoreceptors in plants and animals. *Annu. Rev. Plant Biol.* **2011,** *62*, 335-64.

28. Wang, J.; Du, X.; Pan, W.; Wang, X.; Wu, W., Photoactivation of the cryptochrome/photolyase superfamily. *Journal of Photochemistry and Photobiology C: Photochemistry Reviews* **2015,** *22*, 84-102.

29. Schulz, G. E.; Schirmer, R. H.; Sachsenheimer, W.; Pai, E. F., The structure of the flavoenzyme glutathione reductase. *Nature* **1978,** *273* (5658), 120-4.

30. Meyerhof, O., Uber die Atmung der Froschmuskulatur. *Pfliig. Arch. ges. Physiol.* **1919,** *175*, 20-87.

31. Green, D. E., alpha-Glycerophosphate dehydrogenase. *Biochem. J.* **1936,** *30* (4), 629-44.

32. Tung, T.-C.; Anderson, L.; Lardy, H. A., Studies on the particulate α-glycerophosphate dehydrogenase of muscle. *Arch. Biochem. Biophys.* **1952,** *40* (1), 194-204.

33. Gunsalus, I. C.; Umbreit, W. W., The Oxidation of Glycerol by Streptococcus faecalis. *J. Bacteriol.* **1945,** *49* (4), 347-57.

34. Koditschek, L. K.; Umbreit, W. W., Alpha-glycerophosphate oxidase in Streptococcus faecium F 24. *J. Bacteriol.* **1969,** *98* (3), 1063-8.

35. Mracek, T.; Drahota, Z.; Houstek, J., The function and the role of the mitochondrial glycerol-3-phosphate dehydrogenase in mammalian tissues. *Biochim. Biophys. Acta* **2013,** *1827* (3), 401-10.

36. Chowdhury, S. K.; Gemin, A.; Singh, G., High activity of mitochondrial glycerophosphate dehydrogenase and glycerophosphate-dependent ROS production in prostate cancer cell lines. *Biochem. Biophys. Res. Commun.* **2005,** *333* (4), 1139-45.

37. Yeh, J. I.; Chinte, U.; Du, S., Structure of glycerol-3-phosphate dehydrogenase, an essential monotopic membrane enzyme involved in respiration and metabolism. *Proc Natl Acad Sci U S A* **2008,** *105* (9), 3280-5.

38. Colussi, T.; Parsonage, D.; Boles, W.; Matsuoka, T.; Mallett, T. C.; Karplus, P. A.; Claiborne, A., Structure of alpha-glycerophosphate oxidase from Streptococcus sp.: a template for the mitochondrial alpha-glycerophosphate dehydrogenase. *Biochemistry* **2008,** *47* (3), 965-77.

39. Mahdi, L. K.; Wang, H.; Van der Hoek, M. B.; Paton, J. C.; Ogunniyi, A. D., Identification of a novel pneumococcal vaccine antigen preferentially expressed during meningitis in mice. *J. Clin. Invest.* **2012,** *122* (6), 2208-20.

40. Mahdi, L. K.; Higgins, M. A.; Day, C. J.; Tiralongo, J.; Hartley-Tassell, L. E.; Jennings, M. P.; Gordon, D. L.; Paton, A. W.; Paton, J. C.; Ogunniyi, A. D., The Pneumococcal Alpha-Glycerophosphate Oxidase Enhances Nasopharyngeal Colonization through Binding to Host Glycoconjugates. *EBioMedicine* **2017,** *18*, 236-243.

41. Hames, C.; Halbedel, S.; Hoppert, M.; Frey, J.; Stulke, J., Glycerol metabolism is important for cytotoxicity of Mycoplasma pneumoniae. *J. Bacteriol.* **2009,** *191* (3), 747-53.

42. Grosshennig, S.; Schmidl, S. R.; Schmeisky, G.; Busse, J.; Stulke, J., Implication of glycerol and phospholipid transporters in Mycoplasma pneumoniae growth and virulence. *Infect. Immun.* **2013,** *81* (3), 896-904.

43. Pilo, P.; Vilei, E. M.; Peterhans, E.; Bonvin-Klotz, L.; Stoffel, M. H.; Dobbelaere, D.; Frey, J., A metabolic enzyme as a primary virulence factor of Mycoplasma mycoides subsp. mycoides small colony. *J. Bacteriol.* **2005,** *187* (19), 6824-31.

44. Schmidl, S. R.; Otto, A.; Lluch-Senar, M.; Pinol, J.; Busse, J.; Becher, D.; Stulke, J., A trigger enzyme in Mycoplasma pneumoniae: impact of the glycerophosphodiesterase GlpQ on virulence and gene expression. *PLoS Pathog* **2011,** *7* (9), e1002263.

45. Palioca, C. K. The Crystal Structure of the Potential Drug Target *Mycoplasma pneumoniae* Glycerol 3-Phosphate Oxidase. Oregon State University 2013.

46. Vishniac, W.; Ochoa, S., Photochemical reduction of pyridine nucleotides by spinach grana and coupled carbon dioxide fixation. *Nature* **1951,** *167* (4254), 768-9.

47. Tolmach, L. J., Effects of triphosphopyridine nucleotide upon oxygen evolution and carbon dioxide fixation by illuminated chloroplasts. *Nature* **1951,** *167* (4258), 946-8.

48. Arnon, D. I., Extracellular photosynthetic reactions. *Nature* **1951,** *167* (4260), 1008-10.

49. Arnon, D. I., The discovery of ferredoxin: the photosynthetic path. *Trends Biochem. Sci.* **1988,** *13* (1), 30-3.

50. Avron, M.; Jagendorf, A. T., A TPNH diaphorase from chloroplasts. *Arch. Biochem. Biophys.* **1956,** *65* (2), 475-90.

51. Keister, D. L.; San Pietro, A.; Stolzenbach, F. E., Pyridine nucleotide transhydrogenase from spinach. I. Purification and properties. *J. Biol. Chem.* **1960,** *235*, 2989-96.

52. Forti, G.; Bertole, M. L.; Parisi, B., On the function of cytochrome f in photosynthetic electron transport. *Biochem. Biophys. Res. Commun.* **1963,** *10*, 384-9.

53. Shin, M.; Tagawa, K.; Arnon, D. I., Crystallization of Ferredoxin-Tpn Reductase and Its Role in the Photosynthetic Apparatus of Chloroplasts. *Biochem Z* **1963,** *338*, 84-96.

54. Zanetti, G.; Forti, G., Studies on the triphosphopyridine nucleotide-cytochrome f reductase of chloroplasts. *J. Biol. Chem.* **1966,** *241* (2), 279-85.

55. Karplus, P. A.; Daniels, M. J.; Herriott, J. R., Atomic structure of ferredoxin-NADP+ reductase: prototype for a structurally novel flavoenzyme family. *Science* **1991,** *251* (4989), 60-6.

56. Aliverti, A.; Faber, R.; Finnerty, C. M.; Ferioli, C.; Pandini, V.; Negri, A.; Karplus, P. A.; Zanetti, G., Biochemical and crystallographic characterization of ferredoxin-NADP(+) reductase from nonphotosynthetic tissues. *Biochemistry* **2001,** *40* (48), 14501-8.

57. Aliverti, A. F., C.; Spinola, M.; Raimondi, D.; Zanetti, G.; Finnerty, C. M.; Faber, R.; Karplus, P. A. In *Structural and functional properties of corn root ferredoxin-NADP$^+$-reductase*, Flavins and Flavoproteins, Konstanz, Germany, Ghisla, S.; Kroneck, P.; Macheroux, P.; Sund, H., Eds. Agency for Scientific Publications: Konstanz, Germany, 1999; pp 265-268.

58. Piubelli, L.; Aliverti, A.; Arakaki, A. K.; Carrillo, N.; Ceccarelli, E. A.; Karplus, P. A.; Zanetti, G., Competition between C-terminal tyrosine and nicotinamide modulates pyridine nucleotide affinity and specificity in plant ferredoxin-NADP(+) reductase. *J. Biol. Chem.* **2000,** *275* (14), 10472-6.

59. Deng, Z.; Aliverti, A.; Zanetti, G.; Arakaki, A. K.; Ottado, J.; Orellano, E. G.; Calcaterra, N. B.; Ceccarelli, E. A.; Carrillo, N.; Karplus, P. A., A productive NADP+

binding mode of ferredoxin-NADP + reductase revealed by protein engineering and crystallographic studies. *Nat. Struct. Biol.* **1999,** *6* (9), 847-53.

60. Adak, S.; Sharma, M.; Meade, A. L.; Stuehr, D. J., A conserved flavin-shielding residue regulates NO synthase electron transfer and nicotinamide coenzyme specificity. *Proc Natl Acad Sci U S A* **2002,** *99* (21), 13516-21.

61. Yamada, M.; Tamada, T.; Takeda, K.; Matsumoto, F.; Ohno, H.; Kosugi, M.; Takaba, K.; Shoyama, Y.; Kimura, S.; Kuroki, R.; Miki, K., Elucidations of the catalytic cycle of NADH-cytochrome b5 reductase by X-ray crystallography: new insights into regulation of efficient electron transfer. *J. Mol. Biol.* **2013,** *425* (22), 4295-306.

62. Marohnic, C. C.; Bewley, M. C.; Barber, M. J., Engineering and characterization of a NADPH-utilizing cytochrome b5 reductase. *Biochemistry* **2003,** *42* (38), 11170-82.

63. Shiraishi, N.; Croy, C.; Kaur, J.; Campbell, W. H., Engineering of pyridine nucleotide specificity of nitrate reductase: mutagenesis of recombinant cytochrome b reductase fragment of Neurospora crassa NADPH:Nitrate reductase. *Arch. Biochem. Biophys.* **1998,** *358* (1), 104-15.

64. Hubbard, P. A.; Shen, A. L.; Paschke, R.; Kasper, C. B.; Kim, J. J., NADPH-cytochrome P450 oxidoreductase. Structural basis for hydride and electron transfer. *J. Biol. Chem.* **2001,** *276* (31), 29163-70.

65. Neeli, R.; Roitel, O.; Scrutton, N. S.; Munro, A. W., Switching pyridine nucleotide specificity in P450 BM3: mechanistic analysis of the W1046H and W1046A enzymes. *J. Biol. Chem.* **2005,** *280* (18), 17634-44.

66. Wang, M.; Roberts, D. L.; Paschke, R.; Shea, T. M.; Masters, B. S.; Kim, J. J., Three-dimensional structure of NADPH-cytochrome P450 reductase: prototype for FMN- and FAD-containing enzymes. *Proc Natl Acad Sci U S A* **1997,** *94* (16), 8411-6.

67. Kean, K. M.; Carpenter, R. A.; Pandini, V.; Zanetti, G.; Hall, A. R.; Faber, R.; Aliverti, A.; Karplus, P. A., High-resolution studies of hydride transfer in the ferredoxin:NADP(+) reductase superfamily. *FEBS J.* **2017,** *284* (19), 3302-3319.

68. Edson, N. L., The oxidation of lactic acid by Mycobacterium phlei. *Biochem. J.* **1947,** *41* (2), 145-51.

69. Yamamura, Y.; Kusunose, M.; Kusunose, E., Lactic oxidases of Mycobacterium tuberculosis avium. *Nature* **1952,** *170* (4318), 207-8.

70. Edson, N. L.; Cousins, F. B., Lactate-oxidizing systems in the mycobacteria. *Nature* **1953,** *171* (4355), 702.

71. Sutton, W. B., Isolation and properties of a lactic oxidative decarboxylase from Mycobacterium phlei. *J. Biol. Chem.* **1954,** *210* (1), 309-20.

72. Yorita, K.; Aki, K.; Ohkuma-Soyejima, T.; Kokubo, T.; Misaki, H.; Massey, V., Conversion of L-lactate oxidase to a long chain alpha-hydroxyacid oxidase by site-directed mutagenesis of alanine 95 to glycine. *J. Biol. Chem.* **1996,** *271* (45), 28300-5.

73. Muh, U.; Massey, V.; Williams, C. H., Jr., Lactate monooxygenase. I. Expression of the mycobacterial gene in Escherichia coli and site-directed mutagenesis of lysine 266. *J. Biol. Chem.* **1994,** *269* (11), 7982-8.

74. Muh, U.; Williams, C. H., Jr.; Massey, V., Lactate monooxygenase. II. Site-directed mutagenesis of the postulated active site base histidine 290. *J. Biol. Chem.* **1994,** *269* (11), 7989-93.

75. Muh, U.; Williams, C. H., Jr.; Massey, V., Lactate monooxygenase. III. Additive contributions of active site residues to catalytic efficiency and stabilization of an anionic transition state. *J. Biol. Chem.* **1994,** *269* (11), 7994-8000.

76. Massey, V.; Muller, F.; Feldberg, R.; Schuman, M.; Sullivan, P. A.; Howell, L. G.; Mayhew, S. G.; Matthews, R. G.; Foust, G. P., The reactivity of flavoproteins with sulfite. Possible relevance to the problem of oxygen reactivity. *J. Biol. Chem.* **1969,** *244* (15), 3999-4006.

77. Sun, W.; Williams, C. H., Jr.; Massey, V., Site-directed mutagenesis of glycine 99 to alanine in L-lactate monooxygenase from Mycobacterium smegmatis. *J. Biol. Chem.* **1996,** *271* (29), 17226-33.

78. Kean, K. M.; Porter, J. J.; Mehl, R. A.; Karplus, P. A., Structural insights into a thermostable variant of human carbonic anhydrase II. *Protein Sci.* **2018,** *27* (2), 573-577.

79. Driggers, C. M.; Kean, K. M.; Hirschberger, L. L.; Cooley, R. B.; Stipanuk, M. H.; Karplus, P. A., Structure-Based Insights into the Role of the Cys-Tyr Crosslink and Inhibitor Recognition by Mammalian Cysteine Dioxygenase. *J. Mol. Biol.* **2016,** *428* (20), 3999-4012.

80. Asamizu, S.; Abugreen, M.; Mahmud, T., Comparative metabolomic analysis of an alternative biosynthetic pathway to pseudosugars in Actinosynnema mirum DSM 43827. *ChemBioChem* **2013,** *14* (13), 1548-51.

81. Dell, K. A.; Frost, J. W., Identification and Removal of Impediments to Biocatalytic Synthesis of Aromatics from D-Glucose - Rate-Limiting Enzymes in the Common Pathway of Aromatic Amino-Acid Biosynthesis. *J. Am. Chem. Soc.* **1993,** *115* (24), 11581-11589.

82. Kim, C. G.; Kirschning, A.; Bergon, P.; Zhou, P.; Su, E.; Sauerbrei, B.; Ning, S.; Ahn, Y.; Breuer, M.; Leistner, E.; Floss, H. G., Biosynthesis of 3-amino-5-hydroxybenzoic acid, the precursor of mC(7)N units in ansamycin antibiotics. *J. Am. Chem. Soc.* **1996,** *118* (32), 7486-7491.

83. Mahmud, T., The C7N aminocyclitol family of natural products. **2002**.

84. Dong, H. J.; Mahmud, T.; Tornus, I.; Lee, S.; Floss, H. G., Biosynthesis of the validamycins: Identification of intermediates in the biosynthesis of validamycin A by Streptomyces hygroscopicus var. limoneus. *J. Am. Chem. Soc.* **2001,** *123* (12), 2733-2742.

85. Naganawa, H.; Hashizume, H.; Kubota, Y.; Sawa, R.; Takahashi, Y.; Arakawa, K.; Bowers, S. G.; Mahmud, T., Biosynthesis of the cyclitol moiety of pyralomicin 1a in Nonomuraea spiralis MI178-34F18. *J. Antibiot.* **2002,** *55* (6), 578-584.

86. Choi, W. S.; Wu, X. M.; Choeng, Y. H.; Mahmud, T.; Jeong, B. C.; Lee, S. H.; Chang, Y. K.; Kim, C. J.; Hong, S. K., Genetic organization of the putative salbostatin biosynthetic gene cluster including the 2-epi-5-epi-valiolone synthase gene in Streptomyces albus ATCC 21838. *Appl. Microbiol. Biotechnol.* **2008,** *80* (4), 637-645.

87. Rockser, Y.; Wehmeier, U. F., The gac-gene cluster for the production of acarbose from Streptomyces glaucescens GLA.O-Identification, isolation and characterization. *J. Biotechnol.* **2009,** *140* (1-2), 114-123.

88. Flatt, P. M.; Wu, X.; Perry, S.; Mahmud, T., Genetic insights into pyralomicin biosynthesis in Nonomuraea spiralis IMC A-0156. *J. Nat. Prod.* **2013,** *76* (5), 939-46.

89. Waller, R. F.; Slamovits, C. H.; Keeling, P. J., Lateral gene transfer of a multigene region from cyanobacteria to dinoflagellates resulting in a novel plastid-targeted fusion protein. *Mol. Biol. Evol.* **2006,** *23* (7), 1437-43.

90. Starcevic, A.; Akthar, S.; Dunlap, W. C.; Shick, J. M.; Hranueli, D.; Cullum, J.; Long, P. F., Enzymes of the shikimic acid pathway encoded in the genome of a basal metazoan, Nematostella vectensis, have microbial origins. *Proc Natl Acad Sci U S A* **2008,** *105* (7), 2533-7.

91. Singh, S. P.; Hader, D. P.; Sinha, R. P., Bioinformatics evidence for the transfer of mycosporine-like amino acid core (4-deoxygadusol) synthesizing gene from cyanobacteria to dinoflagellates and an attempt to mutate the same gene (YP_324358) in Anabaena variabilis PCC 7937. *Gene* **2012,** *500* (2), 155-63.

92. Singh, S. P.; Klisch, M.; Sinha, R. P.; Hader, D. P., Genome mining of mycosporine-like amino acid (MAA) synthesizing and non-synthesizing cyanobacteria: A bioinformatics study. *Genomics* **2010,** *95* (2), 120-8.

93. Waditee-Sirisattha, R.; Kageyama, H.; Sopun, W.; Tanaka, Y.; Takabe, T., Identification and upregulation of biosynthetic genes required for accumulation of Mycosporine-2-glycine under salt stress conditions in the halotolerant cyanobacterium Aphanothece halophytica. *Appl. Environ. Microbiol.* **2014,** *80* (5), 1763-9.

94. Pinto-Carbo, M.; Sieber, S.; Dessein, S.; Wicker, T.; Verstraete, B.; Gademann, K.; Eberl, L.; Carlier, A., Evidence of horizontal gene transfer between obligate leaf nodule symbionts. *ISME J* **2016,** *10* (9), 2092-105.

95. Nichols, C. E.; Ren, J.; Lamb, H. K.; Hawkins, A. R.; Stammers, D. K., Ligand-induced conformational changes and a mechanism for domain closure in Aspergillus nidulans dehydroquinate synthase. *J. Mol. Biol.* **2003,** *327* (1), 129-144.

96. Sieber, S.; Carlier, A.; Neuburger, M.; Grabenweger, G.; Eberl, L.; Gademann, K., Isolation and Total Synthesis of Kirkamide, an Aminocyclitol from an Obligate Leaf Nodule Symbiont. *Angew. Chem. Int. Ed. Engl.* **2015,** *54* (27), 7968-70.

97. Carreto, J. I.; Carignan, M. O., Mycosporine-like amino acids: relevant secondary metabolites. Chemical and ecological aspects. *Mar. Drugs* **2011,** *9* (3), 387-446.

98. Miyamoto, K. T.; Komatsu, M.; Ikeda, H., Discovery of gene cluster for mycosporine-like amino acid biosynthesis from Actinomycetales microorganisms and production of a novel mycosporine-like amino acid by heterologous expression. *Appl. Environ. Microbiol.* **2014,** *80* (16), 5028-36.

99. Leite, B.; Nicholson, R. L., Mycosporine-alanine: A self-inhibitor of germination from the conidial mucilage ofColletotrichum graminicola. *Experimental Mycology* **1992,** *16* (1), 76-86.

100. Libkind, D.; Moline, M.; Sommaruga, R.; Sampaio, J. P.; van Broock, M., Phylogenetic distribution of fungal mycosporines within the Pucciniomycotina (Basidiomycota). *Yeast* **2011,** *28* (8), 619-27.

101. Shukla, V.; Kumari, R.; Patel, D. K.; Upreti, D. K., Characterization of the diversity of mycosporine-like amino acids in lichens from high altitude region of Himalaya. *Amino Acids* **2016,** *48* (1), 129-36.

102. Kicklighter, C. E.; Kamio, M.; Nguyen, L.; Germann, M. W.; Derby, C. D., Mycosporine-like amino acids are multifunctional molecules in sea hares and their marine community. *Proc Natl Acad Sci U S A* **2011,** *108* (28), 11494-9.

103. Hartmann, A.; Holzinger, A.; Ganzera, M.; Karsten, U., Prasiolin, a new UV-sunscreen compound in the terrestrial green macroalga Prasiola calophylla (Carmichael ex Greville) Kutzing (Trebouxiophyceae, Chlorophyta). *Planta* **2016,** *243* (1), 161-9.

104. Shick, J. M.; Dunlap, W. C., Mycosporine-like amino acids and related Gadusols: biosynthesis, acumulation, and UV-protective functions in aquatic organisms. *Annu. Rev. Physiol.* **2002,** *64*, 223-62.

105. Plack, P. A.; Fraser, N. W.; Grant, P. T.; Middleton, C.; Mitchell, A. I.; Thomson, R. H., Gadusol, an enolic derivative of cyclohexane-1,3-dione present in the roes of cod and other marine fish. Isolation, properties and occurrence compared with ascorbic acid. *Biochem. J.* **1981,** *199* (3), 741-7.

106. Emerling, C. A., Archelosaurian Color Vision, Parietal Eye Loss, and the Crocodylian Nocturnal Bottleneck. *Mol. Biol. Evol.* **2016**.

107. Shick, J. M.; Dunlap, W. C.; Pearse, J. S.; Pearse, V. B., Mycosporine-like amino acid content in four species of sea anemones in the genus Anthopleura reflects phylogenetic but not environmental or symbiotic relationships. *Biol. Bull.* **2002,** *203* (3), 315-30.

108. Stochaj, W. R.; Dunlap, W. C.; Shick, J. M., 2 New Uv-Absorbing Mycosporine-Like Amino-Acids from the Sea-Anemone Anthopleura-Elegantissima and the Effects of Zooxanthellae and Spectral Irradiance on Chemical-Composition and Content. *Mar. Biol.* **1994,** *118* (1), 149-156.

109. Wehmeier, U. F.; Piepersberg, W., Biotechnology and molecular biology of the alpha-glucosidase inhibitor acarbose. *Appl. Microbiol. Biotechnol.* **2004,** *63* (6), 613-25.

110. Flatt, P. M.; Mahmud, T., Biosynthesis of aminocyclitol-aminoglycoside antibiotics and related compounds. *Nat. Prod. Rep.* **2007,** *24* (2), 358-92.

111. Waditee-Sirisattha, R.; Kageyama, H.; Fukaya, M.; Rai, V.; Takabe, T., Nitrate and amino acid availability affects glycine betaine and mycosporine-2-glycine in response to changes of salinity in a halotolerant cyanobacterium Aphanothece halophytica. *FEMS Microbiol. Lett.* **2015,** *362* (23), fnv198.

112. Singh, S. P.; Klisch, M.; Sinha, R. P.; Hader, D. P., Effects of abiotic stressors on synthesis of the mycosporine-like amino acid shinorine in the Cyanobacterium Anabaena variabilis PCC 7937. *Photochem. Photobiol.* **2008,** *84* (6), 1500-5.

113. Hu, C.; Voller, G.; Sussmuth, R.; Dittmann, E.; Kehr, J. C., Functional assessment of mycosporine-like amino acids in Microcystis aeruginosa strain PCC 7806. *Environ. Microbiol.* **2015,** *17* (5), 1548-59.

114. D'Agostino, P. M.; Javalkote, V. S.; Mazmouz, R.; Pickford, R.; Puranik, P. R.; Neilan, B. A., Comparative Profiling and Discovery of Novel Glycosylated Mycosporine-Like Amino Acids in Two Strains of the Cyanobacterium Scytonema cf. crispum. *Appl. Environ. Microbiol.* **2016,** *82* (19), 5951-9.

115. Llewellyn, C. A.; Airs, R. L., Distribution and abundance of MAAs in 33 species of microalgae across 13 classes. *Mar. Drugs* **2010,** *8* (4), 1273-91.

116. Yuan, Y. V.; Westcott, N. D.; Hu, C.; Kitts, D. D., Mycosporine-like amino acid composition of the edible red alga, Palmaria palmata (dulse) harvested from the west and east coasts of Grand Manan Island, New Brunswick. *Food Chem.* **2009,** *112* (2), 321-328.

117. Carefoot, T. H.; Karentz, D.; Pennings, S. C.; Young, C. L., Distribution of mycosporine-like amino acids in the sea hare Aplysia dactylomela: effect of diet on amounts and types sequestered over time in tissues and spawn. *Comp Biochem Physiol C Toxicol Pharmacol* **2000,** *126* (1), 91-104.

118. Rosic, N. N., Phylogenetic analysis of genes involved in mycosporine-like amino acid biosynthesis in symbiotic dinoflagellates. *Appl. Microbiol. Biotechnol.* **2012,** *94* (1), 29-37.

119. Shick, J. M.; Romaine-Lioud, S.; Ferrier-Pages, C.; Gattuso, J. P., Ultraviolet-B radiation stimulates shikimate pathway-dependent accumulation of mycosporine-like amino acids in the coral Stylophora pistillata despite decreases in its population of symbiotic dinoflagellates. *Limnol. Oceanogr.* **1999,** *44* (7), 1667-1682.

120. Shinzato, C.; Shoguchi, E.; Kawashima, T.; Hamada, M.; Hisata, K.; Tanaka, M.; Fujie, M.; Fujiwara, M.; Koyanagi, R.; Ikuta, T.; Fujiyama, A.; Miller, D. J.; Satoh, N., Using the Acropora digitifera genome to understand coral responses to environmental change. *Nature* **2011,** *476* (7360), 320-3.

121. Paredes-Molina, F. J.; Cubillos, V. M.; Montory, J. A.; Andrade-Villagran, P. A., Are embryonic developing modes determinant in the acquisition and levels of photoprotective compounds in slipper limpets of the Crepipatella genus? *J Photochem Photobiol B* **2016,** *162*, 511-8.

122. Bok, M. J.; Porter, M. L.; Place, A. R.; Cronin, T. W., Biological sunscreens tune polychromatic ultraviolet vision in mantis shrimp. *Curr. Biol.* **2014,** *24* (14), 1636-42.

123. Przeslawski, R.; Benkendorff, K.; Davis, A. R., A quantitative survey of mycosporine-like amino acids (MAAS) in intertidal egg masses from temperate rocky shores. *J. Chem. Ecol.* **2005,** *31* (10), 2417-38.

124. Braun, C.; Reef, R.; Siebeck, U. E., Ultraviolet absorbing compounds provide a rapid response mechanism for UV protection in some reef fish. *J Photochem Photobiol B* **2016,** *160*, 400-7.

125. Eckes, M.; Dove, S.; Siebeck, U. E.; Grutter, A. S., Fish mucus versus parasitic gnathiid isopods as sources of energy and sunscreens for a cleaner fish. *Coral Reefs* **2015,** *34* (3), 823-833.

126. Eckes, M. J.; Siebeck, U. E.; Dove, S.; Grutter, A. S., Ultraviolet sunscreens in reef fish mucus. *Mar. Ecol. Prog. Ser.* **2008,** *353*, 203-211.

127. Rosic, N. N.; Dove, S., Mycosporine-Like Amino Acids from Coral Dinoflagellates. *Appl. Environ. Microbiol.* **2011,** *77* (24), 8478-8486.

128. Stamatakis, K.; Vayenos, D.; Kotakis, C.; Gast, R. J.; Papageorgiou, G. C., The extraordinary longevity of kleptoplasts derived from the Ross Sea haptophyte Phaeocystis antarctica within dinoflagellate host cells relates to the diminished role of the oxygen-evolving Photosystem II and to supplementary light harvesting by mycosporine-like amino acid/s. *Biochim. Biophys. Acta* **2017,** *1858* (2), 189-195.

129. Mahmud, T., The C7N aminocyclitol family of natural products. *Natural Product Reports* **2003,** *20* (1), 137-66.

130. Stratmann, A.; Mahmud, T.; Lee, S.; Distler, J.; Floss, H. G.; Piepersberg, W., The AcbC protein from Actinoplanes species is a C-7-cyclitol synthase related to 3-dehydroquinate synthases and is involved in the biosynthesis of the alpha-glucosidase inhibitor acarbose. *J Biol Chem* **1999,** *274* (16), 10889-10896.

131. Carpenter, E. P.; Hawkins, A. R.; Frost, J. W.; Brown, K. A., Structure of dehydroquinate synthase reveals an active site capable of multistep catalysis. *Nature* **1998,** *394* (6690), 299-302.

132. Nango, E.; Kumasaka, T.; Hirayama, T.; Tanaka, N.; Eguchi, T., Structure of 2-deoxy-scyllo-inososе synthase, a key enzyme in the biosynthesis of 2-deoxystreptamine-containing aminoglycoside antibiotics, in complex with a mechanism-based inhibitor and NAD(+). *Proteins-Structure Function and Bioinformatics* **2008,** *70* (2), 517-527.

133. Asamizu, S.; Xie, P. F.; Brumsted, C. J.; Platt, P. M.; Mahmud, T., Evolutionary Divergence of Sedoheptulose 7-Phosphate Cyclases Leads to Several Distinct Cyclic Products. *J Am Chem Soc* **2012,** *134* (29), 12219-12229.

134. Nichols, C. E.; Hawkins, A. R.; Stammers, D. K., Structure of the 'open' form of Aspergillus nidulans 3-dehydroquinate synthase at 1.7 A resolution from crystals grown following enzyme turnover. *Acta Crystallographica, Section D: Biological Crystallography* **2004,** *60* (Pt 5), 971-3.

135. Nichols, C. E.; Ren, J.; Leslie, K.; Dhaliwal, B.; Lockyer, M.; Charles, I.; Hawkins, A. R.; Stammers, D. K., Comparison of ligand-induced conformational changes and domain closure mechanisms, between prokaryotic and eukaryotic dehydroquinate synthases. *Journal of Molecular Biology* **2004,** *343* (3), 533-46.

136. Sugahara, M.; Nodake, Y.; Sugahara, M.; Kunishima, N., Crystal structure of dehydroquinate synthase from Thermus thermophilus HB8 showing functional importance of the dimeric state. *Proteins: Struct., Funct., Bioinf.* **2005,** *58* (1), 249-252.

137. Liu, J. S.; Cheng, W. C.; Wang, H. J.; Chen, Y. C.; Wang, W. C., Structure-based inhibitor discovery of Helicobacter pylori dehydroquinate synthase. *Biochem Bioph Res Co* **2008,** *373* (1), 1-7.

138. Ruzheinikov, S. N.; Burke, J.; Sedelnikova, S.; Baker, P. J.; Taylor, R.; Bullough, P. A.; Muir, N. M.; Gore, M. G.; Rice, D. W., Glycerol dehydrogenase. structure, specificity, and mechanism of a family III polyol dehydrogenase. *Structure* **2001,** *9* (9), 789-802.

139. Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C., SCOP: a structural classification of proteins database for the investigation of sequences and structures. *Journal of Molecular Biology* **1995,** *247* (4), 536-40.

140. Sillitoe, I.; Cuff, A. L.; Dessailly, B. H.; Dawson, N. L.; Furnham, N.; Lee, D.; Lees, J. G.; Lewis, T. E.; Studer, R. A.; Rentzsch, R.; Yeats, C.; Thornton, J. M.; Orengo, C. A., New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. *Nucleic Acids Research* **2013,** *41* (Database issue), D490-8.

141. Park, A.; Lamb, H. K.; Nichols, C.; Moore, J. D.; Brown, K. A.; Cooper, A.; Charles, I. G.; Stammers, D. K.; Hawkins, A. R., Biophysical and kinetic analysis of wild-type and site-directed mutants of the isolated and native dehydroquinate synthase domain of the AROM protein. *Protein Science* **2004,** *13* (8), 2108-19.

142. Bartlett, G. J.; Borkakoti, N.; Thornton, J. M., Catalysing new reactions during evolution: economy of residues and mechanism. *Journal of Molecular Biology* **2003,** *331* (4), 829-60.

143. Winter, J. M.; Tang, Y., Synthetic biological approaches to natural product biosynthesis. *Curr Opin Biotechnol* **2012,** *23* (5), 736-43.

144. Moore, J. D.; Skinner, M. A.; Swatman, D. R.; Hawkins, A. R.; Brown, K. A., Reactivation of 3-dehydroquinate synthase by lanthanide cations. *J Am Chem Soc* **1998,** *120* (28), 7105-7106.

145. Kudo, F.; Hosomi, Y.; Tamegai, H.; Kakinuma, K., Purification and characterization of 2-deoxy-scyllo-inosose synthase derived from Bacillus circulans. A crucial carbocyclization enzyme in the biosynthesis of 2-deoxystreptarnine-containing aminoglycoside antibiotics. *J Antibiot* **1999,** *52* (2), 81-88.

146. Battye, T. G.; Kontogiannis, L.; Johnson, O.; Powell, H. R.; Leslie, A. G., iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica, Section D: Biological Crystallography* **2011,** *67* (Pt 4), 271-81.

147. Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S., Overview of the CCP4 suite and current developments. *Acta Crystallographica, Section D: Biological Crystallography* **2011,** *67* (Pt 4), 235-42.

148. Evans, P., Scaling and assessment of data quality. *Acta Crystallographica, Section D: Biological Crystallography* **2006,** *62* (Pt 1), 72-82.

149. Terwilliger, T. C.; Dimaio, F.; Read, R. J.; Baker, D.; Bunkoczi, G.; Adams, P. D.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Echols, N., phenix.mr_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta. *Journal of Structural and Functional Genomics* **2012,** *13* (2), 81-90.

150. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallographica, Section D: Biological Crystallography* **2010,** *66* (Pt 4), 486-501.

151. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica, Section D: Biological Crystallography* **2010,** *66* (Pt 2), 213-21.

152. Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P, Sharff, A., Smart, O.S., Vonrhein, C., Womack, T.O. *BUSTER version 2.11.2*, Global Phasing Ltd: Cambridge, United Kingdom, 2011.

153. Kabsch, W.; Sander, C., Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983,** *22* (12), 2577-637.

154. Joosten, R. P.; te Beek, T. A.; Krieger, E.; Hekkelman, M. L.; Hooft, R. W.; Schneider, R.; Sander, C.; Vriend, G., A series of PDB related databases for everyday needs. *Nucleic Acids Research* **2011,** *39* (Database issue), D411-9.

155. Holm, L.; Rosenstrom, P., Dali server: conservation mapping in 3D. *Nucleic Acids Research* **2010,** *38* (Web Server issue), W545-9.

156. Karplus, P. A.; Diederichs, K., Linking crystallographic model and data quality. *Science* **2012,** *336* (6084), 1030-3.

157. Kershaw, N. J.; Murphy, J. M.; Liau, N. P.; Varghese, L. N.; Laktyushin, A.; Whitlock, E. L.; Lucet, I. S.; Nicola, N. A.; Babon, J. J., SOCS3 binds specific receptor-JAK complexes to control cytokine signaling by direct kinase inhibition. *Nat. Struct. Mol. Biol.* **2013,** *20* (4), 469-76.

158. Perkins, A.; Gretes, M. C.; Nelson, K. J.; Poole, L. B.; Karplus, P. A., Mapping the active site helix-to-strand conversion of CxxxxC peroxiredoxin Q enzymes. *Biochemistry-Us* **2012,** *51* (38), 7638-50.

159. Krissinel, E.; Henrick, K., Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* **2007,** *372* (3), 774-797.

160. Brown, K. A.; Carpenter, E. P.; Watson, K. A.; Coggins, J. R.; Hawkins, A. R.; Koch, M. H. J.; Svergun, D. I., Twists and turns: a tale of two shikimate-pathway enzymes. *Biochem Soc T* **2003,** *31*, 543-547.

161. Stokes-Rees, I.; Sliz, P., Protein structure determination by exhaustive search of Protein Data Bank derived databases. *P Natl Acad Sci USA* **2010,** *107* (50), 21476-21481.

162. Mittelstadt, G.; Negron, L.; Schofield, L. R.; Marsh, K.; Parker, E. J., Biochemical and structural characterisation of dehydroquinate synthase from the New Zealand kiwifruit Actinidia chinensis. *Arch Biochem Biophys* **2013,** *537* (2), 185-91.

163. Charmantray, F.; Helaine, V.; Legeret, B.; Hecquet, L., Preparative scale enzymatic synthesis of D-sedoheptulose-7-phosphate from beta-hydroxypyruvate and D-ribose-5-phosphate. *J Mol Catal B-Enzym* **2009,** *57* (1-4), 6-9.

164. Nango, E.; Eguchi, T.; Kakinuma, K., Active site mapping of 2-deoxy-scyllo-inosose synthase, the key starter enzyme for the biosynthesis of 2-deoxystreptamine. Mechanism-based inhibition and identification of lysine-141 as the entrapped nucleophile. *J Org Chem* **2004,** *69* (3), 593-600.

165. Bender, S. L.; Widlanski, T.; Knowles, J. R., Dehydroquinate synthase: the use of substrate analogues to probe the early steps of the catalyzed reaction. *Biochemistry-Us* **1989,** *28* (19), 7560-72.

166. Montchamp, J. G. Mechanism-based carbocyclic inhibitors of dehydroquinate synthase. Ph.D. Dissertation, Purdue University, West Lafayette, IN, 1992.

167. Mahmud, T., Progress in aminocyclitol biosynthesis. *Current opinion in chemical biology* **2009,** *13* (2), 161-70.

168. Nango, E.; Kumasaka, T.; Hirayama, T.; Tanaka, N.; Eguchi, T., Structure of 2-deoxy-scyllo-inosose synthase, a key enzyme in the biosynthesis of 2-deoxystreptamine-containing aminoglycoside antibiotics, in complex with a mechanism-based inhibitor and NAD+. *Proteins* **2008,** *70* (2), 517-27.

169. Floss, H. G., Natural products derived from unusual variants of the shikimate pathway. *Natural product reports* **1997,** *14* (5), 433-52.

170. Yamauchi, N.; Kakinuma, K., Biochemical studies on 2-deoxy-scyllo-inosose, an early intermediate in the biosynthesis of 2-deoxystreptamine. IV. A clue to the similarity of 2-deoxy-scyllo-inosose synthase to dehydroquinate synthase. *The Journal of antibiotics* **1993,** *46* (12), 1916-8.

171. Kim, C. G.; Kirschning, A.; Bergon, P.; Zhou, P.; Su, E.; Sauerbrei, B.; Ning, S.; Ahn, Y.; Breuer, M.; Leistner, E.; Floss, H. G., Biosynthesis of 3-amino-5-hydroxybenzoic acid, the precursor of mC7N units in ansamycin antibiotics. *Journal of the American Chemical Society* **1996,** *118*, 7486-7491.

172. Rascher, A.; Hu, Z.; Buchanan, G. O.; Reid, R.; Hutchinson, C. R., Insights into the biosynthesis of the benzoquinone ansamycins geldanamycin and herbimycin, obtained by gene sequencing and disruption. *Applied and environmental microbiology* **2005,** *71* (8), 4862-71.

173. Mao, Y.; Varoglu, M.; Sherman, D. H., Molecular characterization and analysis of the biosynthetic gene cluster for the antitumor antibiotic mitomycin C from Streptomyces lavendulae NRRL 2564. *Chemistry & biology* **1999,** *6* (4), 251-63.

174. Floss, H. G.; Yu, T. W.; Arakawa, K., The biosynthesis of 3-amino-5-hydroxybenzoic acid (AHBA), the precursor of mC7N units in ansamycin and mitomycin antibiotics: a review. *The Journal of antibiotics* **2011,** *64* (1), 35-44.

175. Mahmud, T.; Tornus, I.; Egelkrout, E.; Wolf, E.; Uy, C.; Floss, H. G.; Lee, S., Biosynthetic studies on the alpha-glucosidase inhibitor acarbose in Actinoplanes sp.:

2-epi-5-epi-valiolone is the direct precursor of the valienamine moiety. *Journal of the American Chemical Society* **1999,** *121* (30), 6973-6983.

176. Dong, H.; Mahmud, T.; Tornus, I.; Lee, S.; Floss, H. G., Biosynthesis of the validamycins: identification of intermediates in the biosynthesis of validamycin A by Streptomyces hygroscopicus var. limoneus. *Journal of the American Chemical Society* **2001,** *123* (12), 2733-42.

177. Osborn, A. R.; Almabruk, K. H.; Holzwarth, G.; Asamizu, S.; LaDu, J.; Kean, K. M.; Karplus, P. A.; Tanguay, R. L.; Bakalinsky, A. T.; Mahmud, T., De novo synthesis of a sunscreen compound in vertebrates. *eLife* **2015,** *4*, e05919.

178. Wu, X.; Flatt, P. M.; Xu, H.; Mahmud, T., Biosynthetic gene cluster of cetoniacytone A, an unusual aminocyclitol from the endosymbiotic Bacterium Actinomyces sp. Lu 9419. *Chembiochem* **2009,** *10* (2), 304-14.

179. Choi, W. S.; Wu, X.; Choeng, Y. H.; Mahmud, T.; Jeong, B. C.; Lee, S. H.; Chang, Y. K.; Kim, C. J.; Hong, S. K., Genetic organization of the putative salbostatin biosynthetic gene cluster including the 2-epi-5-epi-valiolone synthase gene in Streptomyces albus ATCC 21838. *Appl Microbiol Biotechnol* **2008,** *80* (4), 637-45.

180. Hirayama, T.; Kudo, F.; Huang, Z.; Eguchi, T., Role of glutamate 243 in the active site of 2-deoxy-scyllo-inosose synthase from Bacillus circulans. *Bioorganic & medicinal chemistry* **2007,** *15* (1), 418-23.

181. Huang, S.; Chen, Z.; Yan, X.; Yu, T.; Huang, G.; Yan, Q.; Pontarotti, P. A.; Zhao, H.; Li, J.; Yang, P.; Wang, R.; Li, R.; Tao, X.; Deng, T.; Wang, Y.; Li, G.; Zhang, Q.; Zhou, S.; You, L.; Yuan, S.; Fu, Y.; Wu, F.; Dong, M.; Chen, S.; Xu, A., Decelerated genome evolution in modern vertebrates revealed by analysis of multiple lancelet genomes. *Nat Commun* **2014,** *5*, 5896.

182. Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 1), 12-21.

183. Rockser, Y.; Wehmeier, U. F., The gac-gene cluster for the production of acarbose from Streptomyces glaucescens GLA.O: identification, isolation and characterization. *Journal of biotechnology* **2009,** *140* (1-2), 114-23.

184. Asamizu, S.; Yang, J.; Almabruk, K. H.; Mahmud, T., Pseudoglycosyltransferase catalyzes nonglycosidic C-N coupling in validamycin a biosynthesis. *Journal of the American Chemical Society* **2011,** *133* (31), 12124-35.

185. Abuelizz, H. A.; Mahmud, T., Distinct Substrate Specificity and Catalytic Activity of the Pseudoglycosyltransferase VldE. *Chemistry & biology* **2015,** *22* (6), 724-33.

186. Cavalier, M. C.; Yim, Y. S.; Asamizu, S.; Neau, D.; Almabruk, K. H.; Mahmud, T.; Lee, Y. H., Mechanistic insights into validoxylamine A 7'-phosphate synthesis by VldE using the structure of the entire product complex. *PLoS One* **2012,** *7* (9), e44934.

187. Yu, Y.; Bai, L.; Minagawa, K.; Jian, X.; Li, L.; Li, J.; Chen, S.; Cao, E.; Mahmud, T.; Floss, H. G.; Zhou, X.; Deng, Z., Gene cluster responsible for

validamycin biosynthesis in Streptomyces hygroscopicus subsp. jinggangensis 5008. *Applied and environmental microbiology* **2005,** *71* (9), 5066-76.

188. Cheng, K.; Rong, X.; Huang, Y., Widespread interspecies homologous recombination reveals reticulate evolution within the genus Streptomyces. *Mol Phylogenet Evol* **2016,** *102*, 246-254.

189. Mahmud, T.; Flatt, P. M.; Wu, X., Biosynthesis of unusual aminocyclitol-containing natural products. *Journal of natural products* **2007,** *70* (8), 1384-91.

190. Mahmud, T., The C7N aminocyclitol family of natural products. *Natural product reports* **2003,** *20* (1), 137-66.

191. Edgar, R. C., MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **2004,** *32* (5), 1792-7.

192. Price, M. N.; Dehal, P. S.; Arkin, A. P., FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **2010,** *5* (3), e9490.

193. Han, M. V.; Zmasek, C. M., phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics* **2009,** *10*, 356.

194. Battye, T. G.; Kontogiannis, L.; Johnson, O.; Powell, H. R.; Leslie, A. G., iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* **2011,** *67* (Pt 4), 271-81.

195. Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S., Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **2011,** *67* (Pt 4), 235-42.

196. Evans, P. R.; Murshudov, G. N., How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **2013,** *69* (Pt 7), 1204-14.

197. Terwilliger, T. C.; Dimaio, F.; Read, R. J.; Baker, D.; Bunkoczi, G.; Adams, P. D.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Echols, N., phenix.mr_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta. *J Struct Funct Genomics* **2012,** *13* (2), 81-90.

198. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 4), 486-501.

199. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 2), 213-21.

200. Joosten, R. P.; te Beek, T. A.; Krieger, E.; Hekkelman, M. L.; Hooft, R. W.; Schneider, R.; Sander, C.; Vriend, G., A series of PDB related databases for everyday needs. *Nucleic acids research* **2011,** *39* (Database issue), D411-9.

201. Holm, L.; Rosenstrom, P., Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **2010,** *38* (Web Server issue), W545-9.

202. Krissinel, E.; Henrick, K., Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **2004,** *60* (Pt 12 Pt 1), 2256-68.

203. Waites, K. B.; Talkington, D. F., Mycoplasma pneumoniae and its role as a human pathogen. *Clin. Microbiol. Rev.* **2004,** *17* (4), 697-728, table of contents.

204. Prince, O. A.; Krunkosky, T. M.; Krause, D. C., In vitro spatial and temporal analysis of Mycoplasma pneumoniae colonization of human airway epithelium. *Infect. Immun.* **2014,** *82* (2), 579-86.

205. Dym, O.; Eisenberg, D., Sequence-structure analysis of FAD-containing proteins. *Protein Sci.* **2001,** *10* (9), 1712-28.

206. Parsonage, D.; Luba, J.; Mallett, T. C.; Claiborne, A., The soluble alpha-glycerophosphate oxidase from Enterococcus casseliflavus. Sequence homology with the membrane-associated dehydrogenase and kinetic analysis of the recombinant enzyme. *J. Biol. Chem.* **1998,** *273* (37), 23812-22.

207. Charrier, V.; Luba, J.; Parsonage, D.; Claiborne, A., Limited proteolysis as a structural probe of the soluble alpha-glycerophosphate oxidase from Streptococcus sp. *Biochemistry* **2000,** *39* (17), 5035-44.

208. Mortl, M.; Diederichs, K.; Welte, W.; Molla, G.; Motteran, L.; Andriolo, G.; Pilone, M. S.; Pollegioni, L., Structure-function correlation in glycine oxidase from Bacillus subtilis. *J. Biol. Chem.* **2004,** *279* (28), 29718-27.

209. Davis, J. J.; Xia, F.; Overbeek, R. A.; Olsen, G. J., Genomes of the class Erysipelotrichia clarify the firmicute origin of the class Mollicutes. *Int. J. Syst. Evol. Microbiol.* **2013,** *63* (Pt 7), 2727-41.

210. Guindon, S.; Dufayard, J. F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O., New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* **2010,** *59* (3), 307-321.

211. Schryvers, A.; Weiner, J. H., The anaerobic sn-glycerol-3-phosphate dehydrogenase of Escherichia coli. Purification and characterization. *J. Biol. Chem.* **1981,** *256* (19), 9959-65.

212. Settembre, E. C.; Dorrestein, P. C.; Park, J. H.; Augustine, A. M.; Begley, T. P.; Ealick, S. E., Structural and mechanistic studies on ThiO, a glycine oxidase essential for thiamin biosynthesis in Bacillus subtilis. *Biochemistry* **2003,** *42* (10), 2971-81.

213. Umhau, S.; Pollegioni, L.; Molla, G.; Diederichs, K.; Welte, W.; Pilone, M. S.; Ghisla, S., The x-ray structure of D-amino acid oxidase at very high resolution identifies the chemical mechanism of flavin-dependent substrate dehydrogenation. *Proc Natl Acad Sci U S A* **2000,** *97* (23), 12463-8.

214. Maenpuen, S.; Watthaisong, P.; Supon, P.; Sucharitakul, J.; Parsonage, D.; Karplus, P. A.; Claiborne, A.; Chaiyen, P., Kinetic mechanism of L-alpha-glycerophosphate oxidase from Mycoplasma pneumoniae. *FEBS J.* **2015,** *282* (16), 3043-59.

215. Evans, P., Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **2006,** *62* (Pt 1), 72-82.

216. Trapani, S.; Navaza, J., AMoRe: classical and modern. *Acta Crystallogr D Biol Crystallogr* **2008,** *64* (Pt 1), 11-6.

217. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J. Appl. Crystallogr.* **2007,** *40* (Pt 4), 658-674.

218. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **1997,** *53* (Pt 3), 240-55.

219. Aliverti, A.; Pandini, V.; Pennati, A.; de Rosa, M.; Zanetti, G., Structural and functional diversity of ferredoxin-NADP(+) reductases. *Arch. Biochem. Biophys.* **2008,** *474* (2), 283-91.

220. Paladini, D. H.; Musumeci, M. A.; Carrillo, N.; Ceccarelli, E. A., Induced fit and equilibrium dynamics for high catalytic efficiency in ferredoxin-NADP(H) reductases. *Biochemistry* **2009,** *48* (24), 5760-8.

221. Bortolotti, A.; Perez-Dorado, I.; Goni, G.; Medina, M.; Hermoso, J. A.; Carrillo, N.; Cortez, N., Coenzyme binding and hydride transfer in Rhodobacter capsulatus ferredoxin/flavodoxin NADP(H) oxidoreductase. *Biochim. Biophys. Acta* **2009,** *1794* (2), 199-210.

222. Tejero, J.; Perez-Dorado, I.; Maya, C.; Martinez-Julvez, M.; Sanz-Aparicio, J.; Gomez-Moreno, C.; Hermoso, J. A.; Medina, M., C-terminal tyrosine of ferredoxin-NADP+ reductase in hydride transfer processes with NAD(P)+/H. *Biochemistry* **2005,** *44* (41), 13477-90.

223. Bruns, C. M.; Karplus, P. A., Refined crystal structure of spinach ferredoxin reductase at 1.7 A resolution: oxidized, reduced and 2'-phospho-5'-AMP bound states. *J. Mol. Biol.* **1995,** *247* (1), 125-45.

224. Dorowski, A.; Hofmann, A.; Steegborn, C.; Boicu, M.; Huber, R., Crystal structure of paprika ferredoxin-NADP+ reductase. Implications for the electron transfer pathway. *J. Biol. Chem.* **2001,** *276* (12), 9253-63.

225. Kurisu, G.; Kusunoki, M.; Katoh, E.; Yamazaki, T.; Teshima, K.; Onda, Y.; Kimata-Ariga, Y.; Hase, T., Structure of the electron transfer complex between ferredoxin and ferredoxin-NADP(+) reductase. *Nat. Struct. Biol.* **2001,** *8* (2), 117-21.

226. Correll, C. C.; Ludwig, M. L.; Bruns, C. M.; Karplus, P. A., Structural prototypes for an extended family of flavoprotein reductases: comparison of phthalate dioxygenase reductase with ferredoxin reductase and ferredoxin. *Protein Sci.* **1993,** *2* (12), 2112-33.

227. Karplus, P. A.; Bruns, C. M., Structure-function relations for ferredoxin reductase. *J. Bioenerg. Biomembr.* **1994,** *26* (1), 89-99.

228. Bedard, K.; Krause, K. H., The NOX family of ROS-generating NADPH oxidases: physiology and pathophysiology. *Physiol. Rev.* **2007,** *87* (1), 245-313.

229. Sumimoto, H., Structure, regulation and evolution of Nox-family NADPH oxidases that produce reactive oxygen species. *FEBS J.* **2008,** *275* (13), 3249-77.

230. Panday, A.; Sahoo, M. K.; Osorio, D.; Batra, S., NADPH oxidases: an overview from structure to innate immunity-associated pathologies. *Cell Mol Immunol* **2015,** *12* (1), 5-23.

231. Brewer, T. F.; Garcia, F. J.; Onak, C. S.; Carroll, K. S.; Chang, C. J., Chemical approaches to discovery and study of sources and targets of hydrogen peroxide redox signaling through NADPH oxidase proteins. *Annu. Rev. Biochem.* **2015,** *84*, 765-90.

232. Garcin, E. D.; Bruns, C. M.; Lloyd, S. J.; Hosfield, D. J.; Tiso, M.; Gachhui, R.; Stuehr, D. J.; Tainer, J. A.; Getzoff, E. D., Structural basis for isozyme-specific regulation of electron transfer in nitric-oxide synthase. *J. Biol. Chem.* **2004,** *279* (36), 37918-27.

233. Correll, C. C.; Batie, C. J.; Ballou, D. P.; Ludwig, M. L., Phthalate dioxygenase reductase: a modular structure for electron transfer from pyridine nucleotides to [2Fe-2S]. *Science* **1992,** *258* (5088), 1604-10.

234. Dohr, O.; Paine, M. J.; Friedberg, T.; Roberts, G. C.; Wolf, C. R., Engineering of a functional human NADH-dependent cytochrome P450 system. *Proc Natl Acad Sci U S A* **2001,** *98* (1), 81-6.

235. Meints, C. E.; Gustafsson, F. S.; Scrutton, N. S.; Wolthers, K. R., Tryptophan 697 modulates hydride and interflavin electron transfer in human methionine synthase reductase. *Biochemistry* **2011,** *50* (51), 11131-42.

236. Lans, I.; Peregrina, J. R.; Medina, M.; Garcia-Viloca, M.; Gonzalez-Lafont, A.; Lluch, J. M., Mechanism of the hydride transfer between Anabaena Tyr303Ser FNR(rd)/FNR(ox) and NADP+/H. A combined pre-steady-state kinetic/ensemble-averaged transition-state theory with multidimensional tunneling study. *J. Phys. Chem. B* **2010,** *114* (9), 3368-79.

237. Peregrina, J. R.; Lans, I.; Medina, M., The transient catalytically competent coenzyme allocation into the active site of Anabaena ferredoxin NADP+ -reductase. *Eur. Biophys. J.* **2012,** *41* (1), 117-28.

238. Mulo, P.; Medina, M., Interaction and electron transfer between ferredoxin-NADP+ oxidoreductase and its partners: structural, functional, and physiological implications. *Photosynth. Res.* **2017**.

239. Magnani, F.; Nenci, S.; Millana Fananas, E.; Ceccon, M.; Romero, E.; Fraaije, M. W.; Mattevi, A., Crystal structures and atomic model of NADPH oxidase. *Proc Natl Acad Sci U S A* **2017,** *114* (26), 6764-6769.

240. Nogues, I.; Tejero, J.; Hurley, J. K.; Paladini, D.; Frago, S.; Tollin, G.; Mayhew, S. G.; Gomez-Moreno, C.; Ceccarelli, E. A.; Carrillo, N.; Medina, M., Role of the C-terminal tyrosine of ferredoxin-nicotinamide adenine dinucleotide phosphate reductase in the electron transfer processes with its protein partners ferredoxin and flavodoxin. *Biochemistry* **2004,** *43* (20), 6127-37.

241. Batie, C. J.; Kamin, H., Association of ferredoxin-NADP+ reductase with NADP(H) specificity and oxidation-reduction properties. *J. Biol. Chem.* **1986,** *261* (24), 11214-23.

242. Batie, C. J.; Kamin, H., Electron transfer by ferredoxin:NADP+ reductase. Rapid-reaction evidence for participation of a ternary complex. *J. Biol. Chem.* **1984,** *259* (19), 11976-85.

243. Dunford, A. J.; Marshall, K. R.; Munro, A. W.; Scrutton, N. S., Thermodynamic and kinetic analysis of the isolated FAD domain of rat neuronal nitric oxide synthase altered in the region of the FAD shielding residue Phe1395. *Eur. J. Biochem.* **2004,** *271* (12), 2548-60.

244. Orville, A. M.; Buono, R.; Cowan, M.; Heroux, A.; Shea-McCarthy, G.; Schneider, D. K.; Skinner, J. M.; Skinner, M. J.; Stoner-Ma, D.; Sweet, R. M., Correlated single-crystal electronic absorption spectroscopy and X-ray crystallography at NSLS beamline X26-C. *J Synchrotron Radiat* **2011,** *18* (Pt 3), 358-66.

245. Tronrud, D. E.; Berkholz, D. S.; Karplus, P. A., Using a conformation-dependent stereochemical library improves crystallographic refinement of proteins. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 7), 834-42.

246. Young, L.; Post, C. B., Catalysis by entropic guidance from enzymes. *Biochemistry* **1996,** *35* (48), 15129-33.

247. Prasad, G. S.; Kresge, N.; Muhlberg, A. B.; Shaw, A.; Jung, Y. S.; Burgess, B. K.; Stout, C. D., The crystal structure of NADPH : ferredoxin reductase from Azotobacter vinelandii. *Protein Sci.* **1998,** *7* (12), 2541-2549.

248. Lu, G.; Lindqvist, Y.; Schneider, G.; Dwivedi, U.; Campbell, W., Structural studies on corn nitrate reductase: refined structure of the cytochrome b reductase fragment at 2.5 A, its ADP complex and an active-site mutant and modeling of the cytochrome b domain. *J. Mol. Biol.* **1995,** *248* (5), 931-48.

249. Bewley, M. C.; Marohnic, C. C.; Barber, M. J., The structure and biochemistry of NADH-dependent cytochrome b5 reductase are now consistent. *Biochemistry* **2001,** *40* (45), 13574-82.

250. Bruice, T. C.; Lightstone, F. C., Ground state and transition state contributions to the rates of intramolecular and enzymatic reactions. *Acc. Chem. Res.* **1999,** *32* (2), 127-136.

251. Bruice, T. C.; Pandit, U. K., Intramolecular Models Depicting the Kinetic Importance of "Fit" in Enzymatic Catalysis. *Proc Natl Acad Sci U S A* **1960,** *46* (4), 402-4.

252. Rajagopalan, P. T.; Benkovic, S. J., Preorganization and protein dynamics in enzyme catalysis. *Chem Rec* **2002,** *2* (1), 24-36.

253. Almarsson, O.; Bruice, T. C., Evaluation of the Factors Influencing Reactivity and Stereospecificity in Nad(P)H Dependent Dehydrogenase Enzymes. *J. Am. Chem. Soc.* **1993,** *115* (6), 2125-2138.

254. Berkholz, D. S.; Faber, H. R.; Savvides, S. N.; Karplus, P. A., Catalytic cycle of human glutathione reductase near 1 A resolution. *J. Mol. Biol.* **2008,** *382* (2), 371-84.

255. Quaytman, S. L.; Schwartz, S. D., Reaction coordinate of an enzymatic reaction revealed by transition path sampling. *Proc Natl Acad Sci U S A* **2007,** *104* (30), 12253-8.

256. Zhang, J.; Klinman, J. P., Enzymatic methyl transfer: role of an active site residue in generating active site compaction that correlates with catalytic efficiency. *J. Am. Chem. Soc.* **2011,** *133* (43), 17134-7.

257. Hay, S.; Pudney, C. R.; McGrory, T. A.; Pang, J.; Sutcliffe, M. J.; Scrutton, N. S., Barrier compression enhances an enzymatic hydrogen-transfer reaction. *Angew. Chem. Int. Ed. Engl.* **2009,** *48* (8), 1452-4.

258. Tejero, I.; Garcia-Viloca, M.; Gonzalez-Lafont, A.; Lluch, J. M.; York, D. M., Enzyme dynamics and tunneling enhanced by compression in the hydrogen abstraction catalyzed by soybean lipoxygenase-1. *J. Phys. Chem. B* **2006,** *110* (48), 24708-19.

259. Medina, M.; Luquita, A.; Tejero, J.; Hermoso, J.; Mayoral, T.; Sanz-Aparicio, J.; Grever, K.; Gomez-Moreno, C., Probing the determinants of coenzyme specificity in ferredoxin-NADP+ reductase by site-directed mutagenesis. *J. Biol. Chem.* **2001,** *276* (15), 11902-12.

260. Elmore, C. L.; Porter, T. D., Modification of the nucleotide cofactor-binding site of cytochrome P-450 reductase to enhance turnover with NADH in Vivo. *J. Biol. Chem.* **2002,** *277* (50), 48960-4.

261. Zhen, L.; Yu, L.; Dinauer, M. C., Probing the role of the carboxyl terminus of the gp91phox subunit of neutrophil flavocytochrome b558 using site-directed mutagenesis. *J. Biol. Chem.* **1998,** *273* (11), 6575-81.

262. Leslie, A. G. W., Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* **1992,** *26*.

263. Ghisla, S.; Massey, V., L-Lactate Oxidase. In *Chemistry and biochemistry of flavoenzymes*, Müller, F., Ed. CRC Press: 1991.

264. Maeda-Yorita, K.; Aki, K.; Sagai, H.; Misaki, H.; Massey, V., L-lactate oxidase and L-lactate monooxygenase: mechanistic variations on a common structural theme. *Biochimie* **1995,** *77* (7-8), 631-42.

265. Li, S. J.; Umena, Y.; Yorita, K.; Matsuoka, T.; Kita, A.; Fukui, K.; Morimoto, Y., Crystallographic study on the interaction of L-lactate oxidase with pyruvate at 1.9 Angstrom resolution. *Biochem. Biophys. Res. Commun.* **2007,** *358* (4), 1002-7.

266. Lindqvist, Y., Refined structure of spinach glycolate oxidase at 2 Å resolution. *J. Mol. Biol.* **1989,** *209* (1), 151-166.

267. Murray, M. S.; Holmes, R. P.; Lowther, W. T., Active site and loop 4 movements within human glycolate oxidase: implications for substrate specificity and drug design. *Biochemistry* **2008,** *47* (8), 2439-49.

268. Cunane, L. M.; Barton, J. D.; Chen, Z. W.; Welsh, F. E.; Chapman, S. K.; Reid, G. A.; Mathews, F. S., Crystallographic study of the recombinant flavin-binding domain of Baker's yeast flavocytochrome b(2): comparison with the intact wild-type enzyme. *Biochemistry* **2002,** *41* (13), 4264-72.

269. Sukumar, N.; Dewanti, A. R.; Mitra, B.; Mathews, F. S., High resolution structures of an oxidized and reduced flavoprotein. The water switch in a soluble form of (S)-mandelate dehydrogenase. *J. Biol. Chem.* **2004,** *279* (5), 3749-57.

270. Murphy, C. J.; Shepherd, M. G.; Sullivan, P. A., Chemical modification of L-lactate 2-monooxygenase with fluorodinitrobenzene: evidence for two essential histidine residues. *Biochemistry* **1983,** *22* (7), 1665-9.

271. Soon, C. Y.; Shepherd, M. G.; Sullivan, P. A., Inactivation and modification of lactate oxidase with fluorodinitrobenzene. *Biochem. J.* **1978,** *173* (1), 255-62.

272. Soon, C. Y.; Shepherd, M. G.; Sullivan, P. A., Modification of lactate oxidase with diethyl pyrocarbonate. Evidence for an active-site histidine residue. *Biochem. J.* **1977,** *165* (2), 385-93.

273. Peters, R. G.; Jones, W. C.; Cromartie, T. H., Inactivation of L-lactate monooxygenase with 2,3-butanedione and phenylglyoxal. *Biochemistry* **1981,** *20* (9), 2564-71.

274. Stoisser, T.; Brunsteiner, M.; Wilson, D. K.; Nidetzky, B., Conformational flexibility related to enzyme activity: evidence for a dynamic active-site gatekeeper function of Tyr(215) in Aerococcus viridans lactate oxidase. *Sci Rep* **2016,** *6*, 27892.

275. Stoisser, T.; Klimacek, M.; Wilson, D. K.; Nidetzky, B., Speeding up the product release: a second-sphere contribution from Tyr191 to the reactivity of L-lactate oxidase revealed in crystallographic and kinetic studies of site-directed variants. *FEBS J.* **2015,** *282* (21), 4130-40.

276. Sun, W.; Williams, C. H., Jr.; Massey, V., The role of glycine 99 in L-lactate monooxygenase from Mycobacterium smegmatis. *J. Biol. Chem.* **1997,** *272* (43), 27065-76.

277. Deng, Z. Structural Studies of the Active Site and $NADP^+$ Binding of Plant-Type Ferreodxin:$NADP^+$ Reductase. Cornell University, 1999.

278. Renko, M.; Taler-Vercic, A.; Mihelic, M.; Zerovnik, E.; Turk, D., Partial rotational lattice order-disorder in stefin B crystals. *Acta Crystallogr D Biol Crystallogr* **2014,** *70* (Pt 4), 1015-25.

279. Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B., 3rd; Snoeyink, J.; Richardson, J. S.; Richardson, D. C., MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **2007,** *35* (Web Server issue), W375-83.

280. Sullivan, P. A.; Soon, C. Y.; Schreurs, W. J.; Cutfield, J. F.; Shepherd, M. G., The structure of L-lactate oxidase from Mycobacterium smegmatis. *Biochem. J.* **1977,** *165* (2), 375-83.

281. Giegel, D. A.; Williams, C. H., Jr.; Massey, V., L-lactate 2-monooxygenase from Mycobacterium smegmatis. Cloning, nucleotide sequence, and primary structure homology within an enzyme family. *J. Biol. Chem.* **1990,** *265* (12), 6626-32.

282. Reid, G. A.; White, S.; Black, M. T.; Lederer, F.; Mathews, F. S.; Chapman, S. K., Probing the active site of flavocytochrome b2 by site-directed mutagenesis. *Eur. J. Biochem.* **1988,** *178* (2), 329-33.

283. Cunane, L. M.; Barton, J. D.; Chen, Z. W.; Le, K. H.; Amar, D.; Lederer, F.; Mathews, F. S., Crystal structure analysis of recombinant rat kidney long chain hydroxy acid oxidase. *Biochemistry* **2005,** *44* (5), 1521-31.

284. Stoisser, T.; Rainer, D.; Leitgeb, S.; Wilson, D. K.; Nidetzky, B., The Ala95-to-Gly substitution in Aerococcus viridans l-lactate oxidase revisited - structural consequences at the catalytic site and effect on reactivity with O2 and other electron acceptors. *FEBS J.* **2015,** *282* (3), 562-78.

285. Furuichi, M.; Suzuki, N.; Dhakshnamoorhty, B.; Minagawa, H.; Yamagishi, R.; Watanabe, Y.; Goto, Y.; Kaneko, H.; Yoshida, Y.; Yagi, H.; Waga, I.; Kumar, P. K.; Mizuno, H., X-ray structures of Aerococcus viridans lactate oxidase and its complex with D-lactate at pH 4.5 show an alpha-hydroxyacid oxidation mechanism. *J. Mol. Biol.* **2008,** *378* (2), 436-46.

286. Miller, S.; Lesk, A. M.; Janin, J.; Chothia, C., The accessible surface area and stability of oligomeric proteins. *Nature* **1987,** *328* (6133), 834-6.

287. Chothia, C., Principles that determine the structure of proteins. *Annu. Rev. Biochem.* **1984,** *53*, 537-72.

288. Lesk, A. M.; Chothia, C., Solvent accessibility, protein surfaces, and protein folding. *Biophys. J.* **1980,** *32* (1), 35-47.

289. Zhou, H.; Zhou, Y., Quantifying the effect of burial of amino acid residues on protein stability. *Proteins* **2004,** *54* (2), 315-22.

290. Creamer, T. P.; Srinivasan, R.; Rose, G. D., Modeling unfolded states of proteins and peptides. II. Backbone solvent accessibility. *Biochemistry* **1997,** *36* (10), 2832-5.

291. Creamer, T. P.; Srinivasan, R.; Rose, G. D., Modeling unfolded states of peptides and proteins. *Biochemistry* **1995,** *34* (50), 16245-50.

292. Saff, E. B.; Kuijlaars, A. B. J., Distributing many points on a sphere. *Math Intell* **1997,** *19* (1), 5-11.

293. Lee, B.; Richards, F. M., The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **1971,** *55* (3), 379-400.

294. Barka, E. A.; Vatsa, P.; Sanchez, L.; Gaveau-Vaillant, N.; Jacquard, C.; Meier-Kolthoff, J. P.; Klenk, H. P.; Clement, C.; Ouhdouch, Y.; van Wezel, G. P., Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol. Mol. Biol. Rev.* **2016,** *80* (1), 1-43.

295. Lee, M. R.; Sheng, W. H.; Hung, C. C.; Yu, C. J.; Lee, L. N.; Hsueh, P. R., Mycobacterium abscessus Complex Infections in Humans. *Emerg Infect Dis* **2015,** *21* (9), 1638-46.

296. Safaei, S.; Fatahi-Bafghi, M.; Pouresmaeil, O., Role of Tsukamurella species in human infections: first literature review. *New Microbes New Infect* **2018,** *22*, 6-12.

297. Polz, M. F.; Alm, E. J.; Hanage, W. P., Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet.* **2013,** *29* (3), 170-5.

298. Fitzpatrick, D. A., Horizontal gene transfer in fungi. *FEMS Microbiol. Lett.* **2012,** *329* (1), 1-8.

299. Xiao, G.; Ying, S. H.; Zheng, P.; Wang, Z. L.; Zhang, S.; Xie, X. Q.; Shang, Y.; St Leger, R. J.; Zhao, G. P.; Wang, C.; Feng, M. G., Genomic perspectives on the evolution of fungal entomopathogenicity in Beauveria bassiana. *Sci Rep* **2012,** *2*, 483.

300. Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997,** *276*, 307-26.

301. Kabsch, W., Xds. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 2), 125-32.

302. Collaborative Computational Project, N., The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **1994,** *50* (Pt 5), 760-3.

303. Soding, J.; Biegert, A.; Lupas, A. N., The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **2005,** *33* (Web Server issue), W244-8.

304. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **2004,** *60* (Pt 12 Pt 1), 2126-32.

305. Terwilliger, T. C.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Moriarty, N. W.; Zwart, P. H.; Hung, L. W.; Read, R. J.; Adams, P. D., Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* **2008,** *64* (Pt 1), 61-9.

306. Huang, Y.; Niu, B. F.; Gao, Y.; Fu, L. M.; Li, W. Z., CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **2010,** *26* (5), 680-682.

307. Edgar, R. C., MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004,** *32* (5), 1792-1797.

308. Estimate of Unfolded state Accessible Surface Area. http://folding.chemistry.msstate.edu/utils/unfolded.html (accessed Apr 22, 2018).

309. Lans, I.; Medina, M.; Rosta, E.; Hummer, G.; Garcia-Viloca, M.; Lluch, J. M.; Gonzalez-Lafont, A., Theoretical study of the mechanism of the hydride transfer between ferredoxin-NADP+ reductase and NADP+: the role of Tyr303. *J. Am. Chem. Soc.* **2012,** *134* (50), 20544-53.

310. Montelione, G. T., The Protein Structure Initiative: achievements and visions for the future. *F1000 Biol Rep* **2012,** *4*, 7.

311. Jiang, T.; Gao, C.; Ma, C.; Xu, P., Microbial lactate utilization: enzymes, pathogenesis, and regulation. *Trends Microbiol.* **2014,** *22* (10), 589-99.

312. Baloni, P.; Padiadpu, J.; Singh, A.; Gupta, K. R.; Chandra, N., Identifying feasible metabolic routes in Mycobacterium smegmatis and possible alterations under diverse nutrient conditions. *BMC Microbiol.* **2014,** *14*, 276.

313. Roxas, B. A.; Li, Q., Acid stress response of a mycobacterial proteome: insight from a gene ontology analysis. *Int J Clin Exp Med* **2009,** *2* (4), 309-28.

314. Belmouden, A.; Lederer, F., The role of a beta barrel loop 4 extension in modulating the physical and functional properties of long-chain 2-hydroxy-acid oxidase isozymes. *Eur. J. Biochem.* **1996,** *238* (3), 790-8.