CATALYTIC PROMISCUITY AND THE EVOLUTIONARY MECHANISM OF NSAR REACTION SPECIFICITY IN THE NSAR/OSBS SUBFAMILY

A Dissertation

by

DAT PHUOC TRUONG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Margaret E. Glasner
Committee Members,	Frank Raushel
	James Sacchettini
	Paul Straight
Head of Department,	Josh Wand

May 2020

Major Subject: Biochemistry

ABSTRACT

Catalytic promiscuity is the coincidental ability for an enzyme to catalyze non-biological reactions in the same active site as the native biological reaction. Several lines of evidence show that catalytic promiscuity plays an important role in the evolution of new enzyme functions. Studying catalytic promiscuity can help identify structural features that predispose an enzyme to evolve new functions. This dissertation studies the mechanistic basis of catalytic promiscuity in the evolution of *N*-succinylamino acid racemase (NSAR) activity in the NSAR/*o*-succinylbenzoate synthase (NSAR/OSBS) subfamily. The NSAR/OSBS subfamily is a branch of the OSBS family, which belongs to the mechanistically diverse enolase superfamily. We identified a conserved, second-shell residue R266 in all the NSAR/OSBS enzymes while the homologous position is usually hydrophobic in other nonpromiscuous OSBS subfamilies. We found that the R266 residue is an important NSAR reaction specificity determinant because the R266Q mutation in *Amycolatopsis* NSAR/OSBS enzyme profoundly reduces NSAR activity while having a moderate effect on OSBS activity. Mechanistic investigation by hydrogen-deuterium exchange showed that R266 modulates the reactivity of the catalytic base K263, but not the other catalytic base K163. The crystal structure of *Amycolatopsis* NSAR/OSBS R266Q mutant shows that K263 adopts a different conformation, so it is not positioned correctly to act as a general acid/base for catalysis. Further kinetic and mechanistic studies of the R266Q mutation in other NSAR/OSBS members showed that R266 is also important for NSAR activity. However, the specific

phenotypic effects of the R266Q mutation are masked by the sequence and structural contexts in which the mutation occurs (that is, epistatic constraints), making it harder to fully understand the roles of R266 in the NSAR/OSBS subfamily. Up to date, R266 is the first residue that was identified in the NSAR/OSBS enzymes to be pre-adaptive and vital for NSAR activity, and such identification is essential to understand the evolution of NSAR activity in the NSAR/OSBS subfamily. This finding is significant because it can help us understand the reaction specificity determinants in other enzymes. For example, mutating the homologous lysine to a glutamine in members of the dipeptide epimerase family also has a deleterious effect on the activity, further supporting the universal importance of a positively charged amino acid at position 266 for the epimerase/racemase activity in the enolase superfamily. This dissertation provides the mechanistic basis of determining epimerase/racemase reaction specificity in enzymes and can be ultimately used as a predictive tool for functional annotations, the development of protein engineering, and the improvement of protein design methods.

DEDICATION

I dedicate this dissertation to my family. To my mother and father who have always been loving me, believing in me, and supporting me. Mom and Dad, thank you for your patience and all your sacrifice so that I can be where I am now. To my sister who always listens to me as a sister and as a friend. Sis, thank you for taking care of Mom and Dad while I am away all this time. To my uncle, aunt and cousins who all helped and supported me in the early years in the United States. To my wife who has been loving me and always by my side, especially on the toughest days. Ngoc, thank you for your unconditional love and your enormous support and encouragement. Thank you for always being there and accepting my flaws and all. You always inspire me to do better every day.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Glasner, who is a great boss and an awesome mentor. Thank you for all your non-stop guidance, patience, motivation, support, and willingness to share your knowledge throughout my time in the lab and thank you for inspiring me to do and to think better as a scientist. I also would like to thank my committee members, Dr. Raushel, Dr. Sacchettini, and Dr. Straight, for their insightful recommendations, constructive criticisms, and helpful advice throughout the years.

I also would like to thank my labmates in the Glasner Lab who I spent a lot of time "sciencing" with every single day. Especially Mr. Denis Odokonyero, who is not only the best lab manager but also one of my greatest friends. Thank you for your useful insights on some of my experiments, your helpful advice on a lot of scientific and non-scientific things, and for being there for me when I had tough days in lab. I honestly learned a lot from you, as a colleague and as a friend. I also would like to thank Ms. Mariana Lopez, our former lab technician, who personally helped me countlessly during my first year in the Glasner Lab, from showing me how to use all the instruments in the lab, to helping me with some of the experiments. Thank you for making my first year in the lab memorable.

I also would like to thank all my close friends, including those who have been with me on this challenging journey since day 1 and those who became my greatest friends during my time at Texas A&M. Friends, thank you for all the great memories we

have made together and for the lifetime friendships. Without y'all, I wouldn't make this far.

And lastly, I would like to thank the Biochemistry Graduate Association (BGA), our former academic advisor, Mr. Rafael Almanzar, and the faculty and staff of the Department of Biochemistry and Biophysics for making my time at Texas A&M University a great experience.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Margaret E. Glasner (as the committee chair), Dr. James Sacchettini and Dr. Paul Straight of the Department of Biochemistry and Biophysics and Dr. Frank Raushel of the Department of Chemistry.

The NMR instrument used to measure hydrogen-deuterium exchange (k_{ex}) between the enzymes and the substrates shown in Chapter II, III, IV, and V was operated by Dr. Jamison Huddleston from Dr. Frank Raushel Lab of the Department of Chemistry. The NSAR substrates (*N*-succinyl-D/L-phenylglycine) used in this research study were synthesized by Mingzhao Zhu and Kenneth Hull from Dr. Daniel Romo Lab of the Department of Chemistry and Biochemistry, Baylor University.

The crystallographic data collection and structure determination of *Amycolatopsis* NSAR/OSBS R266Q mutant shown in Chapter II were done by Mr. Simon Rousseau from Dr. James Sacchettini Lab of the Department of Biochemistry and Biophysics. Crystallographic results shown in Chapter II are derived from work performed by the staff at beamline 19-ID at Argonne National Laboratory, Structural Biology Center (SBC) at the Advanced Photon Source. SBC-CAT is operated by UChicago Argonne, LLC, for the U.S. Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357.

The phylogenetic tree of the MLE subgroup of the enolase superfamily shown in Chapter III was constructed by Dr. Andrew McMillan (a former postdoc) and Mr. Charles Morey (a former undergraduate student) from Dr. Margaret Glasner Lab of the Department of Biochemistry and Biophysics. The R266Q mutation construction and kinetic data collection of OSBS and NSAR reactions of *Lysinibacillus varians* NSAR/OSBS R266Q shown in Chapter III were done by Ms. Susan Fults from Dr. Margaret Glasner Lab of the Department of Biochemistry and Biophysics. The kinetic data collection of OSBS reaction of *Roseiflexus castenholzii* NSAR/OSBS R266Q shown in Chapter III were done by Mr. Cristian Davila from Dr. Margaret Glasner Lab of the Department of Biochemistry and Biophysics. The L-Ala-D-Glu substrate for the dipeptide epimerase assay shown in Chapter III was synthesized by Dr. Dakota Brock (a former graduate student) from Dr. Jean-Philippe Pellois Lab of the Department of Biochemistry and Biophysics.

The F19A/R20E *Amycolatopsis* NSAR/OSBS double mutant shown in Chapter IV was constructed by Mr. Benjamin Morse from Dr. Margaret Glasner Lab of the Department of Biochemistry and Biophysics. The pH-rate profile for the OSBS activity of *Amycolatopsis* NSAR/OSBS WT shown in Chapter IV was done by Mr. Benjamin Machala (a former undergraduate student) from Dr. Margaret Glasner Lab of the Department of Biochemistry and Biophysics.

All other work conducted for this dissertation was completed by the student independently.

Funding Sources

This work was funded by National Institutes of Health Award R01-GM124409, National Science Foundation CAREER Award 1253975, and the Welch Foundation Grant No. A-1991-20190330 (Principal Investigator, Dr. Margaret E. Glasner).

NOMENCLATURE

OSBS	*o*-Succinylbenzoate Synthase
NSAR	*N*-Succinylamino Acid Racemase
SHCHC	2-Succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate
OSB	*o*-Succinylbenzoate
D/L-NSPG	D/L-*N*-Succinylphenylglycine
AmyNSAR/OSBS	*Amycolatopsis* sp. T-1-60 NSAR/OSBS
EfNSAR/OSBS	*Enterococcus faecalis* NSAR/OSBS
LiNSAR/OSBS	*Listeria innocua* NSAR/OSBS
LvNSAR/OSBS	*Lysinibacillus varians* NSAR/OSBS
RcNSAR/OSBS	*Roseiflexus castenholzii* NSAR/OSBS
ExiOSBS	*Exiguobacterium* sp. AT1b OSBS
EcOSBS	*Escherichia coli* OSBS
TfuscaOSBS	*Thermobifida fusca* OSBS
BsubOSBS	*Bacillus subtilis* OSBS
DE (or AEE)	Dipeptide Epimerase (or L-Ala-L/D-Glu Epimerase)
EcDE	*Escherichia coli* Dipeptide Epimerase
BsDE	*Bacillus subtilis* Dipeptide Epimerase
MLE	Muconate Lactonizing Enzyme
MR	Mandelate Racemase
GlucD	D-Glucarate Dehydratase

ManD	D-Mannonate Dehydratase
GalrD2	Galactarate Dehydratase 2
MAL	3-Methylaspartate Ammonia Lyase
HPBE	4R-Hydroxyproline Betaine 2-Epimerase
HAD	Haloacid Dehalogenase
PMH	Phosphonate Monoester Hydrolase
AP	Alkaline Phosphatase
PAS	Arylsulfatase
PON1	Serum Paraoxonase 1
PTE	Phosphotriesterase
CHMO	Cyclohexanone Monooxygenase
DSF	Differential Scanning Fluorimetry
PDB	Protein Data Bank
k_{ex}	Isotopic exchange rate
η_{rel}	relative viscosity
T_m	melting temperature

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Understanding how an enzyme has evolved a new function can provide detailed insights into proteins' structure-function relationships and mechanisms. Many lines of evidence show that catalytic promiscuity plays an important role in the evolution of new enzyme functions [1-8]. Catalytic promiscuity is the coincidental ability to catalyze non-biological reactions in the same active site as the native biological reaction. This ability is due to many contributing factors including the common intermediate, the similarity in the chemical transformation steps, and/or the similarity in substrate structures between the reactions [7]. However, the molecular mechanism of how catalytic promiscuity can allow enzymes to evolve new functions remains unclear. This dissertation will elucidate some of the mechanistic aspects of how catalytic promiscuity contributes to the evolution of new enzyme functions. In this introduction chapter, I will address the current knowledge about the evolution of our model system, the *N*-succinylamino acid racemase/*o*-succinylbenzoate synthase (NSAR/OSBS) subfamily of the divergent OSBS family. Then I will address our current knowledge about how catalytic promiscuity can contribute to the evolution of new enzymatic functions.

Enzyme evolution and enzyme superfamily

Studying enzyme evolution is enabled by the identifications of homologous enzymes that have evolved to carry out different functions, including those enzymes from functionally diverse superfamilies. Functionally diverse superfamilies are evolutionarily related sets of enzymes that are diverse in sequence and overall reaction. They have structural variations beyond the common overall fold, but share a conserved set of catalytic residues used for a common mechanistic chemical step [9]. They represent a significant proportion of the enzyme universe, making up more than one third (272 out of 704) of all structurally characterized enzyme superfamilies [10]. However, the number of functionally diverse superfamilies may be underestimated because the studies are restricted to only structurally characterized enzymes and there are many superfamilies that have not been characterized. Thus, those functionally diverse superfamilies are an excellent model to study enzyme evolution. Examples of well-characterized functionally diverse enzyme superfamilies include the haloacid dehalogenase, enolase, cytosolic glutathione transferase, metallo-beta-lactamase, and amidohydrolase superfamilies [11].

The enolase superfamily

The enolase superfamily has been extensively studied to understand the evolution of enzymes since it is one of the most mechanistically diverse superfamilies [12].

Members of this superfamily, including mandelate racemase (MR), muconate lactonizing enzyme (MLE), dipeptide epimerase (DE), and *o*-succinyl benzoate synthase (OSBS), catalyze divergent chemical reactions (including racemization, elimination, and cycloisomerization). They span all six EC classes while sharing as little as 15% sequence identity [13]. Conservation of an aspect of catalysis and the overall fold are characteristics of the members of the mechanistically diverse superfamilies [9]. Despite low sequence identity, these enzymes share a common chemical step, in which a base abstracts the alpha proton of the carboxylate substrate to generate an enolate anion intermediate. This enolate intermediate is stabilized by a divalent metal ion in the active site. Subsequently, the enolate can react to yield different products in different enzymes [14]. Furthermore, these enzymes have similarities in their overall structure and a conserved set of catalytic residues. The proteins from this superfamily are composed of a catalytic C-terminal $((\beta/\alpha)_7\beta)$-barrel and a capping domain consisting of components from both N- and C-termini (Figure I.1A) [14]. The catalytic residues reside inside the barrel cavity and at the end of the β-strands. While the catalytic residue at the end of the β2 strand is typically a lysine in a KxK motif, the other catalytic residue on the β6 strand is more divergent (Table I.1). The conserved Mg^{2+} metal ion binding residues are typically glutamate or aspartate, located at the ends of the β3, β4, and β5 strands. However, there are some exceptions, which are described in more detail below. Overall, the evolution of enolase superfamily is constrained by the overall protein structure and the conserved catalytic residues despite the divergent chemical transformations [12].

The members of the enolase superfamily are divided into 7 different subgroups based on either their sequence homology clustering or the identity and positions of the catalytic residues including the ligands of the metal ion (Table I.1) [14-16].

1. The isofunctional enolase subgroup, which converts 2-phosphoglycerate to phosphoenolpyruvate in glycolysis, has a conserved catalytic lysine on the end of the β6 strand and a glutamate on the end of the β2 strand [14].
2. The members of the MR subgroup contain a His-Asp dyad, with the catalytic histidine on the β7 strand and the aspartate on the β6 strand. These enzymes catalyze the MR reaction and several mono- and diacid sugar hydratase reactions (including L-rhamnonate, L-fuconate, D-gluconate, D-galactonate, D-tartrate, and L-talarate/D-galactarate dehydratases) [14].
3. The members of the D-glucarate dehydratase (GlucD) subgroup contain a His-Asp dyad similar to the MR subgroup; however, they also contain an asparagine rather than a glutamate as a ligand for the Mg^{2+} ion at the end of the β5 strand [17].
4. The members of the D-mannonate dehydratase (ManD) subgroup contain an Tyr-Arg dyad with the arginine on the end of the β2 strand and a catalytic histidine on the β3 strand [18]. There is an arginine on the β6 strand and a histidine on the end of the β7 strand; however, they are not involved in catalysis but interact with the substrate instead [18].
5. The members of the galactarate dehydratase 2 (GalrD2) subgroup, which lack the active site loop (the 20s loop), contain a catalytic Arg-Tyr dyad on the end of the

β2 strand and a catalytic tyrosine in the capping domain [19]. They also contain a secondary Mg^{2+} ion and have a histidine on the end of the β5 strand as a ligand for the Mg^{2+} ion stabilizing the enolate intermediate [19].

6. The members of the 3-methylaspartate ammonia lyase (MAL) subgroup, including MAL that catalyzes an anti-deamination reaction, contain a catalytic lysine on the end of the β6 strand and a histidine on the end of the β2 strand [20, 21].
7. And finally, the members of the MLE subgroup contain both catalytic lysines on both β2 and β6 strands; in some cases, lysine on the β6 strand is substituted with an arginine (Figure I.1B). The MLE subgroup, which is the most functionally diverse subgroup, includes the MLE reaction, the OSBS reaction, the NSAR reaction, and the DE reaction (Figure I.2B) [14].

Although the enzymes of the enolase superfamily share the overall fold and the conserved catalytic residues, their sequences and functions are highly divergent, making it difficult to identify which features are responsible for reaction specificity in those enzymes. Thus, detailed structural and biochemical studies of different enzyme families will help us identify important features that determine reaction specificity in these enzyme families.

Table I.1. Conserved residues in the different subgroups of the enolase superfamily*

Number	Subgroup	β2	β3	β4	β5	β6	β7
(1)	enolase	E	D	E	D	K	
(2)	MR	Kx(K,R,H,Y)	D	E	E	D	H
(3)	GlucD	KxK	D	E	N	D	H
(4)	ManD	R	D	E	E	R	H
(5)	GalrD2	RxY	D	E	H		
(6)	MAL	H	D	E	D	K	
(7)	MLE	KxK	D	E	D	K	

*Table is modified from reference [18].

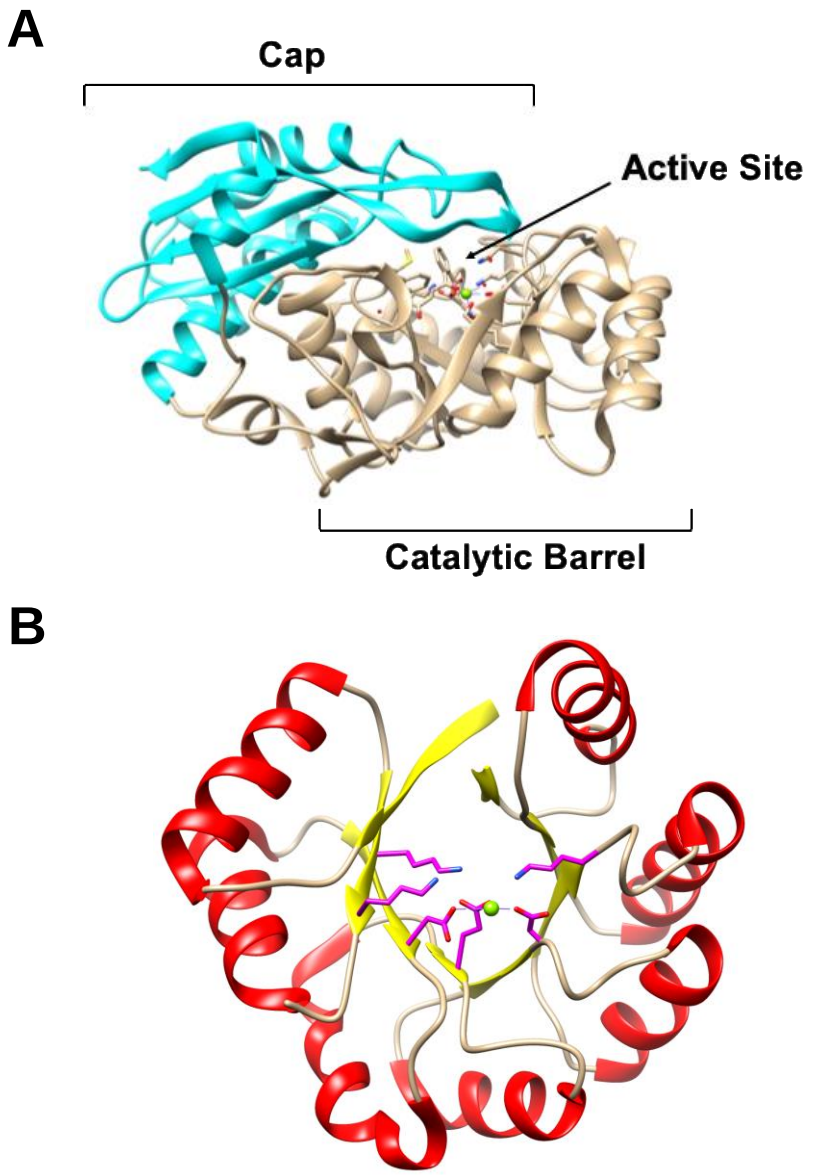


Figure I.1. (A) The overall structure of *Amycolatopsis* NSAR/OSBS (PDB ID 1SJB, [22]), representing the enzymes from the enolase superfamily, consisting the capping and the catalytic barrel domains. (B) The catalytic C-terminal $((\beta/\alpha)_7\beta)$-barrel domain of the *Amycolatopsis* NSAR/OSBS (PDB ID 1SJB, [22]), representing the enzymes of the MLE subgroup. The conserved catalytic residues are shown in magenta. The Mg^{2+} ion is shown in green.

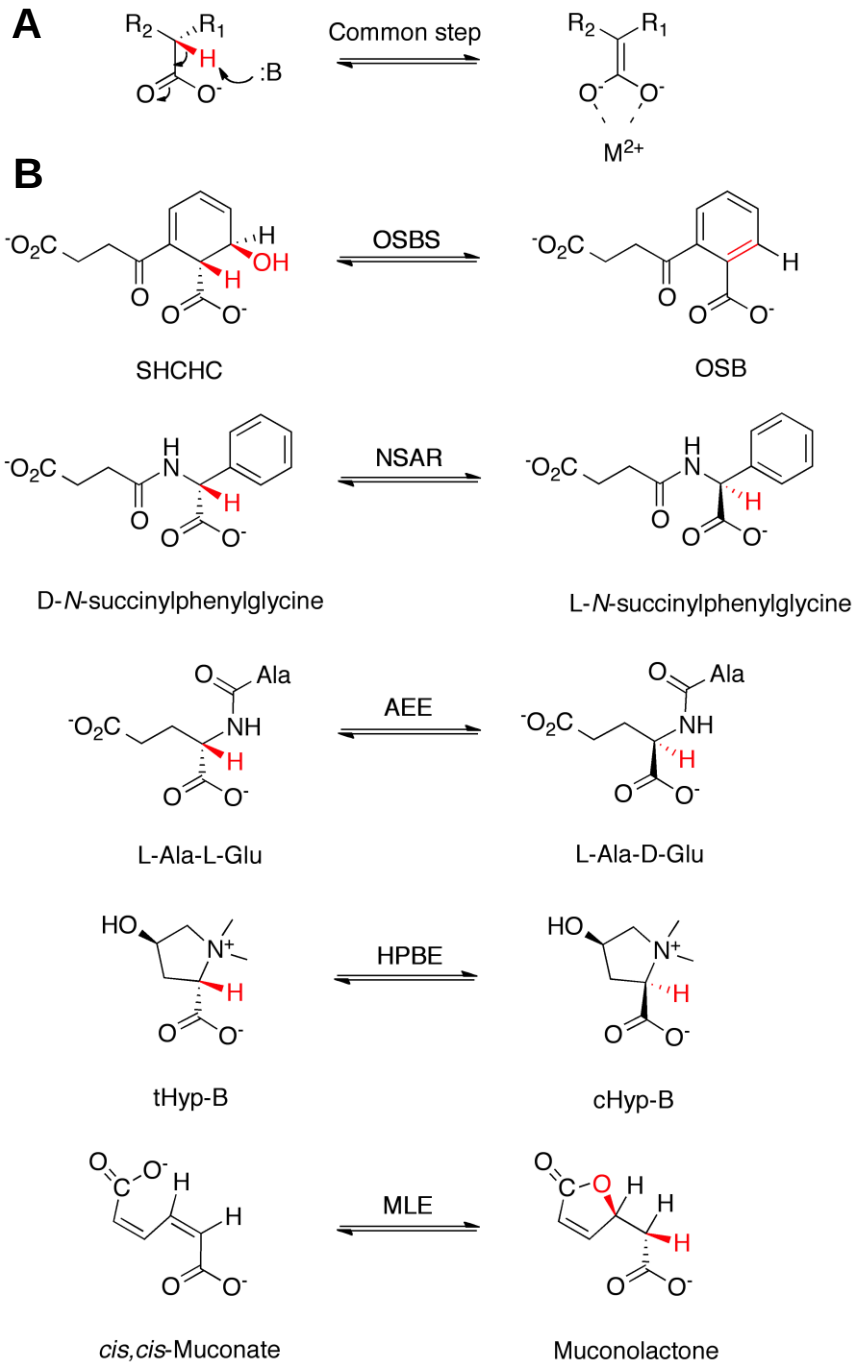


Figure I.2. (A) The common step shared by all members of the enolase superfamily, in which the enolate intermediate is stabilized by a divalent metal ion. (B) Reactions catalyzed by the members of the MLE subgroup, *o*-succinylbenzoate synthase (OSBS), *N*-succinylamino acid racemase (NSAR), L-Ala-L/D-Glu Epimerase (AEE) [23], 4R-hydroxyproline betaine 2-epimerase (HPBE) [24], and muconate lactonizing enzyme (MLE) [25].

The OSBS family and the NSAR/OSBS subfamily

The enzymes in the *o*-succinylbenzoate synthase (OSBS) family of the functionally diverse enolase superfamily catalyze one step in the menaquinone biosynthesis pathway. OSBS catalyzes the conversion of 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate (SHCHC) to OSB (Figure I.2B). Even though the enzymes of the OSBS family have a single evolutionary origin and a conserved function, they can share as little as 15% sequence identity, making it one of the most divergent protein families [13]. The OSBS family consists several large, divergent subfamilies, which correspond to the phylum from which the OSBS enzymes originated [26]. The *N*-succinylamino acid racemase (NSAR)/OSBS subfamily is a branch of the OSBS family in which many members are catalytically promiscuous and catalyze both OSBS and NSAR activities. NSAR catalyzes the conversion of L-succinylamino acid to D-succinylamino acid, and vice versa (Figure I.2B) [27]. Since NSAR activity is only present in the NSAR/OSBS subfamily, but not in other OSBS subfamilies, NSAR activity evolved from an OSBS ancestor.

Historically, NSAR activity was first discovered in a species of *Amycolatopsis* sp. TS-1-60, from a screen of more than 30,000 microorganisms for *N*-acetylamino acid racemase activity [28, 29]. Later, it was found to have a much higher efficiency toward an *N*-succinylamino acid substrate; hence the name NSAR [27]. This enzyme was first identified as a member of the MLE subgroup of the enolase superfamily because the sequence analysis showed that it contains two catalytic lysines like the enzymes from the

MLE subgroup described above [15]. Interestingly, this NSAR enzyme from *Amycolatopsis* shares 43% sequence identity with the OSBS enzyme from *Bacillus subtilis*, which is an intermediate in the menaquinone biosynthesis pathway [30]. The NSAR enzyme from *Amycolatopsis* was assayed for the OSBS activity, and it has a comparable level of OSBS activity to the OSBS enzyme from *B. subtilis*. The Gerlt lab, at the time, assigned OSBS activity as the biological function to this protein from *Amycolatopsis* [30]. The physiological function of the promiscuous NSAR activity from *Amycolatopsis* was still undetermined. Subsequently, the NSAR activity of a promiscuous NSAR/OSBS enzyme from *Geobacillus kaustophilus* was determined to be part of the irreversible conversion pathway of D- to L-amino acids, which is described below [31].

The mechanism of the OSBS reaction was first elucidated in *Escherichia coli* OSBS (EcOSBS) [32], and the mechanism of the NSAR reaction was subsequently elucidated by both functional and structural studies of the promiscuous NSAR/OSBS enzyme from *Amycolatopsis* (AmyNSAR/OSBS) [22, 27]. Both OSBS and NSAR reactions undergo very similar half reaction, requiring two catalytic lysines (numbered K163 and K263 in AmyNSAR/OSBS) and the metal ion-binding ligands (D189, E214, and D239). However, the OSBS reaction utilizes K163 as a single acid/base catalyst whereas the NSAR reaction requires both lysines as acid/base catalysts. In the OSBS reaction, K163 acts as the general base and abstracts the H_{α} of the substrate to create an enolate intermediate, which is stabilized by the Mg^{2+} ion in the active site. Protonated K163 then donates the proton to the hydroxyl leaving group from the intermediate to

form water in a syn-β elimination manner. K263 probably forms a π-cation interaction with the substrate SHCHC and the enolate intermediate. Thus, K263 needs to be protonated in the OSBS reaction. In contrast, the NSAR reaction follows a two-base, 1-1 proton transfer mechanism. K163 is the general base and K263 is the general acid when a D-succinylamino acid is the substrate. The roles of the lysines are reversed when an L-succinylamino acid is the substrate. The crystal structure of AmyNSAR/OSBS shows that the two catalytic lysines are oriented on opposite sides of the active site, pointing to opposite faces of the substrate's C_α and enabling them to act as the general acid and base [22, 27].

This is a striking example showing the functional divergence of a promiscuous enzymes induced by the main and non-canonical substrates. AmyNSAR/OSBS can carry out two different reactions, which share the common metal-stabilized enolate intermediate. Since the identification of NSAR activity in the OSBS family, researchers have longed to understand the evolution of NSAR activity. The example of AmyNSAR/OSBS supports the hypothesis that evolution of a new function evolved from a promiscuity intermediate. However, similarities in reaction mechanisms and the conservation of catalytic residues only partially explain how NSAR activity evolved from an OSBS ancestor. There are several interesting questions about the promiscuous NSAR/OSBS enzymes. First, when did NSAR activity first evolve in the NSAR/OSBS subfamily? Second, when did NSAR activity from the NSAR/OSBS enzymes become biologically relevant and get recruited into the pathway that converts D- to L-amino acids? And third, what mechanistic features determine the NSAR reaction specificity in

the promiscuous NSAR/OSBS enzymes? One could potentially answer these questions by analyzing the genome context, phylogeny, as well as the structure-function relationship of the NSAR/OSBS enzymes.

The evolution of the OSBS family

In order to understand the evolution of NSAR activity in the NSAR/OSBS subfamily, it is important to understand the evolution of the OSBS family. Like all members of the enolase superfamily, OSBS enzymes are composed of a catalytic C-terminal $((\beta/\alpha)_7\beta)$-barrel and a capping domain consisting of components from both N- and C-termini [14]. The conserved set of catalytic residues (typically lysines) and the divalent Mg^{2+} ion binding residues (typically aspartates and glutamates) reside in the barrel domain. Mutations of these conserved catalytic residues in *E. coli* OSBS completely abolish its activity [33, 34]. This set of residues is also conversed among other members of the enolase superfamily, including the MLE and the DE families. Thus, they do not determine OSBS reaction specificity [13]. The OSBS reaction is thermodynamically favorable because an aromatic ring is formed as a product after the proton abstraction step [35]. Mutations of non-catalytic active site residues in EcOSBS have moderate effects on its activity, except R159 and G288, which are subfamily specific and discussed in more detail below. These data support that enzymes within different OSBS subfamilies might have additional non-catalytic residues to determine OSBS reaction specificity [26].

As mentioned above, members of the OSBS family can share as little as 15% sequence identity. This could indicate that residues that determine OSBS reaction specificity could have diverged within the OSBS family. Phylogenetic and sequence conservation analysis of the OSBS subfamilies indicate that enzymes from different OSBS subfamilies possess different residues required for substrate binding and catalysis [26, 36]. For instance, R159 is conserved in all OSBS subfamilies except the NSAR/OSBS subfamily, and G288 is conserved in all OSBS subfamilies except the NSAR/OSBS and Cyanobacteria 2 subfamilies. Mutations of these two residues in EcOSBS decrease efficiency by >100-fold [26]. However, similar mutations only reduce the activity by ~3-fold in *Thermobifida fusca* OSBS (TfuscaOSBS), a non-promiscuous member from the Actinobacteria OSBS subfamily [36]. Furthermore, the substrate in TfuscaOSBS interacts with some additional residues, which are not present in both EcOSBS and AmyNSAR/OSBS [36]. Structural analysis and mutagenesis of TfuscaOSBS revealed some striking differences in the site for substrate entry within the OSBS family [36]. The flexible active site loop (the 20s loop) allows the substrate binding and release in EcOSBS and AmyNSAR/OSBS [22, 32]. In TfuscaOSBS, the conformational flexibility of a loop following the β5 strand (post-β5) in the barrel domain, instead of the 20s loop, allows the substrate entry and binding [36]. Furthermore, it has been shown that the divergence in the OSBS subfamilies was caused by the loss of quaternary structure and accumulation of insertions and deletions which were associated with an increased rate of amino acid substitution [37]. Compared to the monomeric non-promiscuous OSBS proteins, proteins from the NSAR/OSBS subfamily

have dimeric or octameric structures. Non-promiscuous OSBS proteins have higher sequence divergence and structural divergence compared to the promiscuous NSAR/OSBS proteins. OSBS subfamilies have diverged from each other as well as from the rest of the enolase superfamily. OSBS proteins evolved at a faster rate than related families due to sequence divergence, and not to functional divergence. The structures of NSAR/OSBS proteins are much more similar to proteins from other families of the enolase superfamily than to OSBS proteins although their functions are much more diverse. This suggests that the retention of ancestral sequence and structural features could contribute to the evolution of NSAR activity [37].

The evolution of NSAR/OSBS subfamily

First, to answer the question of when did NSAR activity first evolve in the NSAR/OSBS subfamily, phylogenetic and genome context analysis of the NSAR/OSBS subfamily were done. It is important to mention that most species, which will be discussed below, obtained their *NSAR/OSBS* gene via horizontal gene transfer, based on phylogeny and sequence similarity analysis of the OSBS family [13, 38]. Phylogenetic analysis of the OSBS family indicates that earliest branching point of the NSAR/OSBS subfamily from the rest of other OSBS subfamilies is the OSBS from *Staphylococcus aereus* (Figure I.3). The OSBS enzyme from *Staphylococcus aereus* has no detectable NSAR activity [37]. Similarly, the OSBS enzymes from *Bacillus subtilis* and *Alicyclobacillus acidocaldarius* also lack NSAR activity [30, 38].

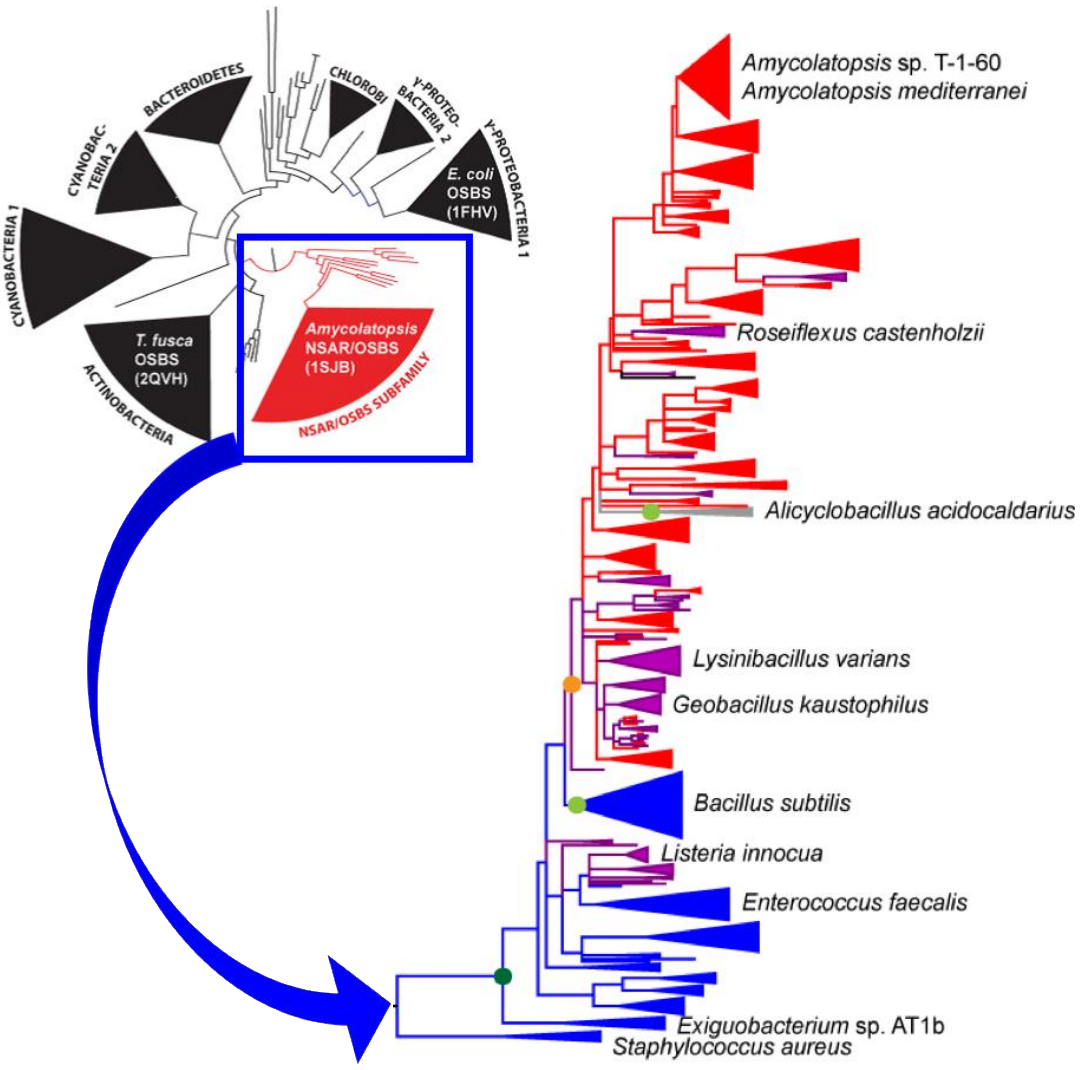


Figure I.3. Phylogenetic tree of the OSBS family (the inset) [26] and the NSAR/OSBS subfamily [38] (Figure is modified with permission from *Biochemistry*). Blue arrow represents the zoomed-in phylogeny of the NSAR/OSBS subfamily from the OSBS family. Blue branches consist of proteins that are encoded in menaquinone operon, indicating that OSBS activity is their biological function. Red branches consist of proteins whose biological function is expected to be NSAR activity because their species do not require OSBS activity to make menaquinone or there is a separate OSBS gene encoded in the menaquinone operon. Purple branches consist of proteins that are expected to be bifunctional, because OSBS activity is required for menaquinone synthesis but the NSAR/OSBS subfamily gene is not in the menaquinone operon. Many of these proteins are encoded in operons with genes from the D-amino acid conversion pathway.

Phylogenetic analysis of the NSAR/OSBS subfamily shows that the earliest branching enzyme that has NSAR activity is the OSBS enzyme from *Exiguobacterium* sp. AT1b (ExiOSBS) (Figure I.3) [39]. ExiOSBS is an efficient OSBS with a very low and inefficient promiscuous NSAR activity. Genome context analysis showed that ExiOSBS is encoded in a menaquinone biosynthesis operon, which requires its OSBS activity. The inefficient promiscuous NSAR activity of ExiOSBS has no biological function. Thus, we could trace the origin of NSAR activity back to the common ancestor of ExiOSBS and the rest of the NSAR/OSBS subfamily. Like ExiOSBS, most OSBS enzymes that diverged near the base of the NSAR/OSBS subfamily phylogeny are encoded in menaquinone biosynthesis operons, indicating that NSAR is the promiscuous, non-biologically relevant activity. This data indicates that NSAR activity first appeared as a promiscuous activity and was later recruited to become biologically relevant.

This brings us to the second question of when did NSAR activity become biologically relevant. As stated above, the origin of NSAR activity is near the base of the NSAR/OSBS subfamily tree where the NSAR/OSBS proteins are encoded in the menaquinone biosynthesis operons. In the absence of natural selection, NSAR activity could have been gained or lost among the NSAR/OSBS proteins, as might be the case for the OSBS enzymes from *B. subtilis* and *A. acidocaldarius* [38, 39]. NSAR activity could then evolve to become biologically relevant due to positive selection after it appeared as a promiscuous activity. For example, the NSAR/OSBS enzyme from *Geobacillus kaustophilus* was discovered to be bifunctional (Figure I.3). NSAR activity

of this enzyme is part of the pathway that irreversibly converts D- to L-phenylalanine in this organism (Figure I.4) [31]. This pathway in *G. kaustophilus* consists of three different enzymes: a succinyl *N*-transferase (GNAT, GCN5-related *N*-acyltransferase), an NSAR, and an L-desuccinylase (M20, a peptidase) [31]. Meanwhile, the OSBS activity is required for menaquinone biosynthesis. A recent study showed that NSAR/OSBS enzymes in other species from different phyla including *Lysinibacillus varians* and *Roseiflexus castenholzii* are also bifunctional [38]. Genome context analysis indicates that the NSAR/OSBS enzyme from *L. varians* is encoded in a similar operon with *G. kaustophilus* NSAR/OSBS. On the other hand, *R. castenholzii* NSAR/OSBS is encoded in an operon consisting a GNAT superfamily member and an α/β hydrolase superfamily member, which might serve the same function as the M20 family L-desuccinylase of *G. kaustophilus*. The OSBS activity of both *L. varians* and *R. castenholzii* NSAR/OSBS enzymes is required for menaquinone biosynthesis [38].

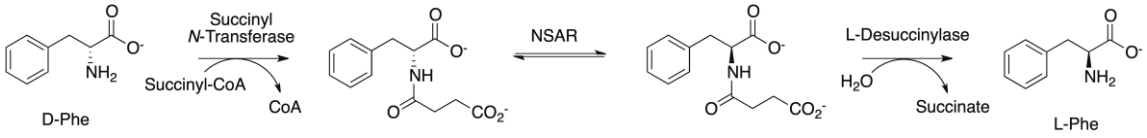


Figure I.4. The irreversible conversion pathway of D- to L-phenylalanine, first discovered in *Geobacillus kaustophilus* [31]

Genome context of *NSAR/OSBS* genes in several *Amycolatopsis* organisms including *A. mediterranei* S699 indicate that NSAR is their only biological function.

These species have an *OSBS* gene encoded in the menaquinone biosynthesis pathway operon and an *NSAR/OSBS* gene encoded in an elaborated *G. kaustophilus* D-amino acid conversion pathway operon. In addition to the GNAT and M20 family enzymes, this operon also encodes a D-glutamate deacylase, 3 different β-lactamases (including prolyl oligopeptidase), and four subunits of a dipeptide/oligopeptide ABC transporter, all of which are predicted to be involved in peptidoglycan degradation [38]. The NSAR/OSBS from *A. mediterranei* S699 shares 83% sequence identity with the well-characterized NSAR/OSBS from *Amycolatopsis* sp. TS-1-60 (AmyNSAR/OSBS) [38]. Although the genome of this organism has not been sequenced, NSAR/OSBS enzymes of several *Amycolatopsis* species share high sequence similarity. Thus, for the NSAR/OSBS enzymes from the *Amycolatopsis* species, the NSAR activity is inferred to be the biological function while the OSBS activity is the promiscuous side reaction. Nevertheless, AmyNSAR/OSBS catalyzes both OSBS and NSAR reactions efficiently.

Ultimately, answering the third and last question of what mechanistic features determine NSAR specificity is essential to fully understand the evolution of NSAR activity and specificity in the NSAR/OSBS subfamily. As mentioned earlier, some of the NSAR/OSBS enzymes can catalyze both OSBS and NSAR reaction efficiently including AmyNSAR/OSBS, while ExiOSBS has a low and inefficient NSAR activity and some NSAR/OSBS enzymes completely lack NSAR activity. The levels of NSAR activity are obviously different among the enzymes within the NSAR/OSBS subfamily. We want to understand what sequence and structural features that are responsible for NSAR reaction specificity. However, because members of the OSBS family, including NSAR/OSBS

enzymes are extraordinarily divergent, such specificity determinants are not always conserved among these divergent homologous enzymes, making it difficult to fully understand NSAR reaction specificity in the NSAR/OSBS enzymes [13].

One way to understand NSAR reaction specificity is to examine the binding and orientation of the substrates in the promiscuous enzymes and non-promiscuous enzymes. The crystal structures of the promiscuous AmyNSAR/OSBS and non-promiscuous EcOSBS and TfuscaOSBS are available for such comparison and can give us some valuable lessons about NSAR reaction specificity. One striking difference is the conformation of OSB in active site of the promiscuous AmyNSAR/OSBS and both non-promiscuous EcOSBS and TfuscaOSBS [22, 32, 36]. The succinyl moiety of OSB is extended in AmyNSAR/OSBS, while it is in a bent conformation in both EcOSBS and TfuscaOSBS. As mentioned above, EcOSBS, TfuscaOSBS, and AmyNSAR/OSBS all use different residues to bind OSB in such different conformations. Likewise, the succinyl moiety of *N*-succinylphenylglycine (an NSAR substrate) is also in an extended conformation in the active site of AmyNSAR/OSBS. A closer look at the active site of AmyNSAR/OSBS indicates that the β8 strand and the loop connecting it with the capping domain are shifted, creating a cavity that can accommodate the extended conformation of the succinyl moiety of each substrate. In contrast, EcOSBS does not have this feature to accommodate the binding of the NSAR substrate [22, 33]. However, the binding accommodation does not appear to be sufficient to explain the reaction specificity in the NSAR/OSBS enzymes. Some NSAR/OSBS subfamily members share this structural feature with AmyNSAR/OSBS, but they have low or undetectable NSAR

activity. This suggests that other factors beyond substrate binding contribute to catalytic promiscuity and bifunctionality of NSAR/OSBS enzymes.

Based on the sequence conservation of the residues in the active site, the Glasner lab identified the first amino acid Y299 in *Alicyclobacillus acidocaldarius* OSBS (AliacOSBS), a non-promiscuous member of the NSAR/OSBS subfamily, that strongly influences the NSAR reaction [38]. The Y299I mutation in AliacOSBS did not affect OSBS activity, while it introduced the NSAR activity into AliacOSBS enzyme. The Y299I mutation increased NSAR activity from undetectable to 1.2×10^2 $M^{-1}s^{-1}$. Replacement of the bulky tyrosine residue with an isoleucine allows the substrate to reorient in such a conformation that both catalytic lysines can be accessible to the alpha proton and enable the racemization reaction to happen. Y299 is an important residue that determines NSAR specificity reaction in AliacOSBS [38]. In Chapter V, I will discuss my contribution in more detail to help elucidate the mechanism of how Y299 determines NSAR reaction specificity in AliacOSBS. However, Y299 is not the sole residue responsible for NSAR specificity determination. For instance, *B. subtilis* OSBS has a leucine at this position, but it has no detectable NSAR activity. On the other hand, *Enterococcus faecalis* NSAR/OSBS has a phenylalanine at the same position, but it can carry out both reactions efficiently. This suggests that there are other residues that can modulate NSAR reaction specificity determination by Y299 residue in the NSAR/OSBS enzymes. The focus of this dissertation is to identify such residues that are required for NSAR reaction specificity. We determined that R266 of AmyNSAR/OSBS is important for NSAR reaction specificity in the NSAR/OSBS enzymes. I will discuss the important

roles of R266 in the evolution of NSAR activity in the NSAR/OSBS subfamily, as well as the homologous residue in the dipeptide epimerase family of the enolase superfamily in more details in Chapter II and Chapter III. It is important to identify structural and sequence features that are responsible for determining NSAR reaction specificity because it will help us understand the evolution of NSAR activity in the NSAR/OSBS subfamily. Ultimately, understanding the reaction specificity in NSAR/OSBS proteins can help us understand the mechanism for catalytic promiscuity in other enzymes.

Catalytic promiscuity and the evolution new enzyme functions

More and more evidence has shown that many (if not most) enzymes can have multiple substrates and activities, contradicting the oversimplified, classroom view of enzymes, which is one enzyme one reaction with absolutely high specificity [1, 7]. Such enzymes with multiple substrates and functions are considered to exhibit promiscuity. There are different ways to define promiscuity [37]. In this dissertation, promiscuity is defined as the coincidental ability of an enzyme to catalyze reactions that are not biologically relevant. Promiscuous enzymes can utilize different substrates for the same type of chemical reaction (substrate promiscuity) or catalyze different chemical transformations with different substrates (catalytic promiscuity). The ability of promiscuous enzymes to carry out different reactions in the same active site is due to many contributing factors including the similarity in the chemical transformation steps, a common intermediate, and/or the similarity in substrate structures between the reactions

[7]. Many lines of evidence show that catalytic promiscuity explains how enzymes can evolve new functions [1-8].

The model of enzyme evolution through promiscuous intermediates helps explain the evolution of new enzyme functions. The model states that the promiscuous, non-biological side reactions of an enzyme, through natural selection, becomes more efficient and functional if the promiscuous reaction confers a selective advantage. This will yield a bifunctional protein that has two biologically functional activities. If gene duplication occurs, mutations can accumulate and result in paralogs of the protein with separate functions. Alternatively, if horizontal gene transfer of this bifunctional gene occurs and a species does not require the original activity, a protein with a new function will arise. The original activity of the protein then can become a promiscuous side reaction with no biological function (Figure I.5) [39].

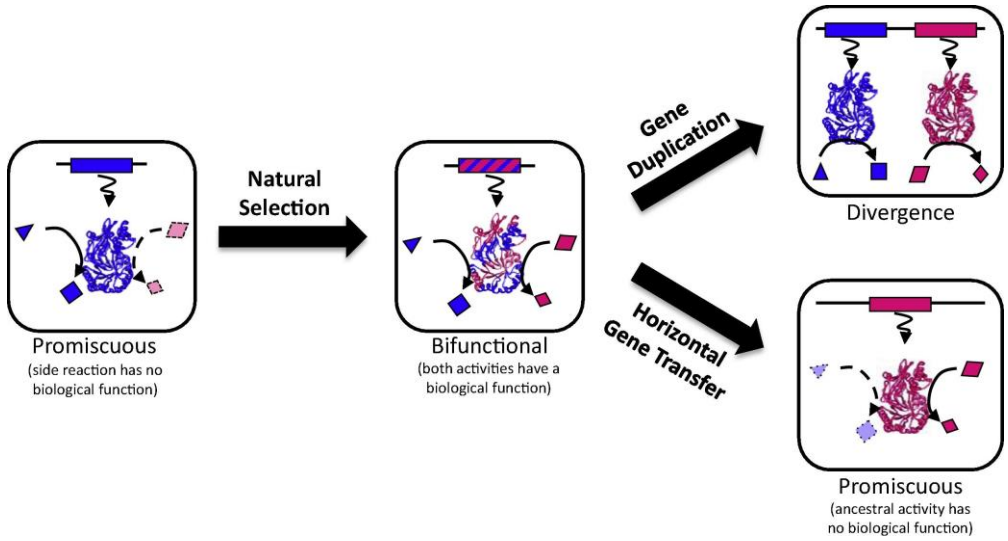


Figure I.5. The model of enzyme evolution through catalytically promiscuous intermediates. The figure is reproduced from reference [39], with permission from Elsevier.

This model of enzyme evolution through promiscuous intermediate explains the evolution of NSAR activity in the NSAR/OSBS subfamily [13, 39]. Extant NSAR/OSBS enzymes that represent each stage of this model have been identified and characterized. It is important to keep in mind that these extant enzymes do not directly represent intermediates in a step-wise evolutionary pathway. Instead, properties of these extant enzymes can be used to infer the functional characteristics of their ancestors [39]. For example, ExiOSBS, which is an enzyme from the first branch off of the NSAR/OSBS subfamily phylogeny that has NSAR activity, represents the early promiscuous stage of the evolution of the NSAR activity [39]. ExiOSBS is encoded in the menaquinone biosynthesis operon, and its low NSAR activity has no biological function. Many NSAR/OSBS enzymes that diverged after ExiOSBS, including *G. kaustophilus*

NSAR/OSBS, are bifunctional with NSAR activity being a part of the pathway that irreversibly converts D- to L-phenylalanine or another uncharacterized pathway [31]. Lastly, NSAR/OSBS enzymes from *Amycolatopsis* species represent the last stage of the evolution of NSAR activity, in which NSAR activity is inferred to be the biological function with OSBS being the side, promiscuous reaction after the horizontal gene transfer event [38].

It is clear that the promiscuous side reaction provides an evolutionary starting point for an enzyme to evolve a new function. However, the underlying molecular mechanism that allows enzymes to be able to catalyze distinct chemical reactions within one active site remains a challenge. More detailed mechanistic and structural studies are necessary to fully understand how such catalytically promiscuous enzymes work. And ultimately, such studies will help us understand the features that determine reaction specificity in those enzymes.

Promiscuity is not an uncommon property in enzymes

As stated above, many enzymes (if not most) are promiscuous. However, generating sufficiently comprehensive libraries to screen for substrate promiscuity is challenging, systematic, and high-throughput screening for catalytically promiscuous activities is not generally feasible. No single detection method is available to identify a complete range of substrates and reactions. Thus, catalytic promiscuity is a lot more difficult to discover than one could imagine. Most characterized catalytically

promiscuous enzymes have come from analysis of mechanistically diverse superfamilies that share a common protein fold and conserved catalytic residues [40-42]. For instance, a study screened a library of different substrates of 10 chemically distinct reactions against 24 enzymes from 15 distinct functional families of the metallo-β-lactamase superfamily [41]. This study revealed that those related enzymes are generally promiscuous and each enzyme can carry out on average 1.5 reactions in addition to its native function. More specifically, five enzymes were able to carry out 3 reactions in addition to their native function, and 20 out of 24 enzymes could catalyze at least one promiscuous hydrolysis reaction [41]. Furthermore, this study also revealed that the crossed reactions between those enzymes generally occur between closely related functional families, in addition to the chemical similarity between the substrates of different reactions [41].

Another example of a well-characterized diversely functional superfamily is the haloacid dehalogenase (HAD) superfamily, which includes enzymes that carry out the cleavage of substrate C-Cl, P-C, and P-OP bonds. Members of the HAD superfamily share a Rossman fold core domain and a fused cap domain. The active site contains a Mg^{2+} ion and a catalytic Asp residue, which mediates different group transfer reactions [43]. Screening over 200 HAD enzymes against a library containing 167 substrates revealed that members of the HAD superfamily exhibit substrate promiscuity, with 75% of the members having the ability to utilize more than 5 different substrates and 23% active with more than 41 substrates [42]. Beside studying the mechanistically diverse superfamilies, promiscuous enzymes could also be identified from screening a library of

substrates with similar chemical properties. For example, screening a library of 9 non-natural aromatic substrates against a phenol hydroxylase revealed that this enzyme can promiscuously hydroxylate different aromatic compounds [44]. Functional characterization of enzymes including those members of mechanistically diverse superfamilies suggest that catalytic promiscuity is an intrinsic property of enzymes and not uncommon in the enzyme universe.

Methods for identifying promiscuity in enzymes

Although most promiscuous enzymes have been identified from *in vitro* studies, some promiscuous enzymes were also identified from *in vivo* analysis of auxotrophic knockout strains that lack an essential enzyme. For example, MetB (cystathione-γ-synthase) was shown to be upregulated and promiscuously catalyze the reaction of IlvA (threonine dehydratase) to overcome the auxotrophic *ΔilvA* strain. The upregulation of MetB, however, was induced by depletion of S-adenosylmethionine (SAM) by a bacteriophage SAM hydrolase [45]. As another example, the promiscuous phosphite-dependent dehydrogenase activity of *E. coli* alkaline phosphatase was identified when the *phn* operon, which utilizes phosphite, was knocked out in *E. coli* [46].

In some other cases, the promiscuous activity of enzymes were also identified when the conditions of the organisms are perturbed. An approach to identify promiscuous activity of an enzyme was to knock out essential genes in an organism. For instance, the Hecht lab studied the rescue of auxotrophs by alteration of gene expression

by non-natural, *de novo* designed proteins, which have no significant sequence similarity and do not necessarily carry out the same functions as the natural proteins. These non-natural proteins were designed from a binary code strategy, in which the sequence pattern of polar and nonpolar residues were specified to coincide with the exposed and buried parts of a structure [47]. The plasmids containing the genes of those non-natural proteins were transformed into different knockout *E. coli* strains. Rescued auxotrophs by those non-natural proteins were then identified and studied further [47]. For example, SerB, which is a phosphoserine phosphatase, catalyzes an essential step of the serine biosynthesis pathway. A *serB* knockout strain was rescued by a non-natural, *de novo* protein SynSerB, which acts as a regulatory protein to upregulate the expression level of HisB (histidinol phosphate phosphatase) [48]. Overexpression compensates for the low rate of the promiscuous phosphoserine phosphatase activity of HisB, which could replace the missing SerB enzyme and rescue the *serB* knockout strain [48]. In another case, knockout of citrate synthase (GltA) is lethal because this enzyme catalyzes an essential step of the TCA cycle to provide a precursor for glutamate biosynthesis. The *gltA* knockout strain was rescued by SynGltA, which upregulates the expression level of methylcitrate synthase, which promiscuously synthesizes citrate from oxaloacetate and acetyl-CoA, rather than its native substrate, propionyl-CoA [49]. These examples demonstrate that catalytic promiscuity is an intrinsic property of enzymes. The promiscuous activity can be manifested and observed when the conditions in the cell are perturbed. However, this *in vivo* approach to identify promiscuity has some limitations.

The promiscuous activity is often inefficient and may be difficult to detect, or even undetectable [40].

Another advanced approach to identify promiscuity in enzymes is high-throughput screening of enzymes with a library of different substrates using robotics and microfluidics. For instance, a study combined a high-throughput screening approach with structural and modeling investigations along with genomic and metabolic context analysis to identify functions and promiscuity of a set of enzymes from an uncharacterized family against a set of candidate substrates. The study showed that 80 out of 124 enzymes from the uncharacterized family DUF849 were able to transform at least one substrate in a β-keto acid condensation reaction, leading to the functional assignment of this family of enzymes as β-keto acid cleavage enzymes [50]. Another study used an "ultrahigh-throughput" screening to identify promiscuous hydrolases of sulfate monoester and phosphotriester substrates against a metagenomic library consisting of 1.25 million sequences [51]. Single *E. coli* cells carrying different expression plasmids from the library were compartmentalized into 2 pL droplets containing lysis agents and fluorogenic sulfate monoesters and phosphotriesters. Droplets with high fluorescence were selected and the plasmid DNA was isolated and sequenced. Two rounds of this approach allowed the identifications of 6 unique arylsulfatase and 8 unique and novel phosphotriesterases, each of which can use more than one substrate [51]. On the other hand, an "absorbance-based ultrahigh-throughput" screening was used against half a million sequences to identify an NAD^+-dependent amino acid dehydrogenase. However, the droplets were relatively larger (180 pL) in

order to observe a UV-Vis sensitive coupled enzymatic reaction, which produced NADH and reduced tetrazolium dye. With a longer incubation period, this approach also lead to identification of hydrogenases that can promiscuously produce NADPH [52]. These advanced high-throughput approaches can allow us to identify promiscuous enzymes. However, they have some limitations and difficulties when the reactions produce a small change in absorbance, a fluorogenic substrate library is not readily available, or they require highly expensive instruments [40].

Another approach that has been used to identify promiscuity in enzymes is *in silico* characterization, involving substrate docking and molecular modeling. For instance, an *in silico* study examined the substrate selectivity and recognition by docking 125 common metabolites into the active sites of 120 key metabolic enzymes in *E. coli.* The study suggested several promiscuous bindings and cross-reactions among 15,000 potential pairs [53]. Another study explored substrate promiscuity of five orthologs of an enoyl-acyl carrier protein reductase against 13 different substrates by calculating quantum mechanical hydride affinity in combination with molecular docking. The physiological function of enoyl-acyl carrier protein reductase is to reduce an enoyl acyl carrier protein to its saturated form. This study suggested that promiscuity in these enzymes toward a wide range of substrates (different in both geometry and oxidation state) can be predicted from the combined *in silico* methods [54]. The *in silico* approach helps broaden the understanding of promiscuity in enzymes by revealing potential promiscuous reactions. However, it has limitations because it alone does not necessarily reflect reality and requires experimental confirmation of the potential hits. Thus,

combinations of different methods, *in vitro*, *in vivo*, and *in silico* is necessary to identify and characterize promiscuity in enzymes.

Mechanisms of enzyme specificity and promiscuity

The conformational flexibility of the active site and promiscuity

There have been many studies to explain promiscuity in enzyme evolution. One of the aspects that has been extensively explored is the relationship between promiscuity and protein dynamics and flexibility, all of which seem to correlate and play essential roles in enzyme evolvability [55, 56]. For example, cytochromes P450, which are ubiquitous heme-containing monooxygenases, have a flexible active site and can bind a wide range of substrates in different active site conformations [57]. Computational studies on cytochromes P450 (CYP), show that the degree of conformational diversity determines their substrate specificity. Highly specific CYP2A6 has the most rigid structure, whereas CYP3A4, the most promiscuous CYP known, exhibits the highest flexibility. However, there is no direct link between the evolvability of CYP enzymes with their dynamics and flexibility [58].

Several studies of reconstructed ancestral enzymes and modern enzymes have demonstrated the direct link between promiscuity and protein dynamics and flexibility. Reconstructions of ancestral enzymes can provide valuable lessons about the evolution trajectory of many modern enzymes [59-61]. However, one needs to keep in mind that

the reconstructed ancestral enzymes are not necessarily the true ancestors but only the best approximation from phylogenetic analysis [62]. It is generally accepted that many highly specific modern enzymes (specialists) have evolved from their promiscuous ancestors (generalists) [1, 7, 63]. One of the most extensively studied enzyme family is the β-lactamases, which are produced by many modern bacteria to become resistant toward β-lactam antibiotics. The resurrected Precambrian β-lactamases have been shown to degrade a wide range of antibiotics, while the modern TEM-1 β-lactamase is more specific for penicillin [64]. Structural and computational analysis of resurrected ancestral and modern β-lactamases showed that the ancestral enzymes have a more flexible active site to accommodate the binding of a wide range of antibiotics with different sizes and shapes. On the other hand, the modern enzymes have comparatively more rigid active sites, reflecting the adaptation to penicillin specificity in modern β-lactamases [64, 65]. These studies show that at least for β-lactamases, conformational dynamics explained the evolution of modern specialists from promiscuous ancestral generalists. That is, within the β-lactamases, evolution from a promiscuous generalist to a specific specialist enzyme is correlated with a loss of conformational diversity.

The conformational diversity observed in enzymes including cytochromes P450 and β-lactamases described above is essential for many enzymes to exhibit promiscuity. Indeed, L.C. James and D.S. Tawfik proposed a model for enzyme evolution that is mediated by conformational diversity and functional promiscuity. The model suggests that a promiscuous enzyme can exist in equilibrium between different conformations. Such an enzyme exists in its predominant conformation to catalyze its native function

and can change its conformation to accommodate the binding and utilize the alternative substrates. Through a gene duplication event, the alternative conformation can be optimized and the catalytic efficiency toward the new substrates can be improved. Thus, a new enzyme can arise [66]. This model was primarily proposed from the studies of antibody SPE7, a monoclonal immunoglobulin E (IgE), which adopts different conformations for different antigens and carries out different functions [67]. The evolution and promiscuity in cytochromes P450 and β-lactamases described above seems to fit fairly well with the conformational diversity model.

Another example demonstrating the direct relationship between active site conformational plasticity and promiscuity is the bifunctional generalist PriA from Actinobacteria. PriA can carry out the isomerization reactions of both N′-[(5′-phosphoribosyl)-formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (ProFAR) for histidine biosynthesis and phosphoribosyl anthranilate (PRA) for tryptophan biosynthesis (Figure I.6A) [68]. In most bacteria, the isomerization of ProFAR for histidine biosynthesis is catalyzed by a HisA specialist, and a TrpF specialist catalyzes the isomerization reaction of PRA for tryptophan biosynthesis. Some organisms in the Actinobacteria phylum are missing the *trpF* gene in the trp operon, and they have a *hisA*-like gene coding for PriA protein, which is bifunctional for both HisA and TrpF activities [68]. The active site of PriA is flexible to accommodate the binding of ProFAR and PRA, which are significantly different in shapes and sizes.

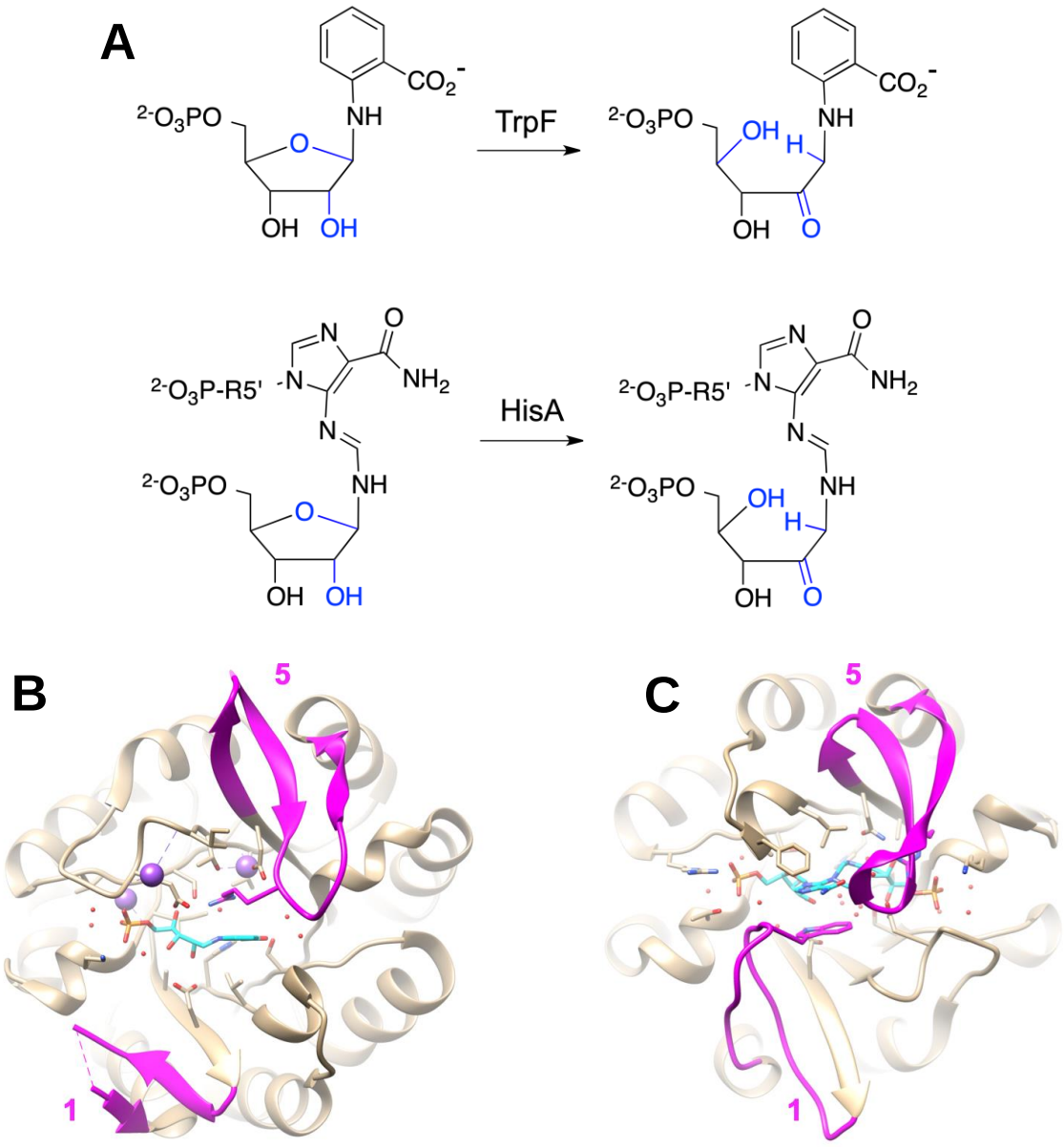


Figure I.6. (A) Reactions carried out by the bifunctional PriA enzyme. The conformations of the active site loops 1 and 5 (numbered and shown in magenta) of PriA for different reactions. (B) In the TrpF state of PriA, loop 5 adopts a β-sheet-like hairpin structure and loop 1 is partly disordered; the reduced product analogue 1-(*o*-carboxyphenylamino)-1-deoxyribulose 5-phosphate (rCdRP) is shown in cyan (PDB ID 2Y85) [69]. (C) In the HisA state of PriA, loop 5 twists into a knot-like conformation and loop 1 completely covers the active site; the product N′-[(5′-phosphoribulosyl)formimino]- 5-aminoimidazole-4-carboxamide ribonucleotide (PrFAR) is shown in cyan (PDB ID 2Y88) [69].

Structural analysis of the bifunctional PriA from *Mycobacterium tuberculosis* shows that the PriA active site has different conformations to catalyze the isomerization reactions of different substrates [69]. Specifically, active site loop 5 adopts a β-sheet-like hairpin structure for PRA isomerization reaction (tryptophan biosynthesis) and twists into a knot-like conformation for ProFAR isomerization reaction (histidine biosynthesis) (Figure I.6B&C). Furthermore, active site loop 1 is partly disordered and does not cover the active site in PRA isomerization reaction (tryptophan biosynthesis) (Figure I.6B), whereas it completely wraps over the active site of PriA in ProFAR isomerization reaction (histidine biosynthesis) (Figure I.6C). The different conformations of PriA active site rearrange the hydrogen bond networks and critical interactions required for isomerization reactions of different substrates [69]. The efficiency of the HisA and TrpF specialists also depends on the dynamics of these two active site loops. Mutations in a TrpF specialist appeared to stabilize the loops for TrpF activity. Introducing a small duplication in loop 1 of a HisA specialist caused the loop to have a TrpF-like conformation and enabled TrpF activity in a HisA specialist [70].

In other cases, active site plasticity and flexibility do not play an important role in facilitating promiscuity in enzymes. The well-characterized members of the alkaline phosphatase (AP) superfamily are highly promiscuous and able to hydrolyze the cleavage of P-O, S-O, and P-C bonds [71]. The hallmark of this family is the crosswise promiscuity, in which the native substrate of one member of the superfamily is often a promiscuous substrate of another member. Examples include phosphonate monoester hydrolase (PMH) from *Burkholderia caryophilli* PG2952, which can remarkably

hydrolyze six different substrate classes. In addition to the four classes of substrates of the AP superfamily (phosphate monoesters, phosphate diesters, phosphonate monoesters, and sulfate monoesters), PMH can also hydrolyze phosphate triesters and sulfonate monoesters, which have never been the substrates of other members of the AP superfamily (Figure I.7) [72].

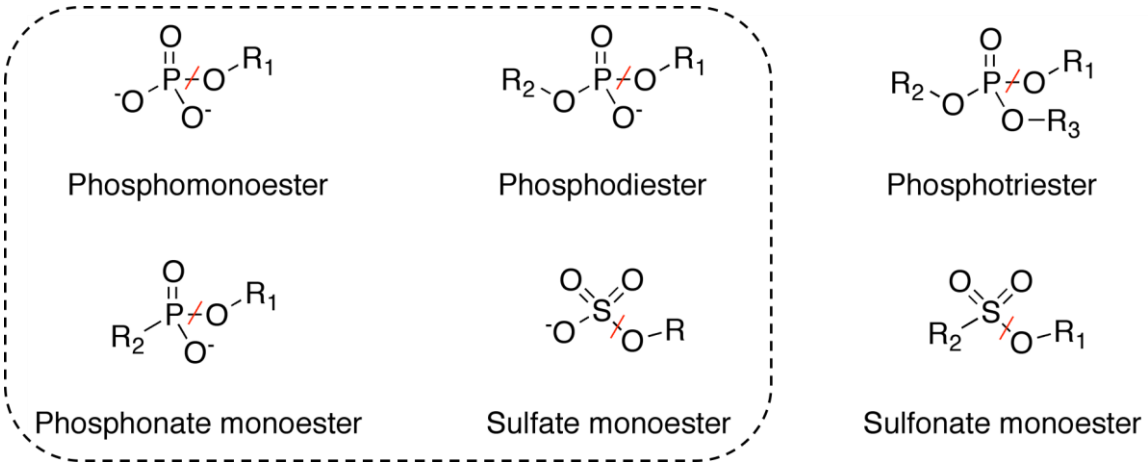


Figure I.7. Generic structures of 6 different classes of substrates that undergo hydrolysis by the phosphonate monoester hydrolase from *Burkholderia caryophilli* PG2952 [72]. Structures shown in the box are known substrates of the alkaline phosphatase superfamily. Red lines show the bond cleavage during hydrolysis.

In some cases, some of the AP superfamily members can catalyze their native and promiscuous reactions with comparable efficiencies [71]. Experimental studies of AP superfamily members demonstrated that their active sites are fairly large and rigid to accommodate the binding of different substrates [73]. Computational studies of PMH

demonstrated no significant change in the active site along the reaction coordinate in different reactions catalyzed by PMH, suggesting that conformational diversity does not facilitate promiscuity in PMH. Instead, active site structural comparisons of different members of the AP superfamily indicate a direct correlation between active site volume, polar solvent accessible surface area and the number of known activities [74]. That is, the more promiscuous enzymes have a larger active site volume and more polar surface area. A large active site allows the binding of a wide range of substrates with different sizes and shapes; however, a large active site is insufficient for promiscuity because non-productive binding events can happen. Furthermore, a large active site with a large polar surface area provides the electrostatic requirements to bind different chemically distinct substrates, and/or to stabilize a variety of transition states [74, 75]. These studies suggest that the electrostatic flexibility, in addition to a large active site, is a common feature for the functional evolution in the AP superfamily.

The ultimate question that we are interested in is what are the mechanistic and structural features that determine the specificity in a family of enzymes. Answering this question helps us not only to understand the basis of how enzymes work, but also to understand structural and mechanistic factors that allow an enzyme to be able to catalyze different reactions. Ultimately, this will help us rationally pick a proper starting scaffold for protein engineering. Although promiscuity provides a selective advantage, it does not ensure the development of a new activity. Mutations that are required for evolution and optimization of new function might not be selected because they might negatively affect the protein's native activity or the overall protein stability [7]. Thus, information about

the physical and mechanistic features that determine specificity and promiscuity will help us understand how enzymes have emerged throughout evolution.

Specificity or substrate discrimination

Specificity (or substrate discrimination) does not depend solely on the relative affinity (K_M values) of different substrates for an enzyme; it rather depends on the ratio of k_{cat}/K_M values of different reactions carried out by an enzyme. The basis of k_{cat}/K_M values determining reaction specificity can be explained through application of transition state theory, which was first proposed by Eyring in 1935 to explain the minimum energy requirement for an enzymatic reaction [76]. Accordingly, the logarithm of k_{cat}/K_M is proportional to the free energy difference ($\Delta G^{\ddagger}$) between the free enzyme and substrate and the transition state complex (Figure I.8A) [77]. Thus, differences in k_{cat}/K_M between different substrates can reflect the different binding energies in their transition states ($\Delta\Delta G^{\ddagger}$). The differential interactions between alternative substrates and the enzyme can be formed in the initial enzyme-substrate complex (which contribute mostly to K_M; Figure I.8B) or in the transition state complex (which contribute to k_{cat}; Figure I.8C). However, where a substrate forms the interactions with the enzyme along the reaction coordinate does not affect the total binding energy associated with that substrate [76, 77]. Discrimination between different substrates can arise from how well the enzyme-transition state complementarity for different substrates can be achieved [78].

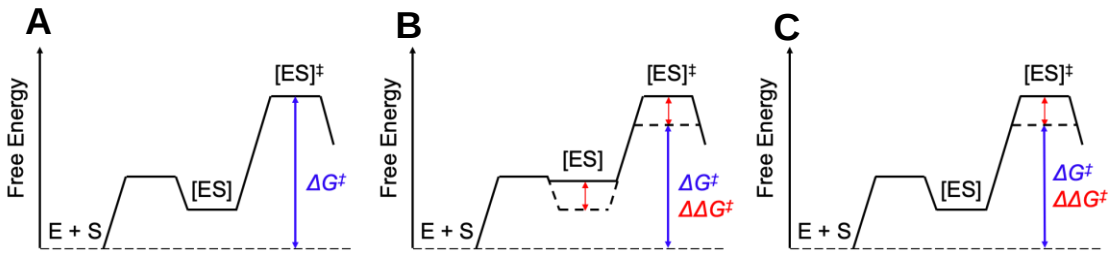


Figure I.8. The application of Eyring transition state theory for substrate discrimination in enzymes. Figure is modified from references [76, 77]. (A) The free energy difference ($\Delta G^{\ddagger}$) between the free enzyme and substrate and the transition state complex is proportional to the logarithm of k_{cat}/K_M of an enzyme. The differences in k_{cat}/K_M between different substrates reflect the different binding energies in their transition states ($\Delta\Delta G^{\ddagger}$). (B) The differential interactions between alternative substrates and the enzyme can be formed in the initial enzyme-substrate complex [ES] which contribute mostly on K_M, or (C) in the transition state complex $[ES]^{\ddagger}$ which contribute to k_{cat}.

Substrate binding and specificity

Understanding how a promiscuous enzyme can recognize different substrates and utilize different active site features to allow the evolution of a new function is always intriguing to evolutionary biochemists. A catalytically promiscuous enzyme can catalyze different reactions utilizing a common subset of catalytic residues. For substrate promiscuity, additional residues that are required to interact with and accommodate the binding of different substrates would not affect catalysis. For catalytic promiscuity, the same catalytic residues might have different roles in different reactions and the enzyme might require additional residues to stabilize different transition states and to facilitate different reactions. C.M. Miton and coworkers have attempted to address this problem by studying an arylsulfatase (PAS), which is a member of the AP superfamily and has a

low level of promiscuity toward phosphonate monoesters [79]. They examined the Michaelis complex (enzyme substrate complex) stabilization in both wild type PAS and a laboratory-evolved sulfatase with a more efficient phosphonate monoesterase activity. The authors found that the suboptimal, less-productive binding of the promiscuous phosphonate substrate to the WT enzyme active site creates a misalignment of the catalytic residues, resulting in a Michaelis complex with higher energy for both ground state and transition state. Mutations in the evolved sulfatase reshape the active site to allow a more productive binding of the phosphonate substrate, with an enhanced enzyme-substrate complementarity. This results in a lower energy of the newly evolved Michaelis complex with better transition state stabilization, thus increasing the promiscuous phosphonate monoesterase activity, while the native sulfatase activity loses efficiency [79].

In some catalytically promiscuous enzymes, reaction specificity does not depend on binding alternative substrates in different conformations or positions. For instance, the promiscuous AmyNSAR/OSBS binds both OSBS and NSAR substrates in similar conformations [13, 22]. Similarly, the heme-dependent peroxygenase from *Agrocybe aegerita* can catalyze both peroxidative activity (one-electron oxidation) and peroxygenative activity (two-electron oxidation and oxygen-transfer with hydrogen peroxide as the oxygen source) with a wide range of substrates [80]. Structural studies of this highly promiscuous peroxygenase revealed that the enzyme maintains the same binding mode for different substrates. Regardless of the nature of the substrates, they bind in the active site in the same conformations and form hydrophobic interactions with

a conserved phenylalanine triad [81]. This phenylalanine triad is directly involved in binding different substrates despite the low specificity of the hydrophobic interactions. The reaction specificity of this promiscuous peroxygenase rather comes from the nature of different substrates and hydrophobic interactions with residues in a channel that directs different substrates for different reactions toward the heme in the active site [81].

The flexibility of active site loops

In some enzymes, active site loops play an important role in determining reaction specificity. For instance, serum paraoxonase 1 (PON1), a Ca^{2+} dependent lactonase with promiscuous paraoxonase activity, contains active site loops that are responsible for binding different substrates for different reactions (Figure I.9A&B). Some alanine mutations in one of the loops decreased the promiscuous paraoxonase activity, without affecting the native lactonase activity [82]. However, those mutations decreased k_{cat} while having a minimal effect on K_M of the paraoxonase activity, suggesting that the loop is critical for stabilizing the transition state for catalysis and/or the loop conformation is crucial for catalytic turnover [82]. Further kinetic and computational studies indicate this active site loop determines specificity by contributing to the hydrophobicity of the active site [83]. Substitutions of a key residue Y71 in this loop expand the active site volume and allowed more water to enter the active site, changing its electrostatic environment and disrupting the position of K192 (which has an important role in modulating paraoxonase activity). This effect is more pronounced on

the paraoxonase activity than the native lactonase activity, resulting in the differential effects on the two activities [83].

Another study of phosphotriesterase (PTE) from the amidohydrolase superfamily, which has a weak, promiscuous lactonase activity, showed that the paraoxon specificity in PTE is due to some additional insertions in the active site loop 7 (Figure I.9A&C) [84]. Other related lactonases with promiscuous PTE activity do not have those insertions in loop 7, suggesting that PTE has evolved from a lactonase. Indeed, removing these insertions in loop 7 significantly improved the lactonase activity with a minimal effect on the native paraoxonase activity in PTE. However, this effect totally depends on a single H254R mutation on the same loop, suggesting the His at this position has an epistatic effect on the lactonase activity. Structural modelling of loop 7-truncated, H254R PTE indicated that the loop 7 truncation allows the optimal binding of the lactonase substrate while maintaining the proper binding of the paraoxon substrate. Thus, this explains the observed effects on the two activities when removing the insertions in loop 7 in PTE [84].

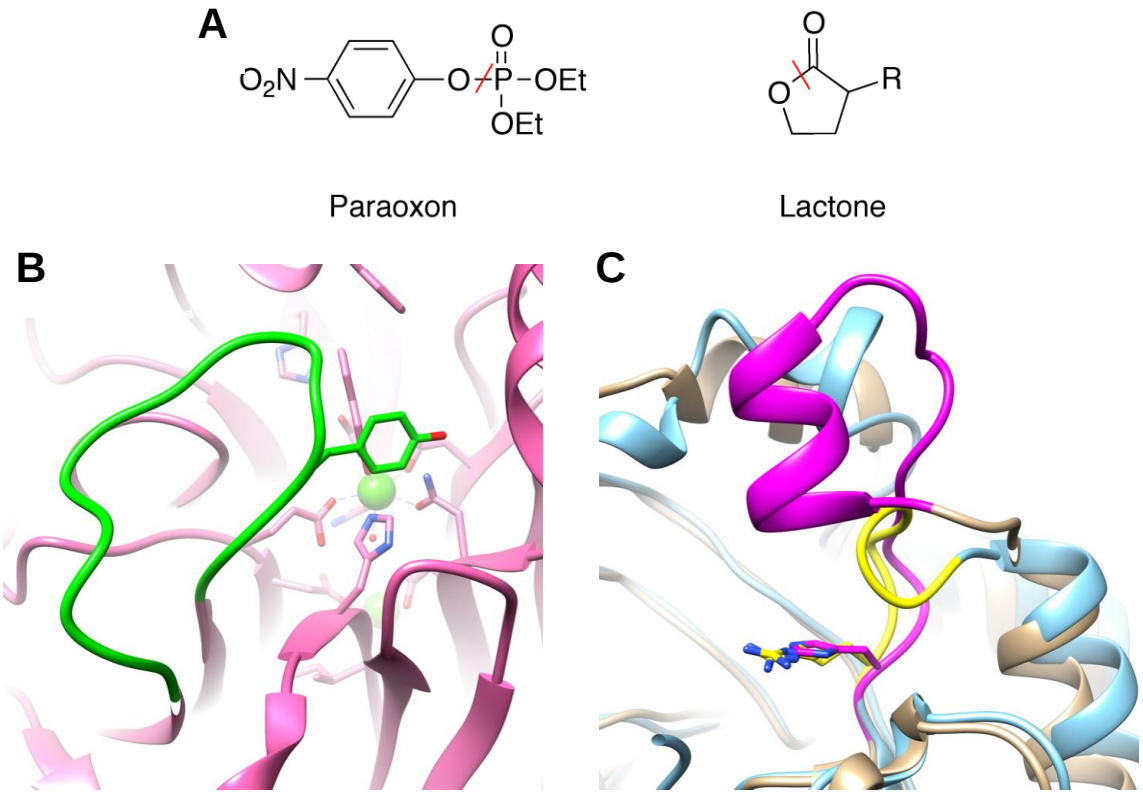


Figure I.9. Active site loops mediate promiscuity in phosphotriesterase (PTE) and serum paraoxonase 1 (PON1) (A) The paraoxon and lactone substrates for the hydrolysis reactions carried out by PTE and PON1. (B) The active site of PON1 (shown in hot pink, PDB ID 3SRG [82]) showing the flexible loop containing the key residue Y71 (highlighted in green). (C) The active sites of the PTE from *Pseudomonas diminuta* (shown in khaki, PDB ID 1HZY [85]) and the lactonase from *Sulfolobus solfataricus* (shown in cyan, PDB ID 2VC7 [86]). The active site loop 7 insertion is highlighted in magenta for the PTE; the active site loop 7 in the lactonase is highlighted in yellow. The residue H254 of the PTE and R254 of the lactonase are also shown.

In other enzymes, active site loops do not contribute to the proper binding of the substrates nor the stabilization of the transition state. Instead, rearrangement of active site loops brings the catalytic residues into proximity of different substrates. An example includes the catalytically promiscuous, bifunctional enzyme fructose-1,6-bisphosphate

aldolase (FBPA)/FBP phosphatase (FBPP) from *Sulfolobus tokodaii* [87, 88]. This enzyme carries out the reversible condensation of dihydroxyacetone phosphate (DHAP) and glyceraldehyde-3-phosphate to yield FBP and the dephosphorylation of FBP to fructose-6-phosphate. The lysine residue required for Schiff base formation in the aldo condensation/cleavage reaction interacts with DHAP. However, in the phosphatase reaction, the region containing this lysine residue is further away from the active site and does not contact FBP [87, 88].

Flexible loops covering the active site of an enzyme have been generally proposed to play important roles in determining reaction specificity, such as in the L-Ala-D/L-Glu epimerase (AEE) from the enolase superfamily [89]. Specifically, mutations in this loop, in combination with a mutation in the barrel domain, completely switches its epimerase specificity to the OSBS reaction [90-92]. Furthermore, the active site loop in AEE determine its substrate specificity by forming different interactions with different dipeptide substrates [93]. However, the role of this active site loop in substrate specificity is not conserved in the enolase superfamily, because mutations of this active site loop in other members, including the OSBS and MR, only had small effects on their activities [26, 36, 94]. On the other hand, the active site loop in the promiscuous NSAR/OSBS enzyme from *Amycolatopsis* plays a very limited role in determining specificity because mutations in this loop had similar effects on both OSBS and NSAR activities, mainly by increasing the K_M values [95]. This suggests that in promiscuous AmyNSAR/OSBS, the active site is responsible for the optimal binding of both OSBS and NSAR substrates for catalysis, mainly by hydrophobic packing, and other residues

in and/or beyond the active site might determine specificity for the OSBS and NSAR reaction [95].

In some enzymes, reaction specificity determinants located in the active site have direct interactions with the substrates. For example, a study on prenyltransferases, which carry out the regiospecific transfer of C5 dimethylallyl donors to the side chain of macrocyclic acceptor substrates, revealed that a single amino acid substitution in the active site, F222G, was sufficient to completely switch the donor specificity from a C5 to a C10 geranyl prenyltransferase. The single mutation remarkably allows the binding of the extended C10 substrate [96].

A recent study of the highly promiscuous flavin-containing cyclohexanone monooxygenase (CHMO), which is part of the cyclohexanol degradation pathway in some bacteria, revealed that not all enzymes have a mechanism to discriminate among their substrates [97]. Structural studies of the promiscuous CHMO from *Thermocrispum municipale* showed that, unlike cytochrome P450 mentioned above, this CHMO has no conformational changes in its active site during catalysis, yet this enzyme can utilize a broad substrate range [97]. Furthermore, CHMO contains a hydrophobic, highly conserved active site pocket, which was hypothesized to be important for substrate specificity. When up to eight highly conserved residues on the surface of this tunnel were mutated to alanine, the enzyme linearly decreased its activity rather than switching its substrate selectivity. The authors proposed that some enzymes, including CHMO, have no structural features responsible for their substrate specificity and selectivity;

rather the accessibility to the intermediate stabilization by the cofactor in the active site appears to be important for their substrate preferences [97].

Loop and domain insertions

As mentioned earlier, some members of the HAD superfamily are highly promiscuous and share a Rossman fold core domain and a fused cap domain [42]. The HAD enzymes with broad substrate ranges also contain an active site loop insertion, which might play a role in interacting with different substrates [98]. Furthermore, the HAD members that have no cap or insertion domains besides the Rossman core domain are highly specific, while the HAD members that have multiple insertions tend to have broader substrate ranges [42]. For example, most members of the HAD superfamily catalyze the hydrolysis of P-O bonds. Some members gain phosphonate hydrolase activity (P-C bond cleavage) by a domain insertion, which introduces a lysine residue and allows for Schiff base formation for P-C cleavage [99]. Thus, loop and domain insertions allow the expansion to new substrates and reactions, either by providing new interactions with new substrates or new catalytic residues for new chemistry in the HAD superfamily [100].

Similarly, loop and domain insertions to the shared active site allow the evolution of reaction specificity in the AP superfamily. Sunden and coworkers examined the structural features that contribute to reaction specificity to study the evolution of the AP superfamily, which includes many promiscuous members [101, 102]. As mentioned

above, the members of the AP superfamily are highly promiscuous. One of two main subgroups of the AP family, including the phosphate mono- and diesterases, contains a Zn^{2+} bimetallo core [73]. The authors created "pruned" versions of the phosphate mono- and diesterase enzymes, in which the same Zn^{2+} bimetallo core was maintained and the residues that are in contact with the substrates were mutated to alanine [101]. The "pruned" enzymes sharing the same bimetallo core were shown to have highly similar phosphate mono- and diesterase, sulfatase, and phosphotriesterase activities, suggesting the electrostatic effects of the Zn^{2+} bimetallo core did not specialize this bimetallo site for different reactions [101, 102]. The authors suggested that "pruned" Zn^{2+} bimetallo core generalists would need to gain other structural requirements to evolve into specialists [101]. The authors determined that all phosphomonoesterases, but not phosphodiesterases, in the AP supefamily contain a "phosphatase helix" which forms interactions with the phosphomonoester substrates. On the other hand, the phosphodiesterases have a "diester pocket" to facilitate the binding and stabilize the phosphodiester substrates [102]. Furthermore, this Zn^{2+} bimetallo core also had a very low sulfatase activity, which is normally catalyzed by the one-metal ion sulfatase subgroup of the AP superfamily. The authors suggested that the AP sulfatases might have evolved by further remodeling the Zn^{2+} bimetallo core, maybe due to metal ligand mutations leading to the loss of the second metal ion. However, it is unclear if the ancestor of those AP enzymes is a sulfatase or a phosphoesterase [101]. Interestingly, this Zn^{2+} bimetallo core also has phosphate triesterase activity, which is a biological reaction catalyzed by the PTE members of amidohydrolase superfamily [103]. However,

the members of the AP superfamily are unable to catalyze phosphotriesterase activity because they do not have structural features to accommodate the binding of phosphate triesters, resulting in the absence of AP superfamily triesterases [101]. These studies on the AP superfamily show how enzymes adapt during their evolution to specialize their catalytic activity using a common core scaffold.

Understanding how enzyme specificity evolves is a fundamental problem in the world of biochemistry because it will help scientists determine protein structure-function relationship, predict and assign protein functions, as well as engineer proteins to catalyze desired chemistry. Understanding the molecular mechanism of how a catalytically promiscuous enzyme evolves its reaction specificity will help us understand how it evolves a new function. This dissertation discusses the fundamental basis of how an enzyme evolves a new function and reaction specificity by studying the catalytically promiscuous NSAR/OSBS subfamily of the extraordinarily divergent OSBS family. The promiscuous NASR/OSBS enzyme is a very intriguing and interesting model system because of several reasons. First, unlike some other enzymes mentioned above, the active site of NSAR/OSBS enzymes is fairly rigid and the active site loop does not contribute to the reaction specificity. Second, unlike the enzymes from the AP superfamily, PTE or PON1, which catalyzes different types of hydrolysis reactions, the promiscuous NSAR/OSBS enzymes catalyze two chemically distinct reactions (one is a syn-β elimination dehydration reaction and the other is a 1,1-proton transferring racemization reaction). And third, while the catalytic residues in the enzymes mentioned

above serve the same function in different reactions, the conserved catalytic lysines of the NSAR/OSBS enzymes have different roles in catalysis in each reaction, making it intriguing to identify the factors that can modulate the reactivity of these two catalytic lysines. Thus, understanding how the NSAR/OSBS enzymes evolve its reaction specificity will help us generally identify factors that could affect the reaction specificity in other enzymes, including other enzymes in the enolase superfamily. Ultimately, this dissertation can be used as a fundamental framework to improve reaction specificity in protein engineering and protein design methods.

Overall, this dissertation explores the fundamental mechanism of how NSAR reaction specificity evolves in the NSAR/OSBS subfamily. As mentioned earlier, reaction specificity of catalytically promiscuous NSAR/OSBS enzymes are beyond substrate binding. There must be additional residues beside the catalytic residues that are involved in enzymatic catalysis. We identified a non-active site, conserved residue R266 that is mechanistically required to contribute to NSAR reaction specificity in the promiscuous NSAR/OSBS enzymes. I will discuss the role of R266, which is mechanistically important for NSAR reaction in the catalytically promiscuous AmyNSAR/OSBS in Chapter II. I will then discuss the role of the residue R266 in other NSAR/OSBS members and in the dipeptide epimerase (DE) family of the enolase superfamily in Chapter III. I will then discuss further mechanistic characterizations of the promiscuous AmyNSAR/OSBS enzyme to understand how it is able to catalyze both reactions efficiently in Chapter IV. In Chapter V, I will discuss my contribution to the understanding of how the single mutation Y299I could introduce NSAR activity in the

non-promiscuous *Alicyclobacillus acidocaldarius* OSBS enzyme. And lastly, I will discuss the future directions of studying the evolution of the NSAR/OSBS subfamily in Chapter VI. Even though this dissertation discusses some essential residues for NSAR activity, not all of the NSAR reaction specificity determinants are identified because some NSAR/OSBS members contain those residues, yet they have little or completely lack NSAR activity. This suggests that other residues also contribute to NSAR activity. The identities of those residues remain unknown, however. Thus, it is essential to identify those unknown residues so we can completely understand the evolution of NSAR reaction specificity in the NSAR/OSBS subfamily.

References

1. O'Brien, P.J. and D. Herschlag, *Catalytic promiscuity and the evolution of new enzymatic activities.* Chem Biol, 1999. **6**(4): p. R91-R105.
2. Matsumura, I. and A.D. Ellington, *In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates.* J Mol Biol, 2001. **305**(2): p. 331-9.
3. Copley, S.D., *Enzymes with extra talents: moonlighting functions and catalytic promiscuity.* Curr Opin Chem Biol, 2003. **7**(2): p. 265-72.
4. Glasner, M.E., J.A. Gerlt, and P.C. Babbitt, *Evolution of enzyme superfamilies.* Curr Opin Chem Biol, 2006. **10**(5): p. 492-7.

5. Patrick, W.M., et al., *Multicopy suppression underpins metabolic evolvability.* Mol Biol Evol, 2007. **24**(12): p. 2716-22.

6. Kim, J., et al., *Three serendipitous pathways in E. coli can bypass a block in pyridoxal-5'-phosphate synthesis.* Mol Syst Biol, 2010. **6**: p. 436.

7. Khersonsky, O. and D.S. Tawfik, *Enzyme promiscuity: a mechanistic and evolutionary perspective.* Annu Rev Biochem, 2010. **79**: p. 471-505.

8. Baas, B.J., et al., *Recent advances in the study of enzyme promiscuity in the tautomerase superfamily.* Chembiochem, 2013. **14**(8): p. 917-26.

9. Gerlt, J.A. and P.C. Babbitt, *Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies.* Annu Rev Biochem, 2001. **70**: p. 209-46.

10. Almonacid, D.E. and P.C. Babbitt, *Toward mechanistic classification of enzyme functions.* Curr Opin Chem Biol, 2011. **15**(3): p. 435-42.

11. Baier, F., J.N. Copp, and N. Tokuriki, *Evolution of Enzyme Superfamilies: Comprehensive Exploration of Sequence-Function Relationships.* Biochemistry, 2016. **55**(46): p. 6375-6388.

12. Glasner, M.E., J.A. Gerlt, and P.C. Babbitt, *Mechanisms of protein evolution and their application to protein engineering.* Adv Enzymol Relat Areas Mol Biol, 2007. **75**: p. 193-239, xii-xiii.

13. Glasner, M.E., et al., *Evolution of structure and function in the o-succinylbenzoate synthase/N-acylamino acid racemase family of the enolase superfamily.* J Mol Biol, 2006. **360**(1): p. 228-50.

14. Gerlt, J.A., P.C. Babbitt, and I. Rayment, *Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity.* Arch Biochem Biophys, 2005. **433**(1): p. 59-70.

15. Babbitt, P.C., et al., *The enolase superfamily: a general strategy for enzyme-catalyzed abstraction of the alpha-protons of carboxylic acids.* Biochemistry, 1996. **35**(51): p. 16489-501.

16. Akiva, E., et al., *The Structure-Function Linkage Database.* Nucleic Acids Res, 2014. **42**(Database issue): p. D521-30.

17. Gulick, A.M., et al., *Evolution of enzymatic activities in the enolase superfamily: crystallographic and mutagenesis studies of the reaction catalyzed by D-glucarate dehydratase from Escherichia coli.* Biochemistry, 2000. **39**(16): p. 4590-602.

18. Rakus, J.F., et al., *Evolution of enzymatic activities in the enolase superfamily: D-Mannonate dehydratase from Novosphingobium aromaticivorans.* Biochemistry, 2007. **46**(45): p. 12896-908.

19. Rakus, J.F., et al., *Computation-facilitated assignment of the function in the enolase superfamily: a regiochemically distinct galactarate dehydratase from Oceanobacillus iheyensis.* Biochemistry, 2009. **48**(48): p. 11546-58.

20. Asuncion, M., et al., *The structure of 3-methylaspartase from Clostridium tetanomorphum functions via the common enolase chemical step.* J Biol Chem, 2002. **277**(10): p. 8306-11.

21. Levy, C.W., et al., *Insights into enzyme evolution revealed by the structure of methylaspartate ammonia lyase.* Structure, 2002. **10**(1): p. 105-13.

22. Thoden, J.B., et al., *Evolution of enzymatic activity in the enolase superfamily: structural studies of the promiscuous o-succinylbenzoate synthase from Amycolatopsis.* Biochemistry, 2004. **43**(19): p. 5716-27.

23. Schmidt, D.M., B.K. Hubbard, and J.A. Gerlt, *Evolution of enzymatic activities in the enolase superfamily: functional assignment of unknown proteins in Bacillus subtilis and Escherichia coli as L-Ala-D/L-Glu epimerases.* Biochemistry, 2001. **40**(51): p. 15707-15.

24. Zhao, S., et al., *Discovery of new enzymes and metabolic pathways by using structure and genome context.* Nature, 2013. **502**(7473): p. 698-702.

25. Sakai, A., et al., *Evolution of enzymatic activities in the enolase superfamily: stereochemically distinct mechanisms in two families of cis,cis-muconate lactonizing enzymes.* Biochemistry, 2009. **48**(7): p. 1445-53.

26. Zhu, W.W., et al., *Residues required for activity in Escherichia coli o-succinylbenzoate synthase (OSBS) are not conserved in all OSBS enzymes.* Biochemistry, 2012. **51**(31): p. 6171-81.

27. Taylor Ringia, E.A., et al., *Evolution of enzymatic activity in the enolase superfamily: functional studies of the promiscuous o-succinylbenzoate synthase from Amycolatopsis.* Biochemistry, 2004. **43**(1): p. 224-9.

28. Tokuyama, S. and K. Hatano, *Cloning, DNA sequencing and heterologous expression of the gene for thermostable N-acylamino acid racemase from*

Amycolatopsis sp. TS-1-60 in Escherichia coli. Appl Microbiol Biotechnol, 1995. **42**(6): p. 884-9.

29. Tokuyama, S. and K. Hatano, *Purification and properties of thermostable N-acylamino acid racemase from Amycolatopsis sp. TS-1-60.* Appl Microbiol Biotechnol, 1995. **42**(6): p. 853-9.

30. Palmer, D.R., et al., *Unexpected divergence of enzyme function and sequence: "N-acylamino acid racemase" is o-succinylbenzoate synthase.* Biochemistry, 1999. **38**(14): p. 4252-8.

31. Sakai, A., et al., *Evolution of enzymatic activities in the enolase superfamily: N-succinylamino acid racemase and a new pathway for the irreversible conversion of D- to L-amino acids.* Biochemistry, 2006. **45**(14): p. 4455-62.

32. Thompson, T.B., et al., *Evolution of enzymatic activity in the enolase superfamily: structure of o-succinylbenzoate synthase from Escherichia coli in complex with Mg2+ and o-succinylbenzoate.* Biochemistry, 2000. **39**(35): p. 10662-76.

33. Klenchin, V.A., et al., *Evolution of enzymatic activity in the enolase superfamily: structural and mutagenic studies of the mechanism of the reaction catalyzed by o-succinylbenzoate synthase from Escherichia coli.* Biochemistry, 2003. **42**(49): p. 14427-33.

34. Nagatani, R.A., et al., *Stability for function trade-offs in the enolase superfamily "catalytic module".* Biochemistry, 2007. **46**(23): p. 6688-95.

35. Taylor, E.A., D.R. Palmer, and J.A. Gerlt, *The lesser "burden borne" by o-succinylbenzoate synthase: an "easy" reaction involving a carboxylate carbon acid.* J Am Chem Soc, 2001. **123**(24): p. 5824-5.

36. Odokonyero, D., et al., *Divergent evolution of ligand binding in the o-succinylbenzoate synthase family.* Biochemistry, 2013. **52**(42): p. 7512-21.

37. Odokonyero, D., et al., *Loss of quaternary structure is associated with rapid sequence divergence in the OSBS family.* Proc Natl Acad Sci U S A, 2014. **111**(23): p. 8535-40.

38. Odokonyero, D., et al., *Comparison of Alicyclobacillus acidocaldarius o-Succinylbenzoate Synthase to Its Promiscuous N-Succinylamino Acid Racemase/ o-Succinylbenzoate Synthase Relatives.* Biochemistry, 2018. **57**(26): p. 3676-3689.

39. Brizendine, A.M., et al., *Promiscuity of Exiguobacterium sp. AT1b o-succinylbenzoate synthase illustrates evolutionary transitions in the OSBS family.* Biochem Biophys Res Commun, 2014. **450**(1): p. 679-84.

40. Copley, S.D., *Shining a light on enzyme promiscuity.* Curr Opin Struct Biol, 2017. **47**: p. 167-175.

41. Baier, F. and N. Tokuriki, *Connectivity between catalytic landscapes of the metallo-beta-lactamase superfamily.* J Mol Biol, 2014. **426**(13): p. 2442-56.

42. Huang, H., et al., *Panoramic view of a superfamily of phosphatases through substrate profiling.* Proc Natl Acad Sci U S A, 2015. **112**(16): p. E1974-83.

43. Burroughs, A.M., et al., *Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes.* J Mol Biol, 2006. **361**(5): p. 1003-34.

44. Wang, J., et al., *Exploring the Promiscuity of Phenol Hydroxylase from Pseudomonas stutzeri OX1 for the Biosynthesis of Phenolic Compounds.* ACS Synth Biol, 2018. **7**(5): p. 1238-1243.

45. Jerlstrom Hultqvist, J., et al., *A bacteriophage enzyme induces bacterial metabolic perturbation that confers a novel promiscuous function.* Nat Ecol Evol, 2018. **2**(8): p. 1321-1330.

46. Yang, K. and W.W. Metcalf, *A new activity for an old enzyme: Escherichia coli bacterial alkaline phosphatase is a phosphite-dependent hydrogenase.* Proc Natl Acad Sci U S A, 2004. **101**(21): p. 7919-24.

47. Fisher, M.A., et al., *De novo designed proteins from a library of artificial sequences function in Escherichia coli and enable cell growth.* PLoS One, 2011. **6**(1): p. e15364.

48. Digianantonio, K.M. and M.H. Hecht, *A protein constructed de novo enables cell growth by altering gene regulation.* Proc Natl Acad Sci U S A, 2016. **113**(9): p. 2400-5.

49. Digianantonio, K.M., M. Korolev, and M.H. Hecht, *A Non-natural Protein Rescues Cells Deleted for a Key Enzyme in Central Metabolism.* ACS Synth Biol, 2017. **6**(4): p. 694-700.

50. Bastard, K., et al., *Revealing the hidden functional diversity of an enzyme family.* Nat Chem Biol, 2014. **10**(1): p. 42-9.

51. Colin, P.Y., et al., *Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics.* Nat Commun, 2015. **6**: p. 10008.

52. Gielen, F., et al., *Ultrahigh-throughput-directed enzyme evolution by absorbance-activated droplet sorting (AADS).* Proc Natl Acad Sci U S A, 2016. **113**(47): p. E7383-E7389.

53. Macchiarulo, A., I. Nobeli, and J.M. Thornton, *Ligand selectivity and competition between enzymes in silico.* Nat Biotechnol, 2004. **22**(8): p. 1039-45.

54. Freund, G.S., et al., *Elucidating Substrate Promiscuity within the FabI Enzyme Family.* ACS Chem Biol, 2017. **12**(9): p. 2465-2473.

55. Tokuriki, N. and D.S. Tawfik, *Protein dynamism and evolvability.* Science, 2009. **324**(5924): p. 203-7.

56. Pabis, A., et al., *Cooperativity and flexibility in enzyme evolution.* Curr Opin Struct Biol, 2018. **48**: p. 83-92.

57. Muralidhara, B.K., et al., *Thermodynamic fidelity of the mammalian cytochrome P450 2B4 active site in binding substrates and inhibitors.* J Mol Biol, 2008. **377**(1): p. 232-45.

58. Skopalik, J., P. Anzenbacher, and M. Otyepka, *Flexibility of human cytochromes P450: molecular dynamics reveals differences between CYPs 3A4, 2C9, and 2A6, which correlate with their substrate preferences.* J Phys Chem B, 2008. **112**(27): p. 8165-73.

59. Ortlund, E.A., et al., *Crystal structure of an ancient protein: evolution by conformational epistasis.* Science, 2007. **317**(5844): p. 1544-8.

60. Thomson, J.M., et al., *Resurrecting ancestral alcohol dehydrogenases from yeast.* Nat Genet, 2005. **37**(6): p. 630-5.

61. Voordeckers, K., et al., *Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication.* PLoS Biol, 2012. **10**(12): p. e1001446.

62. Thornton, J.W., *Resurrecting ancient genes: experimental analysis of extinct molecules.* Nat Rev Genet, 2004. **5**(5): p. 366-75.

63. Jensen, R.A., *Enzyme recruitment in evolution of new function.* Annu Rev Microbiol, 1976. **30**: p. 409-25.

64. Risso, V.A., et al., *Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian beta-lactamases.* J Am Chem Soc, 2013. **135**(8): p. 2899-902.

65. Zou, T., et al., *Evolution of conformational dynamics determines the conversion of a promiscuous generalist into a specialist enzyme.* Mol Biol Evol, 2015. **32**(1): p. 132-43.

66. James, L.C. and D.S. Tawfik, *Conformational diversity and protein evolution--a 60-year-old hypothesis revisited.* Trends Biochem Sci, 2003. **28**(7): p. 361-8.

67. James, L.C., P. Roversi, and D.S. Tawfik, *Antibody multispecificity mediated by conformational diversity.* Science, 2003. **299**(5611): p. 1362-7.

68. Barona-Gomez, F. and D.A. Hodgson, *Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis.* EMBO Rep, 2003. **4**(3): p. 296-300.

69. Due, A.V., et al., *Bisubstrate specificity in histidine/tryptophan biosynthesis isomerase from Mycobacterium tuberculosis by active site metamorphosis.* Proc Natl Acad Sci U S A, 2011. **108**(9): p. 3554-9.

70. Newton, M.S., et al., *Structural and functional innovations in the real-time evolution of new (betaalpha)8 barrel enzymes.* Proc Natl Acad Sci U S A, 2017. **114**(18): p. 4727-4732.

71. Mohamed, M.F. and F. Hollfelder, *Efficient, crosswise catalytic promiscuity among enzymes that catalyze phosphoryl transfer.* Biochim Biophys Acta, 2013. **1834**(1): p. 417-24.

72. van Loo, B., et al., *An efficient, multiply promiscuous hydrolase in the alkaline phosphatase superfamily.* Proc Natl Acad Sci U S A, 2010. **107**(7): p. 2740-5.

73. Jonas, S. and F. Hollfelder, *Mapping catalytic promiscuity in the alkaline phosphatase superfamily.* Pure and Applied Chemistry, 2009. **81**(4): p. 731-742.

74. Barrozo, A., et al., *Cooperative Electrostatic Interactions Drive Functional Evolution in the Alkaline Phosphatase Superfamily.* J Am Chem Soc, 2015. **137**(28): p. 9061-76.

75. Pabis, A. and S.C. Kamerlin, *Promiscuity and electrostatic flexibility in the alkaline phosphatase superfamily.* Curr Opin Struct Biol, 2016. **37**: p. 14-21.

76. Peracchi, A., *The Limits of Enzyme Specificity and the Evolution of Metabolism.* Trends Biochem Sci, 2018. **43**(12): p. 984-996.

77. Fersht, A., *Structure and mechanism in protein science : a guide to enzyme catalysis and protein folding*. 1999, New York: W.H. Freeman. xxi, 631 p.

78. Fersht, A.R., *Catalysis, binding and enzyme-substrate complementarity.* Proc R Soc Lond B Biol Sci, 1974. **187**(1089): p. 397-407.

79. Miton, C.M., et al., *Evolutionary repurposing of a sulfatase: A new Michaelis complex leads to efficient transition state charge offset.* Proc Natl Acad Sci U S A, 2018. **115**(31): p. E7293-E7302.

80. Wang, Y., et al., *Peroxygenases en route to becoming dream catalysts. What are the opportunities and challenges?* Curr Opin Chem Biol, 2017. **37**: p. 1-9.

81. Ramirez-Escudero, M., et al., *Structural Insights into the Substrate Promiscuity of a Laboratory-Evolved Peroxygenase.* ACS Chem Biol, 2018. **13**(12): p. 3259-3268.

82. Ben-David, M., et al., *Catalytic versatility and backups in enzyme active sites: the case of serum paraoxonase 1.* J Mol Biol, 2012. **418**(3-4): p. 181-96.

83. Blaha-Nelson, D., et al., *Active Site Hydrophobicity and the Convergent Evolution of Paraoxonase Activity in Structurally Divergent Enzymes: The Case of Serum Paraoxonase 1.* J Am Chem Soc, 2017. **139**(3): p. 1155-1167.

84. Afriat-Jurnou, L., C.J. Jackson, and D.S. Tawfik, *Reconstructing a missing link in the evolution of a recently diverged phosphotriesterase by active-site loop remodeling.* Biochemistry, 2012. **51**(31): p. 6047-55.

85. Benning, M.M., et al., *High resolution X-ray structures of different metal-substituted forms of phosphotriesterase from Pseudomonas diminuta.* Biochemistry, 2001. **40**(9): p. 2712-22.

86. Elias, M., et al., *Structural basis for natural lactonase and promiscuous phosphotriesterase activities.* J Mol Biol, 2008. **379**(5): p. 1017-28.

87. Fushinobu, S., et al., *Structural basis for the bifunctionality of fructose-1,6-bisphosphate aldolase/phosphatase.* Nature, 2011. **478**(7370): p. 538-41.

88. Nishimasu, H., et al., *The first crystal structure of the novel class of fructose-1,6-bisphosphatase present in thermophilic archaea.* Structure, 2004. **12**(6): p. 949-59.

89. Gerlt, J.A., et al., *Divergent evolution in enolase superfamily: strategies for assigning functions.* J Biol Chem, 2012. **287**(1): p. 29-34.

90. Schmidt, D.M., et al., *Evolutionary potential of (beta/alpha)8-barrels: functional promiscuity produced by single substitutions in the enolase superfamily.* Biochemistry, 2003. **42**(28): p. 8387-93.

91. Vick, J.E., D.M. Schmidt, and J.A. Gerlt, *Evolutionary potential of (beta/alpha)8-barrels: in vitro enhancement of a "new" reaction in the enolase superfamily.* Biochemistry, 2005. **44**(35): p. 11722-9.

92. Vick, J.E. and J.A. Gerlt, *Evolutionary potential of (beta/alpha)8-barrels: stepwise evolution of a "new" reaction in the enolase superfamily.* Biochemistry, 2007. **46**(50): p. 14589-97.

93. Lukk, T., et al., *Homology models guide discovery of diverse enzyme specificities among dipeptide epimerases in the enolase superfamily.* Proc Natl Acad Sci U S A, 2012. **109**(11): p. 4122-7.

94. Bourque, J.R. and S.L. Bearne, *Mutational analysis of the active site flap (20s loop) of mandelate racemase.* Biochemistry, 2008. **47**(2): p. 566-78.

95. McMillan, A.W., et al., *Role of an active site loop in the promiscuous activities of Amycolatopsis sp. T-1-60 NSAR/OSBS.* Biochemistry, 2014. **53**(27): p. 4434-44.

96. Estrada, P., et al., *A Single Amino Acid Switch Alters the Isoprene Donor Specificity in Ribosomally Synthesized and Post-Translationally Modified Peptide Prenyltransferases.* J Am Chem Soc, 2018. **140**(26): p. 8124-8127.

97. Furst, M., et al., *Side-Chain Pruning Has Limited Impact on Substrate Preference in a Promiscuous Enzyme.* ACS Catal, 2018. **8**(12): p. 11648-11656.

98. Lahiri, S.D., et al., *Analysis of the substrate specificity loop of the HAD superfamily cap domain.* Biochemistry, 2004. **43**(10): p. 2812-20.

99. Morais, M.C., et al., *The crystal structure of bacillus cereus phosphonoacetaldehyde hydrolase: insight into catalysis of phosphorus bond cleavage and catalytic diversification within the HAD enzyme superfamily.* Biochemistry, 2000. **39**(34): p. 10385-96.

100. Pandya, C., et al., *Enzyme promiscuity: engine of evolutionary innovation.* J Biol Chem, 2014. **289**(44): p. 30229-36.

101. Sunden, F., et al., *Differential catalytic promiscuity of the alkaline phosphatase superfamily bimetallo core reveals mechanistic features underlying enzyme evolution.* J Biol Chem, 2017. **292**(51): p. 20960-20974.

102. Sunden, F., et al., *Mechanistic and Evolutionary Insights from Comparative Enzymology of Phosphomonoesterases and Phosphodiesterases across the Alkaline Phosphatase Superfamily.* J Am Chem Soc, 2016. **138**(43): p. 14273-14287.

103. Seibert, C.M. and F.M. Raushel, *Structural and catalytic diversity within the amidohydrolase superfamily.* Biochemistry, 2005. **44**(17): p. 6383-91.

CHAPTER II

ROLE OF THE SECOND-SHELL AMINO ACID R266 IN DETERMINING *N*-SUCCINYLAMINO ACID RACEMASE REACTION SPECIFICITY

Introduction

Second-shell amino acids are residues that do not contact the substrates, but rather interact with the residues that are in direct contact with the substrates. Several studies showed that non-active site amino acids are often important for enzymes to achieve optimal catalysis via their structural or electrostatic effects on the active site [1-5]. For example, studies of the metallo-β-lactamase enzyme showed that non-active site amino acids could impact the enzyme's activity by modulating the protein dynamics, flexibility and stability [3-5]. In other enzymes, second-shell amino acids could contribute to enzyme catalysis by modulating the binding and reactivity of a catalytic metal ion [6-9] or the reactivity of a coenzyme [10]. Other studies on trypsin and chymotrypsin show that non-active site amino acids are responsible for substrate recognition and substrate specificity determination in those enzymes [11-13]. In those studies, the second-shell amino acid substitutions that alter the enzyme functions often have no effect on the positions of the catalytic residues. Identifying non-active site residues that are responsible for enzyme activity and substrate specificity can be valuable for fully understanding the mechanistic bases for reaction specificity and

evolution new functions. However, determining which second-shell amino acids or non-active site residues are important for enzyme function and specificity is not an easy task.

One strategy for identifying non-active site residues that influence catalysis is to examine effects of mutations on specificity of catalytically promiscuous enzymes. Catalytic promiscuity refers to the ability of an enzyme to catalyze alternative reactions in the same active site that the enzyme uses for its biological function [14, 15]. Because all reactions of promiscuous enzymes generally rely on the same structural scaffold, non-active site mutations that affect overall structure and stability are likely to have similar effects on all activities. But mutations that affect the catalytic mechanism or ligand binding are more likely to alter the relative specificity for each reaction. Many lines of evidence show that enzymes often transition through catalytically promiscuous intermediates to evolve a new function [14-21]. Thus, identifying sites that affect specificity provides information about the mechanisms and potential for evolving new enzyme activities.

This study investigates the role of a second-shell amino acid in a catalytically promiscuous enzyme from the *N*-succinylamino acid racemase/*o*-succinylbenzoate synthase (NSAR/OSBS) subfamily (Figure II.1). This subfamily is a branch of a larger family of OSBS enzymes, which belongs to the functionally diverse enolase superfamily [22, 23]. Most enzymes in the OSBS family catalyze a step in the menaquinone biosynthesis pathway and are divided into several large, divergent subfamilies, which correspond to the phylum from which the OSBS enzymes originated (Figure II.2) [24]. Phylogenetic and genome context analysis in combination with biochemical

characterization of NSAR/OSBS enzymes indicated that NSAR activity in the NSAR/OSBS subfamily evolved through promiscuous intermediates [25-27]. While the OSBS enzyme from *Staphylococcus aureus*, the earliest branch of the NSAR/OSBS subfamily that diverges from other OSBS subfamilies, completely lacks NSAR activity [28], the OSBS enzyme from *Exiguobacterium* sp. AT1b, which diverged slightly later, inefficiently and promiscuously catalyzes NSAR activity [26]. Most NSAR/OSBS enzymes that diverged after *Exiguobacterium* sp. AT1b OSBS catalyze the NSAR reaction with hihger efficiency, and NSAR activity is a biological function in many of them. For example, the NSAR/OSBS from *Geobacillus kaustophilus* is bifunctional, requiring OSBS activity for the menaquinone biosynthesis and NSAR activity for a pathway that converts D-amino acid to L-amino acid [29]. For many of the later diverging NSAR/OSBS subfamily enzymes, genome context indicates that NSAR or another activity is their biological function, but they retain OSBS activity as a promiscuous side reaction [23]. Because NSAR activity evolved through promiscuous intermediates and is only known to occur in one branch of the OSBS subfamily, we hypothesized that this subfamily has unique, pre-adaptive sequence or structural features that enabled NSAR activity to evolve.

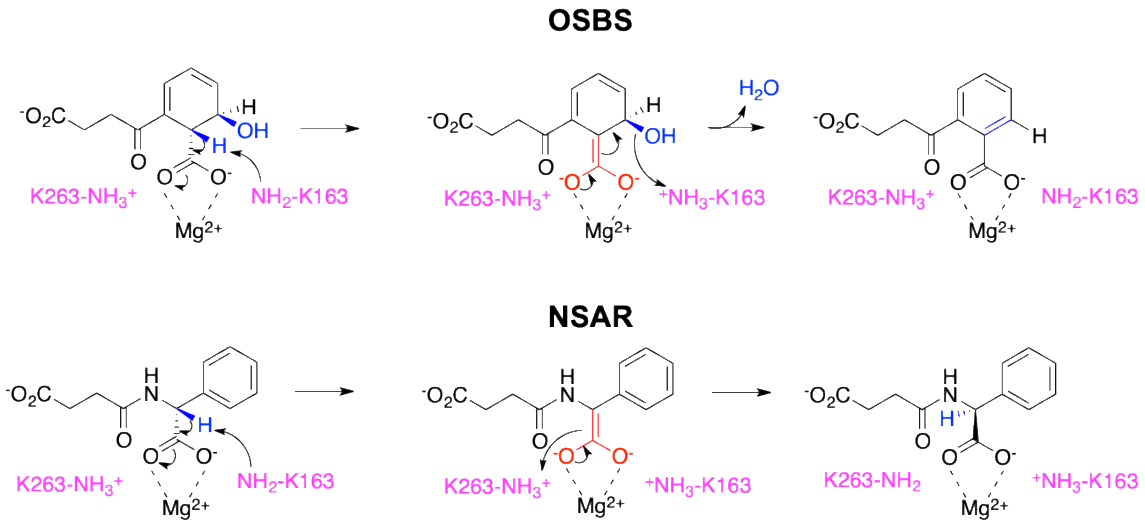


Figure II.1. The mechanisms of *o*-succinylbenzoate synthase (OSBS) and *N*-succinylamino acid racemase (NSAR) reactions. The divalent metal ion-stabilized enolate intermediate shared by the two reactions is shown in red. The atoms that are rearranged or lost during catalysis are shown in blue. The catalytic lysines shared by the two reactions are shown in magenta (numbering is relative to AmyNSAR/OSBS).

Comparison of the promiscuous *Amycolatopsis* sp. T-1-60 NSAR/OSBS (AmyNSAR/OSBS) to two OSBSs from other subfamilies that lack NSAR activity suggests that orientation of ligand binding is one of the pre-adaptive structural features [30, 31]. Specifically, in AmyNSAR/OSBS, the β8 strand and the loop connecting it to the capping domain are shifted, creating a cavity that can accommodate the extended conformation of the succinyl group of the substrates of the OSBS and NSAR reactions. In contrast, *Escherichia coli* OSBS (EcOSBS), which shares <25% sequence identity with AmyNSAR/OSBS, does not have this feature to accommodate the binding of the NSAR substrate [30, 31]. Both EcOSBS and *Thermobifida fusca* OSBS (TfuscaOSBS) have OSB bound with the succinyl moiety in a bent conformations [31, 32]. On the other

hand, the succinyl moiety of OSB and *N*-succinylphenylglycine (NSPG, an NSAR substrate) are in an extended conformation in AmyNSAR/OSBS [30]. However, the binding accommodation does not appear to be sufficient to explain the ability of AmyNSAR/OSBS to efficiently catalyze both reactions. Some NSAR/OSBS subfamily members share this structural feature with AmyNSAR/OSBS, but they have low or undetectable NSAR activity. For example, both docking and inhibition studies of *Alicyclobacillus acidocaldarius* OSBS, which is a nonpromiscuous member of the NSAR/OSBS subfamily, show that NSPG can bind to the active site of the enzyme even though the enzyme is unable to carry out the racemization reaction [27]. This suggests that other factors beyond substrate binding contribute to catalytic promiscuity and bifunctionality of NSAR/OSBS enzymes.

The OSBS and NSAR reactions require two catalytic lysine residues in the active site (Figure II.1). In the OSBS reaction, only K163 acts as the general acid and base. K163 abstracts the H_α of 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate (SHCHC) to create an enolate intermediate, which is stabilized by the metal ion in the active site. Protonated K163 then donates the proton to the hydroxyl leaving group from the intermediate to form water in a syn-β elimination manner, producing OSB. K263 probably forms a π-cation interaction with the SHCHC and the enolate intermediate. Thus, K263 needs to be protonated in the OSBS reaction. This π-cation interaction in the OSBS reaction is similar to the π-cation interaction between the catalytic residues and the substrate/enolate intermediate in mandelate racemase, another member of the enolase superfamily [33]. In contrast, the NSAR reaction follows a two-base, 1-1 proton transfer

mechanism. K163 is the general base and K263 is the general acid when a D-succinyl amino acid is the substrate. The roles of the lysines are reversed when an L-succinyl amino acid is the substrate. The crystal structure of *Amycolatopsis* NSAR/OSBS shows that the two catalytic lysines are oriented on opposite sides of the active site, pointing to opposite faces of the substrate's C_α and enabling them to act as the general acid and base [30, 34].

Because K263 has different roles in the OSBS and NSAR reactions, we hypothesized that the local environment modulates the reactivity of this catalytic lysine so it can participate in both OSBS and NSAR reactions. To identify amino acids that could affect the local environment around K263, we compared the sequence conservation of the NSAR/OSBS subfamily to that of other OSBS subfamilies whose members lack NSAR activity. The sequence logos constructed by WebLogo indicate that two second-shell amino acids, N261 and R266, are conserved only in the NSAR/OSBS subfamily, but nonpolar residues are typically at these positions in other OSBS subfamilies (Figure II.2) [35]. N261 and R266 are located near K263 and are not in contact with the substrate. Thus, they might have an effect on the reaction mechanism and/or the reaction specificity, without affecting the binding of the substrates. We hypothesized that these conserved amino acids in the NSAR/OSBS subfamily were pre-adaptive, enabling the evolution of NSAR activity. In this work, we elucidated the roles of N261 and R266 in AmyNSAR/OSBS. We discovered that while N261 primarily influences structural stability, R266 directly influences the reactivity of K263. Thus, R266, which is conserved in all members of the NSAR/OSBS subfamily, regardless of

their ability to catalyze the NSAR reaction, is likely to be a pre-adaptive structural feature that enabled the emergence and evolution of NSAR activity in the NSAR/OSBS subfamily.

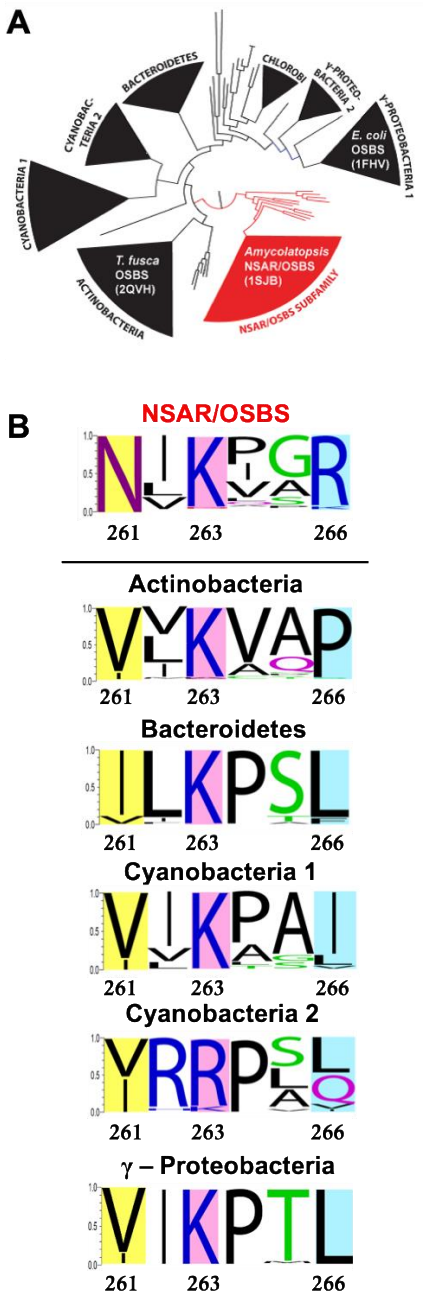


Figure II.2. (A) Phylogenetic tree of the OSBS family, illustrating the division of the family into several subfamilies, which primarily correspond to the phylum from which the OSBS enzymes originated [24]. (B) Sequence logos showing the conservation of amino acids at positions 261 and 266 in different OSBS subfamilies [35]. The letter size is proportional to the frequency of the amino acid at that position in the sequence alignment. Sequence numbering is relative to the AmyNSAR/OSBS. N261 and R266 are highlighted in yellow and cyan, respectively. The catalytic K263 is highlighted in pink.

Materials and Methods

Mutagenesis

Site-directed mutagenesis was performed using the Q5 mutagenesis protocol (New England BioLabs). The template for mutagenesis was the gene encoding *Amycolatopsis* sp. T-1-60 NSAR/OSBS (UniProt entry: Q44244), which was cloned into a pET17b vector (a gift from J.A Gerlt, University of Illinois, Urbana, IL). Mutations were confirmed by sequencing in both directions (Eurofins Genomics LLC). The primers used for *Amycolatopsis* NSAR/OSBS mutagenesis were designed using NEBaseChanger, NEB's online design software (NEBasechanger.com) and are shown in Table II.1.

Table II.1. Primers used for mutagenesis at position 266 and 261 in *Amycolatopsis* NSAR/OSBS

Mutation	Primer sequence
R266Q	Forward: CAAACCGGGCcagGTCGGCGGGT Reverse: ATGTTCACGATTTGGACCGC
R266K	Forward: CAAACCGGGCaaaGTCGGCGGGT Reverse: ATGTTCACGATTTGGACCG

Table II.1. Continued.

Mutation	Primer sequence
N261L	Forward: CCAAATCGTGctgATCAAACCGGGCCGC Reverse: ACCGCGCCCAGCTTGATG
All DNA primers are shown in the 5' to 3' direction. The underlined bases designate the codons where mutations were introduced.	

Protein Production

Proteins were expressed in *E. coli* strain BW25113 (*menC::kan*, DE3) to ensure that OSBS from the host cell would not contaminate the purified proteins expressed from the plasmid [24]. Cultures were grown for 48 h at 30 °C in LB media containing carbenicillin and kanamycin at a final concentration of 50 μg/mL, then harvested by centrifugation. The cell pellet was resuspended in 20 mM Tris (pH 8.0), 5 mM $MgCl_2$, 0.4 mM phenylmethylsulfonyl fluoride (PMSF), and 10 μg/mL DNase I. Resuspended cells were lysed by sonication. The supernatant was collected after centrifugation and filtered using a 0.22 μm Steriflip filter (Millipore). The protein was loaded onto a 20 mL HiPrep 16/10 DEAE FF column (GE Healthcare). The protein was eluted using a buffer containing 20 mM Tris (pH 8.0), 5 mM $MgCl_2$, and 500 mM NaCl with an initial step at 30% elution buffer for 6 column volumes to elute loosely bound proteins, followed by a linear gradient from 30% to 65% elution buffer over 10 column volumes. Fractions

containing *Amycolatopsis* NSAR/OSBS variants were identified by SDS-PAGE. The fractions were combined with $(NH_4)_2SO_4$, at a final concentration of 0.4 M and then applied to three 5 mL HiTrap Phenyl FF (low sub) columns (GE Healthcare) attached in tandem and equilibrated with a buffer containing 20 mM Tris (pH 8.0), 5 mM $MgCl_2$, and 0.4 M $(NH_4)_2SO_4$. The protein was eluted with 20 mM Tris (pH 8.0) and 5 mM $MgCl_2$ using a linear gradient from 0% to 100% elution buffer over 15 column volumes. Fractions containing AmyNSAR/OSBS and variants were identified by SDS-PAGE, exchanged into storage buffer (20 mM Tris (pH 8.0) and 5 mM $MgCl_2$), and concentrated using a Vivaspin Turbo 15 centrifuge filter with a 10 kDa molecular weight cutoff (Sartorius). Glycerol was added to a final concentration of 25%, and the purified proteins were stored at -20 °C.

OSBS Assay

2-Succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate (SHCHC) was synthesized from chorismate and α-ketoglutarate as described previously [24]. The enzymes were assayed in 50 mM Tris (pH 8.0) and 0.1 mM $MnCl_2$ with various SHCHC concentrations. The reactions were monitored by a SpectraMax Plus384 plate reader (Molecular Devices) at 310 nm and at 25 °C. The disappearance of SHCHC ($\Delta\varepsilon$ = -2400 M^{-1} cm^{-1}) was measured as a function of time [33, 35]. The initial rates were determined by fitting the linear portion of the data in Microsoft Excel, and the initial rates at

different substrate concentrations were fit to the Michaelis-Menten equation using Prism (GraphPad).

NSAR Assay

L- and D-*N*-Succinylphenylglycine (L- and D-NSPG) were synthesized as described previously [36]. The enzyme was assayed in 200 mM Tris (pH 8.0) and 0.1 mM $MnCl_2$ with various L- or D-*N*-succinylphenylglycine concentrations. The reactions were carried out in a sample cell with a 5 cm path length. The change in the optical rotation of the substrate was monitored by a Jasco P-2000 polarimeter at 405 nm and 25 °C. Measurements were taken using a 10 s integration time and reading every 30 seconds. The specific rotation value at 405 nm of L-NSPG is 6.54 deg M^{-1} cm^{-1} and of D-NSPG is 6.22 deg M^{-1} cm^{-1} [36]. The rates were determined as described above.

Isotopic Exchange Experiments Using ^1H NMR Spectroscopy

AmyNSAR/OSBS variants were exchanged into 20 mM Tris buffer (pD 8.0 with NaOD) using a Vivaspin Turbo 15 centrifuge filter (Sartorius). A 10 mL aliquot of protein was concentrated to 1 mL; 9 mL of 20 mM Tris buffer (pD 8.0) was added, and the protein solution was again concentrated to 1 mL. This process was repeated three times to maximize the exchange. Each reaction contained 20 mM L- or D-NSPG (pD 8.0 with NaOD), 50 mM Tris (pD 8.0), 0.1 mM $MnCl_2$, and AmyNSAR/OSBS variants in

D_2O. The intensity of the α-proton was monitored as it was exchanged with deuterium over time by ^1H NMR (500 MHz Bruker NMR spectrometer). The peak of the α-proton (δ = 5.15 ppm) was integrated relative to that of the five aromatic protons (δ = 7.40 ppm). The relative peak area was converted to concentration based on the initial substrate concentration. The slope of the plots of the NSPG substrate concentration as a function of time was fit to a line to obtain the isotopic exchange rates (k_{ex}).

Differential Scanning Fluorimetry (DSF)

Thermal stability of *Amycolatopsis* NSAR/OSBS variants were determined by DSF using a CFX96 real-time PCR (RT-PCR, Bio-Rad). All samples contained SYPRO Orange (Sigma-Aldrich) at a dilution of 1:1250, 0.5 mg/mL of protein, 25 mM HEPES (pH 7.5), 300 mM NaCl, 10 mM DTT, and 10 mM EDTA in a final volume of 50 µL. SHCHC or NSPG were added to some samples. All samples were run in quadruplicate in a 96-well RT-PCR plate (VWR). The RT-PCR machine was programmed to raise temperature from 25 °C to 99 °C every 1 °C/min and the fluorescent intensity was measured every 1 °C (excitation at 470 nm/ emission at 570 nm, [39]). Data analysis and the unfolding transition (T_m) were determined by using the DSF analysis protocol as described in ref [39]. Briefly, the raw data was fitted to a 4th polynomial equation, and the T_m value was then calculated by solving the 2nd derivative by using the quadratic equation. T_m measurements from quadruplicate samples were averaged and the standard error was calculated.

Crystallization, Data Collection, and Structure Determination

Diffraction quality crystals were obtained by mixing 3 μL of 10 mg/mL protein and 20 mM NSPG with 3 μL of 28% PEG 4000 and 0.1 M Tris (pH 8.0) and equilibrating by vapor diffusion against 1 mL of the same precipitant at 16 °C. The resulting crystals were cryoprotected and flash-cooled in liquid nitrogen. The diffraction data was collected at the 19ID beamline of the Advanced Photon Source at a wavelength of 0.979 Å. The data was indexed using HKL 2000 software suite in $H3_2$ space group. HKL 2000 was also used for integration and scaling. A maximum resolution of 2.9 Å was used during scaling. The structure was phased by molecular replacement using Phenix Phaser. The template for molecular replacement was a single protein chain of AmyNSAR/OSBS WT with OSB bound (PDB 1SJB, [30]). The model was then refined with alternating cycles of refinement using Phenix Refine and manual model building in COOT. The substrate was subsequently added to the structure based on observed density in the F_0-F_c map. Crystallographic data collection and refinement statistics for AmyNSAR/OSBS R266Q are listed in Table II.2.

Table II.2. Crystallographic data collection and refinement statistics for AmyNSAR/OSBS R266Q*

Data Collection	
space group	$R3_2$:H
cell dimensions	a = 213.74 Å b = 213.74 Å c = 253.078 Å $\alpha = 90^{\circ}$ $\beta = 90^{\circ}$ $\gamma = 120^{\circ}$
no. of molecules per ASU	4
wavelength (Å)	0.97918
resolution (Å)	48.82 - 2.904 (3.008 - 2.904)
no. of unique reflections	49028 (4865)
completeness (%)	99.90 (100.00)
R_{merge}	0.162
$\langle I/\sigma I \rangle$	11.33
redundancy	3.5
Refinement	
R	0.1867 (0.2741)
R_{free}	0.2247 (0.3202)

Table II.2. Continued.

Refinement (continued)	
no. of non-hydrogen atoms	11002
protein	10886
ligand	76
solvent	40
average *B* factor (Å^2)	62.56
protein	62.53
ligand	74.25
solvent	49.42
Ramachandran statistics (%)	
favored	96.50
allowed	3.50
outlier	0.00
root-mean-square deviation from ideal	
bonds (Å)	0.003
angles (deg)	0.57
*Statistics for the highest-resolution shell are shown in parentheses.	

Results

Because N261 and R266 are conserved in the NSAR/OSBS subfamily but not in other OSBS subfamilies, mutating N261 and R266 was predicted to decrease NSAR activity and have less effect on OSBS activity of AmyNSAR/OSBS. Initially, we constructed the mutations N261V and R266L mutations which mimic the hydrophobicity of homologous positions in *Escherichia coli* OSBS and other OSBS members of the γ-Proteobacteria 1 subfamily. Both N261V and R266L mutations severely decreased the protein expression levels (~120-fold and ~40-fold less than the WT enzyme, respectively). Therefore, we constructed the mutations N261L and R266Q, which had better protein yields.

Effects of mutating of N261 on enzyme activity and stability

The N261L mutation unexpectedly decreased OSBS activity ~320-fold and had a marginal effect on NSAR activity (Table II.3). Thus, the N261 residue is not an NSAR reaction specificity determinant, contradicting our prediction.

Table II.3. Kinetic constants for AmyNSAR/OSBS wild type and mutants

	OSBS			**NSAR** [b]			
	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)	Relative specificity $(k_{cat}/K_M)^{OSBS}/(k_{cat}/K_M)^{NSAR}$
WT	365 ± 60	42 ± 3	(1.2 ± 0.1) × 10^5	1500 ± 100	99 ± 5	(6.4 ± 0.3) × 10^4	1.8
N261L	n.d.[a]	n.d.[a]	(3.8 ± 0.2) × 10^2	2200 ± 200	37 ± 2	(1.7 ± 0.1) × 10^4	0.02
R266Q	359 ± 30	6 ± 0.1	(1.6 ± 0.1) × 10^4	1300 ± 10	0.21 ± 0.01	(1.6 ± 0.1) × 10^2	99
R266K	182 ± 10	4.2 ± 0.01	(2.3 ± 0.1) × 10^4	*n.d.*	*n.d.*	(9.3 ± 0.2) × 10^3	2.5

[a]Not determined because substrate saturation could not be achieved.

[b]D-NSPG was the substrate

Because the protein yield of the AmyNSAR/OSBS N261L is ~60-fold lower than the wild type, we measured the thermostability of the AmyNSAR/OSBS variants using differential scanning fluorimetry (DSF). DSF was previously described as a good method to measure the melting temperature (T_m) of proteins [40]. AmyNSAR/OSBS is an octameric protein, and although we did not observe multiple unfolding transitions with the DSF experiments, the T_m should be considered apparent. The T_m value of the

N261L mutant is 5 °C lower than the WT protein's T_m, suggesting that the mutation affects the stability of the protein (Table II.4). However, it still has substantial NSAR activity, even though its OSBS activity is severely decreased. A possible explanation for this is that NSPG binding stabilize the protein. The binding of SHCHC and L-NSPG stabilizes the WT enzyme by 1.3 and 1.1 °C, respectively. The binding of L-NSPG (up to 10 mM) has no effect on the melting temperature of AmyNSAR/OSBS N261L. In contrast, the T_m value dropped about 2 °C in the presence of 2 mM SHCHC, suggesting SHCHC is destabilizing, potentially explaining why the N261L mutation decreased OSBS activity more than NSAR activity. Even though the genome of *Amycolatopsis* sp. TS-1-60 has not been sequenced, AmyNSAR/OSBS shares 83% sequence identity with the NSAR/OSBS from *A. mediterranei* S699, in which the genome context indicates that its NSAR activity is the only biological function [27]. Because of the high sequence similarity of the NSAR/OSBS enzymes of the *Amycolatopsis* species, the NSAR activity is inferred to be the biological function while the OSBS activity is the promiscuous side reaction. We speculate that NSAR activity was less affected in the AmyNSAR/OSBS N261L because NSAR is inferred to be the biological function of this enzyme and the enzyme is optimized to bind *N*-succinylamino acids.

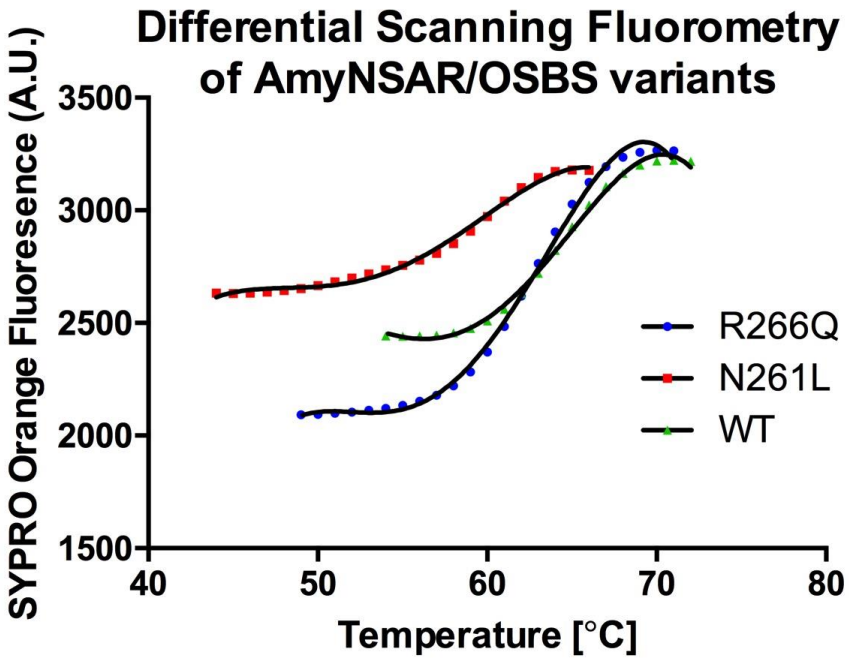


Figure II.3. Differential scanning fluorimetry of AmyNSAR/OSBS variants.

Table II.4. T_m (°C) determined by DSF for AmyNSAR/OSBS variants

	+ 0 substrate	+ 2 mM SHCHC	+ 10 mM L-NSPG
WT	64.9 ± 0.1	66.2 ± 0.1	66.0 ± 0.1
R266Q	63.5 ± 0.1	63.4 ± 0.1	64.0 ± 0.1
N261L	59.7 ± 0.1	57.9 ± 0.1	59.7 ± 0.1

Effects of mutating of R266 on enzyme activity and stability

As expected, the R266Q mutation significantly decreased NSAR activity and had only a small effect on OSBS activity (Table II.3). Interestingly, AmyNSAR/OSBS R266Q does not change K_M^{OSBS} and K_M^{NSAR}, suggesting that the mutation minimally affects binding affinity of either SHCHC or NSPG. In contrast, the R266Q mutation decreased k_{cat}^{NSAR} by ~470-fold while only decreasing k_{cat}^{OSBS} by 7-fold, resulting in a strong preference toward the OSBS reaction. Retaining the positively charge with the R266K mutation decreased both OSBS and NSAR activities equally (~5-7-fold), suggesting that, in general, mutations at R266 slightly perturb the active site structure, which affects both activities, but a positive charge at position 266 is vital for the NSAR reaction. Thus, R266 is an NSAR reaction specificity determinant, as predicted. Furthermore, AmyNSAR/OSBS R266Q has a minor effect on stability, given that its calculated T_m value is slightly lower (1.4 °C lower) than that of AmyNSAR/OSBS WT protein (Figure II.3, Table II.4). This may be insignificant *in vivo*, given that the protein expression level of the R266Q mutant was the same as the wild type enzyme.

Mechanism for R266Q's effect on NSAR activity

To gain an insight into how R266Q decreased the NSAR activity of AmyNSAR/OSBS, we used ^{1}H NMR spectroscopy to measure the exchange rate (k_{ex}) between the alpha proton of D- or L-NSPG and the catalytic lysines, after pre-

equilibrating the enzyme in D_2O. This experiment allows us to observe the rate of the first step of the NSAR reaction (Figure II.4A).

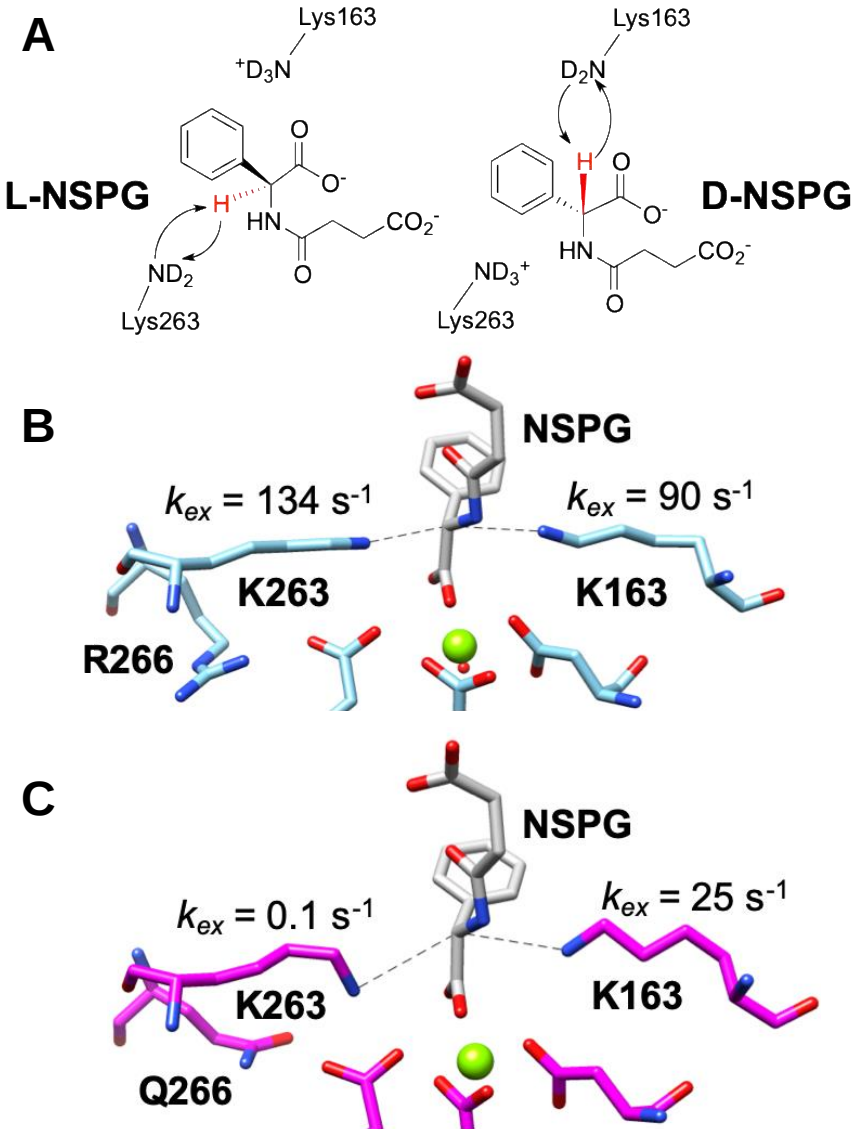


Figure II.4. R266Q mutation specifically decreases the reactivity of K263. (A) Experimental scheme to measure the deuterium-hydrogen exchange rate, k_{ex}. (B) k_{ex} values measured for AmyNSAR/OSBS WT and NSPG. (C) k_{ex} values measured for AmyNSAR/OSBS R266Q and NSPG.

Table II.5. Deuterium-hydrogen exchange rate (k_{ex}) of AmyNSAR/OSBS WT and R266Q

	D-NSPG and K163	**L-NSPG and K263**
AmyNSAR/OSBS WT	90 s^{-1}	134 s^{-1}
AmyNSAR/OSBS R266Q	25 s^{-1}	0.1 s^{-1}
Fold Change	3.6	1340

In AmyNSAR/OSBBS WT, the k_{ex} values between the two catalytic lysines, K163 and K263, and the alpha proton of D- and L-NSPG are similar to each other (Table II.5 and Figure II.4B). Their similarity was expected because NSAR reaction rates with D- and L-NSPG are similar [34]. In AmyNSAR/OSBS R266Q, the mutation drastically decreases k_{ex} between K263 and L-NSPG by 1340-fold, but only reduces k_{ex} between K163 and D-NSPG by 3.6-fold (Table II.5, Figure II.4C). The drop of the k_{ex} value between K163 and D-NSPG is similar to the drop of k_{cat}^{OSBS} in the R266Q mutant. This data indicates that R266 primarily affects the reactivity of K263. The mechanistic role of K263 is not the same between the OSBS and NSAR reactions. In the OSBS reaction, K263 probably forms a π-cation interaction with the substrate SHCHC and/or the enolate intermediate and is thus required to be protonated [30]. On the other hand, K263 acts as the general acid or base to protonate or deprotonate the *N*-succinylamino acids in the NSAR reaction. Because K263 is directly involved in the chemical steps of the NSAR reaction but not the OSBS reaction, the R266Q mutation has a greater effect on the NSAR reaction. Similar K_M^{NSAR} and K_M^{OSBS} values between AmyNSAR/OSBS WT and

R266Q suggest that the substrates bind with similar affinity to the active sites of both AmyNSAR/OSBS WT and R266Q. Given that R266 is a second-shell amino acid and does not contact the substrates, the binding of the substrates was not expected to be affected by the R266Q mutation. Because R266 is proximal to the catalytic K263, R266Q mutation might reposition K263 so that it can no longer abstract the α proton as sufficiently as in the WT enzyme.

The crystal structure of AmyNSAR/OSBS R266Q

To better understand how the R266Q mutation affects NSAR activity, we determined the crystal structure of AmyNSAR/OSBS R266Q bound to NSPG. The crystal structure of AmyNSAR/OSBS R266Q was determined at a resolution of 2.9 Å. The R266Q mutation does not alter the overall structure of the protein. The root-mean-square deviations (RMSD) between the crystal structure of AmyNSAR/OSBS R266Q and those of AmyNSAR/OSBS WT (PDB IDs 1SJA, 1SJB, and 1SJC) are 0.31-0.39 Å. Importantly, K263 adopts a different conformation in the wild type and R266Q mutant. The electron density map of the AmyNSAR/OSBS R266Q structure shows that this part of the structure is well-resolved, indicating that the interaction between K263 and D239 is not an artifact (Figure II.5B). In AmyNSAR/OSBS WT, R266 forms a salt bridge with D239, a conserved metal binding residue in the enolase superfamily. This interaction is missing in AmyNSAR/OSBS R266Q, and K263 now forms a salt bridge with D239. This interaction shifts the amino functional group of K263 away from the α proton of

NSPG and toward D239. This shift explains the large decrease in k_{ex} between K263 and L-NSPG, which impairs K263's ability to act as a general acid/base in the NSAR reaction. In the R266Q mutant, the π-cation interaction between K263 and the SHCHC substrate and/or the enolate intermediate might not be disrupted completely by this shift, limiting the effect of the mutation on OSBS activity.

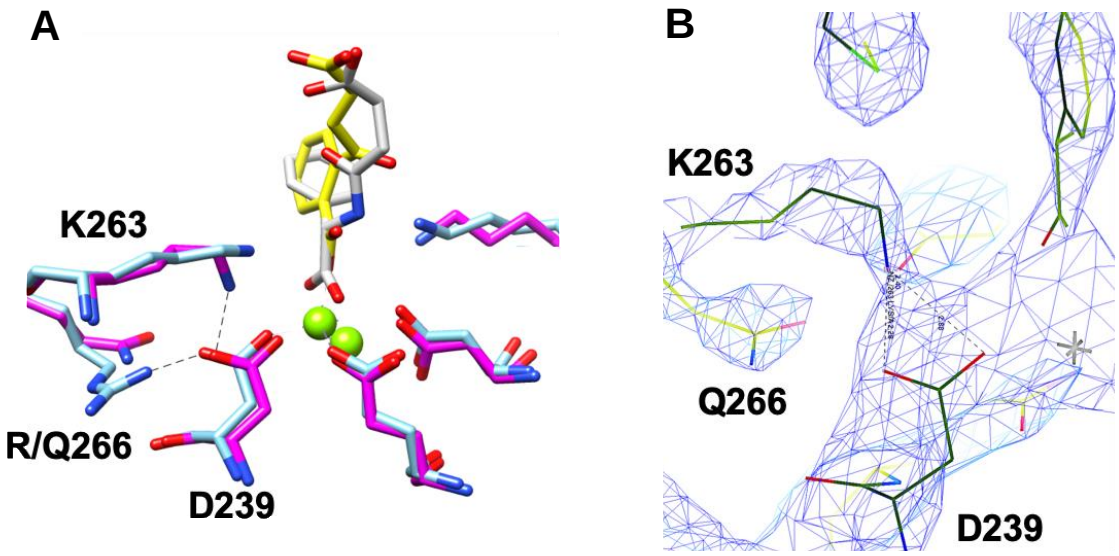


Figure II.5. The R266Q mutation allows a new interaction between K263 and D239. (A) Superimposed structures of AmyNSAR/OSBS WT and R266Q. AmyNSAR/OSBS WT is shown in cyan (PDB ID 1SJB, [30]); AmyNSAR/OSBS R266Q is shown in magenta; OSB is shown yellow; NSPG is shown in white. The salt bridges between R266 and D239 in AmyNSAR/OSBS WT and between K263 and D239 in AmyNSAR/OSBS R266Q are shown as dashed lines. (B) Electron density map showing the resolution of K263, Q266, D239 and other active site residues in AmyNSAR/OSBS R266Q.

Discussion

Roles of second-shell amino acids in enzyme catalysis

During the course of evolution of an enzyme, amino acid conservation is correlated with the distance from the active site. That is, first-shell amino acids are more conserved than second shell amino acids, which are more conserved than the third-shell amino acids, and so on [41]. Thus, in principal, non-active site residues that are more conserved than expected could be essential for catalysis. Mutations at those conserved non-catalytic positions, in any shell, might give an enzyme an opportunity to evolve a new function. Several studies have shown that second-shell amino acids are important for the enzymes to achieve optimal catalysis via their structural, dynamic or electrostatic effects on the active site [1-5]. Some studies determined that second-shell residues are responsible for substrate recognition and substrate specificity and often have no effect on the positions of the catalytic residues [11-13]. In this study, we determined that the residue R266, which is a conserved second-shell amino acid in the NSAR/OSBS subfamily, affects the enzyme's mechanism, rather than substrate discrimination. Here, we showed that the R266Q mutation has a deleterious effect on the NSAR activity but has a minimal effect on OSBS activity. Thus, R266 is vital for the NSAR activity by positioning the catalytic residue for optimal catalysis, while having a less important role on the OSBS activity in the NSAR/OSBS subfamily.

The roles of second-shell amino acids in enzyme catalysis have also been investigated in other members of the enolase superfamily, including mandelate racemase (MR). MR enzymes reversibly catalyze the conversion of *R*- to *S*-mandelate, following a similar mechanism to the NSAR reaction. Previously, a study demonstrated the important role of the second-shell residue D270 in MR enzymes [42]. The D270N mutation in MR decreased the k_{cat} values by ~10^4-fold with both *R*- and *S*-mandelate substrates, suggesting that D270 is essential for catalysis. This study showed that the second-shell amino acid D270 forms a hydrogen bond with the catalytic H297, which assists H297 to act as the general base to abstract the α proton of the *R*-mandelate substrate in the MR reaction [42]. The identifications of D270 in the MR enzymes and R266 in the NSAR/OSBS enzymes demonstrate the essential contributions of second-shell amino acids in catalysis in the members of the enolase superfamily. However, unlike D270, which is important for the only known activity of MR, R266 contributes to the NSAR reaction and has a lesser effect on the OSBS reaction of the catalytically promiscuous AmyNSAR/OSBS. In the NSAR/OSBS enzymes, and other members of the MLE, MR, and D-glucarate dehydratase subgroups of the enolase superfamily, the second lysine (K163) found in the conserved KxK motif at the end the β2 strand is the other catalytic base [22, 43]. The first lysine (K161) in the KxK motif is a first-shell amino acid and directly contacts the substrates. K161 is essential to modulate the activity of the catalytic K163 because the K161A mutation in *E. coli* OSBS, for instance, abolished its activity [44]. R266 is the first residue that was identified to modulate the activity of the catalytic K263 in the NSAR/OSBS subfamily. However, contrasting with

the conserved KxK motif, which is present in many families in the enolase superfamily, we showed that R266 is not conserved throughout the OSBS family. This second-shell amino acid can fine-tune the reactivity of the catalytic K263 to specifically participate in the NSAR reaction.

Pre-adaptation and evolution of new enzyme functions

Pre-adaptive mutations can be considered neutral mutations that have no effect on the primary function, but potentially can contribute to the emergence and evolution of the new function [45]. Gaining pre-adaptive mutations can allow enzymes to evolve a new function. An example, which demonstrated that pre-adaptation via permissive mutations is advantageous in the evolution of new enzyme functions, includes the well-characterized group I RNA enzyme (or ribozyme) derived from *Azoarcus* pre-tRNA^{Ile}. The study showed that while the cleavage activity toward the native substrate, the RNA phosphate bonds, stayed unchanged with accumulated permissive mutations, ribozymes with these accumulated pre-adaptive mutations evolved cleavage activity toward the promiscuous phosphorothioate substrates. Furthermore, this study showed that the ribozymes that contain the accumulated pre-adaptive mutations evolved more rapidly toward the new substrate than the ribozymes without the pre-adaptive mutations [46]. In catalytically promiscuous enzymes, pre-adaptive mutations can be considered permissive mutations that facilitate catalysis of a promiscuous activity with marginal effects on the native activity [47]. We hypothesized that the enzymes from the NSAR/OSBS subfamily

have pre-adaptive features to enabled them to evolve racemase activity from the OSBS activity. In this study, we identified a residue that is important for the NSAR activity which was conserved in enzymes from the NSAR/OSBS subfamily prior to the evolution of NSAR activity as a promiscuous activity or a biological function. Here, we showed that co-opting a conserved arginine at position 266 enabled AmyNSAR/OSBS to carry out the NSAR reaction. However, gaining R266 does not guarantee the enzymes to be able to carry out the NSAR activity because some of the members of the NSAR/OSBS subfamily have R266 but completely lack detectable NSAR. This indicates that other structural factors contribute to NSAR activity.

Enzyme evolvability

In a catalytically promiscuous enzyme, adaptation towards a new activity has been shown to involve a weak trade-off with the original activity [15]. In terms of evolvability, gaining a conserved arginine at position 266 in the NSAR/OSBS subfamily agreed well with this idea. Gaining R266 is essential for the evolution of NSAR activity, while maintaining its efficient OSBS activity. However, as stated before, NSAR activity in the NSAR/OSSB enzymes from many *Amycolatopsis* species including *Amycolatopsis* sp. T-1-60 is inferred to be their biological, native function, and the OSBS activity is the side promiscuous reaction. The phenotypic effects on the NSAR and OSBS activities by the R266Q mutation in AmyNSAR/OSBS seem to contradict the stated idea. The R266Q mutation significantly decreased the biological racemase activity, while having little

effect on the side promiscuous OSBS activity. A possible reason for the phenotypic effects in the R266Q mutant is that OSBS is the ancestral function of the NSAR/OSBS enzymes and such mutations like R266Q have a little effect on its ancestral OSBS activity. Our findings here agree with another study on a bacterial phosphotriesterase (PTE) that showed its native PTE activity was more sensitive to mutations than its promiscuous arylesterase activity [48].

In summary, in this study, we identified a second-shell amino acid R266 which is important for NSAR reaction specificity in the NSAR/OSBS subfamily. The R266Q mutation in AmyNSAR/OSBS profoundly decreases NSAR activity but only moderately reduces OSBS activity. R266 is required to appropriately position the catalytic K263 to act as the general acid/base during catalysis. Here, we report that R266 is the first residue that was identified to assist the catalytic K263 for NSAR reaction catalysis in the NSAR/OSBS subfamily. Gaining R266 is a pre-adaptive feature to allow the emergence and evolution of NSAR activity in the NSAR/OSBS subfamily. However, NSAR activity might or might not be the biological function in different NSAR/OSBS enzymes. Thus, further kinetic and mechanistic investigation of this second-shell residue R266 in other members of the NSAR/OSBS subfamily is essential to fully understand the evolution of NSAR activity in the NSAR/OSBS subfamily.

References

1. Brodkin, H.R., et al., *Evidence of the participation of remote residues in the catalytic activity of Co-type nitrile hydratase from Pseudomonas putida.* Biochemistry, 2011. **50**(22): p. 4923-35.
2. Zimmerman, S., et al., *Role of Trp19 and Tyr200 in catalysis by the gamma-class carbonic anhydrase from Methanosarcina thermophila.* Arch Biochem Biophys, 2013. **529**(1): p. 11-7.
3. Zou, T., et al., *Evolution of conformational dynamics determines the conversion of a promiscuous generalist into a specialist enzyme.* Mol Biol Evol, 2015. **32**(1): p. 132-43.
4. Gonzalez, M.M., et al., *Optimization of Conformational Dynamics in an Epistatic Evolutionary Trajectory.* Mol Biol Evol, 2016. **33**(7): p. 1768-76.
5. Tomatis, P.E., et al., *Adaptive protein evolution grants organismal fitness by improving catalysis and flexibility.* Proc Natl Acad Sci U S A, 2008. **105**(52): p. 20605-10.
6. Karlin, S., Z.Y. Zhu, and K.D. Karlin, *The extended environment of mononuclear metal centers in protein structures.* Proc Natl Acad Sci U S A, 1997. **94**(26): p. 14225-30.
7. Karlin, S. and Z.Y. Zhu, *Classification of mononuclear zinc metal sites in protein structures.* Proc Natl Acad Sci U S A, 1997. **94**(26): p. 14231-6.

8. Xu, X., X.Q. Qin, and E.R. Kantrowitz, *Probing the role of histidine-372 in zinc binding and the catalytic mechanism of Escherichia coli alkaline phosphatase by site-specific mutagenesis.* Biochemistry, 1994. **33**(8): p. 2279-84.

9. Stone, E.M., L. Chantranupong, and G. Georgiou, *The second-shell metal ligands of human arginase affect coordination of the nucleophile and substrate.* Biochemistry, 2010. **49**(49): p. 10582-8.

10. Fesko, K., D. Suplatov, and V. Svedas, *Bioinformatic analysis of the fold type I PLP-dependent enzymes reveals determinants of reaction specificity in l-threonine aldolase from Aeromonas jandaei.* FEBS Open Bio, 2018. **8**(6): p. 1013-1028.

11. Graf, L., et al., *Selective alteration of substrate specificity by replacement of aspartic acid-189 with lysine in the binding pocket of trypsin.* Biochemistry, 1987. **26**(9): p. 2616-23.

12. Perona, J.J., et al., *Structural origins of substrate discrimination in trypsin and chymotrypsin.* Biochemistry, 1995. **34**(5): p. 1489-99.

13. Venekei, I., et al., *Attempts to convert chymotrypsin to trypsin.* FEBS Lett, 1996. **379**(2): p. 143-7.

14. Copley, S.D., *Enzymes with extra talents: moonlighting functions and catalytic promiscuity.* Curr Opin Chem Biol, 2003. **7**(2): p. 265-72.

15. Khersonsky, O. and D.S. Tawfik, *Enzyme promiscuity: a mechanistic and evolutionary perspective.* Annu Rev Biochem, 2010. **79**: p. 471-505.

16. O'Brien, P.J. and D. Herschlag, *Catalytic promiscuity and the evolution of new enzymatic activities.* Chem Biol, 1999. **6**(4): p. R91-R105.

17. Matsumura, I. and A.D. Ellington, *In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates.* J Mol Biol, 2001. **305**(2): p. 331-9.

18. Glasner, M.E., J.A. Gerlt, and P.C. Babbitt, *Evolution of enzyme superfamilies.* Curr Opin Chem Biol, 2006. **10**(5): p. 492-7.

19. Patrick, W.M., et al., *Multicopy suppression underpins metabolic evolvability.* Mol Biol Evol, 2007. **24**(12): p. 2716-22.

20. Kim, J., et al., *Three serendipitous pathways in E. coli can bypass a block in pyridoxal-5'-phosphate synthesis.* Mol Syst Biol, 2010. **6**: p. 436.

21. Baas, B.J., et al., *Recent advances in the study of enzyme promiscuity in the tautomerase superfamily.* Chembiochem, 2013. **14**(8): p. 917-26.

22. Gerlt, J.A., P.C. Babbitt, and I. Rayment, *Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity.* Arch Biochem Biophys, 2005. **433**(1): p. 59-70.

23. Glasner, M.E., et al., *Evolution of structure and function in the o-succinylbenzoate synthase/N-acylamino acid racemase family of the enolase superfamily.* J Mol Biol, 2006. **360**(1): p. 228-50.

24. Zhu, W.W., et al., *Residues required for activity in Escherichia coli o-succinylbenzoate synthase (OSBS) are not conserved in all OSBS enzymes.* Biochemistry, 2012. **51**(31): p. 6171-81.

25. Sakai, A., et al., *Evolution of enzymatic activities in the enolase superfamily: stereochemically distinct mechanisms in two families of cis,cis-muconate lactonizing enzymes.* Biochemistry, 2009. **48**(7): p. 1445-53.

26. Brizendine, A.M., et al., *Promiscuity of Exiguobacterium sp. AT1b o-succinylbenzoate synthase illustrates evolutionary transitions in the OSBS family.* Biochem Biophys Res Commun, 2014. **450**(1): p. 679-84.

27. Odokonyero, D., et al., *Comparison of Alicyclobacillus acidocaldarius o-Succinylbenzoate Synthase to Its Promiscuous N-Succinylamino Acid Racemase/ o-Succinylbenzoate Synthase Relatives.* Biochemistry, 2018. **57**(26): p. 3676-3689.

28. Odokonyero, D., et al., *Loss of quaternary structure is associated with rapid sequence divergence in the OSBS family.* Proc Natl Acad Sci U S A, 2014. **111**(23): p. 8535-40.

29. Sakai, A., et al., *Evolution of enzymatic activities in the enolase superfamily: N-succinylamino acid racemase and a new pathway for the irreversible conversion of D- to L-amino acids.* Biochemistry, 2006. **45**(14): p. 4455-62.

30. Thoden, J.B., et al., *Evolution of enzymatic activity in the enolase superfamily: structural studies of the promiscuous o-succinylbenzoate synthase from Amycolatopsis.* Biochemistry, 2004. **43**(19): p. 5716-27.

31. Klenchin, V.A., et al., *Evolution of enzymatic activity in the enolase superfamily: structural and mutagenic studies of the mechanism of the reaction catalyzed by*

o-succinylbenzoate synthase from Escherichia coli. Biochemistry, 2003. **42**(49): p. 14427-33.

32. Odokonyero, D., et al., *Divergent evolution of ligand binding in the o-succinylbenzoate synthase family*. Biochemistry, 2013. **52**(42): p. 7512-21.

33. Nagar, M. and S.L. Bearne, *An additional role for the Bronsted acid-base catalysts of mandelate racemase in transition state stabilization.* Biochemistry, 2015. **54**(44): p. 6743-52.

34. Taylor Ringia, E.A., et al., *Evolution of enzymatic activity in the enolase superfamily: functional studies of the promiscuous o-succinylbenzoate synthase from Amycolatopsis.* Biochemistry, 2004. **43**(1): p. 224-9.

35. Crooks, G.E., et al., *WebLogo: a sequence logo generator*. Genome Res, 2004. **14**(6): p. 1188-90.

36. Palmer, D.R., et al., *Unexpected divergence of enzyme function and sequence: "N-acylamino acid racemase" is o-succinylbenzoate synthase.* Biochemistry, 1999. **38**(14): p. 4252-8.

37. McMillan, A.W., et al., *Role of an active site loop in the promiscuous activities of Amycolatopsis sp. T-1-60 NSAR/OSBS.* Biochemistry, 2014. **53**(27): p. 4434-44.

38. Cleland, W.W., *Statistical analysis of enzyme kinetic data.* Methods Enzymol, 1979. **63**: p. 103-38.

39. Cook, P.C., WW, *Enzyme Kinetics and Mechanism*. 2007, USA, UK: Garland Science Publishing, Taylor & Francis Group, LLC.

40. Niesen, F.H., H. Berglund, and M. Vedadi, *The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability.* Nat Protoc, 2007. **2**(9): p. 2212-21.

41. Jack, B.R., et al., *Functional Sites Induce Long-Range Evolutionary Constraints in Enzymes.* PLoS Biol, 2016. **14**(5): p. e1002452.

42. Schafer, S.L., et al., *Mechanism of the reaction catalyzed by mandelate racemase: structure and mechanistic properties of the D270N mutant.* Biochemistry, 1996. **35**(18): p. 5662-9.

43. Gulick, A.M., et al., *Evolution of enzymatic activities in the enolase superfamily: crystallographic and mutagenesis studies of the reaction catalyzed by D-glucarate dehydratase from Escherichia coli.* Biochemistry, 2000. **39**(16): p. 4590-602.

44. Nagatani, R.A., et al., *Stability for function trade-offs in the enolase superfamily "catalytic module".* Biochemistry, 2007. **46**(23): p. 6688-95.

45. Wagner, A., *Robustness, evolvability, and neutrality.* FEBS Lett, 2005. **579**(8): p. 1772-8.

46. Hayden, E.J., E. Ferrada, and A. Wagner, *Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme.* Nature, 2011. **474**(7349): p. 92-5.

47. Aharoni, A., et al., *The 'evolvability' of promiscuous protein functions.* Nat Genet, 2005. **37**(1): p. 73-6.

48. Kaltenbach, M., et al., *Functional Trade-Offs in Promiscuous Enzymes Cannot Be Explained by Intrinsic Mutational Robustness of the Native Activity.* PLoS Genet, 2016. **12**(10): p. e1006305.

CHAPTER III

ROLES OF THE SECOND-SHELL AMINO ACID R266 IN OTHER MEMBERS OF THE MLE SUBGROUP OF THE ENOLASE SUPERFAMILY

Introduction

The evolution of new enzyme functions may require the accumulation of several adaptive mutations. Functional mutations can be found in the first shell amino acids which directly interact with the substrate or second shell amino acids, which interact with the first shell amino acids and so on. While mutations in the first shell amino acids in the active site can be directly responsible for the evolution of enzyme ligand interactions, remote mutations beyond the first shell, including second or third shell and other non-active site residues can contribute to functional adaptation toward new enzymatic functions [1, 2]. Identification of these remote mutations could help us understand the fundamental question in structure-function relationships, which is how do residues throughout the structure interact with each other? Such identifications could also help us understand how new enzymatic activities evolve. However, epistasis limits the ability to correctly identify such mutations. Epistasis occurs when the same mutation has different phenotypic effects in different genetic backgrounds, resulting in the non-addictive effects of combinations of mutations [3, 4]. While catalytic residues in the first shell seem unlikely to exhibit epistasis because they make direct interactions with the substrates that are required for the chemical mechanism, non-active site residues (i.e.,

second- and third-shell residues) are more likely to be prone to epistasis because they could make multiple interactions with different sets of coevolving residues due to the intertwined nature of the amino acid network within the enzyme [2].

Previously, we identified a point mutation, R266Q, in the catalytically promiscuous *Amycolatopsis* sp. T-1-60 *N*-succinylamino acid racemase/*o*-succinylbenzoate synthase (AmyNSAR/OSBS) that has a deleterious effect on NSAR activity with a lesser effect on OSBS acitivity (Truong *et al.*, in preparation). AmyNSAR/OSBS belongs to the NSAR/OSBS subfamily, in which many members are catalytically promiscuous and can catalyze both OSBS and NSAR reactions efficiently. The R266Q mutation in AmyNSAR/OSBS profoundly reduces NSAR activity, but only moderately reduces OSBS activity. The second-shell amino acid R266 is close to the catalytic acid/base K263, but it does not contact the substrate. The crystal structure of AmyNSAR/OSBS R266Q shows that K263 adopts an alternate conformation and is not positioned correctly for proton abstraction or donation, decreasing the rate of proton exchange between the alpha proton of the NSAR substrate (*N*-succinylphenylglycine, NSPG) and the general acid/base K263. This mutation is less deleterious for the OSBS reaction because K263 forms a cation-π interaction with the OSBS substrate (2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate, SHCHC) and/or the intermediate, rather than acting as a general acid/base. We showed that R266 was a conserved second-shell residue in the NSAR/OSBS subfamily and not present in other non-promiscuous OSBS subfamilies. We demonstrated that R266 was a pre-adaptive feature that enabled the emergence and evolution of NSAR activity in AmyNSAR/OSBS. However, if R266 is

truly pre-adaptive, we expect that mutations at this position will have the same phenotypic effects on the NSAR and OSBS activities in other NSAR/OSBS members.

Phylogenetic and genome context analysis of other members of the NSAR/OSBS subfamily indicates that NSAR activity evolved through promiscuous intermediates (Figure III.1). For example, the biological function of the NSAR/OSBS enzymes from many species of *Amycolatopsis*, including *Amycolatopsis* sp. T-1-60, is expected to be NSAR activity because these species do not require OSBS activity to make menaquinone or there is a separate OSBS gene encoded in the menaquinone operons [5]. The *Enterococcus faecalis* NSAR/OSBS (EfNSAR/OSBS) gene is encoded in a menaquinone operon, indicating that OSBS activity is its biological function [6]. The NSAR/OSBS enzymes from *Lysinibacillus varians* and *Roseiflexus castenholzii* appear to be bifunctional, based on analysis of genome context. Their OSBS activity is required for menaquinone synthesis, but the NSAR/OSBS gene is in a different operon than other the menaquinone synthesis genes. Genome context analysis indicates that the NSAR/OSBS enzyme from *L. varians* is encoded in a similar operon with *Geobacillus kaustophilus* NSAR/OSBS, which consists of a succinyltranferase (GNAT superfamily), an NSAR/OSBS, and an L-desuccinylase (M20 family) [5, 7]. On the other hand, *R. castenholzii* NSAR/OSBS is encoded in an operon consisting a GNAT superfamily member and an α/β hydrolase superfamily member, which might serve the same function as the M20 family L-desuccinylase of *G. kaustophilus* [5].

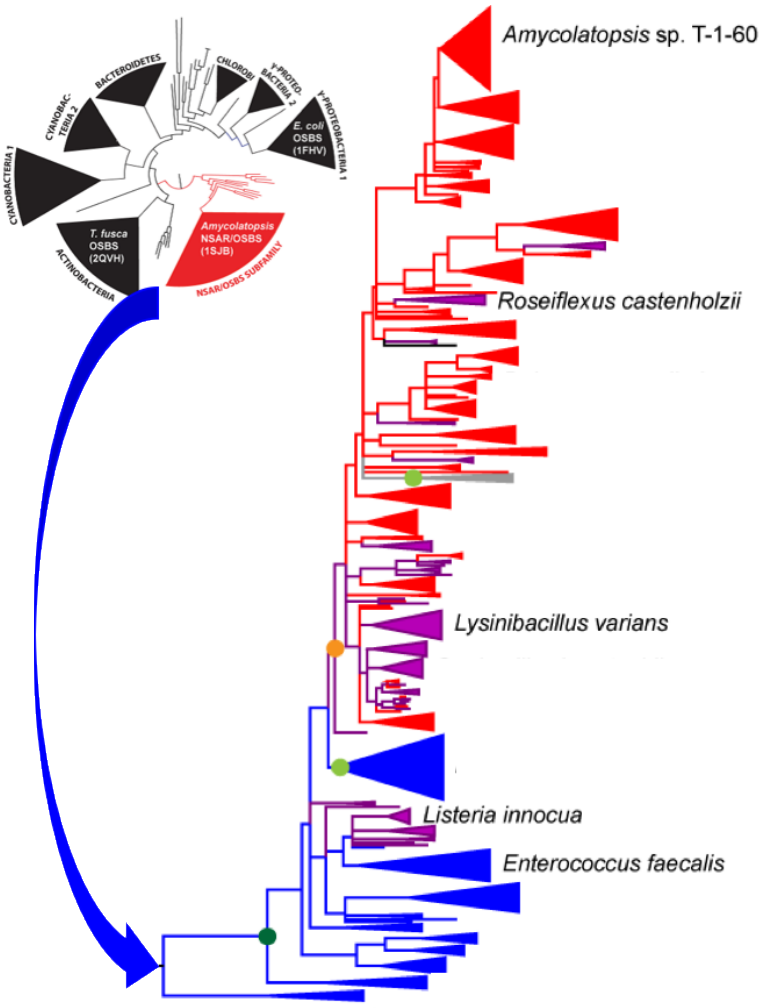


Figure III.1. Phylogenetic distribution of the NSAR/OSBS enzymes used in this study. Figure is modified from references [5, 11], with permission from *Biochemistry*. The inset shows that the OSBS family is subdivided into several large, divergent subfamilies, which corresponding to the phylum from which the OSBS originated [11]. The blue arrow represents the zoomed-in phylogenetic tree of the NSAR/OSBS subfamily [5]. Blue branches indicate proteins that are encoded in menaquinone operons, indicating that OSBS activity is their biological function. Red branches indicate proteins whose biological function is expected to be NSAR activity because their species do not require OSBS activity to make menaquinone or there is a separate OSBS gene encoded in the menaquinone operon. Purple branches indicate proteins that are expected to be bifunctional, because OSBS activity is required for menaquinone synthesis but the NSAR/OSBS subfamily gene is not in the menaquinone operon. Many of these proteins are encoded in operons with genes from the D-amino acid conversion pathway.

Sequence divergence of homologous enzymes has an important effect on evolvability of promiscuous and native activities [8]. For example, a study on eight homologs of the L-gamma-glutamyl phosphate (GP) reductase (ProA), which promiscuously catalyzes N-acetyl-L-glutamyl phosphate (NAGP) reduction, showed that the degree of improvement of the promiscuous NAGP reductase activity achieved by the single mutation E383A varied dramatically and did not correlate with the starting level of the promiscuous activity [8]. This single mutation in ProA also has differential effects on the native activity of the eight homologs. The authors speculated that the E383A mutation might remove some steric conflicts so the enzyme could accommodate the binding of NAGP, which is slightly larger than GP [9]. This study illustrates how the effect of a mutation depends on the sequence and structural contexts in which the mutation occurs (that is, epistatic constraints). Unlike ProA, which exhibits substrate promiscuity, NSAR/OSBS enzymes are catalytically promiscuous. Furthermore, while the mechanism of E383's effect on substrate specificity of ProA was speculative, the effect of R266 on specificity determination in AmyNSAR/OSBS enzymes has been demonstrated (Truong *et al.*, in preparation). However, members of the NSAR/OSBS subfamily are moderately divergent, generally sharing >40% sequence identity, while having variation in their relative OSBS and NSAR activities and differences in their biological function [5]. This raises the question that R266 might not have the same role in different sequence backgrounds.

Here, we examine the role of the residue R266 in the evolution of NSAR activity by examining the effects of the single substitution R266Q in other members of the

NSAR/OSBS subfamily. We made an arginine-to-glutamine substitution at the homologous position of EfNSAR/OSBS, *Roseiflexus castenholzii* NSAR/OSBS (RcNSAR/OSBS), *Lysinibacillus varians* NSAR/OSBS (LvNSAR/OSBS), and *Listeria innocua* NSAR/OSBS (LiNSAR/OSBS), which have been previously characterized. These enzymes efficiently carry out both OSBS and NSAR activities [5, 10]. RcNSAR/OSBS, LvNSAR/OSBS, EfNSAR/OSBS, and LiNSAR/OSBS are 49, 48, 40, and 37% identical, respectively, to AmyNSAR/OSBS. We found that while the R266Q mutation decreases NSAR activity more than OSBS activity, as expected, in most NSAR/OSBS members, the differential effects of the R266Q substitution on NSAR and OSBS activities are not as striking as observed in AmyNSAR/OSBS. In some homologs, the R266Q mutation has very deleterious effects on both OSBS and NSAR activities. Furthermore, the mutation unexpectedly decreases OSBS activity more than NSAR activity in LiNSAR/OSBS. Thus, the effects of R266Q on NSAR and OSBS activities depend on differences in sequence and structural contexts between members of the NSAR/OSBS subfamily, demonstrating the complex role of epistasis in the evolution of NSAR activity in the NSAR/OSBS subfamily.

Materials and Methods

Mutagenesis

Site-directed mutagenesis was performed using the Q5 mutagenesis protocol (New England BioLabs). The templates for mutagenesis included: the gene encoding *Enterococcus faecalis* NSAR/OSBS (EfNSAR/OSBS) (UniProt entry: Q838J7) and the gene encoding for *Listeria innocua* NSAR/OSBS (LiNSAR/OSBS) (UniProt entry: Q927X3), which were cloned into a pET15b vector (Novagen) (gifts from J.A Gerlt, University of Illinois, Urbana, IL); the gene encoding *Lysinibacillus varians* GY32 NSAR/OSBS (LvNSAR/OSBS) (UniProt entry: X2GR01) and the gene encoding *Roseiflexus castenholzii* HLO8 NSAR/OSBS (RcNSAR/OSBS) (UniProt entry: A7NLX0) which were cloned into a modified pET21a vector pMCSG7 or pMSCSG8, which encodes an N-terminal His6 tag, via ligation-independent cloning [12]; the gene encoding for *E. coli* Dipeptide Epimerase (EcDE) (UniProt entry: P51981) and *B. subtilis* Dipeptide Epimerase (BsDE) (UniProt entry: O34508) which were cloned into a modified pET15b vector, which encodes an N-terminal His10 tag (gifts from J.A Gerlt, University of Illinois, Urbana, IL). Mutations were confirmed by sequencing in both directions (Eurofins Genomics LLC). The primers used for mutagenesis were designed using NEBaseChanger, NEB's online design software (NEBasechanger.com) and are shown in the table.

Table III.1. Primers used for mutagenesis of different NSAR/OSBS and DE enzymes

Mutation	Template	Primer sequence
R266Q	EfNSAR/OSBS	Forward: GAAGATTCCTcagGTAGGTGGGATTC Reverse: AAATTGATACTACGGCAAC
R266Q	LiNSAR/OSBS	Forward: GAAGCTGGCAcagGTTGGAGGTATG Reverse: AAATTAATAGCCCGACAAC
R266Q	LvNSAR/OSBS	Forward: TAAAATTGGAcagGTAGGCGGCATAAC Reverse: ATATTAATTACACCGCAGC
R266Q	RcNSAR/OSBS	Forward: TAAGATCGGGcagGTCGGCGGGC Reverse: ATGTTGATCACCCGGCAGGC
K250Q	EcDE	Forward: TAAGCTCGATcagACCGGGGGTC Reverse: ATATTAACCATCTCATAGCGC
K271Q	BsDE	Forward: TAAATTGATGcagGCGGGCGGCA Reverse: ATATTGATCAAGTCTGCGC
All DNA primers are shown in the 5' to 3' direction. The underlined bases designate the codons where mutations were introduced.		

Protein Production

Proteins were expressed in *E. coli* strain BW 25113 (*menC::kan*, DE3) to ensure that OSBS from the host cell would not contaminate the purified proteins expressed from the plasmid [11]. Cultures were grown overnight at 37 °C in LB media containing carbenicillin and kanamycin at a final concentration of 50 μg/mL each with no induction, then harvested by centrifugation. The cell pellet was resuspended in 10 mM Tris (pH 8.0), 500 mM NaCl, 5 mM imidazole, 0.4 mM phenylmethylsulfonyl fluoride (PMSF), and 10 μg/mL DNase I. The supernatant was collected after centrifugation and filtered using a 0.22 μm Steriflip filter (Millipore). The protein was loaded into a 5 mL HisTrap FF column charged with Ni^{2+} (GE Healthcare). The protein was eluted using a buffer containing 10 mM tris (pH 8.0), 500 mM NaCl, and 500 mM imidazole with a step to 15% elution buffer to elute loosely bound proteins, followed by a linear gradient to 100% elution buffer over 20 column volumes. Fractions containing the proteins were identified by SDS-PAGE, exchanged in storage buffer (20 mM Tris (pH 8.0) and 5 mM $MgCl_2$), and concentrated using a Vivaspin Turbo 15 centrifuge filter with a 10 kDa molecular weight cutoff (Sartorius). Glycerol was added to a final concentration of 25%, and the purified proteins were stored at -20 °C.

OSBS Assay

2-Succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate (SHCHC) was synthesized from chorismate and α-ketoglutarate as described previously [11]. The enzyme was assayed in 50 mM Tris (pH 8.0) and 0.1 mM $MnCl_2$ with various SHCHC concentrations. The reactions were monitored by a SpectraMax Plus384 plate reader (Molecular Devices) at 310 nm and at 25 °C. The disappearance of SHCHC ($\Delta\varepsilon$ = -2400 M^{-1} cm^{-1}) was measured as a function of time [13, 14]. The initial rates were determined by fitting the linear portion of the data in Microsoft Excel, and the initial rates at different substrate concentrations were fit to the Michaelis-Menten equation using Prism (GraphPad).

NSAR Assay

L- and D-*N*-Succinylphenylglycine (L- and D-NSPG) were synthesized as described previously [12]. The enzymes were assayed in 200 mM Tris (pH 8.0) and 0.1 mM $MnCl_2$ with various L- or D-*N*-succinylphenylglycine concentrations. The reactions were carried out in a cell with a 5 cm path length. The change in the optical rotation of the substrate was monitored by a Jasco P-2000 polarimeter at 405 nm and at 25 °C. Measurements were taken using a 10 s integration time and read every 30 seconds. The specific rotation value at 405 nm of L-NSPG is 6.54 deg M^{-1} cm^{-1} and of D-NSPG is 6.22 deg M^{-1} cm^{-1} [15]. The rates were determined as described above.

Dipeptide Epimerase Assay

L-Ala-L-Glu was commercially purchased (Chem Impex Int'l, Inc) and L-Ala-D-Glu was synthesized as described previously [16]. The dipeptide epimerase enzyme was assayed in 50 mM Tris (pH 8.0), 10 mM $MgCl_2$ with various L-Ala-L-Glu concentrations (pH 8.0). The reactions were carried out in a cell with a 5 cm path length in a total volume of 1.4 mL. The change in the optical rotation of the substrate was monitored by a Jasco P-2000 polarimeter at 365 nm and at 25 °C. Measurements were taken using a 10 s integration time and reading every 30 seconds for a total of 30 minutes. The specific rotation value at 365 nM of L-Ala-L-Glu is 7.07 deg M^{-1} cm^{-1}. The initial rates were determined by fitting the linear portion of the data in Microsoft Excel, and the initial rates at different substrate concentrations were fit to the Michaelis-Menten equation using Prism (GraphPad).

Isotopic Exchange Experiments Using ^1H NMR Spectroscopy

The NSAR/OSBS and dipeptide epimerase variants were exchanged into 20 mM Tris buffer (pD 8.0 with NaOD) using a Vivaspin Turbo 15 centrifuge filter (Sartorius). A 10 mL aliquot of protein was concentrated to 1 mL; 9 mL of 20 mM Tris buffer (pD 8.0) was added, and the protein solution was again concentrated to 1 mL. This process was repeated three times to maximize the exchange. Each NSAR reaction contained 20 mM L- or D-NSPG (pD 8.0 with NaOD), 50 mM Tris (pD 8.0 with NaOD), 0.1 mM

$MnCl_2$, and NSAR/OSBS variants in D_2O. The intensity of the α-proton was monitored as it was exchanged with deuterium over time by ^1H NMR (500 MHz Bruker NMR spectrometer). The peak of the α-proton (δ = 5.15 ppm) was integrated relative to that of the five aromatic protons (δ = 7.40 ppm). The relative peak area was converted to concentration based on the initial substrate concentration. The slopes of the plots of the NSPG substrate concentration as a function of time were fit to a line to obtain the isotopic exchange rates (k_{ex}).

Each dipeptide epimerase reaction contained 10 mM L-Ala-L-Glu or L-Ala-D-Glu (pD 8.0 with NaOD), 20 mM Tris (pD 8.0 with NaOD), 0.1 mM $MnCl_2$, and dipeptide epimerase enzymes in D_2O. The intensity of the α-proton was monitored as it was exchanged with deuterium over time by ^1H NMR (500 MHz Bruker NMR spectrometer). The peak of the H_α of the Glu moiety (δ = 4.08 ppm) was integrated relative to that of the H_α of the Ala moiety (δ = 3.95 ppm). The relative peak area was converted to concentration based on the initial substrate concentration. The slopes of the plots the L-Ala-L/D-Glu substrate concentration as a function of time were fit to a line to obtain the isotopic exchange rates (k_{ex}).

Results

The roles of R266 in other members of the NSAR/OSBS subfamily

Previously, we demonstrated that R266 was a pre-adaptive feature that enabled the emergence and evolution of NSAR activity in AmyNSAR/OSBS (Truong *et al.*, in preparation). However, if R266 is truly pre-adaptive, we expect that the R266Q mutation will have the same phenotypic effects on the NSAR and OSBS activities in other NSAR/OSBS members. To determine if R266 has the same effect on the activities in other members of the NSAR/OSBS subfamily, we made an arginine-to-glutamine substitution at the homologous position 266 in *Enterococcus faecalis* NSAR/OSBS (EfNSAR/OSBS), *Roseiflexus castenholzii* NSAR/OSBS (RcNSAR/OSBS), *Lysinibacillus varians* NSAR/OSBS (LvNSAR/OSBS), and *Listeria innocua* NSAR/OSBS (LiNSAR/OSBS). These enzymes efficiently carry out both OSBS and NSAR activities, with a much stronger preference toward OSBS activity (Table III.2). We expected that the R266Q mutation would decrease NSAR activity more than OSBS activity, as observed in AmyNSAR/OSBS. Overall, the R266Q mutation decreases NSAR activity more than OSBS activity in all variants except LiNSAR/OSBS (Table III.2, Figure III.2). However, the R266Q mutation in those enzymes does not affect the relative specificity as dramatically as in AmyNSAR/OSBS. In AmyNSAR/OSBS, while the R266Q mutation has no effect on either K_M^{OSBS} and K_M^{NSAR}, it only decreased k_{cat}^{OSBS} by 7-fold and significantly decreased k_{cat}^{NSAR} by ~470-fold (Truong *et al.*, in

preparation). In contrast, the R266Q mutation in other NSAR/OSBS enzymes decreased k_{cat}^{OSBS} by ~7-379-fold and decreased k_{cat}^{NSAR}, ranging from ~10-fold to undetectable, while generally having lesser or no effect on either K_M^{OSBS} and K_M^{NSAR} (Table III.2, Figure III.2)

To determine if the R266Q mutation decreases the reactivity of the catalytic K263 as observed in AmyNSAR/OSBS R266Q, we used ^{1}H NMR spectroscopy to measure the exchange rate (k_{ex}) between the alpha proton of D- or L-NSPG and the catalytic lysines. This experiment allows us to observe the rate of the first step of the NSAR reaction. Previously, we showed that the R266Q mutation slightly decreased the k_{ex} value of K163, while it markedly decreased the k_{ex} value of K263 by 1340-fold in AmyNSAR/OSBS (Truong *et al.*, in preparation). We expected that the R266Q mutation would decrease the k_{ex} value of K263 with a lesser effect on the k_{ex} value of K163 in other NSAR/OSBS enzymes, as observed in AmyNSAR/OSBS. Overall, the R266Q mutation decreased the k_{ex} value of K263, as expected (Table III.3). However, in several cases, the mutation also significantly decreased the k_{ex} values of K163 (EfNSAR/OSBS, LvNSAR/OSBS, and LiNSAR/OSBS), contrasting with our expectation. We will discuss the effects of the R266Q mutation on OSBS and NSAR activities and the reactivity of the catalytic lysines in each individual enzyme below.

Table III.2. Kinetic constants of WT and R266Q variants of different members of the NSAR/OSBS subfamily

	OSBS			**NSAR**[a]			
	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)	Relative specificity $(k_{cat}/K_M)^{OSBS}/(k_{cat}/K_M)^{NSAR}$
AmyNSAR/OSBS WT *	365 ± 60	42.2 ± 3	(1.16 ± 0.2) × 10^5	1500 ± 100	99 ± 5	(6.37 ± 0.5) × 10^4	1.8
AmyNSAR/OSBS R266Q *	359 ± 30	5.86 ± 0.1	(1.63 ± 0.1) × 10^4	1300 ± 10	0.21 ± 0.01	(1.64 ± 0.1) × 10^2	99
RcNSAR/OSBS WT	116 ± 15	54.6 ± 2	(4.74 ± 0.1) × 10^5	900 ± 100	12.7 ± 0.4	(1.41 ± 0.1) × 10^4	34
RcNSAR/OSBS R266Q	298 ± 14	7.89 ± 0.1	(2.66 ± 0.1) × 10^4	3600 ± 300	1.21 ± 0.04	(3.33 ± 0.3) × 10^2	80
LvNSAR/OSBS WT	70.5 ± 10	186 ± 15	(2.64 ± 0.4) × 10^6	1000 ± 70	15.9 ± 0.4	(1.59 ± 0.1) × 10^4	166
LvNSAR/OSBS R266Q	719 ± 130	20.9 ± 1.5	(2.91 ± 0.6) × 10^4	1600 ± 250	0.122 ± 0.01	(0.76 ± 0.1) × 10^2	382
EfNSAR/OSBS WT	17 ± 3	160 ± 7	(9.53 ± 2) × 10^6	1500 ± 150	34 ± 1	(2.27 ± 0.2) × 10^4	419
EfNSAR/OSBS R266Q	19 ± 2	0.08 ± 0.002	(4.21 ± 0.5) × 10^3	*1500*[b]	*<0.0025*[c]	*<1.67*	*est. >2500*
LiNSAR/OSBS WT	155 ± 16	216 ± 7	(1.39 ± 0.2) × 10^6	1000 ± 270	3.74 ± 0.5	(3.74 ± 0.1) × 10^3	371
LiNSAR/OSBS R266Q	163 ± 15	0.57 ± 0.02	(3.49 ± 0.3) × 10^3	1800 ± 275	0.109 ± 0.009	61 ± 10	57

Table III.2. Continued.

	OSBS	**NSAR[a]**	
*Kinetic parameters measured by Truong DP and shown in Chapter II [a] D-NSPG was used as the substrate [b] This value was estimated assuming that K_M^{NSAR} is the same in the mutant and wild type, as observed for OSBS activity [c] This is the lower limit of detection. NSAR activity was measured using 10 μM enzyme and 20 mM D-NSPG			

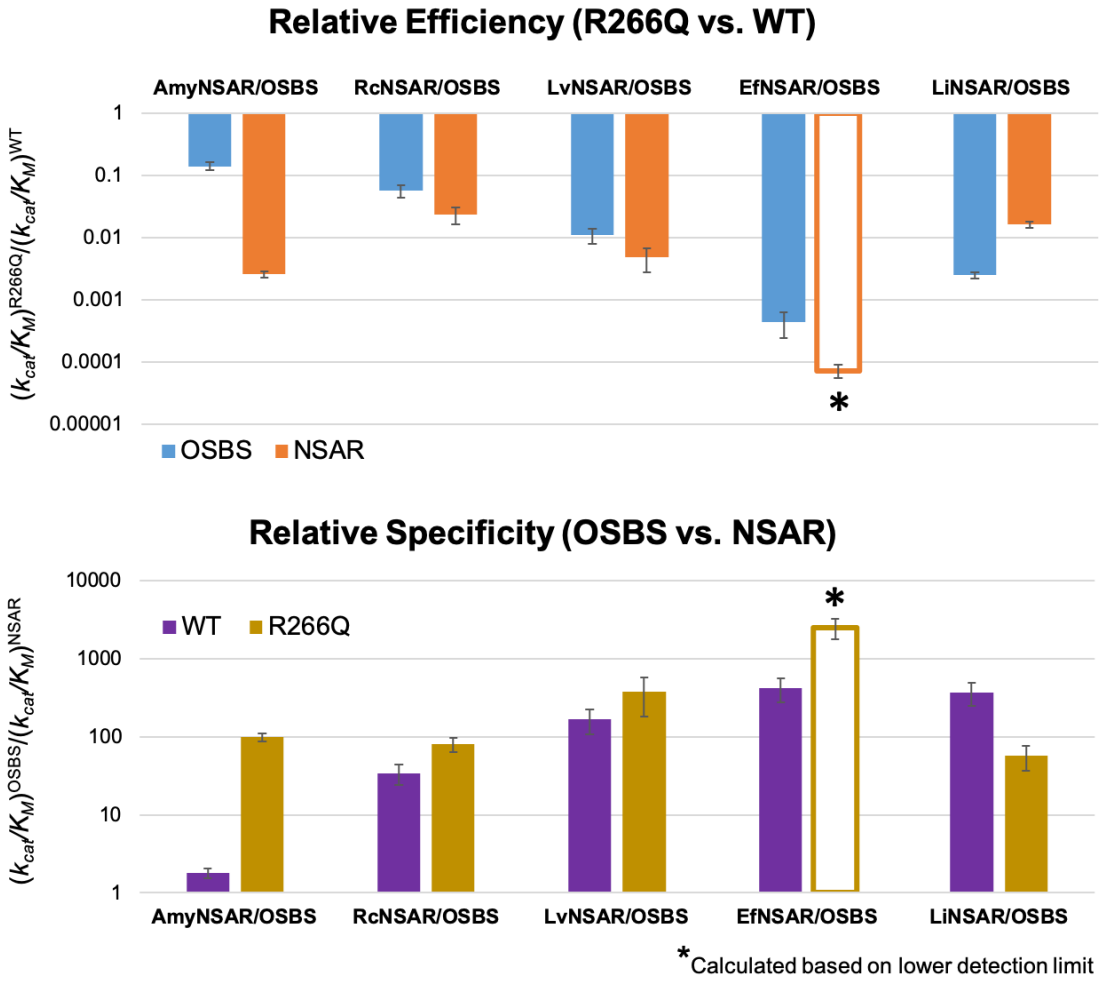


Figure III.2. The effects of R266Q in several members of the NSAR/OSBS subfamily. (A) Relative efficiency ratios of R266Q versus WT variants for OSBS (cyan) and NSAR activities (orange). (B) Relative specificity ratio of OSBS activity versus NSAR activity for WT (purple) and R266Q variants (gold). The asterisk indicates that the NSAR activity of this variant was below the detection limit, so K_M^{NSAR} value for EfNSAR/OSBS R266Q was estimated assuming that K_M^{NSAR} is the same in the mutant and wild type, as observed for OSBS activity, and k_{cat} was estimated as the lower limit of detection.

Table III.3. Deuterium-hydrogen exchange rate (k_{ex}) of other NSAR/OSBS variants

	AmyNSAR/OSBS WT	**AmyNSAR/OSBS R266Q**	**Fold change**
D-NSPG and K163	90 s^{-1}	25 s^{-1}	3.6
L-NSPG and K263	134 s^{-1}	0.1 s^{-1}	1340
	RcNSAR/OSBS WT	**RcNSAR/OSBS R266Q**	**Fold change**
D-NSPG and K163	11 s^{-1}	10 s^{-1}	1.1
L-NSPG and K263	22 s^{-1}	0.65 s^{-1}	34
	LvNSAR/OSBS WT	**LvNSAR/OSBS R266Q**	**Fold change**
D-NSPG and K163	156 s^{-1}	2 s^{-1}	78
L-NSPG and K263	43 s^{-1}	0.4 s^{-1}	108
	LiNSAR/OSBS WT	**LiNSAR/OSBS R266Q**	**Fold change**
D-NSPG and K163	68 s^{-1}	5 s^{-1}	14
L-NSPG and K263	6 s^{-1}	0.2 s^{-1}	30
	EfNSAR/OSBS WT	**EfNSAR/OSBS R266Q**	**Fold change**
D-NSPG and K163	50 s^{-1}	0.02 s^{-1}	2500
L-NSPG and K263	2.7 s^{-1}	<0.002 s^{-1} [a]	>1350

[a] This value is the lower detection limit.

RcNSAR/OSBS

RcNSAR/OSBS has a lower relative specificity than the other enzymes and is the most similar to AmyNSAR/OSBS in that respect. Likewise, the effect of the R266Q mutation in RcNSAR/OSBS is also the most similar to its effect on AmyNSAR/OSBS.

RcNSAR/OSBS R266Q slightly increased K_M^{OSBS} and lowered k_{cat}^{OSBS} by 7-fold, resulting in a ~17-fold decrease in k_{cat}/K_M^{OSBS}, while it increased K_M^{NSAR} by 4-fold and lowered k_{cat}^{NSAR} by 10-fold, resulting in a ~43-fold decrease in k_{cat}/K_M^{NSAR}. This agrees with our expectation that R266Q will be more deleterious for NSAR activity than OSBS activity, although the differential effect is not as striking as observed in AmyNSAR/OSBS. However, in RcNSAR/OSBS, the R266Q mutation did not affect the k_{ex} value of K163, but it decreased the k_{ex} value of K263 by 34-fold. The decrease of the k_{ex} value of K263 is the same with order of magnitude as the decrease in the k_{cat}/K_M^{NSAR} in RcNSAR/OSBS, suggesting that the deleterious effect of R266Q primarily arises from the effect of R266 on the reactivity of K263. This result is in good agreement with our hypothesis that R266 is a pre-adaptive feature that enabled the evolution of NSAR activity in the NSAR/OSBS activity.

LvNSAR/OSBS

In LvNSAR/OSBS, the R266Q mutation unexpectedly increased K_M^{OSBS} by 10-fold and decreased k_{cat}^{OSBS} by 9-fold, resulting in a ~90-fold drop in k_{cat}/K_M^{OSBS}, while it had a minimal effect on K_M^{NSAR} and decreased k_{cat}^{NSAR} by 130-fold, resulting in a ~209-fold decrease in k_{cat}/K_M^{NSAR}. The unexpected increase in K_M^{OSBS} combined with an approximately equivalent decrease in k_{cat}^{OSBS} suggests that R266Q in LvNSAR/OSBS increases the dissociation constant of SHCHC by distorting the active site. Active site distortion is also evident in the k_{ex} values of K163 and K263 in LvNSAR/OSBS, which

decreased by 78- and 108-fold, respectively. This decrease is approximately the same order of magnitude as the decrease in k_{cat}/K_M^{OSBS} and k_{cat}/K_M^{NSAR} in LvNSAR/OSBS R266Q. Thus, structural perturbation by R266Q on LvNSAR/OSBS's active site supersedes a specific effect on K263, if there is any, which would have differentially affected NSAR and OSBS activities.

EfNSAR/OSBS

Of the four additional NSAR/OSBS enzymes, EfNSAR/OSBS has the highest relative specificity, with a much stronger preference toward OSBS activity. In EfNSAR/OSBS, while the R266Q mutation dramatically decreased k_{cat}^{OSBS} by ~2000-fold without affecting K_M^{OSBS}, the EfNSAR/OSBS R266Q mutant had no detectable NSAR activity even with 10 μM enzyme. The k_{cat}/K_M^{NSAR} value decreased $>10^4$-fold, assuming K_M^{NSAR} was unchanged from the WT enzyme, as observed in the OSBS reaction. This suggests that R266Q is more deleterious for NSAR activity than OSBS in EfNSAR/OSBS, as expected. In EfNSAR/OSBS, the R266Q mutation also had the strongest deleterious effect on the k_{ex} values. While the k_{ex} value of K163 dropped dramatically by 2500-fold, the k_{ex} value of K263 was below the detection limit, which explained why we could not detect any NSAR activity in EfNSAR/OSBS R266Q. Thus, the fold change of the k_{ex} value of K263 by the R266Q substitution in this enzyme was estimated based on the lower limit of detection, resulting in an uncertainty in the differential effect of the R226Q on the catalytic lysines. However, the decrease in the k_{ex}

value of K163 is on the same order of magnitude as the decrease of k_{cat}^{OSBS}, explaining the decrease of OSBS activity in EfNSAR/OSBS R266Q. Thus, while R266Q is potentially more deleterious for NSAR activity based on differential effects on k_{cat}/K_M^{NSAR}, the primary effect of R266Q appears to be structural distortion of the active site, which affects both activities, as seen in LvNSAR/OSBS.

LiNSAR/OSBS

Unlike other enzymes, LiNSAR/OSBS R266Q unexpectedly decreased OSBS activity more than NSAR activity, contrasting with our prediction. The R266Q mutation significantly decreased k_{cat}^{OSBS} by ~379-fold with no effect on K_M^{OSBS}, while it had minimal effect on K_M^{NSAR} and decreased k_{cat}^{NSAR} by only 37-fold, resulting in only a 61-fold drop in k_{cat}/K_M^{NSAR}. However, it is also worth noting that LiNSAR/OSBS shares the least sequence identity with AmyNSAR/OSBS, and LiNSAR/OSBS WT already has a much lower level of NSAR activity compared to its OSBS activity. Because LiNSAR/OSBS was already less optimal for carrying out NSAR activity than other NSAR/OSBS enzymes, perhaps the less dramatic effect of R266Q on NSAR activity is less surprising. Consistent with the mutation's effect on NSAR activity, the R266Q mutation decreased the k_{ex} value of K263 by the same order of magnitude as the decrease in k_{cat}^{NSAR}. LiNSAR/OSBS R266Q altered the k_{ex} value of K163 to a slightly lower extent (14-fold) but whether the differential effects on K163 and K263 is statistically significant is uncertain. Combined with a more dramatic loss of OSBS activity, however,

similar effects of R266Q on both K163 and K263 suggest perturbation of the active site, which affects both activities. This perturbation could partially or wholly mask a specific effect on K263 in the NSAR reaction, if one exists.

Our kinetic data show that while the R266Q mutation generally decreased NSAR activity more than OSBS activity in the NSAR/OSBS enzymes other than LiNSAR/OSBS, the mutation was also deleterious for the OSBS reaction in some enzymes (especially EfNSAR/OSBS and LiNSAR/OSBS). The differential effects of the R266Q mutation on OSBS and NSAR activities in different NSAR/OSBS enzymes is likely due to the structural contexts in which the mutation occurs. While the R266Q mutation marginally decreases OSBS activity in some sequence and structural contexts (like AmyNSAR/OSBS and RcNSAR/OSBS), it has a bigger and more deleterious effect on OSBS activity in some other contexts (EfNSAR/OSBS and LiNSAR/OSBS). Previous studies have demonstrated that effects mediated by epistasis could manifest on protein folding and stability; that is, the same mutation in different homologs might affect the protein stability differently [3, 17]. This observation partially agrees with our estimation of the effects of R266Q on protein folding and stability, based on protein yields from heterologous protein expression in *E. coli*. While the R266Q mutation did not decrease protein expression compared to wild type in all examined NSAR/OSBS enzymes, the R266Q mutation significantly decreased the protein yields of RcNSAR/OSBS (~40-fold less than the wild type enzyme), suggesting that the mutation could have a destabilizing effect in RcNSAR/OSBS. On the other hand, the R266Q

mutation did not affect the protein yields of EfNSAR/OSBS and LiNSAR/OSBS, even though the OSBS activity decreased more dramatically due to the mutation in these proteins. The differential phenotypic effects of R266Q on activity and protein expression is an example of epistasis. While we showed that R266 is important for NSAR activity in the NSAR/OSBS enzymes, epistatic constraints on the phenotypic effects of the R266Q mutation complicate efforts to determine whether its ability to modify the reactivity of K263 (and enable evolution of NSAR activity) was a pre-adaptive, ancestral feature.

The observed effects of R266Q on OSBS and NSAR activities in different members of the NSAR/OSBS subfamily are due to the different sequence contexts of each NSAR/OSBS enzyme. We could observe some correlation between sequence divergence of the NSAR/OSBS enzymes and the phenotypic effects on NSAR and OSBS by the single point mutation R266Q. RcNSAR/OSBS, LvNSAR/OSBS, EfNSAR/OSBS, and LiNSAR/OSBS are 49, 48, 40, and 37% identical, respectively, to AmyNSAR/OSBS. RcNSAR/OSBS shares the highest sequence identity with AmyNSAR/OSBS, and the effects of the R266Q mutation have the most similar phenotypic effects with AmyNSAR/OSBS R266Q. On the other hand, LiNSAR/OSBS shares the least identity with AmyNSAR/OSBS, and LiNSAR/OSBS R266Q has the opposite phenotypic effects on NSAR and OSBS activities compared with AmyNSAR/OSBS R266Q. However, LvNSAR/OSBS is as similar to AmyNSAR/OSBS as RcNSAR/OSBS, but the mutation is more deleterious on the OSBS reaction and decreases the k_{ex} values of both catalytic lysines. Here, we can see that epistasis plays an

important role in the evolution of NSAR activity in the NSAR/OSBS subfamily. Due to the complexity of epistasis, the phenotypic effects on activities of the same amino acid substitution and the roles of the conserved residues, such as R266, cannot be always assumed to be the same in different NSAR/OSBS homologs.

We attempted to understand the effects of the R266Q mutation by examining the environment surrounding the R266 residue in the crystal structures of EfNSAR/OSBS and LiNSAR/OSBS and comparing that to the structure of AmyNSAR/OSBS [18]. However, only the apo- structure of EfNSAR/OSBS (PDB ID 1WUE) and Mg^{2+} ion-bound structure of LiNSAR/OSBS (PDB ID 1WUF) are available [10]. Overall, the environments within 5 Å surrounding the R266 residue are very similar (Figure III.3). Only two positions differ significantly in the aligned structures. First, position 243 in AmyNSAR/OSBS is a valine, while it is substituted with an arginine in both EfNSAR/OSBS and LiNSAR/OSBS. Second, position 97 in AmyNSAR/OSBS is an arginine while it is a tryptophan in EfNSAR/OSBS and a glutamate in LiNSAR/OSBS. R243 and E97 form a salt bridge in LiNSAR/OSBS while in EfNSAR/OSBS, R243 and W97 could form a cation-π interaction. These interactions electrostatically contribute to the active site, potentially making the area around R266 more rigid, and less able to tolerate the R266Q substitution in these enzymes. These interactions are not present in AmyNSAR/OSBS and could account for the different effects of the R266Q mutation. The crystal structures of RcNSAR/OSBS and LvNSAR/OSBS are not available but sequence alignment analysis revealed that the position 97 and 243 in RcNSAR/OSBS are a proline and a histidine, respectively; while in LvNSAR/OSBS, they are a glycine

and a threonine, respectively. These interactions might be non-specific as in AmyNSAR/OSBS. These observations indicate that small differences surrounding R266 in different NSAR/OSBS homologs might contribute to the different effects of the R266Q mutation. Because of sequence divergence of the NSAR/OSBS enzymes in this study, other residues that are more remote from R266 could also potentially influence the effects of the R266Q mutation. Identifications of these residues that influence R266 remotely are necessary to fully understand the evolution of NSAR activity in the NSAR/OSBS subfamily.

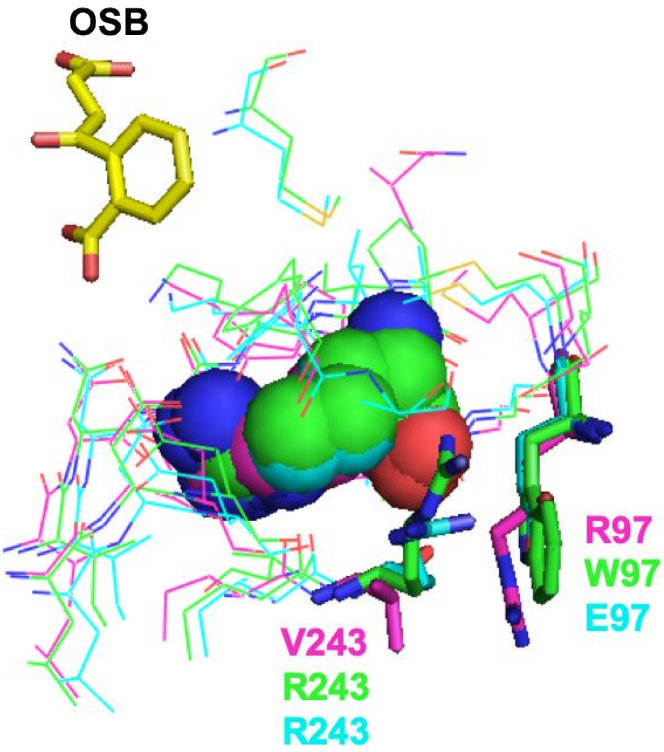


Figure III.3. The local environments (within 5 Å) surrounding R266 in EfNSAR/OSBS (PDB 1WUE, shown in green) [10], LiNSAR/OSBS (PDB 1WUF, shown in cyan) [10], and AmyNSAR/OSBS (PDB 1SJB, shown in magenta) [18]. R266 is shown in spheres. Residues 97 and 243 are labeled corresponding to the colors of the structures. OSB from 1SJB is shown in yellow.

The roles of R266 in the Dipeptide Epimerase family of the MLE subgroup

The enolase superfamily includes several other protein families with racemization activity. For example, the muconate lactonizing enzyme (MLE) subgroup of the enolase superfamily contains the OSBS family, dipeptide epimerase (DE) family, 4R-Hydroxyproline Betaine 2- Epimerase family, and MLE 1 and 2 families (Figure

III.4). The members of the MLE subgroup contain the conserved catalytic lysine residues on the end of the second and sixth β-strands [19].

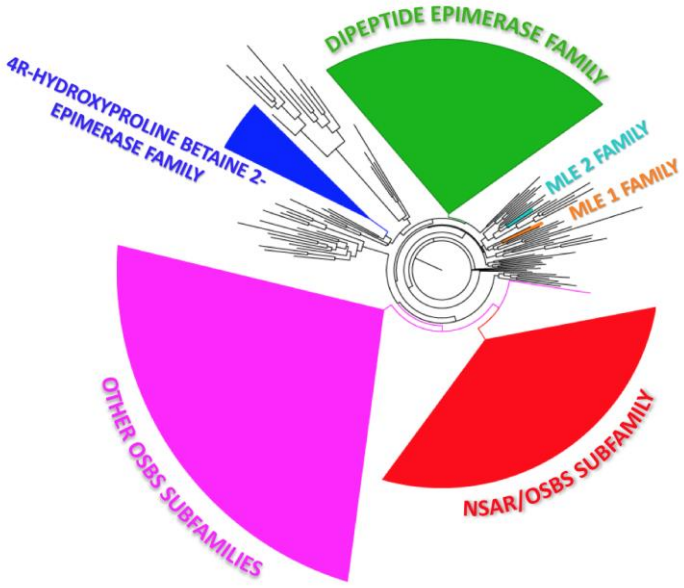


Figure III.4. Phylogenetic tree of the MLE subgroup of the enolase superfamily. The MLE subgroup of the enolase superfamily contains the OSBS subfamilies (shown in magenta), the NSAR/OSBS subfamily (shown in red), the dipeptide epimerase (DE) family (shown in green), the 4R-Hydroxyproline Betaine 2- Epimerase family (shown in blue), and the muconate lactonizing enzyme (MLE) 1 and 2 families (shown in orange and cyan, respectively).

For example, the DE family follows a two-base, 1-1 proton transfer mechanism similar to the NSAR reaction (Figure III.5C). Sequence logo analysis indicates that enzymes from this family also have a positively charged residue at position 266, but it is a lysine instead of an arginine (Figure III.6). The structure of *Bacillus subtilis* DE (BsDE) with the bound Ala-Glu substrate shows that the homologous K271 forms a

hydrogen bond with the homologous metal binding residue D244 (Figure III.5B) (PDB ID 1TKK, [20]). Conservation of a positively charged residue at position 266 in the NSAR/OSBS subfamily and the dipeptide epimerase family suggests that R266/K271 is important for 2-base, 1-1 proton transfer reaction in the members of the MLE subgroup.

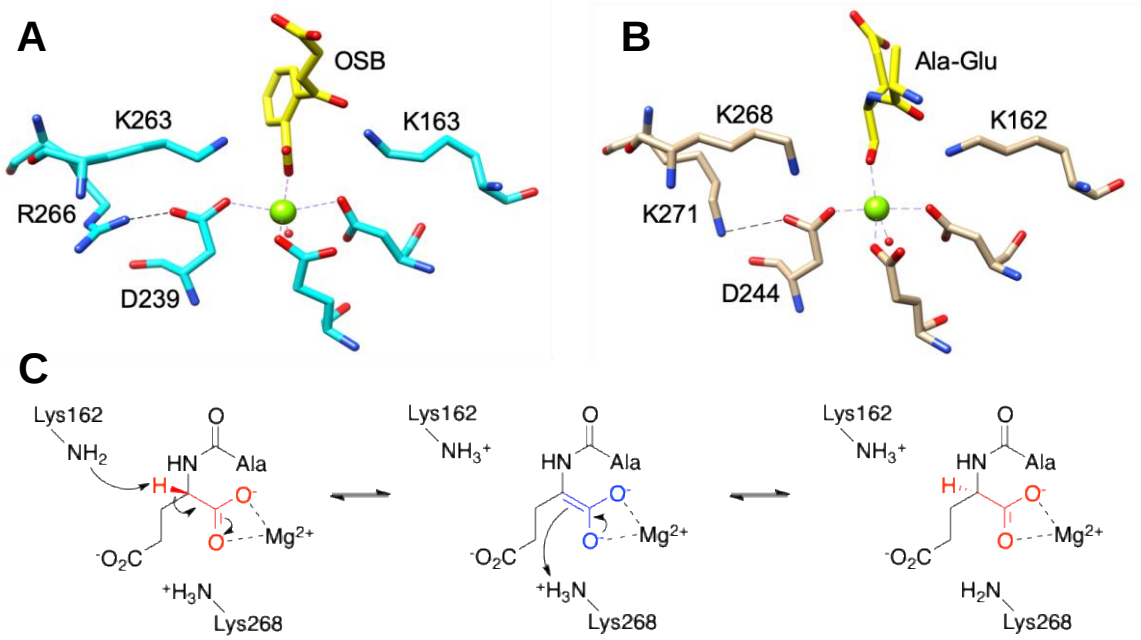


Figure III.5. Comparison of the active sites of (A) AmyNSAR/OSBS (PDB 1SJB, [18]) and (B) BsDE (PDB 1TKK, [20]). (C) The mechanism of L-Ala-L/D-Glu dipeptide epimerase [20]. The divalent metal ion-stabilized enolate intermediate is shown in blue. The atoms that are rearranged during catalysis are shown in red. The numbering of the catalytic lysines are based on the sequence of *Bacillus subtilis* DE.

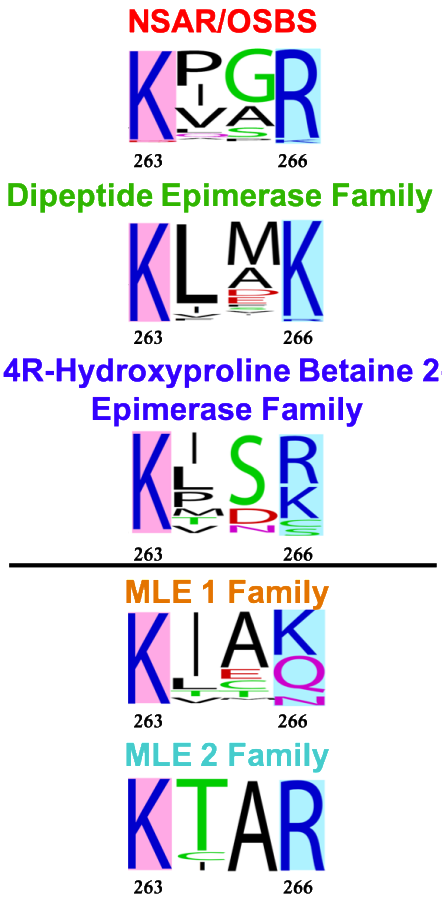


Figure III.6. Sequence logos showing the conservation at positions 266 in some characterized families of the MLE subgroup, including the dipeptide epimerase family, the 4R-hydroxyproline betaine 2-epimerase family, and the MLE 1 and 2 families [22]. The letter size is proportional to the frequency at which each amino acid is found in the sequence alignment. The sequence numbering is relative to the *Amycolatopsis* NSAR/OSBS protein. Position 266 is highlighted in cyan. The catalytic K263 is highlighted in pink.

To test the hypothesis that the positively charged residue at the position equivalent to 266 is universally important for the 2-base, 1-1 proton transfer reaction, we also made the K250Q mutation in *E. coli* DE (EcDE) and K271Q mutation in *B. subtilis* DE (BsDE), which are homologous to R266 in AmyNSAR/OSBS. Like R266 in

AmyNSAR/OSBS, K250 in EcDE and K271 in BsDE are second-shell amino acids and are not in contact with the substrate, suggesting they will not affect the binding of the substrate.

Table III.4. Kinetic data for DE variants*

	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)
EcDE WT	340 ± 60	11.4 ± 0.7	$(3.37 \pm 0.6) \times 10^4$
EcDE K250Q	385 ± 85	0.26 ± 0.02	$(6.72 \pm 1.6) \times 10^2$
BsDE WT	452 ± 120	21.4 ± 2	$(4.73 \pm 1.3) \times 10^4$
BsDE K271Q	508 ± 165	0.23 ± 0.03	$(4.53 \pm 1.6) \times 10^2$
*L-Ala-L-Glu was used as the substrate			

The K to Q mutation decreases k_{cat} of EcDE and BsDE by 44-fold and 93-fold, respectively, with no change in the K_M value in either enzyme (Table III.4). This is consistent with the hypothesis that the positively charged residue at this position has the same effect in the NSAR and dipeptide epimerase reactions. We also measured the k_{ex} values between the catalytic lysines and the alpha proton of the substrates to determine if this mutation impacts epimerase activity by affecting the reactivity of one lysine but not the other, as observed in AmyNSAR/OSBS (Table III.5).

Table III.5. Deuterium-Hydrogen exchange rate (k_{ex}) of DE variants

	EcDE WT	**EcDE K250Q**	**Fold change**
Ala-D-Glu and K151	70.4 s^{-1}	67.5 s^{-1}	1
Ala-L-Glu and K247	34.5 s^{-1}	15.2 s^{-1}	2
	BsDE WT	**BsDE K271Q**	**Fold change**
Ala-D-Glu and K162	25.8 s^{-1}	0.78 s^{-1}	33
Ala-L-Glu and K268	27.6 s^{-1}	0.13 s^{-1}	212

In BsDE, the K to Q mutation decreases the k_{ex} values of the lysine of the 6th β-strand, K268, more than those of the lysine on the 2nd β-strand, K162, agreeing with our hypothesis. The reactivity of K162 drops by only 33-fold while the reactivity of K268 drops by 212-fold, consistent with the drop of k_{cat} by the K271Q mutation. However, in EcDE, the mutation only decreases the reactivity of K247 only by 2-fold while has no effect on the reactivity of K151. One possible explanation for the observed k_{ex} values is that the *E. coli* strain used for DE protein expression also contains the native dipeptide epimerase, which could have co-purified with the mutant during purification. As a result, this experiment will be repeated by expressing and purifying the mutant in a DE knockout strain to ensure no contamination by native enzyme.

Discussion

In this study, we examined the role of the second-shell amino acid R266 in the evolution of NSAR activity by examining the effects of the single substitution R266Q in other members of the NSAR/OSBS subfamily. We found that while the R266Q mutation decreases NSAR activity more than OSBS activity, as expected, in most NSAR/OSBS members, the differential effects of the R266Q substitution on NSAR and OSBS activities depend on differences in sequence and structural contexts of between members of the NSAR/OSBS activities. Our study on the second-shell amino acid R266 agreed with the idea that non-active site residues are more prone to epistasis than the catalytic residues [2]. In the NSAR/OSBS and DE families, mutation of the second-shell amino acid R266 has multiple phenotypic roles. The R266Q mutation has an important role in catalysis in most cases, but in some other cases it also has a role on protein folding and stability (as in RcNSAR/OSBS), on K_M^{OSBS} (as in LvNSAR/OSBS), and the overall structure and conformation of the active site (as in LiNSAR/OSBS and EfNSAR/OSBS). The effect of the R266Q mutation is quite specific in AmyNSAR/OSBS, and the phenotypic effects almost solely came from decreasing the reactivity of the catalytic K263. In contrast, the same mutation in RcNSAR/OSBS has similar effects on the reactivity of K263 as in AmyNSAR/OSBS, but it also significantly affects protein stability. This agrees with previous observations that the same mutation will not always have the same effects on stability of the homologous proteins [3, 17]. In LvNSAR/OSBS, the mutation significantly affected K_M^{OSBS} and the reactivity of both

catalytic lysines, but it also decreased OSBS activity. In other cases (EfNSAR/OSBS and LiNSAR/OSBS), the R266Q mutation has deleterious effects on both activities, significantly affecting k_{cat} and the reactivity of both catalytic lysines, suggesting the overall structures and/or the conformation of the active sites of these enzymes are strongly perturbed. In the latter three cases, the effects on the reactivity of the catalytic lysines by the R266Q mutation suggest that if R266 is required to modulate the reactivity of K263, its effects on K263 are masked by other major, more fundamental defects in structure, conformation, and stability of the enzymes. It is also possible that the R266Q mutation in the latter enzymes is not the most suitable substitution to allow the observation of only R266's effect on K263. Saturated mutagenesis at this position could potentially identify another amino acid with a more specific phenotypic effect on NSAR reaction specificity.

We demonstrated that both NSAR/OSBS and DE enzymes require a positively charged residue at the position equivalent to R266 in AmyNSAR/OSBS to carry out the 2-base, 1-1 proton transfer racemase/epimerase reactions because glutamine mutations at this site significantly decreased their activities. In most cases, the mutation primarily affects k_{cat}, and in several cases, there is a convincing deleterious effect on the reactivity of K263, supporting the idea of pre-adaptation. Gaining R/K266 is a pre-adaptive feature to enable the emergence and evolution of racemase/epimerase activity. However, gaining R/K266 does not guarantee the enzymes will carry out racemase/epimerase activity. Interestingly, most enzymes from the MLE 1 and 2 families also have a positively charged residue at position 266, although they are not known to have

racemase or epimerase activity (Figure III.6). In MLE, the lysine on the 2nd β-strand acts as an acid catalyst to protonate the enolate intermediate to yield muconolactone product. On the other hand, the lysine on the 6th β-strand is closer to the carboxylic moiety of the product and presumably assists the stabilization of the enolate intermediate [23]. The role of this lysine on the 6th β-strand is similar to that in the enzymes of the non-promiscuous OSBS subfamilies, which do not have a positively charged residue at position 266. It would be interesting to explore why the enzymes that do not carry out racemase/epimerase activity as a biological function, like MLE, have a positively charged residue at position 266.

Examining the structure of mandelate racemase (MR) enzyme, another member of the enolase superfamily, showed that MR enzymes also have a positively charged residue at the position homologous to 266, which is K273 (Figure III.7). MR reversibly catalyzes the racemization of *R*- to *S*-mandelate, in a mechanism similar to NSAR and DE. In MR, while the conserved K166 on the end of the β2 strand is one catalytic acid/base, H297 in the H297-D270 dyad is the other catalytic base. The second-shell residue D270 has been shown to be important to assist H297 in catalysis [24]. The homologous K273 is in close proximity to D270 and a little further away from H297, but it could remotely interact with and potentially assist H297 in catalysis. The role of this lysine in MR activity has not been explored, and more evidence, including experimental characterization and mechanistic determination are necessary to test the role of this lysine in MR. However, a recent study showed that Y137, which is remote from H297, is

essential for catalysis because it can modulate the pK_a of both K166 and H297 by influencing the electrostatic environment of the whole active site [25].

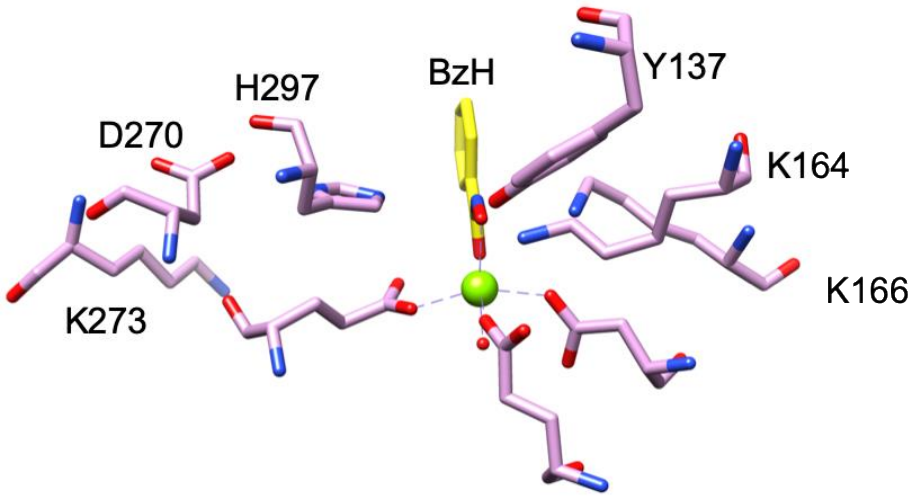


Figure III.7. The active site of mandelate racemase (PDB ID 3UXK, [26]). The catalytic H297-D270 dyad is in close proximity to K273, which is homologous to R266 in AmyNSAR/OSBS. The catalytic triad Y137-K164-K166 is also shown. The intermediate/transition state analog benzohydroxamate (BzH) is shown in yellow.

An intriguing question is that whether an arginine or a lysine at this position can be used as a predictive tool for racemase/epimerase activity. Functional prediction and assignment for uncharacterized enzyme families are important yet difficult tasks. Many sequences with similar structures have similar functions, but some proteins with very similar structures have completely different functions [27]. On the other hand, enzymes with the same functions can have very different structures [28]. Functional predictions and annotations of enzyme superfamilies still remain a challenge even though there are

powerful and advanced tools involving genomics, proteomics and metabolomics [29]. We showed that a positively charged amino acid at the position equivalent to 266 is important for racemase/epimerase activity in two members of the MLE subgroup of the enolase superfamily. Furthermore, mandelate racemase, a member of a different subgroup in the enolase superfamily, also has a lysine at this position, although its function is unknown. All characterized families with racemase/epimerase activity have a positively charged residue at the position homologous to 266 of AmyNSAR/OSBS. However, not all the members of the MLE subgroup (and the enolase superfamily) have been experimentally characterized. With the data presented here, we can potentially use a positively charged amino acid at the position equivalent to 266 as a foothold for experimental characterization and mechanistic determination in unassigned members of the enolase superfamily. Our data indicate that a positively charged residue at position 266 points to racemase/epimerase activity is a likely biological function, narrowing the breadth of chemical reaction space to experimentally explore.

References

1. Lee, J. and N.M. Goodey, *Catalytic contributions from remote regions of enzyme structure.* Chem Rev, 2011. **111**(12): p. 7595-624.
2. Halabi, N., et al., *Protein sectors: evolutionary units of three-dimensional structure.* Cell, 2009. **138**(4): p. 774-86.

3. Starr, T.N. and J.W. Thornton, *Epistasis in protein evolution.* Protein Sci, 2016. **25**(7): p. 1204-18.
4. Miton, C.M. and N. Tokuriki, *How mutational epistasis impairs predictability in protein evolution and design.* Protein Sci, 2016. **25**(7): p. 1260-72.
5. Odokonyero, D., et al., *Comparison of Alicyclobacillus acidocaldarius o-Succinylbenzoate Synthase to Its Promiscuous N-Succinylamino Acid Racemase/ o-Succinylbenzoate Synthase Relatives.* Biochemistry, 2018. **57**(26): p. 3676-3689.
6. Glasner, M.E., et al., *Evolution of structure and function in the o-succinylbenzoate synthase/N-acylamino acid racemase family of the enolase superfamily.* J Mol Biol, 2006. **360**(1): p. 228-50.
7. Sakai, A., et al., *Evolution of enzymatic activities in the enolase superfamily: N-succinylamino acid racemase and a new pathway for the irreversible conversion of D- to L-amino acids.* Biochemistry, 2006. **45**(14): p. 4455-62.
8. Khanal, A., et al., *Differential effects of a mutation on the normal and promiscuous activities of orthologs: implications for natural and directed evolution.* Mol Biol Evol, 2015. **32**(1): p. 100-8.
9. McLoughlin, S.Y. and S.D. Copley, *A compromise required by gene sharing enables survival: Implications for evolution of new enzyme activities.* Proc Natl Acad Sci U S A, 2008. **105**(36): p. 13497-502.

10. Odokonyero, D., et al., *Loss of quaternary structure is associated with rapid sequence divergence in the OSBS family.* Proc Natl Acad Sci U S A, 2014. **111**(23): p. 8535-40.

11. Zhu, W.W., et al., *Residues required for activity in Escherichia coli o-succinylbenzoate synthase (OSBS) are not conserved in all OSBS enzymes.* Biochemistry, 2012. **51**(31): p. 6171-81.

12. Eschenfeldt, W.H., et al., *A family of LIC vectors for high-throughput cloning and purification of proteins.* Methods Mol Biol, 2009. **498**: p. 105-15.

13. Taylor Ringia, E.A., et al., *Evolution of enzymatic activity in the enolase superfamily: functional studies of the promiscuous o-succinylbenzoate synthase from Amycolatopsis.* Biochemistry, 2004. **43**(1): p. 224-9.

14. Palmer, D.R., et al., *Unexpected divergence of enzyme function and sequence: "N-acylamino acid racemase" is o-succinylbenzoate synthase.* Biochemistry, 1999. **38**(14): p. 4252-8.

15. McMillan, A.W., et al., *Role of an active site loop in the promiscuous activities of Amycolatopsis sp. T-1-60 NSAR/OSBS.* Biochemistry, 2014. **53**(27): p. 4434-44.

16. Schmidt, D.M., B.K. Hubbard, and J.A. Gerlt, *Evolution of enzymatic activities in the enolase superfamily: functional assignment of unknown proteins in Bacillus subtilis and Escherichia coli as L-Ala-D/L-Glu epimerases.* Biochemistry, 2001. **40**(51): p. 15707-15.

17. Gong, L.I., M.A. Suchard, and J.D. Bloom, *Stability-mediated epistasis constrains the evolution of an influenza protein.* Elife, 2013. **2**: p. e00631.

18. Thoden, J.B., et al., *Evolution of enzymatic activity in the enolase superfamily: structural studies of the promiscuous o-succinylbenzoate synthase from Amycolatopsis.* Biochemistry, 2004. **43**(19): p. 5716-27.

19. Gerlt, J.A., P.C. Babbitt, and I. Rayment, *Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity.* Arch Biochem Biophys, 2005. **433**(1): p. 59-70.

20. Gulick, A.M., et al., *Evolution of enzymatic activities in the enolase superfamily: crystal structures of the L-Ala-D/L-Glu epimerases from Escherichia coli and Bacillus subtilis.* Biochemistry, 2001. **40**(51): p. 15716-24.

21. Klenchin, V.A., et al., *Evolution of enzymatic activities in the enolase superfamily: structure of a substrate-liganded complex of the L-Ala-D/L-Glu epimerase from Bacillus subtilis.* Biochemistry, 2004. **43**(32): p. 10370-8.

22. Crooks, G.E., et al., *WebLogo: a sequence logo generator.* Genome Res, 2004. **14**(6): p. 1188-90.

23. Sakai, A., et al., *Evolution of enzymatic activities in the enolase superfamily: stereochemically distinct mechanisms in two families of cis,cis-muconate lactonizing enzymes.* Biochemistry, 2009. **48**(7): p. 1445-53.

24. Schafer, S.L., et al., *Mechanism of the reaction catalyzed by mandelate racemase: structure and mechanistic properties of the D270N mutant.* Biochemistry, 1996. **35**(18): p. 5662-9.

25. Fetter, C.M., et al., *Altering the Y137-K164-K166 triad of mandelate racemase and its effect on the observed pKa of the Bronsted base catalysts.* Arch Biochem Biophys, 2019. **666**: p. 116-126.

26. Lietzan, A.D., et al., *Structure of mandelate racemase with bound intermediate analogues benzohydroxamate and cupferron.* Biochemistry, 2012. **51**(6): p. 1160-70.

27. Gerlt, J.A. and P.C. Babbitt, *Can sequence determine function?* Genome Biol, 2000. **1**(5): p. REVIEWS0005.

28. Galperin, M.Y. and E.V. Koonin, *Divergence and convergence in enzyme evolution.* J Biol Chem, 2012. **287**(1): p. 21-8.

29. Pearson, W.R., *Protein Function Prediction: Problems and Pitfalls.* Curr Protoc Bioinformatics, 2015. **51**: p. 4 12 1-4 12 8.

CHAPTER IV

MECHANISTIC INVESTIGATIONS OF THE ENZYMES FROM THE DIVERGENT OSBS FAMILY

Introduction

Understanding mechanistic aspects of enzyme reaction specificity helps us understand not only how an enzyme works but also how an enzyme evolves its function. Focusing these studies on a catalytically promiscuous enzyme provides information to improve reaction specificity in protein engineering and protein design methods. Our model system is the NSAR/OSBS subfamily, in which some members are catalytically promiscuous and efficiently carry out both OSBS and NSAR reactions. The NSAR/OSBS subfamily is a branch to the divergent OSBS family, which belongs to the mechanistically diverse enolase superfamily. The OSBS family also consists of several large, divergent subfamilies, which correspond to the phylum from which the OSBS enzymes originated [1].

Previous studies in our lab have explored sequence, structural and mechanistic aspects to understand how catalytic promiscuity allowed the evolution of NSAR activity. For example, one study investigated the effects of quaternary structure and rates of sequence evolution on changes in structure and activity. Enzymes from the promiscuous NSAR/OSBS subfamily were found to exist in multimeric quaternary structures, like other enzymes in the enolase superfamily, while enzymes from other non-promiscuous

OSBS subfamilies are primarily monomers [2]. The paper showed that loss of quaternary structure in combinations of accumulation of insertions and deletions and high rate of amino acid substitutions resulted in the sequence and structural divergence in other OSBS subfamilies. Retention of ancestral sequence and structural features in the NSAR/OSBS subfamily suggests that the rate of protein evolution is not always proportional to the capacity to evolve new protein functions [2].

Our lab also turned to the promiscuous NSAR/OSBS enzyme from *Amycolatopsis* sp. T-1-60 (AmyNSAR/OSBS) to understand what structural features might determine NSAR reaction specificity. Specifically, the lab examined the roles of the active site loop (the 20s loop) in AmyNSAR/OSBS [3]. The paper found that the active site loop is required for both NSAR and OSBS activities. Deleting the loop decreased OSBS efficiency by 4500-fold and NSAR efficiency by 25000-fold. Most point mutations had small effects on both activities, but the most dramatic mutation is F19A, which decreased OSBS efficiency by 200-fold and NSAR efficiency by 120-fold. F19 is responsible for the hydrophobic packing around the ligand and thus is required for enzymatic activity. R20 on the 20s loop is an important residue because it forms a salt bridge with the catalytic barrel domain to help close the active site and contribute to the catalytic efficiency. The R20E mutation decreased OSBS efficiency by 32-fold and NSAR efficiency by 8-fold. Furthermore, the K_M of NSAR is more sensitive to the loop mutations than the K_M of OSBS. The data presented in this paper showed that the active site loop is required for catalysis via a structural role rather than as reaction specificity determinant.

The studies on NSAR/OSBS enzymes mentioned above helped us understand the evolution of NSAR specificity to a certain extent. We do not fully understand the factors that determine NSAR reaction specificity in the promiscuous NSAR/OSBS enzymes, however. This chapter will address some experiments that I attempted to fully understand NSAR reaction specificity determinants. I carried out different experiments on the promiscuous AmyNSAR/OSBS enzyme such as viscosity experiments, inhibition assays, pH-rate profiles, and differential scanning fluorometry with the presence of different substrates. These experiments will help illuminate the enzymatic reaction mechanism of the promiscuous AmyNSAR/OSBS enzyme. Furthermore, this chapter also includes some of the additional experiments to explain the phenotypic effects on OSBS and NSAR activities of AmyNSAR/OSBS variants discussed in Chapters II and III.

Materials and Methods

Mutagenesis

Site-directed mutagenesis was performed by the Q5 mutagenesis protocol (New England BioLabs). The template for mutagenesis was the gene encoding *Amycolatopsis* sp. T-1-60 NSAR/OSBS (UniProt entry: Q44244), which was cloned into a pET17b vector (a gift from J.A Gerlt, University of Illinois, Urbana, IL). Mutations were confirmed by sequencing in both directions (Eurofins Genomics LLC). The primers used

for *Amycolatopsis* NSAR/OSBS mutagenesis were designed using NEBaseChanger, NEB's online design software (NEBasechanger.com) and are shown in Table IV.1. AmyNSAR/OSBS F19A/R20E double mutation was previously made by Mr. Benjamin Morse.

Table IV.1. Primers used for mutagenesis in *Amycolatopsis* NSAR/OSBS in this study

Mutation	Primer sequence
R266L	Forward: CAAACCGGGCctgGTCGGCGGGT Reverse: ATGTTCACGATTTGGACCGC
F119G	Forward: CGAGAGGTCGggcGCCGCCGAACTCG Reverse: TGCGCGCGGAGTTCGGCG
L123G	Forward: CGCCGCCGAAggcGGATCGGTGC Reverse: AACGACCTCTCGTGCGCG
G124T	Forward: CGCCGAACTCaccTCGGTGCGCGATTCTG Reverse: GCGAACGACCTCTCGTGC
F119G/L123G/G123T	Forward: gaaggcaccTCGGTGCGCGATTCTGTGC Reverse: ggcggcgccCGACCTCTCGTGCGCGCG
All DNA primers are shown in the 5' to 3' direction. The underlined bases designate the codons where mutations were introduced.	

Protein Production

Proteins were expressed in *E. coli* strain BW25113 (*menC::kan*, DE3) to ensure that OSBS from the host cell would not contaminate the purified proteins expressed from the plasmid [24]. Cultures were grown for 48 h at 30 °C in LB media containing carbenicillin and kanamycin at a final concentration of 50 μg/mL, then harvested by centrifugation. The cell pellet was resuspended in 20 mM Tris (pH 8.0), 5 mM $MgCl_2$, 0.4 mM phenylmethylsulfonyl fluoride (PMSF), and 10 μg/mL DNase I. Resuspended cells were lysed by sonication. The supernatant was collected after centrifugation and filtered using a 0.22 μm Steriflip filter (Millipore). The protein was loaded onto a 20 mL HiPrep 16/10 DEAE FF column (GE Healthcare). The protein was eluted using a buffer containing 20 mM Tris (pH 8.0), 5 mM $MgCl_2$, and 500 mM NaCl with an initial step at 30% elution buffer for 6 column volumes to elute loosely bound proteins, followed by a linear gradient from 30% to 65% elution buffer over 10 column volumes. Fractions containing *Amycolatopsis* NSAR/OSBS variants were identified by SDS-PAGE. The fractions were combined with $(NH_4)_2SO_4$, at the final concentration of 0.4 M and then applied to three 5 mL HiTrap Phenyl FF (low sub) columns (GE Healthcare) attached in tandem and equilibrated with a buffer containing 20 mM Tris (pH 8.0), 5 mM $MgCl_2$, and 0.4 M $(NH_4)_2SO_4$. The protein was eluted with 20 mM Tris (pH 8.0) and 5 mM $MgCl_2$ using a linear gradient from 0% to 100% elution buffer over 15 column volumes. Fractions containing AmyNSAR/OSBS and its variants were identified by SDS-PAGE, exchanged into storage buffer (20 mM Tris (pH 8.0) and 5 mM $MgCl_2$), and

concentrated using a Vivaspin Turbo 15 centrifuge filter with a 10 kDa molecular weight cutoff (Sartorius). Glycerol was added to a final concentration of 25%, and the purified proteins were stored at -20 °C.

OSBS Assay

2-Succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate (SHCHC) was synthesized from chorismate and α-ketoglutarate as described previously [1]. The enzyme was assayed in 50 mM Tris (pH 8.0) and 0.1 mM $MnCl_2$ with various SHCHC concentrations. The reactions were monitored by a SpectraMax Plus384 plate reader (Molecular Devices) at 310 nm and at 25 °C. The disappearance of SHCHC ($\Delta\varepsilon$ = -2400 M^{-1} cm^{-1}) was measured as a function of time [4, 5]. The initial rates were determined by fitting the linear portion of the data in Microsoft Excel, and the initial rates at different substrate concentrations were fit to the Michaelis-Menten equation using Prism (GraphPad).

NSAR Assay

L- and D-*N*-Succinylphenylglycine (L- and D-NSPG) were synthesized as described previously [3]. The enzyme was assayed in 200 mM Tris (pH 8.0) and 0.1 mM $MnCl_2$ with various L- or D-*N*-succinylphenylglycine concentrations. The reactions were carried out in a cell with a 5 cm path length. The change in the optical rotation of the

substrate was monitored by a Jasco P-2000 polarimeter at 405 nm and at 25 °C. Measurements were taken using a 10 s integration time and reading every 30 seconds. The specific rotation value at 405 nm of L-NSPG is 6.54 deg M^{-1} cm^{-1} and of D-NSPG is 6.22 deg M^{-1} cm^{-1} [3]. The rates were determined as described above.

Measurements of K_M values for Metal Ion

The K_M values of the metal ion were measured for both OSBS and NSAR reactions as described above. Instead of varying the concentrations of SHCHC or L/D-NSPG, various concentrations of the $MnCl_2$ or $MgCl_2$ was used at the fixed concentration of SHCHC (1200 μM) or L/D-NSPG (10 mM). The rates were calculated as described above.

Inhibition Assay

Inhibition of OSBS activity was measured by adding to the OSBS reaction mixture 0 - 2.1 mM D-NSPG under the conditions described above. K_M^{app} values at different D-NSPG concentrations were calculated by fitting to the Michaelis-Menten equation using Prism (GraphPad). The plot between K_M^{app} vs. D-NSPG concentration was constructed. K_I can be calculated from the slope of the linear relationship (Eq. IV.1).

$$K_M^{app} = K_M (1 + [I]/K_I) \qquad \text{(Eq. IV.1)}$$

Viscosity Measurement

The final concentrations of the microviscogen sucrose in the reactions were 10%, 20%, and 30%. Solutions of the macroviscogen polyethylene glycol (PEG) 8000 were prepared at a final concentration of 6%. The viscosities of the solutions were measured at 25 ^{o}C using a Brookfield LV model viscometer (Ametek). The relative viscosity (η_{rel}) is calculated from the ratio of the viscosity of the solution with viscogen (η) to the viscosity of the solution without viscogen (η^{o}). For the microviscogen sucrose, the relative viscosities were 1.23, 1.40, and 2.05 for 10%, 20%, and 30% sucrose, respectively. For the macroviscogen PEG 8000, the relative viscosity was 1.83 for 6% PEG 8000.

Calculations of the rate constants in the kinetic mechanism

The kinetic mechanism of the OSBS reaction is shown in Scheme IV.1.

$$E \underset{k_{-1}*}{\overset{k_1*A}{\rightleftharpoons}} EA \xrightarrow{k_2} EP \xrightarrow{k_3*} E + P$$

(Scheme IV.1)

The asterisk indicates the dependence of rates on the solvent viscosity. Whereas the rate of the actual chemical transformation (k_2) is independent of solvent viscosity, the association and dissociation rates of the substrate and the dissociation rate of the product (k_1, k_{-1}, and k_3, respectively) are inversely proportional to the solvent viscosity (η_{rel}) [6, 7].

The following relationships can be derived between the values of k_{cat} and k_{cat}/K_M and different rate constants in the kinetic mechanism:

$$\frac{k_{cat}}{K_M} = \frac{k_1 k_2}{k_2 + k_{-1}} \rightarrow \frac{K_M}{k_{cat}} = \frac{k_2 + k_{-1}}{k_1 k_2}$$

With viscogen: $$\frac{K_M}{k_{cat}} = \frac{k_2 + k_{-1}/\eta_{rel}}{k_1 k_2/\eta_{rel}} = \frac{\eta_{rel} k_2 + k_{-1}}{k_1 k_2}$$

$$\frac{K_M}{k_{cat}} = \frac{\eta_{rel}}{k_1} + \frac{k_{-1}/k_2}{k_1} \qquad \text{(Eq. IV.2)}$$

$$k_{cat} = \frac{k_2 k_3^{\circ}}{k_2 + k_3^{\circ}}$$

With viscogen: $$k_{cat} = \frac{k_2 k_3/\eta_{rel}}{k_2 + k_3/\eta_{rel}}$$

$$\text{Relative } k_{cat} = \frac{k_3^{\circ}}{k_2 + k_3^{\circ}} + \frac{k_2}{k_2 + k_3^{\circ}} \eta_{rel} \qquad \text{(Eq. IV.3)}$$

The naught (°) indicates no viscogen is present in the reaction. The rates of each step of the reactions were calculated as describe in reference [7]. The values of k_1 and k_2 can be determined from Eq. IV.2 and the value of k_3 can be determined from Eq. IV.3.

Differential Scanning Fluorimetry (DSF)

Thermal stability of *Amycolatopsis* NSAR/OSBS variants were determined by DSF using a CFX96 real-time PCR (RT-PCR, Bio-Rad). All samples contained SYPRO Orange (Sigma-Aldrich) at a dilution of 1:1250, 0.5 mg/mL of protein, 25 mM HEPES (pH 7.5), 300 mM NaCl, 10 mM DTT, and 10 mM EDTA in a final volume of 50 µL.

SHCHC or NSPG at different concentrations were added to some sample. All samples were run in quadruplicates in a 96-well RT-PCR plate (VWR). The RT-PCR machine was programmed to raise temperature from 25 °C to 99 °C every 1 °C and the fluorescent intensity was collected every 1°C (excitation at 470 nm/ emission at 570 nm, [8]). Data analysis and the unfolding transition (T_m) of proteins were determined by using the DSF analysis protocol as described in ref [8]. Briefly, the raw data was fitted to a 4th polynomial equation, T_m was then calculated from solving the 2nd derivative by using quadratic equation. T_m from quadruplicate samples were averaged and standard error were calculated.

Isotopic Exchange Experiments Using ^1H NMR Spectroscopy

AmyNSAR/OSBS variants were exchanged into D_2O using a Vivaspin Turbo 15 centrifuge filter (Sartorius). A 10 mL aliquot of protein was concentrated to 1 mL; 9 mL of D_2O was added, and the protein solution was again concentrated to 1 mL. This process was repeated three times to maximize the exchange. Each reaction contained 20 mM L- or D-NSPG (pD 8.0), 50 mM Tris (pD 8.0), 0.1 mM $MnCl_2$, and AmyNSAR/OSBS variants in D_2O. The intensity of the α-proton was monitored as it was exchanged with deuterium over time by ^1H NMR (500 MHz Bruker NMR spectrometer). The peak for the α-proton (δ = 5.15 ppm) was integrated relative to that of the five aromatic protons (δ = 7.40 ppm). The relative peak area was converted to concentration based on the initial substrate concentration. The slopes of the plots the

NSPG substrate concentration as a function of time were fit to a line to obtain the isotopic exchange rates (k_{ex}).

pH-rate Profiles

The buffers for the pH profiles included MES (pH 5.5-6.5), Tris (pH 7.0-9.0), and CAPS (pH 9.25-10.0). The buffers were titrated to the desired pH by the addition of NaOH. The pH of the buffers was determined by an Accumet Basic AB15 pH meter (Fisher Scientific). The final concentration of the buffer at a desired pH was 50 mM and 200 mM for the OSBS and NSAR reactions, respectively.

The pH-rate dependence profiles were measured and fit to Eq. IV.4, 5, or 6 using Prism (GraphPad) to calculate the pK_a values [6, 9].

$$\log y = \log[c/(1 + \mathbf{H}/K_1 + K_2/\mathbf{H})] \quad \text{(Eq. IV.4)}$$

$$\log y = \log[c/(1 + \mathbf{H}/K_1)] \quad \text{(Eq. IV.5)}$$

$$\log y = \log[c/(1+ K_2/\mathbf{H})] \quad \text{(Eq. IV.6)}$$

where $y = k_{cat}$ or k_{cat}/K_M, c = pH-independent value of y, K_1 and K_2 = dissociation constants of ionizable groups, and **H** = hydrogen ion concentration. The pK_a was then calculated from the log base 10 of K_1 and K_2 values.

Results and Discussion

Viscosity Experiments on AmyNSAR/OSBS

The general mechanisms of both OSBS and NSAR reactions have been established, [4, 10] but there are significant differences in Michaelis-Menten kinetic parameters between the two reactions in some previously characterized members of the NSAR/OSBS subfamily (Table IV.2) [2, 11]. This suggests that there will be differences between the rates of the individual steps, including the rate-limiting step in the mechanisms of the two reactions. How will the rates of each step of the reactions contribute to the catalytic promiscuity of NSAR/OSBS subfamily?

Table IV.2. Kinetic parameters of previously characterized NSAR/OSBS enzymes

	OSBS			**NSAR**		
Species and reference	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)
Amycolatopsis sp. T-1-60 *	365	42.2	1.2×10^5	1500	99	6.4×10^4
Exiguobacterium sp. AT1b [11]	20	51	2.6×10^6	1700	0.07	41
Listeria innocua [2]	59	170	2.9×10^6	640	1.6	2.6×10^3
Deinococcus radiodurans [2]	26	8.1	3.1×10^5	1400	520	3.7×10^5
Thermus thermophilus [2]	9.4	6.1	6.5×10^5	440	33	7.5×10^4
Enterococcus faecalis [2]	72	120	1.6×10^6	170	24	1.4×10^5
*Kinetic parameters were measured by Truong, DP and originally shown in Chapter II.						

To examine the differences among the rates of each step of the OSBS and NSAR reactions, the viscosity dependence of both reactions was measured. The NSAR and OSBS reactions of AmyNSAR/OSBS WT were assayed in the presence of 0%, 10%, 20%, and 30% sucrose. The reactions of AmyNSAR/OSBS WT were also assayed in 0% and 6% of polyethylene glycol 8000 (a macroviscogen) as a control to ensure that the viscosity effects produced by sucrose are associated with diffusional processes [7]. The plots of K_M/k_{cat} vs. relative viscosity (η_{rel}) and the plots of relative k_{cat} vs. relative viscosity (η_{rel}) of the OSBS and NSAR reactions for AmyNSAR/OSBS enzyme are shown in Figure IV.1. The rates of each step of the reactions were calculated as describe in reference [7], and are shown in Table IV.3.

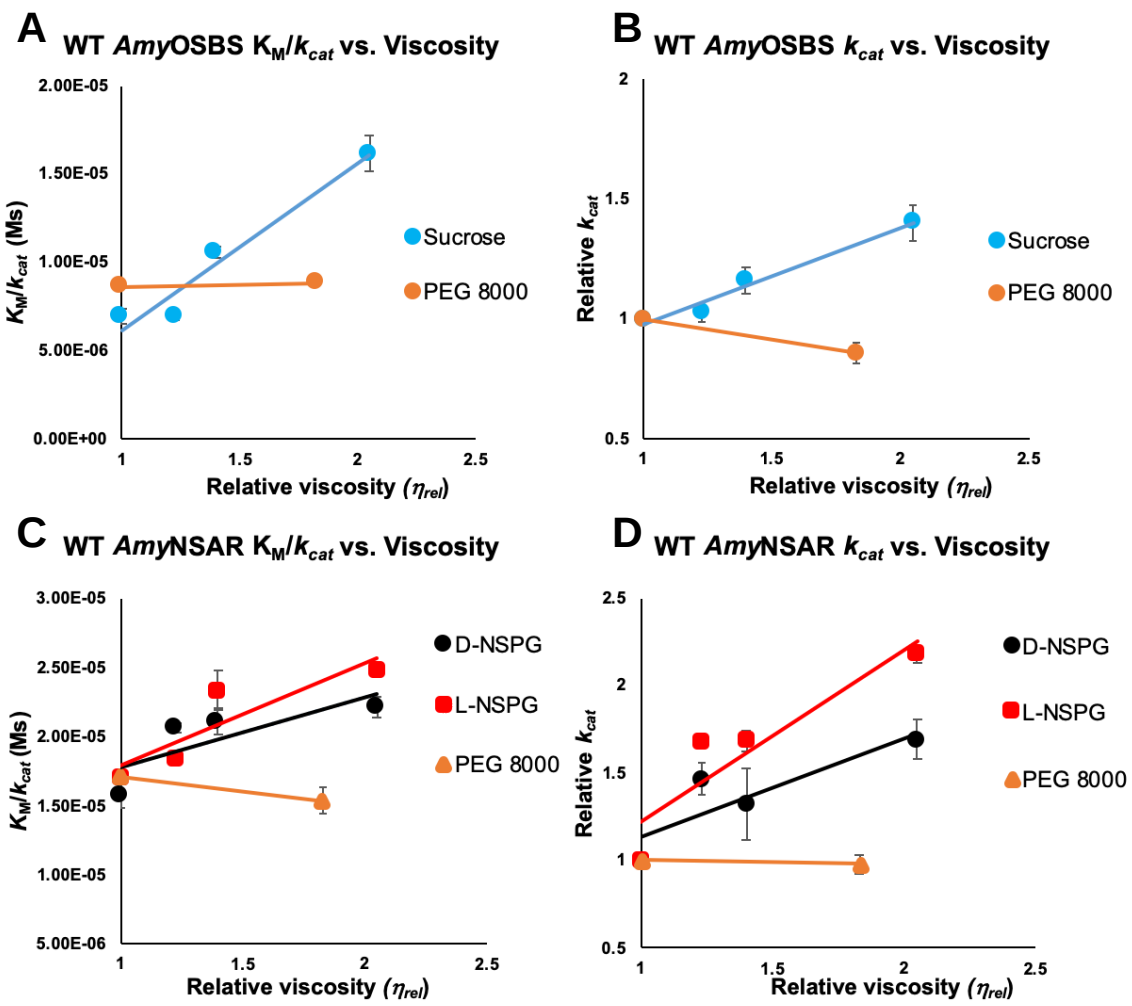
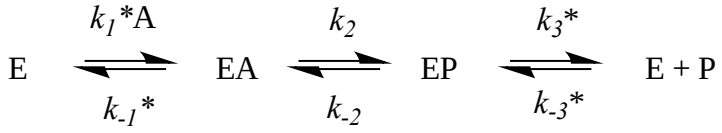


Figure IV.1. Plots of kinetic parameters of AmyNSAR/OSBS WT vs. relative viscosity *(η_{rel}).* (A). The viscosity dependence of K_M/k_{cat} of the OSBS reaction. (B) The viscosity dependence of k_{cat} of the OSBS reaction. (C) The viscosity dependence of K_M/k_{cat} of the NSAR reaction. (D) The viscosity dependence of k_{cat} of the NSAR reaction. NSAR activity with D-SPG substrate is shown in black circle. NSAR activity with L-SPG substrate is shown in red square.

The kinetic values k_{cat} and k_{cat}/K_M of both OSBS and NSAR reactions of AmyNSAR/OSBS WT are dependent on solvent viscosity, suggesting that k_1, k_2, and/or k_3 all are partially rate limiting (Table IV.3). The stickiness ratios, k_2/k_{-1}, is the tendency for the substrate to be committed to the chemical transformation rather than to dissociate from the active site. The substrate of the OSBS reaction is more likely to be committed for catalysis than to dissociate from the active site. The kinetic mechanism shown in Scheme IV.1 cannot be applied for NSAR reaction because of its reversibility. The kinetic mechanism for the reversible NSAR reaction is modified and shown in Scheme IV.2, which contains additional rates representing the reversible steps (k_{-2} and k_{-3}), so Eq IV.2 and IV.3 have too many rate parameters to fit simultaneously. Thus, the kinetic rates for the NSAR reaction could not be calculated.



(Scheme IV.2)

Table IV.3. Kinetic rates of AmyNSAR/OSBS WT calculated from the viscosity dependence experiments

	OSBS	**NSAR (D-NSPG)**	**NSAR (L-NSPG)**
k_{cat}/K_M	1.16×10^5 $M^{-1}s^{-1}$	6.37×10^4 $M^{-1}s^{-1}$	5.83×10^4 $M^{-1}s^{-1}$
k_1	2.31×10^5 $M^{-1}s^{-1}$	2.02×10^5 $M^{-1}s^{-1}$	1.35×10^5 $M^{-1}s^{-1}$
k_{cat}	42 s^{-1}	99 s^{-1}	157 s^{-1}
k_3	104 s^{-1}	*n.d.*	*n.d.*
k_2	70 s^{-1}	*n.d.*	*n.d.*
k_{-1}	24 s^{-1}	*n.d.*	*n.d.*
Stickiness ratio k_2/k_{-1}	2.9	*n.d.*	*n.d.*
n.d.: not determined because NSAR reaction is reversible			

Contribution of the Active Site 20s Loop to Kinetic Rates

Previously, the residues F19 and R20 in the active site loop (the 20s loop) were found to be important for catalysis in AmyNSAR/OSBS (Figure IV.2, Table IV.4) [3]. F19 contributes to the hydrophobic packing of the ligand, while R20 forms salt bridges with D140 and E165 in the barrel domain to help stabilize the 20s loop in the closed conformation [3]. To determine the contribution of the 20s loop to the reaction kinetics,

we measured the activity of AmyNSAR/OSBS in the presence of different viscogen concentrations when the active site loop is disrupted.

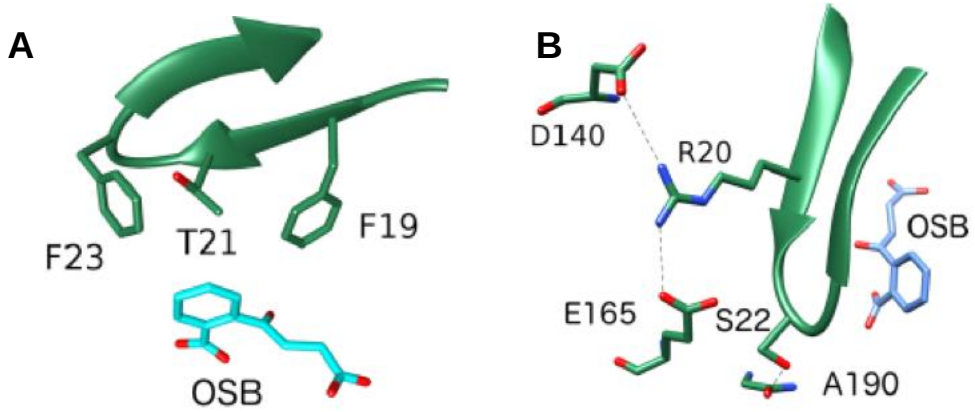


Figure IV.2. The active site loop (the 20s loop) in AmyNSAR/OSBS (figure modified from reference [3], with permission from *Biochemistry*), containing residue F19 (shown in A) and R20 (shown in B). The double mutation F19A/R20E is used to determine the contribution of the 20s loop to the rate constants.

Table IV.4. Enzymatic activities of AmyNSAR/OSBS with the 20s loop variants

	OSBS			NSAR [b]		
	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)
WT	365 ± 60	42 ± 3	$(1.2 \pm 0.1) \times 10^5$	1500 ± 100	99 ± 5	$(6.4 \pm 0.3) \times 10^4$
F19A [a]	n.d.	n.d.	$(4.1 \pm 0.2) \times 10^2$	9100 ± 1500	3.2 ± 0.2	$(3.5 \pm 0.7) \times 10^2$
R20E [a]	n.d.	n.d.	$(2.6 \pm 0.2) \times 10^3$	8100 ± 1400	45 ± 1	$(5.5 \pm 0.1) \times 10^3$
F19A/R20E	n.a. [c]	n.a. [c]	n.a. [c]	5100 ± 600	0.18 ± 0.01	35 ± 6

[a]OSBS and NSAR activities were measured in reference [3]

[b]D-NSPG was the substrate

[c]No activity detected at 10 μM of enzyme

We found that the AmyNSAR/OSBS F19A/R20E double mutant completely abolished OSBS activity while decreasing NSAR activity by ~1800-fold (Table IV.4). The F19A/R20E double mutant significantly decreased k_{cat}^{NSAR} (~550-fold decrease) with a lesser effect on K_M^{NSAR} (~3-fold increase). We used this double mutant to determine the contribution of the active site loop to the reaction kinetics. We measured the NSAR activity of the AmyNSAR/OSBS F19A/R20E double mutant in the presence of different viscogen concentrations to determine if the 20s loop contributes to the rate-limiting step. If the 20s loop contributes to the reaction kinetics, the activity of the double mutant will no longer be dependent on solvent viscosity. If it does not contribute

to the reaction kinetics, the activity of the double mutant will be dependent on solvent viscosity regardless of the disruption of the 20s loop.

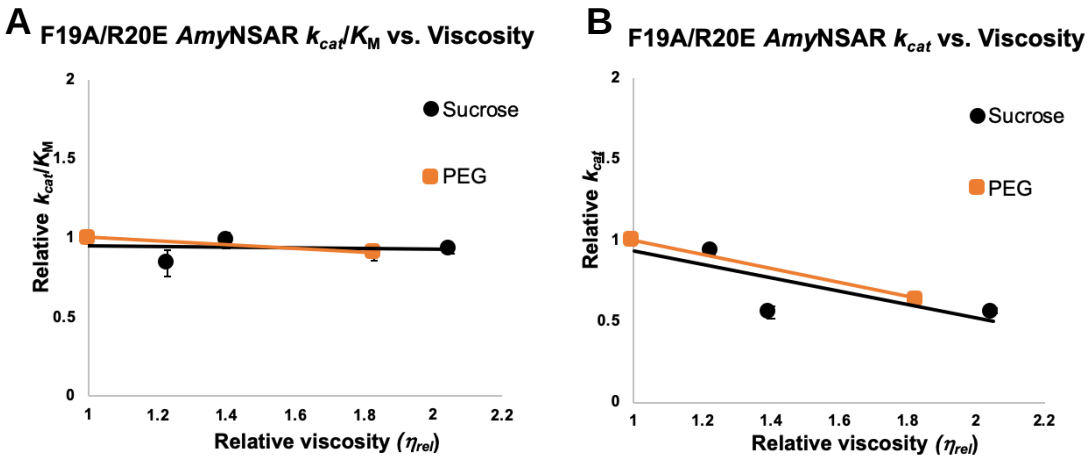


Figure IV.3. Plots of kinetic parameters of AmyNSAR/OSBS F19A/R20E vs. relative viscosity *(η_{rel})*. (A) The relative K_M/k_{cat} of the NSAR reaction vs. relative viscosity *(η_{rel})* of AmyNSAR/OSBS F19A/R20E. (B) The relative k_{cat} of the NSAR reaction vs. relative viscosity *(η_{rel})* of AmyNSAR/OSBS F19A/R20E.

The plots of relative k_{cat}/K_M vs. relative viscosity *(η_{rel})* and relative k_{cat} vs. relative viscosity *(η_{rel})* of the NSAR reaction of AmyNSAR/OSBS F19A/R20E are shown in Figure IV.3. The disruption of the 20s loop in AmyNSAR/OSBS by the F19A/R20E double mutation results in the independence on solvent viscosity. Disruption of the 20s loop completely shifts the rate limiting step to chemistry because the diffusional processes no longer contribute to the rate limiting step. Thus, k_2 (and k_{-2}) solely contributes to k_{cat}^{NSAR} in the F19A/R20E double mutant. The data shows that the

proper opening and closing of the 20s loop is required to commit the substrate for catalysis.

Structural Effects of the Linker Region on AmyNSAR/OSBS

Like the proteins from the enolase superfamily, OSBS enzymes are composed of a catalytic C-terminal $((\beta/\alpha)_7\beta)$-barrel and a capping domain consisting of components from both N- and C-termini [12]. The structure of the linker region between the capping and catalytic barrel domains is conserved in promiscuous NSAR/OSBS subfamily while it is more divergent in other nonpromiscuous OSBS subfamilies (Figure IV.4A&B) [2]. Thus, the linker region between different members of NSAR/OSBS is a potential NSAR specificity determinant. The linker region might have effects on the overall structures and conformational changes upon the substrate binding, rather than direct interactions with the substrate in the active site. The sequence logo identifying the conserved residues of the linker region of the NSAR/OSBS subfamily was constructed using WebLogo [13]. The three most conserved residues among the NSAR/OSBS subfamily are Phe119, Leu123, and Gly124 (following the AmyNSAR/OSBS sequence numbers) (Figure IV.4C).

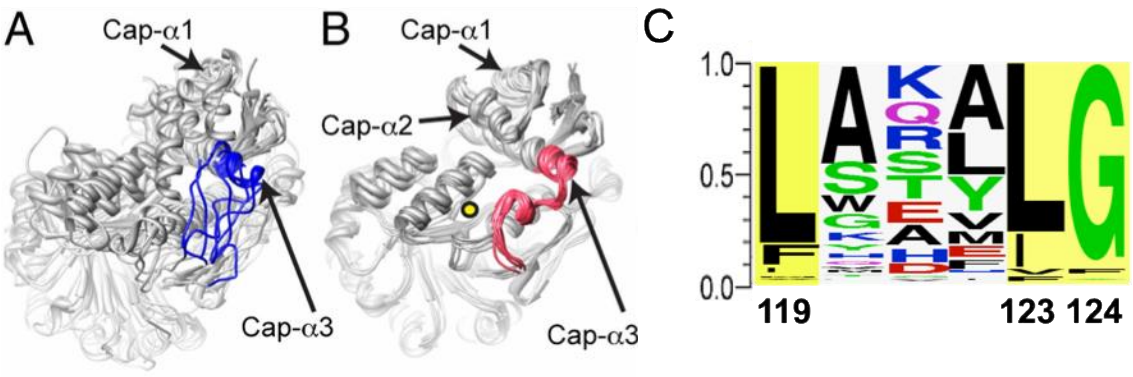


Figure IV.4. The structural and sequence conservation of the linker regions in the NSAR/OSBS subfamily. The structural alignment showing the linker region between the barrel and capping domains (figure is reproduced from reference [2] , with permission from PNAS). (A) OSBS proteins excluding the NSAR/OSBS subfamily (the linker region is shown in blue; PDB IDs 1FHV, 2OKT, 2OZT, 2PGE, 2QVH, and 3CAW). (B) The NSAR/OSBS subfamily (the linker region is shown in pink, PDB IDs 1SJB, 1WUE, 1WUF, 1XS2, and 2ZC8). (C) Sequence logo showing the conservation of the linker region between the barrel and capping domains in the NSAR/OSBS subfamily [13]. The letter size is proportional to the frequency at which an amino acid is found in the sequence alignment. Sequence numbering is relative to AmyNSAR/OSBS. The most conserved residues are highlighted in yellow.

To determine if the linker region is important for catalysis, the conserved sites highlighted in Figure IV.4C were mutated to a glycine or threonine, which would completely disrupt the structure of the linker region. Single mutations F11G, L123G, G124T and triple mutation F11G/L123G/G124T yielded insoluble proteins and were found in the cell pellet fractions. The mutations might induce major structural changes to the proteins and cause aggregation in the pellets. These data suggest that the linker plays a structural role rather than helping to determine reaction specificity in AmyNSAR/OSBS.

Inhibition Assays of AmyNSAR/OSBS, EcOSBS and TfuscaOSBS with NSPG

Our central hypothesis is that the ability of a catalytically promiscuous enzyme to catalyze different reactions is beyond the ability to accommodate the binding of different substrates. To understand the evolution of NSAR activity in the divergent OSBS family, I attempted to determine if the active sites of some non-promiscuous OSBS enzymes can accommodate the binding of the NSAR substrate, NSPG. Those enzymes include *Escherichia coli* OSBS (EcOSBS), which belongs to the γ-Probacteria 1 subfamily and *Thermobifida fusca* OSBS (TfuscaOSBS), which belongs to the Actinobacteria subfamily of the OSBS family [1, 14]. EcOSBS and TfuscaOSBS are 22% and 28% identical to AmyNSAR/OSBS, respectively. They both are efficient OSBS enzymes and have no detectable NSAR activity [2]. The OSBS reaction of EcOSBS is not inhibited in the presence of D-NSPG (Figure IV.5A&B). Both k_{cat} and K_M values of EcOSBS are unchanged when D-SPG is present up to 2.1 mM. EcOSBS was expected not to bind D-NSPG because this enzyme lacks NSAR activity. The structure of EcOSBS reveals that the active site of this enzyme cannot accommodate the extended conformation of D-SPG [15].

Surprisingly, TfuscaOSBS was inhibited by D-NSPG (Figure IV.5B&C). Increasing D-NSPG concentration up to 0.94 mM affects K_M values with minimal effect on k_{cat} value of TfuscaOSBS, suggesting that it competitively binds to the enzyme. This is very surprising because TfuscaOSBS has no detectable NSAR activity. The $K_I^{D\text{-}NSPG}$ value was determined by plotting K_M^{app} vs. D-NSPG concentrations (Figure IV.5B). $K_I^{D\text{-}}$

^{NSPG} of TfuscaOSBS is about 780 (±110) μM, which is close to the $K_I^{D\text{-}NSPG}$ value that was determined from a global fit, which is 457 (±60) μM. This data suggests that TfuscaOSBS has a higher binding affinity for SHCHC than for D-NSPG because its K_M^{OSBS} value (187 (± 27) μM) is smaller than its $K_I^{D\text{-}NSPG}$ value.

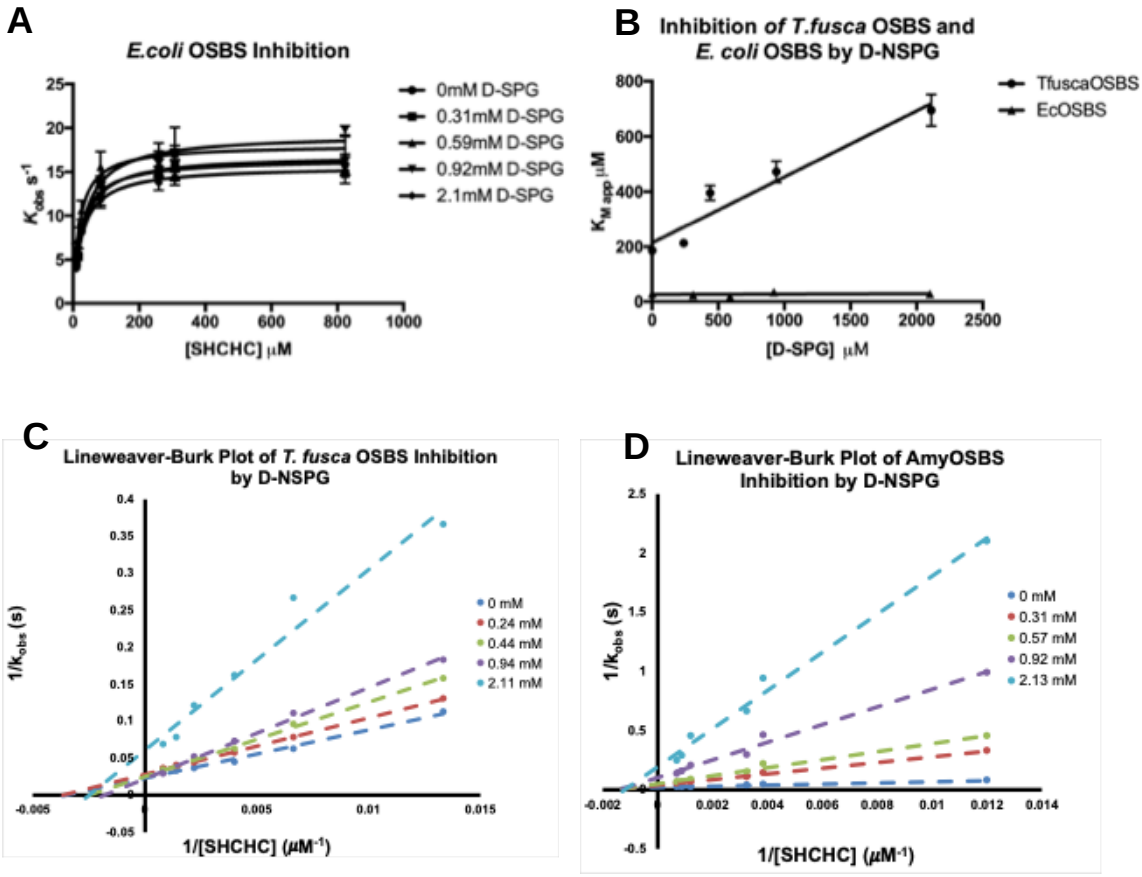


Figure IV.5. Inhibition studies of the enzymes from the OSBS family. (A) EcOSBS is not inhibited by D-NSPG. (B) Plot of K_M^{app} vs. D-NSPG concentration showing inhibition of *T. fusca* OSBS and *E. coli* OSBS by D-NSPG. (C) Lineweaver-Burk plot shows that D-NSPG competitively inhibits TfuscaOSBS. (D) Lineweaver-Burk plot shows that D-NSPG noncompetitively inhibits AmyNSAR/OSBS.

We then measured the $K_I^{D\text{-}NSPG}$ of the OSBS reaction of the promiscuous AmyNSAR/OSBS enzyme and compared it to that of TfuscaOSBS (Figure IV.5D). The OSBS activity of AmyNSAR/OSBS was inhibited noncompetitively by D-NSPG. The noncompetitive inhibition of the OSBS activity of AmyNSAR/OSBS by D-NSPG is not surprising because AmyNSAR/OSBS is multimeric. The binding of D-NSPG in one or more subunits affects the OSBS activity of the other subunits, resulting in a noncompetitive inhibition as observed. The $K_I^{D\text{-}NSPG}$ value was calculated to be 205 (± 10) μM, which is smaller than the K_M^{NSAR} value, which is 1500 (± 100) μM. It is better to compare the two K_I values than to compare $K_I^{D\text{-}NSPG}$ of TfuscaOSBS with the K_M^{NSAR} of AmyNSAR/OSBS, since the K_M value also consists of the chemical step and tends to be larger than the actual K_I value. Furthermore, comparing $K_I^{D\text{-}NSPG}$ of TfuscaOSBS and $K_I^{D\text{-}NSPG}$ of AmyNSAR/OSBS is based on the assumption that both K_I values represent the $K_d^{D\text{-}NSPG}$ values binding to the active site of the enzymes. The $K_I^{D\text{-}NSPG}$ value of TfuscaOSBS is only 2-fold higher than that of AmyNSAR/OSBS, indicating that D-NSPG binds to the active site of TfuscaOSBS with a fairly high affinity. A recent study showed that *Alicyclobacillus acidocaldarius* OSBS (AliacOSBS), a non-promiscuous member of NSAR/OSBS subfamily, can also bind D-NSPG but with a much lower affinity ($K_I^{D\text{-}NSPG}$ is 2600 (± 400) μM) [16]. Both TfuscaOSBS and AliacOSBS are unable to catalyze the conversion of D-NSPG to L-NSPG. It would be very interesting to determine why TfuscaOSBS does not have the ability to catalyze NSAR activity even though it can bind D-NSPG with a relatively high affinity.

Metal Binding Affinity of AmyNSAR/OSBS WT and R266Q

Because R266 forms a salt bridge with D239 in the active site of AmyNSAR/OSBS, I first hypothesized that R266 might affect the ability of D239 to coordinate the metal ion. To determine if the R266Q mutation affects the ability to bind the metal ion in the active site, K_M values for the metal ion of the NSAR and OSBS reactions were measured for both AmyNSAR/OSBS WT and R266Q. The K_M values of the metal ions could be sufficient to estimate the binding affinity of the metal ions to the enzymes, assuming chemistry is rate-limiting (Table IV.5).

Overall, the R266Q mutation does not significantly increase the K_M^{M2+} value for the metal ion in the NSAR and OSBS reactions. This suggests that the R266Q mutation does not decrease the binding affinity of the metal ion. A possible explanation for a lower apparent K_M for Mn^{2+} in the NSAR reaction is that C_α of NSPG is in an sp^3 hybridization state so that the electrons could be localized at the carboxyl functional group and interact more strongly with Mn^{2+}. On the other hand, C_α of the transition state of the OSBS reaction is in an sp^2 hybridization state, and the electrons of the carboxyl functional group could be delocalized throughout the phenyl ring and have weaker interactions with Mn^{2+}.

The R266Q mutant has a much higher apparent affinity for Mn^{2+} than the WT enzyme when the OSBS activity was measured. A possible explanation for this observation is that D239 no longer interacts with position 266 in R266Q and solely coordinates the metal ion, thus enhancing the metal affinity in the R266Q mutant. While

this result suggests that metal affinity is not disrupted by the mutation, the data does not rule out the possibility that the position of metal ion is slightly altered. The crystal structure of AmyNSAR/OSBS R266Q with bound Mg^{2+} shows that Mg^{2+} is slightly in a different position from AmyNSAR/OSBS WT (0.85 Å shift). However, Mn^{2+} was used in our enzymatic assays, thus it is uncertain whether Mn^{2+} is also in an altered position.

Table IV.5. K_M values of the metal ion for AmyNSAR/OSBS variants

Reaction	Metal ion used	K_M^{M2+} of WT	K_M^{M2+} of R266Q
OSBS	Mn^{2+}	23.5 (± 3) μM	0.39 (± 0.06) μM
OSBS	Mg^{2+}	743 (± 120) μM	369 (± 70) μM
NSAR (with D-SPG)	Mn^{2+}	1.04 (± 0.08) μM	1.58 (± 0.16) μM
NSAR (with L-SPG)	Mn^{2+}	1.18 (± 0.13) μM	0.75 (± 0.08) μM

pH-rate Profiles of AmyNSAR/OSBS Variants

As described in Chapter II, N261L and R266 are second-shell amino acids that are conserved in the NSAR/OSBS subfamily. The positions homologous to N261 and R266 in other non-promiscuous OSBS subfamilies are usually hydrophobic and are often leucine. Thus, we first hypothesized that N261 and R266 would be NSAR reaction specificity determinants and would be important for NSAR catalysis. We initially

attempted to determine the roles of N261 and R266 in AmyNSAR/OSBS by constructing the N261L and R266L mutations. We expected that those mutations would decrease NSAR activity with marginal effects on the OSBS reaction. However, N261L mutation decreased OSBS activity more than NSAR activity, and R266L mutation decreased both OSBS and NSAR activities to similar extents (Table IV.6).

Table IV.6. Enzymatic activities of AmyNSAR/OSBS variants

	OSBS			**NSAR** [a]		
	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)
WT	365 ± 60	42 ± 3	$(1.2 \pm 0.1) \times 10^5$	1500 ± 100	99 ± 5	$(6.4 \pm 0.3) \times 10^4$
N261L	n.d.[b]	n.d.[b]	$(3.8 \pm 0.2) \times 10^2$	2200 ± 200	37 ± 2	$(1.7 \pm 0.1) \times 10^4$
R266L	356 ± 50	1.8 ± 0.1	$(5.0 \pm 0.3) \times 10^3$	2000 ± 300	3 ± 0.2	$(1.50 \pm 0.1) \times 10^3$

[a]D-NSPG was the substrate

[b]Not determined because substrate saturation could not be achieved.

Because K263 serves different roles in the OSBS and NSAR reactions, the protonation states of K263 are presumably different in the two reactions. We initially hypothesized that N261 and R266 could affect the pK_a value of the catalytic K263 but not K163. To test this hypothesis, I determined the pH-rate profiles of AmyNSAR/OSBS

variants for both OSBS and NSAR reactions (Figure IV.6). In the OSBS reaction of AmyNSAR/OSBS WT, the pK_a values are 7.6 and 9.2, which are expected to correspond to K163 and K263, respectively. The pK_a values for the OSBS activity of the wild type enzyme are consistent with computationally predicted pK_a's of K163 and K263, (7.15 and 9.14, respectively) [17]. The pK_a values are slightly increased in the N261L mutant (8.1 and 10, which are expected to be K163 and K263, respectively). Our hypothesis had predicted that N261 would affect the pK_a of K263 but not that of K163, so our data does not support with the proposed hypothesis because the N261L mutation seemed to affect the pK_a of both catalytic lysines (Figure IV.6A). A possible explanation for the slight increase in the pK_a values in AmyNSAR/OSBS N261L is that the N261L mutation has a destabilizing effect on the protein as shown in Chapter II. Such destabilizing effects on the overall structure and/or the active site of the enzyme could slightly change the pK_a values in the N261L mutation.

In the NSAR reaction of AmyNSAR/OSBS WT, the ascending limb that represents the general base is missing, and the pH-rate profile exhibits only one pK_a value, which is 9.4 with both L- and D-NSPG substrates. There are several possible explanations for this unexpected pH-rate profile for the NSAR reaction of AmyNSAR/OSBS WT. One possibility is that one of the carboxylates of NSPG could act as a general acid/base and potentially assist in catalysis. Substrate assisted-catalysis has been observed in other, natural and engineered enzymes [18, 19]. Alternatively, the missing ascending limb in the NSAR reaction suggests that within the analyzed pH range, the general acid, which protonates the enolate intermediate, could catalyze the

rate-limiting step. Thus, the ascending limb representing the general base, which abstracts the alpha proton from the substrate, could be masked by the rate-limiting step of the general acid and be invisible in the pH-profile of the NSAR reaction. This would be surprising, since the general base must break a C-H bond starting from the ground state, while the general acid protonates the higher-energy transition state. Last, the pK_a's of the general acid/base catalysts could be masked if a step other than catalysis is rate-limiting. For example, a salt bridge mediated by R20 helps mediate closure of the 20s loop, and the viscosity data described above suggests that closure of this loop is at least partially rate-limiting (although it should be noted that a pK_a of 9.4 would be very low for a solvent exposed arginine). In support of the possibility that the pK_a represents a rate-limiting step other than catalysis, k_{cat}^{NSAR} of AmyNSAR/OSBS WT is independent of pH. Further mechanistic investigations would be needed to assess these possibilities. For the present study, however, uncertainty about which residues determine the shape of the pH-rate profile for NSAR activity of AmyNSAR/OSBS WT makes it difficult to determine whether R266 or N261 alters the pK_a of K263.

In contrast to the pH-rate profile of AmyNSAR/OSBS WT NSAR activity, the NSAR activity of AmyNSAR/OSBS N261L exhibits a bell-shaped pH-rate profile, with two pK_a values of 8.0 and 8.5 (the values are actually too close to each other to state that the values are different from each other, Figure IV.6B). The best explanation for this behavior is that N261L changes the rate-limiting step of the reaction, since the pK_a values from the log(k_{cat}) vs. pH plot are different between AmyNSAR/OSBS WT and

N261L (Figure IV.6C). Consequently, it is not possible to definitively say whether N261L alters the pK_a of K263 in the NSAR reaction.

The OSBS and NSAR activities of AmyNSAR/OSBS R226Q seem to be pH-independent within the analyzed pH range and could not be fit to the equations to yield pK_a values (Figure IV.6D&E). Likewise, k_{cat}^{NSAR} of AmyNSAR/OSBS R266Q is independent of pH (Figure IV.6C). The missing of both ascending and descending limbs suggest that the proton abstraction and donation steps by the base and acid, respectively, are not rate-limiting. In contrast, the pH-rate profile and the pK_a values for the OSBS reaction of AmyNSAR/OSBS R266L are similar to the pK_a values of AmyNSAR/OSBS WT (Figure IV.6F). The pK_a values are 7.7 and 9.6, which are expected to correspond to K163 and K263, respectively, in AmyNSAR/OSBS R266L. This did not support the hypothesis that R266 modulates the pK_a value of K263 because the pK_a values of R266L mutant are very similar with those of the WT enzyme. Overall, while the N261L, R266Q and R266L mutations have effects on the shape and/or pK_a values determined from the pH-rate profiles, changes in the rate-limiting steps due to these mutations and the unexpected pH-rate profile of the wild type enzyme's NSAR activity, made evaluating the hypothesis that N261 and R266 could affect the pK_a value of the catalytic K263 residue unfeasible.

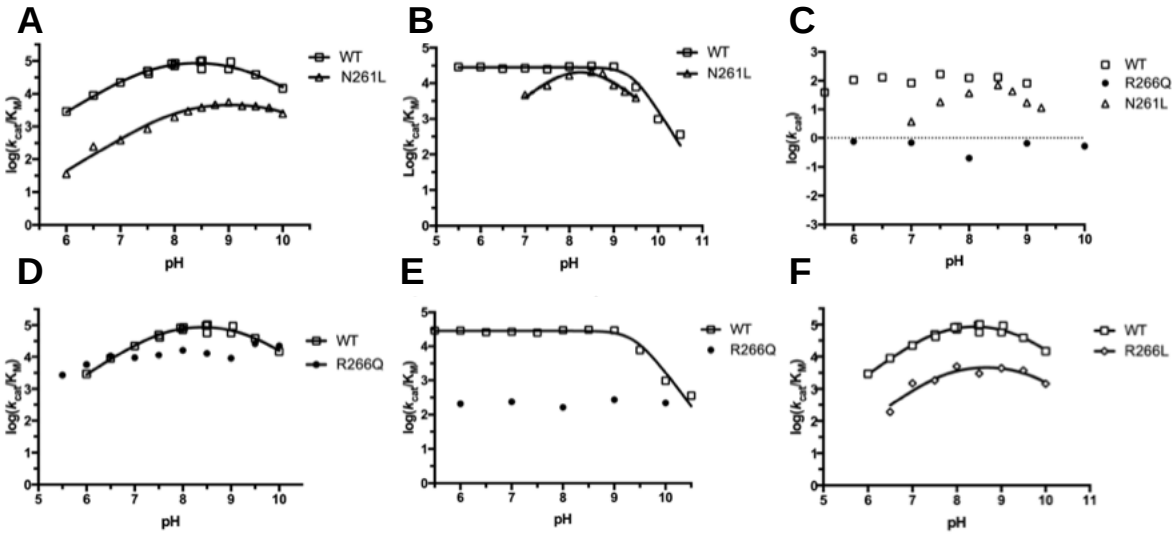


Figure IV.6. pH-rate profiles of AmyNSAR/OSBS variants. (A) The pH-rate profiles of AmyNSAR/OSBS WT and N261L for OSBS reaction. (B) The pH-rate profiles of AmyNSAR/OSBS WT and N261L for NSAR reaction. (C) The plot shows log(k_{cat}) vs. pH of the NSAR reaction for different AmyNSAR/OSBS variants. (D) The pH-rate profiles of AmyNSAR/OSBS WT and R266Q for OSBS reaction. (E) The pH-rate profiles of AmyNSAR/OSBS WT and R266Q for NSAR reaction. (F) The pH-rate profiles of the OSBS reaction of AmyNSAR/OSBS WT and R266L.

Effects of Substrate Binding on Stability in AmyNSAR/OSBS Variants

As described in Chapter II, we identified N261, which is a conserved residue in the NSAR/OSBS subfamily but not in other OSBS subfamilies. We first hypothesized that N261 would be an NSAR reaction specificity determinant. We constructed AmyNSAR/OSBS N261L to test the hypothesis that N261 is important for the NSAR reaction. The N261L mutation unexpectedly decreased the OSBS reaction more than the NSAR reaction. However, we observed a low protein expression level with the N261L mutation (~60-fold lower compared to the WT protein) so we hypothesized that the

mutation could affect the protein folding and stability. To determine if the N261L mutation affects the overall protein folding, we used differential scanning fluorometry (DSF) to measure the melting temperature (T_m) of AmyNSAR/OSBS variants. The measured T_m values by DSF for octameric AmyNSAR/OSBS proteins should be considered approximate because multiple transitions would expected due to dissociation of subunits accompanied by unfolding [8]. In practice, however, we only observed one transition and were able to fit the data to a single apparent T_m. As shown in Chapter II, the T_m value of N261L mutant is 5 °C lower than that of the WT protein, suggesting that the mutation affects the overall folding of the protein. This hypothesis was further supported by the proton exchange rate (k_{ex}) measurements by NMR (Table IV.7). The N261L mutation decreased the exchange rates between the catalytic lysines and the NSAR substrates equally.

Table IV.7. k_{ex} values between NSAR substrates and catalytic lysines of AmyNSAR/OSBS WT and N261L

	D-SPG and K163	L-SPG and K263
WT	90 s^{-1}	134 s^{-1}
N261L	4.9 s^{-1}	5.5 s^{-1}

As shown in Chapter II, the binding of both SHCHC and L-NSPG stabilized the WT enzyme. As a result, we hypothesized that the N261L mutation decreased the OSBS activity more than the NSAR activity because binding NSPG stabilized the protein more than binding SHCHC. We again used DSF method to measure the T_m values of AmyNSAR/OSBS N261L with different concentrations of SHCHC and NSPG (Table IV.8). While the binding of L-NSPG has no effect on the melting temperature of AmyNSAR/OSBS N261L, the T_m value decreased slightly with the increasing concentrations of SHCHC up to 6 mM, suggesting that SHCHC binding could be destabilizing, resulting in a larger decrease on the OSBS activity than the NSAR activity in AmyNSAR/OSBS N261L. However, the mutant has a similar T_m with no SHCHC and at 10 mM SHCHC and it is unclear why I observed this effect.

Table IV.8. T_m (°C) of AmyNSAR/OSBS N261L with different substrate concentrations determined by DSF

OSBS substrate		NSAR substrate	
+ SHCHC	T_m	+ L-SPG	T_m
+ 0 mM	59.7 ± 0.1	+ 0 mM	59.7 ± 0.1
+ 1 mM	58.3 ± 0.1	+ 1 mM	59.7 ± 0.1
+ 2 mM	57.9 ± 0.1	+ 2 mM	59.6 ± 0.1
+ 4 mM	57.4 ± 0.1	+ 5 mM	59.7 ± 0.1
+ 6 mM	57.8 ± 0.1	+ 10 mM	59.7 ± 0.1
+ 10 mM	59.8 ± 0.1	+ 20 mM	59.8 ± 0.1

In summary, I attempted to investigate the mechanistic basis of activity and specificity of AmyNSAR/OSBS by carrying out various biochemical assays including inhibition assays, viscosity experiments, pH-rate profiles, and differential scanning fluorometry. These experiments helped illuminate the mechanisms of the promiscuous AmyNSAR/OSBS enzyme to a certain extent. However, these experiments revealed that dissecting the basis of activity and specificity is complex, raising questions that require additional experiments.

References

1. Zhu, W.W., et al., *Residues required for activity in Escherichia coli o-succinylbenzoate synthase (OSBS) are not conserved in all OSBS enzymes.* Biochemistry, 2012. **51**(31): p. 6171-81.
2. Odokonyero, D., et al., *Loss of quaternary structure is associated with rapid sequence divergence in the OSBS family.* Proc Natl Acad Sci U S A, 2014. **111**(23): p. 8535-40.
3. McMillan, A.W., et al., *Role of an active site loop in the promiscuous activities of Amycolatopsis sp. T-1-60 NSAR/OSBS.* Biochemistry, 2014. **53**(27): p. 4434-44.
4. Taylor Ringia, E.A., et al., *Evolution of enzymatic activity in the enolase superfamily: functional studies of the promiscuous o-succinylbenzoate synthase from Amycolatopsis.* Biochemistry, 2004. **43**(1): p. 224-9.
5. Palmer, D.R., et al., *Unexpected divergence of enzyme function and sequence: "N-acylamino acid racemase" is o-succinylbenzoate synthase.* Biochemistry, 1999. **38**(14): p. 4252-8.
6. Cook, P.C., WW, *Enzyme Kinetics and Mechanism*. 2007, USA, UK: Garland Science Publishing, Taylor & Francis Group, LLC.
7. Wood, B.M., et al., *Mechanism of the orotidine 5'-monophosphate decarboxylase-catalyzed reaction: effect of solvent viscosity on kinetic constants.* Biochemistry, 2009. **48**(24): p. 5510-7.

8. Niesen, F.H., H. Berglund, and M. Vedadi, *The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability.* Nat Protoc, 2007. **2**(9): p. 2212-21.

9. Cleland, W.W., *Statistical analysis of enzyme kinetic data.* Methods Enzymol, 1979. **63**: p. 103-38.

10. Thoden, J.B., et al., *Evolution of enzymatic activity in the enolase superfamily: structural studies of the promiscuous o-succinylbenzoate synthase from Amycolatopsis.* Biochemistry, 2004. **43**(19): p. 5716-27.

11. Brizendine, A.M., et al., *Promiscuity of Exiguobacterium sp. AT1b o-succinylbenzoate synthase illustrates evolutionary transitions in the OSBS family.* Biochem Biophys Res Commun, 2014. **450**(1): p. 679-84.

12. Gerlt, J.A., P.C. Babbitt, and I. Rayment, *Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity.* Arch Biochem Biophys, 2005. **433**(1): p. 59-70.

13. Crooks, G.E., et al., *WebLogo: a sequence logo generator.* Genome Res, 2004. **14**(6): p. 1188-90.

14. Odokonyero, D., et al., *Divergent evolution of ligand binding in the o-succinylbenzoate synthase family.* Biochemistry, 2013. **52**(42): p. 7512-21.

15. Thompson, T.B., et al., *Evolution of enzymatic activity in the enolase superfamily: structure of o-succinylbenzoate synthase from Escherichia coli in complex with Mg2+ and o-succinylbenzoate.* Biochemistry, 2000. **39**(35): p. 10662-76.

16. Odokonyero, D., et al., *Comparison of Alicyclobacillus acidocaldarius o-Succinylbenzoate Synthase to Its Promiscuous N-Succinylamino Acid Racemase/ o-Succinylbenzoate Synthase Relatives.* Biochemistry, 2018. **57**(26): p. 3676-3689.

17. Sanchez-Tarin, M., et al., *Enzyme promiscuity in enolase superfamily. Theoretical study of o-succinylbenzoate synthase using QM/MM methods.* J Phys Chem B, 2015. **119**(5): p. 1899-911.

18. Reisinger, B., et al., *A sugar isomerization reaction established on various (betaalpha)(8)-barrel scaffolds is based on substrate-assisted catalysis.* Protein Eng Des Sel, 2012. **25**(11): p. 751-60.

19. Dall'Acqua, W. and P. Carter, *Substrate-assisted catalysis: molecular basis and biological significance.* Protein Sci, 2000. **9**(1): p. 1-9.

CHAPTER V

EFFECT OF Y299I MUTATION ON PROTON ABSTRACTION IN *Alicyclobacillus acidocaldarius* OSBS [1]

Introduction

Alicyclobacillus acidocaldarius NSAR/OSBS (AaOSBS) is a member of the NSAR/OSBS subfamily. It shares 34% identity with the well-characterized AmyNSAR/OSBS. *A. acidocaldarius* has a menaquinone synthesis operon, but the gene that encodes AaOSBS is over 1300 kilobases away from the rest of the menaquinone synthesis genes. Also, the AaOSBS enzyme clusters in the phylogeny with other proteins whose biological functions are predicted to be NSAR activity, based on genome context. This raised the possibility that AaOSBS could be bifunctional, like *Geobacillus kaustophilus* NSAR/OSBS. However, the *aaOSBS* gene in *A. acidocaldarius* is between some ribosomal RNA genes and an operon comprised of putative cell division (*ftsE*, *ftsX* and *minJ*) and flagellar motor genes (*motA* and *motB*). The *aaOSBS* gene is not predicted to be co-transcribed with any of these genes [1-3]. Furthermore, no homolog of the *G. kaustophilus* succinyltransferase could be identified, and the two M20 family proteins encoded by *A. acidocaldarius* share only 28% and 33% identity with the *G. kaustophilus* L-desuccinylase. Thus, unless AaOSBS has an undiscovered activity, OSBS activity is

[1] Reprinted with permission from *Biochemistry* 2018, 57, 26, 3676-3689. Copyright 2018, American Chemical Society.

likely to be its only biological function. Assaying the enzyme from *Alicyclobacillus acidocaldarius* for OSBS and NSAR activities confirmed its predicted activities (Table V.1).

Table V.1. Enzymatic activities of AaOSBS variants

	OSBS activity			NSAR activity [a]		
Variant	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($M^{-1}s^{-1}$)	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($M^{-1}s^{-1}$)
WT	64 ± 3	118 ± 18	5.4×10^5	<0.0025	-	-
Y299I	91 ± 3	220 ± 18	4.2×10^5	0.22 ± 0.02	1800 ± 330	1.2×10^2

[a] *N*-succinyl-D-phenylglycine was the substrate.

AaOSBS is an efficient OSBS enzyme, but NSAR activity was below the limit of detection, using the substrates L-NSPG, D-NSPG, *N*-succinyl-L-phenylalanine, *N*-succinyl-D-phenylalanine, *N*-succinyl-L-valine, or *N*-succinyl-L-tryptophan. Thus, given the position of AaOSBS in the phylogeny, we suspect that it lost its ancestral NSAR activity. To determine if AaOSBS can bind D- or L-NSPG, we measured the inhibition of the OSBS reaction by these compounds. Lineweaver-Burk and Dixon plots are consistent with competitive inhibition. The K_I of L-NSPG and D-NSPG are 2600 ± 400 μM and 3600 ± 400 μM, respectively. The K_I values are ~20-fold higher than the K_M of SHCHC, suggesting that the affinity for succinylamino acids is much lower than that for

SHCHC. However, we would have expected to detect activity if only binding affinity were affected.

Based on the structure of AaOSBS, K170 abstracts a proton from the substrate, while K269 presumably stabilizes the transition state through a cation-π interaction [5]. Structural analysis of AaOSBS and docking study with NSAR substrate revealed that Y299 could sterically conflict with the NSAR substrate and limit the ability to properly position it for racemization (Figure V.1).

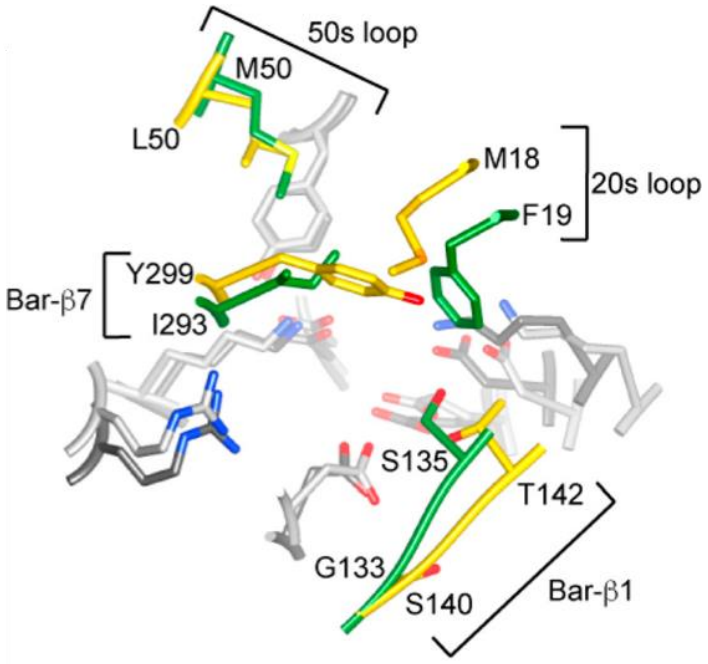


Figure V.1. Superposition of the active sites of AaOSBS and AmyNSAR (PDB entry 1SJB, chain B) [6]. Conserved amino acids are colored light gray (AaOSBS) or dark gray (AmyNSAR). Nonconserved residues (labeled) are colored gold (AaOSBS) or green (AmyNSAR). Bar-β1 and Bar-β7 are strands in the barrel domain. Alternate conformations of M18, M298, and Y299 in AaOSBS and some conserved active site residues have been omitted for the sake of clarity.

Based on sequence analysis, Y299 was mutated to isoleucine because 8 out of 9 known promiscuous NSAR/OSBS subfamily enzymes have leucine or isoleucine at this position. Kinetic data for AaOSBS Y299I is shown in Table V.1. Remarkably, Y299I increased NSAR activity with D-succinylphenylglycine from undetectable to 1.2×10^2 $M^{-1}s^{-1}$ without reducing OSBS activity. The increase in NSAR activity appears to be primarily from increasing k_{cat}. Because the $K_M^{D\text{-}NSPG}$ of Y299I is only marginally lower than the $K_I^{D\text{-}NSPG}$, replacing a bulky tyrosine with an isoleucine appears to allow the substrate to adopt a more appropriate orientation for catalysis, rather than merely increasing its binding affinity. This chapter will discuss the mechanism of how Y299I mutation introduces NSAR activity in AaOSBS.

Materials and Methods

Isotopic Exchange Experiments Using ^1H NMR Spectroscopy

AaOSBS variants were exchanged into D_2O using a Vivaspin Turbo 15 centrifugal filter (Sartorius). A 10 mL aliquot of protein was concentrated to 1 mL; 4 mL of 50 mM Tris (pD 8.0) was added, and the protein solution was again concentrated to 1 mL. This process was repeated three times to maximize the exchange. Samples for ^1H NMR (600 μL) contained 20 mM L- or D-*N*-succinylphenylglycine, 50 mM Tris (pD 8.0), 0.1 mM $MnCl_2$, and 0.5 mg/mL (11.7 μM) AaOSBS WT or AaOSBS Y299I in D_2O. The exchange was monitored by ^1H NMR (500 MHz Bruker NMR), following the

change in intensity of the α-proton as it was exchanged with deuterium over time. The peak for the α-proton (δ = 5.15 ppm) was integrated relative to that of the five aromatic protons (δ = 7.40 ppm). The relative peak area was converted to concentration based on the initial substrate concentration. The slopes of the plots of the NSPG substrate concentration as a function of time were used to obtain the isotopic exchange rates (k_{ex}).

Results and Discussion

To gain insight into the effect of Y299I, we measured the exchange (k_{ex}) of the proton on the α-carbon of D- or L-NSPG with deuterated solvent, which corresponds to the first step in the reaction mechanism (Table 4). In wild type AaOSBS, we observed slow proton-deuterium exchange with L-NSPG only. In contrast, a previous study reported that k_{ex} of AmyNSAR was >380 s^{-1} for both isomers.[5] AaOSBS Y299I catalyzes slow exchange with both D- and L-NSPG. These results indicate that AaOSBS binds L-NSPG in a position proximal to K269, which abstracts the α-proton from L-NSPG. The Y299I mutation does not alter k_{ex} when L-NSPG is the substrate, suggesting that the substrate is in similar positions relative to K269 in wild type and Y299I OSBS. Thus, the question is what prevents K170 from getting into the right position for the NSAR reaction?

Table V.2. Rate of proton-deuterium exchange (k_{ex}) for AaOSBS

	D-NSPG	L-NSPG
AaOSBS WT	-	0.29 s^{-1}
AaOSBS Y299I	0.59 s^{-1}	0.28 s^{-1}

The structural analysis above proposed several models to explain AaOSBS's lack of NSAR activity. The first model suggested that the active site is too wide for both catalytic lysines to effectively participate in acid-base catalysis. If this model were correct, the bulky tyrosine, in combination with L- or D-NSPG, which is slightly larger than the substrate of the OSBS reaction, could limit a conformation change that narrows the active site. Replacement of tyrosine with a smaller isoleucine could permit this conformation change and thus allow catalysis, as supported by the observed proton exchange and NSAR activity by AaOSBS Y299I. The second model suggested that, due to a larger conformation change upon substrate binding, Y299 sterically interferes with movement of the phenyl ring during catalysis. If this were correct, reaching the transition state would be impeded, resulting in low or undetectable proton exchange rates with both isomers in wild type AaOSBS. The Y299I mutation would also be expected to increase k_{ex} of both isomers, if it makes the transition state more accessible. Our results are not consistent with this model, because k_{ex} was detectable with only one isomer, and its rate did not increase. We also speculated that the electrostatic environment could limit the ability of K269 to act as a base. If that were the only limitation preventing

NSAR activity, we would have expected to observe proton exchange by K170, but not K269 in wild type AaOSBS. Because our results showed the opposite, the electrostatic environment cannot play a primary role in AaOSBS's lack of NSAR activity. We cannot rule out a secondary role for electrostatics, because the Y299I mutation would have a marginal effect on the electrostatic environment.

References

1. Dam, P., et al., *Operon prediction using both genome-specific and general genomic information.* Nucleic Acids Res, 2007. **35**(1): p. 288-98.
2. Mao, F., et al., *DOOR: a database for prokaryotic operons.* Nucleic Acids Res, 2009. **37**(Database issue): p. D459-63.
3. Mao, X., et al., *DOOR 2.0: presenting operons and their functions through dynamic and integrated views.* Nucleic Acids Res, 2014. **42**(Database issue): p. D654-9.
4. Odokonyero, D., et al., *Comparison of Alicyclobacillus acidocaldarius o-Succinylbenzoate Synthase to Its Promiscuous N-Succinylamino Acid Racemase/ o-Succinylbenzoate Synthase Relatives.* Biochemistry, 2018. **57**(26): p. 3676-3689.
5. Taylor Ringia, E.A., et al., *Evolution of enzymatic activity in the enolase superfamily: functional studies of the promiscuous o-succinylbenzoate synthase from Amycolatopsis.* Biochemistry, 2004. **43**(1): p. 224-9.

6. Thoden, J.B., et al., *Evolution of enzymatic activity in the enolase superfamily: structural studies of the promiscuous o-succinylbenzoate synthase from Amycolatopsis.* Biochemistry, 2004. **43**(19): p. 5716-27.

CHAPTER VI

CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions

In conclusion, this dissertation aids the understanding the evolution of NSAR activity and specificity in the NSAR/OSBS subfamily. We identified the first residue, R266, which is essential to determine NSAR reaction specificity in the catalytically promiscuous AmyNSAR/OSBS. Gaining an arginine at position 266 is a pre-adaptive feature enabling the emergence and evolution of NSAR activity in the NSAR/OSBS subfamily. The R266 residue modulates the reactivity of K263 but not K163, both of which serve different roles in catalysis in the OSBS and NSAR reactions. The R266Q mutation in AmyNSAR/OSBS significantly decreased NSAR activity with a marginal effect on OSBS activity. R266 is also important for the NSAR activity in other members of the NSAR/OSBS subfamily and important for the epimerase activity in the DE enzymes, another member of the MLE subgroup. The phenotypic effects of the R266Q substitution in those NSAR/OSBS enzymes are not as dramatic as observed in AmyNSAR/OSBS. Although the R266Q mutation decreased the NSAR activity much more than the OSBS activity in some members as expected, the mutation was also deleterious for OSBS activity. Furthermore, the R266Q mutation unexpectedly decreases the OSBS activity more than the NSAR activity in one member (LiOSBS/NSAR). Mechanistic investigations revealed that overall, the R266 residue decreased the

reactivity of the catalytic K263 more than that of K163. In some cases, the reactivity of both K163 and K263 decreased significantly, suggesting that the mutation might have a structural effect on the overall protein or the conformation of the active site. Thus, the phenotypic effects on the reactivity of the catalytic K263 are masked by the structural effects caused by the R266Q mutation in some NSAR/OSBS enzymes. The differential phenotypic effects of R266Q mutation are not the same in different enzymes, suggesting the contribution of epistasis in the evolution of NSAR activity. Because the effects of the R266Q mutation depend on the sequence backgrounds and structural contexts in which the mutation occurs, it is hard and complex to fully understand the evolution of NSAR activity in the NSAR/OSBS subfamily. However, gaining the R266 residue is essential to evolve NSAR activity. Some NSAR/OSBS enzymes also have R266 yet they have no detectable NSAR activity, suggesting that there are other structural features that contribute to NSAR specificity. It is essential to identify and characterize those residues to fully understand the evolution of NSAR activity in the NSAR/OSBS subfamily.

Catalytic promiscuity is important for the evolution of new enzymatic functions. However, the mechanism of how it allows the evolution of new functions in enzymes is complex, due to the contribution of epistasis, as in the case of NSAR activity in the NSAR/OSBS subfamily. This dissertation helps us understand that catalytic promiscuity is important for an enzyme to evolve a new function and is beyond the ability of the active site to bind different substrates. Other factors need to contribute to catalysis and allow the enzymes to be catalytically promiscuous, and ultimately to evolve a new function. Overall, this dissertation can be used to help explore the reaction specificity

determinants in other enzymes, to predict and assign functions for uncharacterized families in the enolase superfamily, and ultimately to aid the development of protein engineering and design methods.

Future Directions

Saturated mutagenesis at position 266 in other NSAR/OSBS and AEE enzymes

Investigating the effects of other mutations at the position homologous to R266 in other members of the NSAR/OSBS subfamily is important to determine whether specific effects of R266 on reactivity of the adjacent catalytic lysine can be disentangled from its other roles in catalysis, structure, and protein folding. The sequence backgrounds and structural contexts in these enzymes might hinder the observation of the specific effects of the R266Q mutation on K263. Thus, saturated mutagenesis at R266 in LvNSAR/OSBS, for instance, might help us identify other mutations that allow more specific effects on K263, as observed in AmyNSAR/OSBS. Eighteen other mutations at position 266 in LvNSAR/OSBS will be constructed and the mutants will then be expressed, purified, and assayed. Mutants that significantly decrease NSAR activity but marginally affect the OSBS activity will be selected for further characterization including k_{ex} measurements by NMR. Identifying such mutations at R266 is important to completely understand the evolution of NSAR activity in those enzymes. Furthermore, DE proteins were expressed from the *E. coli* strain that also

contains the native dipeptide epimerase and could have been co-purified during purification. It is essential to express and purify the mutant in a DE knockout strain to ensure no contamination of native enzyme. Saturated mutagenesis at position R266 in DE enzyme will also be carried out. Mutations in DE enzymes will be screened for the decreased DE activity and the k_{ex} values will be then measured by NMR. The mutations with a decreased reactivity of K263 but not K163 will be selected for further characterization.

Further characterization and directed evolution of TfuscaOSBS for NSAR activity

As shown in Chapter IV, D-NSPG, an NSAR substrate, can competitively bind and inhibit TfuscaOSBS, a non-promiscuous OSBS enzyme from the Actinobacteria subfamily. Further characterization of TfuscaOSBS is required to understand why it can bind D-NSPG but cannot catalyze the racemase conversion.

1. The first step is to re-examine the active site of TfuscaOSBS because the crystal structures of TfuscaOSBS are available. We can then compare the active site of TfuscaOSBS and the promiscuous AmyNSAR/OSBS to identify significant differences. Furthermore, the residues surrounding K263 in TfuscaOSBS should be closely examined because this lysine has different roles in each reaction. We can then mutate these residues in TfuscaOSBS to those in AmyNSAR/OSBS and test for NSAR activity.

2. We can also re-assay TfuscaOSBS with other succinyl- and acylamino acids because NSPG might not be the correct substrate for TfuscaOSBS.
3. The crystal structure of TfuscaOSBS with NSAR substrates can help us understand the factors that prevent NSAR activity in TfuscaOSBS. Alternatively, molecular modeling and docking experiments of NSAR substrates into the TfuscaOSBS active site could help to determine factors that might prevent NSAR activity. Those factors might include potential steric conflicts or the non-productive binding conformation of D-NSPG in the active site. Furthermore, the openness of the active site of TfuscaOSBS might not be suitable for NSAR activity.
4. Then, we can introduce an arginine mutation at position 266, which is a proline in TfuscaOSBS, to see if a positively charged residue at this position is sufficient for TfuscaOSBS to gain NSAR activity. Alternatively, the P266K mutation can also be made. The P266R (or P266K) TfuscaOSBS will then be expressed, purified, and assayed for both OSBS and NSAR activity. If it gains NSAR activity, k_{ex} values of the catalytic lysines will be determined. If it does not gain NSAR activity, we can further mutate other residues in the active site to mimic the active site of AmyNSAR/OSBS to determine how many mutations are required for TfuscaOSBS to gain NSAR activity.

Further characterizations of non-promiscuous NSAR/OSBS members

Not all members of the NSAR/OSBS subfamily are catalytically promiscuous, including *B.subtilis* OSBS. It is important to understand why they cannot catalyze NSAR reaction. Identification of the substitutions that are needed to allow NSAR activity to occur in those enzymes is essential to fully understand the evolution of NSAR activity.

1. First, inhibition studies of the OSBS reaction of BsubOSBS with NSAR substrates are required to test if it can bind D/L-NSPG (or even with other succinyl- and acylamino acids). If BsubOSBS can bind the NSAR substrates, we can then carry out similar experiments as proposed for TfuscaOSBS.
2. If it cannot bind the NSAR substrates, we then will compare the active sites of BsubOSBS and AmyNSAR/OSBS to identify differences that may be required for NSAR activity. We can then mutate the active site residues in BsubOSBS to those in AmyNSAR/OSBS. These mutations might allow BsubOSBS to bind and catalyze NSAR substrates.

Identification and characterization of other active site and non-active site residues that are important for NSAR activity

Thus far, we have only identified R266 in AmyNSAR/OSBS as important for NSAR activity and the Y299I mutation in AliacOSBS as required for gaining NSAR activity. Other additional residues must contribute to NSAR activity besides these two

residues. Thus, it is important to identify and characterize other active site and non-active site residues that are essential for NSAR reaction specificity. Sequence conservation and structural analysis in the active site and outside the active site of promiscuous and non-promiscuous enzymes will be done. The conserved residues of the NSAR/OSBS that are not present in other OSBS subfamilies will be focused and explored because they might contribute to NSAR catalysis. Furthermore, additional NSAR/OSBS genes from species of which the whole genomes are sequenced can be added to expand the phylogenetic analysis of the NSAR/OSBS subfamily. Thus, new additional conserved residues might arise and can be identified.

Attempt to identify an NSAR with no OSBS activity

Thus far, we have not identified an NSAR enzyme with no OSBS activity. Identification of such an enzyme will help us fully understand the evolution of NSAR activity. Such an enzyme might not exist because OSBS activity is an easy and thermodynamically favorable reaction. However, not all of the members of the NSAR/OSBS subfamily have been biochemically characterized. Thus, we need to characterize them all in hope of finding an NSAR with no OSBS activity. Because NSAR activity first appeared at the base of the NSAR/OSBS subfamily phylogenetic tree and became the biological function with OSBS activity as the promiscuous activity, the enzyme that are the latest to diverge or which are most divergent should be prioritized. Furthermore, genome context analysis and comparison of the expanding

sequence database will help us identify such enzymes. However, seeking the answer for this question might require a lot of effort. Hopefully with the available tools (and a little bit of luck), we can identify an NSAR with no OSBS activity to fully understand the evolution of NSAR activity.

Attempt to engineer NSAR/OSBS to catalyze "novel" reactions with unnatural substrates

The ultimate goal of studying enzyme evolution and reaction specificity evolution is to design and engineer "novel" enzymatic reactions. We understand the evolution of NSAR activity and NSAR specificity to a certain extent, at the current stage. Because L/D-NSPG and SHCHC are structurally similar, the remaining question is if we can engineer and manipulate an NSAR/OSBS enzyme to catalyze a novel racemase reaction on a chemically modified SHCHC that mimics L/D-NSPG and to catalyze a novel β-syn elimination reaction a chemically modified NDPG mimicking SHCHC.

1. Because SHCHC and L/D-NSPG are similar in structures, chemically modified SHCHC will potentially be good substrates for novel racemase activity by an NSAR/OSBS, and vice versa, chemically modified NSPG will also be good substrates for β-syn elimination reaction by an NSAR/OSBS (Figure VI.1).
2. SHCHC derivatives including 2-succinyl-6*R/S*-methyl-2,4-cyclohexadiene-1*S/R*-carboxylate (SMCHC) and 2-succinyl-6,6-dimethyl-2,4-cyclohexadiene-1*S/R*-

carboxylate (SDMCHC) can be chemically synthesized. These two SHCHC derivatives, having one or two methyl functional groups, do not have the γ-hydroxyl leaving group; thus, the β-syn elimination reaction will not happen. Instead, after the proton abstraction by one lysine, the other acidic lysine can re-protonate the enolate intermediate to assist the racemization reaction. Alternatively, we can modify the C_γ with different functional groups and test which substrates are best for the racemase activity. Another option is to modify NSPG to carry a β-hydroxyl leaving group, since the double bonds in the ring might help facilitate the departure of water because the conjugated product might be favorable. Alternatively, we can modify the C_ε of the ring with different functional groups to make the leaving group departure more favorable.

3. We can then test for racemase activity by a polarimeter with AmyNSAR/OSBS, which is the best candidate enzyme. On the other hand, to test for the β-elimination activity, we can test for UV-Vis absorbance for the NSPG derivatives and use UV-Vis spectrometer.
4. We then can carry out the proton-deuterium exchange between the H_α of these substrate derivatives and the catalytic lysines and measure k_{ex} values by NMR spectroscopy.

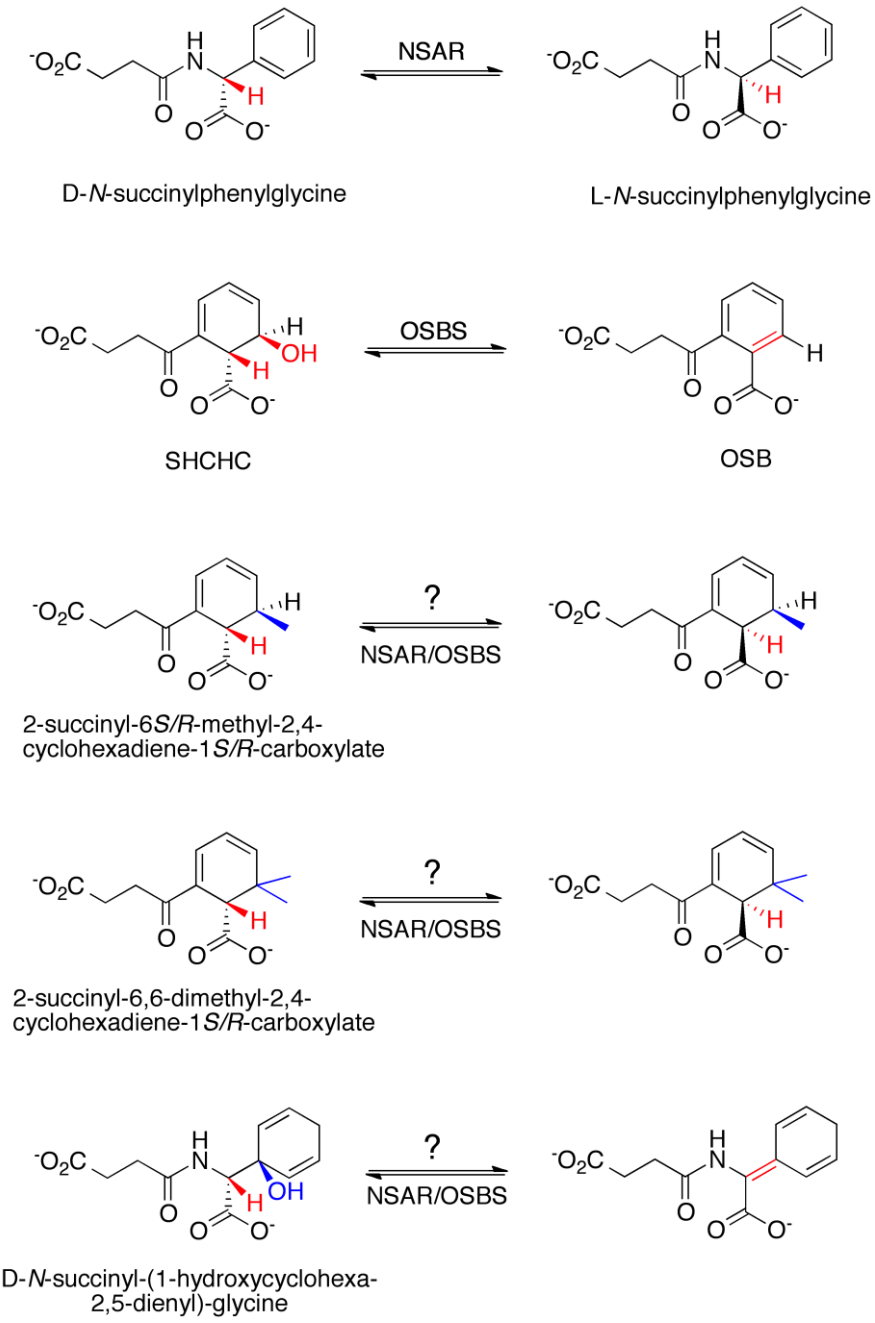


Figure VI.1. SHCHC (OSBS substrate) and NSPG (NSAR substrate) derivatives for "novel" reactions by NSAR/OSBS enzymes.

In summary, we are beginning to understand the mechanistic basis and the essential roles of promiscuity in the evolution of the new enzyme functions, especially those from mechanistically diverse enzyme superfamilies. However, even with the established sequence-structure-function relationships, reaction specificity determinants in such enzymes, including our model system (NSAR/OSBS subfamily), have not been fully explored. Thanks to the powerful and diverse biochemical methods, evolutionary biochemists have extraordinary opportunities to explore and understand the vast and diverse universe of enzymes. Greater efforts are required to further develop more powerful, sensitive, and applicable high-throughput genomic screening methods to identify and to determine reaction promiscuity and specificity. Doing so will not only help us detect but also optimize useful promiscuous activities for designing and engineering novel enzymes.