THE ROLE OF EPISTATIC INTERACTIONS IN THE EVOLUTION OF NEW ENZYME FUNCTIONS IN THE ENOLASE SUPERFAMILY

A Thesis

by

SUSAN CAMPBELL FULTS

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

MASTER OF SCIENCE

Chair of Committee,	Margaret Glasner
Committee Members,	Vishal Gohil
	Frank Raushel
Head of Department,	Josh Wand

August 2020

Major Subject: Biochemistry

ABSTRACT

The factors that enable enzymes to evolve new functions and specificities are not well understood. Two factors that influence enzyme evolution are enzyme promiscuity and epistasis. Enzymes are said to be promiscuous if their active site has the ability to catalyze more than one reaction. Epistasis occurs when amino acid substitutions have different effects when placed in different sequence contexts. Understanding the role that factors like enzyme promiscuity and epistasis play in the evolution of new enzyme functions will enable scientists to design enzymes with new functions. Here, I seek to understand how promiscuity and epistasis have influenced the evolution of enzymes within the enolase superfamily. I examine the differing roles that the conserved residues G254 and R128 play in two enzymes in the OSBS family, *T. fusca* and *E. coli* OSBS. I also examine the role that a conserved second shell amino acid, R266, played in the evolution of racemase and epimerase activities in the enolase superfamily.

DEDICATION

I dedicate this work to my dear husband, Matthew Glen Fults, who loves me unconditionally and pushes me to make my dreams become realities.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Glasner, for her guidance and support during my time as a member of her lab. Thank you for helping me to expand my knowledge and realize my potential.

Thank you to all of my colleagues from the Glasner lab who have become very dear friends and mentors. I would not have been able to complete the work presented here without you.

Finally, thanks to all of the members of Bunny Buddies. Our time spent together helping rescue rabbits provided me with much-needed direction and purpose during this challenging period of my life.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Dr. Margaret E. Glasner (as the committee chair), Dr. Vishal Gohil, and Dr. Frank Raushel of the Department of Biochemistry and Biophysics.

The NMR instrument used to measure proton exchange between the enzymes and the substrates shown in Chapter III was operated by Dr. Jamison Huddleston of Dr. Frank Raushel's lab in the Department of Chemistry. The NSAR substrates (*N*-succinyl-D/L-phenylglycine) used in this thesis were synthesized by Mingzhao Zhu and Kenneth Hull of Dr. Daniel Romo's lab in the Department of Chemistry and Biochemistry, Baylor University.

The N73A/T75L variant of *Thermobifida fusca* used to construct the N73/T75L/R128 variant of *T. fusca* used in Chapter II was constructed by Mr. Denis Odokonyero of Dr. Margaret Glasner's lab in the Department of Biochemistry and Biophysics.

The K250Q variant of the *Escherichia coli* dipeptide epimerase shown in Chapter III was constructed by Dr. Dat Truong, a former graduate student of Dr. Margaret Glasner's lab of the Department of Biochemistry and Biophysics. The rest of the K250 *E. coli* dipeptide epimerase variants used in Chapter III were constructed by Kyle Glockzin of Dr. Margaret Glasner's lab of the Department of Biochemistry and Biophysics and Abhinav Swaminathan, a former rotation student off Dr. Margaret Glasner's 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 of Dr. Jean-Philippe Pellois's lab of the Department of Biochemistry and Biophysics.

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

Funding Sources

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

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Factors that Influence Enzyme Evolution

Enzymes can be engineered to have new functions and specificities through rational design approaches, in which directed mutations based on structural and sequence data are made to an enzyme to alter function or specificity[1]. The effects of mutations on function and specificity are often unpredictable, complicating rational enzyme design approaches. The inability to consistently make accurate predictions about how mutations will alter function and specificity limits the current methods used to engineer enzymes[2, 3]. Developing a better understanding of how function and reaction specificity are determined will enable scientists to more readily design enzymes with new functions and specificities. Enzymes have been able to evolve many new and complex functions and specificities naturally, so understanding how these natural evolutionary processes occur may enable scientists to replicate the same processes using artificial means.

One natural phenomenon that evolution is thought to exploit on the path to new enzyme functions is the phenomenon known as enzyme promiscuity. Contrary to what is taught in entry-level science classes, not all enzymes are highly specific to only one reaction and/or substrate; some enzymes are promiscuous and catalyze multiple reactions and/or utilize multiple substrates. A promiscuous enzyme is an enzyme whose active site has the ability to catalyze a native biologically relevant reaction as well as

other reactions that are not biological functions[4, 5]. Some promiscuous enzymes display promiscuity by utilizing multiple substrates to catalyze similar reactions, while other promiscuous enzymes display promiscuity by catalyzing chemically distinct reactions[6]. Biological functions are under natural selection pressures because they contribute to the fitness of the organism, and promiscuous functions do not contribute to fitness unless a change in natural selection pressures occurs that results in the promiscuous function conferring increased organismal fitness[4, 7, 8].

When promiscuous functions undergo natural selection, they can become the biological function of the enzyme. It has been proposed that all extant highly specific enzymes have evolved from primordial enzymes that were highly promiscuous[8]. Enzyme promiscuity is not only a trait of primordial enzymes; growing evidence points to enzyme promiscuity being a widespread phenomenon among modern enzymes[6, 9]. If all enzymes began as promiscuous enzymes that later evolved to be more specific, then it raises the possibility that existing promiscuous activities can be exploited in the laboratory to gain new desired enzyme functions. Several experiments exploited naturally occurring promiscuous reactions to produce new enzyme functions or become better at catalyzing existing promiscuous activities[7, 10, 11] . Studying enzymes with these promiscuous activities can help advance knowledge of how enzymes evolve new functions and specificities, and in turn help advance the ability to design enzymes with new functions and specificities.

Another property that influences the evolution of new enzyme functions is epistasis[12-14]. Epistasis occurs when amino acid substitutions have differential effects

when placed in different sequence contexts[13-15]. The biophysical basis of epistatic interactions is poorly understood[13, 16]. Epistatic interactions can be said to be local, meaning they are short range and occur between neighboring amino acids, or global, meaning they are long range and occur between distant amino acids within the protein[13]. Long range epistatic interactions can lend increased stability to an enzyme that can allow amino acid changes that destabilize the enzyme but lead to innovations in reactivity or substrate binding[13]. The work presented here seeks to understand the types of epistatic interactions at play in different sequence and structural contexts and how they impact the evolution of new enzyme functions.

The Enolase Superfamily

This thesis will investigate the epistatic interactions at play in the evolution of specificity by investigating enzymes in the mechanistically diverse enolase superfamily. Enzymes in the enolase superfamily are grouped together because they share a similar partial reaction (Fig. 1). A base abstracts an alpha-proton from a carboxylate and generates an enolate intermediate that is stabilized by coordination to a Mg^{2+} ion[17, 18]. Using these shared steps, the enzymes in the enolase superfamily carry out a vast array of different reactions utilizing different substrates and producing different products[17, 18]. The enolase superfamily contains enzymes that catalyze many different reactions, including: enolase, madelate racemase, muconate lactonizing enzyme I, muconate lactonizing enzyme II, D-glucarate dehydratase, D-glucarate dehydratase, *o*-succinylbenzoate synthases, L-Ala-D/L-Glu epimerases, and 3-methylaspartate ammonia

lyase[17]. The enolase superfamily has been divided into subgroups of enzyme families based on their sequences and their catalytic residues[18]. One of the subgroups is the muconate lactonizing enzyme (MLE) subgroup. In members of the MLE subgroup, a lysine residue on the end of the second or sixth β-strand of the TIM-barrel domain abstracts the proton of the carbon acid substrate and the ligand at the end of the fifth β-strand is always an aspartic acid followed by a glutamic acid[17]. The MLE subgroup is subdivided into the *o*-succinylbenzoate synthase (OSBS) family, dipeptide epimerase family, 4R-hydroxyproline betaine 2-epimerase family, and the MLE 1 and 2 families[17, 19]. Two enzyme families in the MLE subgroup, the OSBS family and the dipeptide epimerase family, are the focus of this thesis[17, 20, 21].

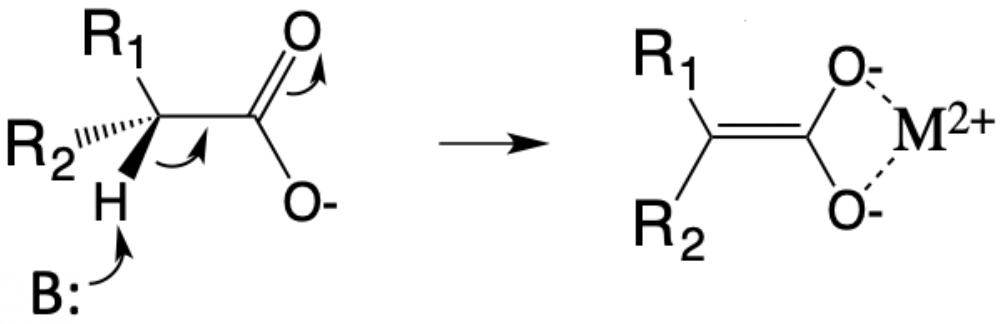

Figure 1: The common partial reaction shared by members of the enolase superfamily.

One of the families of the MLE subgroup is the OSBS family. Enzymes in this family carry out *o*-succinylbenzoate synthesis, which is a step in the menaquinone synthesis pathway (vitamin K_2) (Fig. 2). Two catalytic lysine residues, K163 and K263, in the active site are required for the OSBS reaction. K163 acts as a general acid/base in the OSBS reaction by abstracting the alpha proton of 2-succinyl-6-hydroxy-2,4-

cyclohexadiene-1-carboxylate (SHCHC) to create an enolate intermediate that is stabilized by the Mg^{2+} ion. The K163 residue then donates the proton to the intermediate's hydroxyl leaving group in order to form water and produce OSB. The K263 residue likely forms a π-cation interaction with the intermediate and SHCHC. Despite all carrying out the same reaction, members of the OSBS family are extremely divergent, with some members sharing less than fifteen percent sequence identity with one another[20, 22]. The structures of these enzymes also vary widely[20, 23].

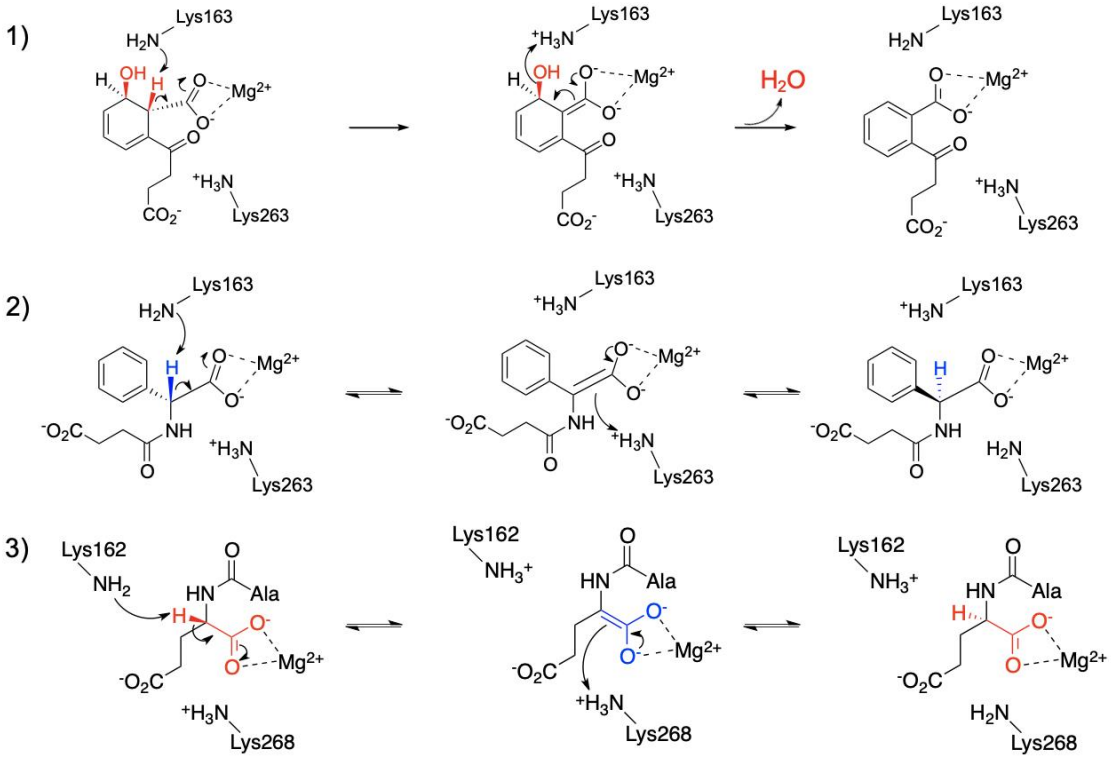

Figure 2: 1) The OSBS reaction mechanism. 2) The NSAR reaction mechanism. 3) The dipeptide epimerase reaction mechanism when utilizing L-Ala-D/L-Glu as the substrate. Residue numbering is shown for *Amycolatopsis* sp. T-1-60 NSAR/OSBS for the OSBS and NSAR reactions and for *B. subtilis* dipeptide epimerase for the dipeptide epimerase reaction.

The OSBS family of enzymes is divided into several divergent subfamilies, one of which is the *N*-succinylamino acid racemase/*o*-succinylbenzoate synthase (NSAR/OSBS) subfamily. Based on experimentally determined activities and the phylogeny of the subfamily, NSAR activity appeared early on in the subfamily's divergence[24]. Most of the enzymes within the NSAR/OSBS subfamily carry out *o*-succinylbenzoate synthesis. Some of the enzymes within the NSAR/OSBS branch of the OSBS subfamily are promiscuous and also carry out *N*-succinyl amino acid racemization, which, in some species, is involved in a pathway to convert D-amino acids to L-amino acids (Fig. 2)[20, 25]. The NSAR reaction utilizes the same two catalytic lysine residues, K163 and K263, as the OSBS reaction (numbered according to the *Amycolatopsis* sp. T-1-60 NSAR/OSBS)[26]. In the NSAR reaction, when a D-succinyl amino acid is used as the substrate, K163 acts as a general base and K263 acts as a general acid in a two-base, 1-1 proton transfer mechanism. When an L-succinyl amino acid is used as the substrate, K163 acts as a general acid and K263 acts as a general base. Within the NSAR/OSBS subfamily, for some enzymes, the OSBS reaction is the biological function and the NSAR reaction is a promiscuous side reaction, and for other enzymes, the OSBS reaction is a promiscuous side reaction and the NSAR reaction is the biological function of the enzyme[24, 25, 27].

The dipeptide epimerase family is another enzyme family in the MLE subgroup. Enzymes in this family follow a similar mechanism to the NSAR reaction by utilizing a two-base, 1-1 proton transfer (Fig. 2)[21]. Dipeptide epimerase enzymes catalyze an epimerase reaction utilizing a diverse range of dipeptide substrates[21, 28, 29]. Dipeptide

epimerase enzymes have been shown to utilize several substrates including L-Ala-D/L-Glu, L-Ala-D/L-Phe, and L-Ser-D/L-Glu, and most of the characterized dipeptide enzymes utilize L-Ala-D/L-Glu[21, 28, 29]. As the dipeptide epimerase family is a member of the MLE subgroup, it shares some sequence similarities with the OSBS family. Members of these enzyme families likely share a common ancestry and evolved over time to be specific to one type of reaction[30]. The dipeptide epimerase from *Escherichia coli* was able to carry out the OSBS reaction after only one mutation, which decreased the dipeptide epimerase's affinity for its native substrate and allowed it to bind SHCHC, pointing to an ability for related enzymes to easily evolve new specificities with few needed substitutions[31].

Context Dependence of Conserved Residues in the OSBS Family

Because of epistatic interactions, conserved amino acids do not always play the same role in homologous proteins due to sequence divergence and structural differences. The OSBS family has high sequence divergence—only eight amino acids are completely conserved within the family[32]. These eight amino acids are located in the active site. They are not sufficient to confer the ability to carry out the OSBS reaction, as they are conserved in other members of the enolase superfamily, including the dipeptide epimerase family[20]. How are these enzymes able to carry out the same reaction when they differ so much in sequence and structure? One possibility is that the OSBS reaction may not require additional conserved residues because it is a relatively easier reaction than other reactions in the enolase superfamily[33]. Another possibility is that additional

residues contribute to substrate binding and orientation, but these residues have diverged within different lineages in the OSBS family[34].

Two additional amino acids located in the active site are conserved in all OSBS enzyme subfamilies except for the NSAR/OSBS subfamily (Fig. 3). In *E. coli* OSBS, these conserved residues are R159 and G288. R159 forms a water-mediated salt-bridge with the substrate, and if G288 is mutated, there is a steric clash with the substrate (Fig. 4)[34]. Mutating R159 and G288 in *E. coli* OSBS was detrimental to reaction efficiency[34]. When the corresponding sites, R128 and G254, were mutated in *Thermobifida fusca* OSBS, the mutations only slightly impacted activity[32]. Although the two sites are conserved in both *E. coli* OSBS and *T. fusca* OSBS, the other residues within the two enzymes are highly divergent[32, 34]. The structures of the two enzymes also greatly differ due to many insertions and deletions[32]. In *T. fusca* OSBS, there are two residues, N73 and T75, that form hydrogen bonds with the succinyl carboxylate[32]. The residues at the homologous positions in *E. coli* OSBS, A107 and L109, do not have these same hydrogen bonds[32]. To examine if amino acids on β-strand 1 in the barrel make significant contributions to function in the two enzymes, an N73A/T75L variant was made of *T. fusca* OSBS, and an A107N/L109T variant was made of *E. coli* OSBS[32]. These mutations in *T. fusca* OSBS reduced activity 33-fold, while the mutations in *E. coli* OSBS reduced activity 385-fold[32]. This data shows that the amino acids on β-strand 1 make significant contributions to function and are not interchangeable between different enzymes in the OSBS family. It is possible that epistatic interactions that differ between *E. coli* OSBS and *T. fusca* OSBS allow conserved residues to play different

roles. These epistatic interactions will be further explored in Chapter 2 of this thesis by examining how the R128M behaves in the N73A/T75L background in *T. fusca* OSBS. The way that the G254A mutations impacts *T. fusca* OSBS activity is also further explored in Chapter 2.

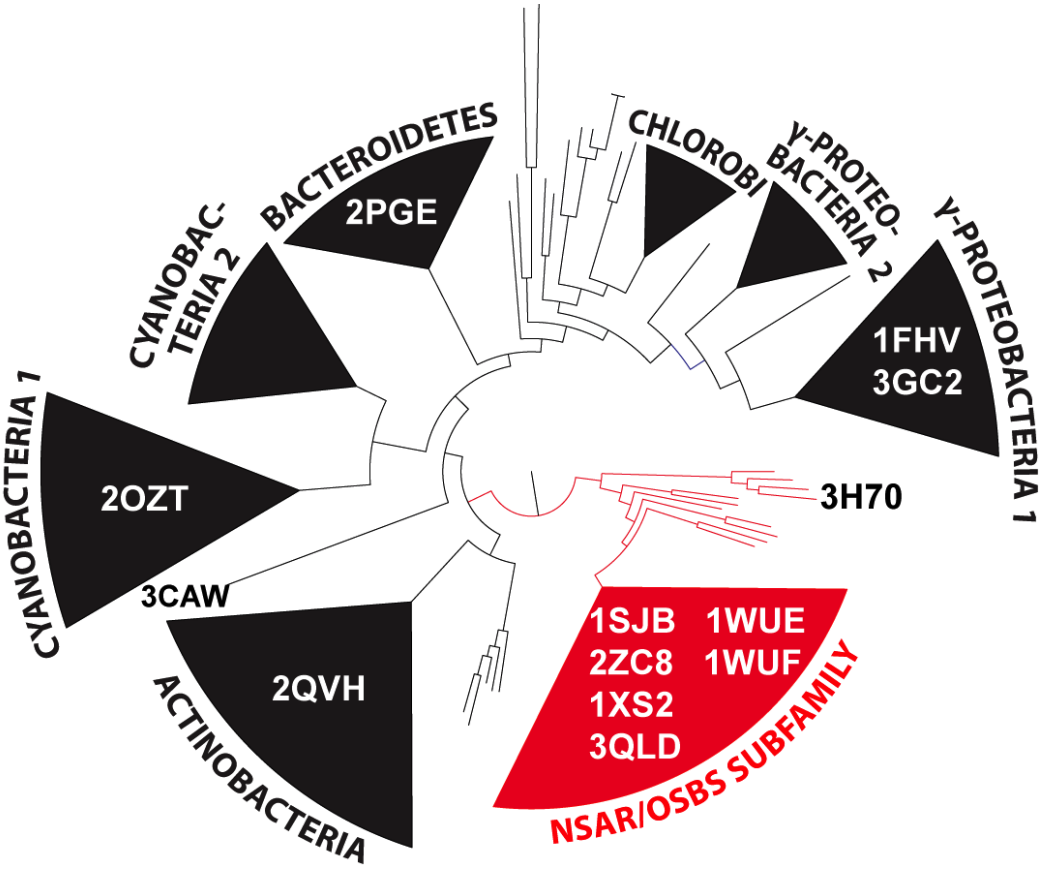

Figure 3: A phylogenetic tree of the OSBS family which illustrates the division of the family into subfamilies[34]. Reprinted with permission from Zhu, W. W., Wang, C., Jipp, J., Ferguson, L., Lucas, S. N., Hicks, M. A., and Glasner, M. E. (2012) Residues required for activity in Escherichia coli o-succinylbenzoate synthase (OSBS) are not conserved in all OSBS enzymes, *Biochemistry 51*, 6171-6181. Copyright 2012 American Chemical Society.

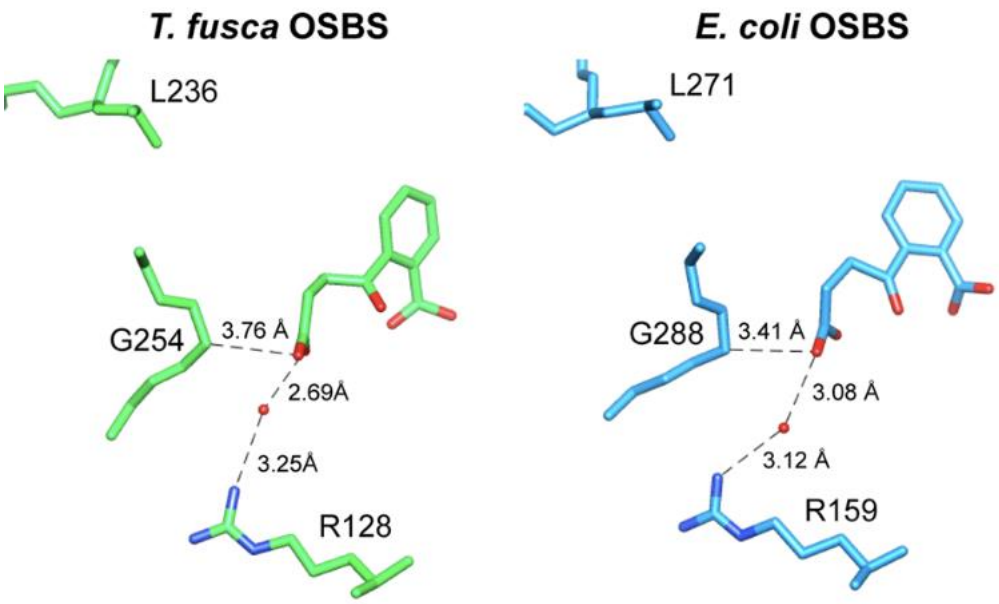

Figure 4: Comparison of the position of the ligand relative to R159 and G288 in the *E. coli* OSBS enzyme and R128 and G254 in *T. fusca* OSBS. Reprinted with permission from Odokonyero, D., Ragumani, S., Lopez, M. S., Bonanno, J. B., Ozerova, N. D., Woodard, D. R., Machala, B. W., Swaminathan, S., Burley, S. K., Almo, S. C., and Glasner, M. E. (2013) Divergent evolution of ligand binding in the o-succinylbenzoate synthase family, *Biochemistry 52*, 7512-7521. Copyright 2013 American Chemical Society.

Role of a Conserved Second Shell Amino Acid in Racemase Activity of NSAR/OSBS and Dipeptide Epimerase Enzymes

In an effort to determine what enabled some of the enzymes within the NSAR/OSBS subfamily to gain the ability to catalyze the NSAR reaction, sequence conservation between other subfamilies of the OSBS family and the NSAR/OSBS subfamily was compared[35]. Examining sequence conservation revealed that arginine is highly conserved at the 266 position in the NSAR/OSBS subfamily (red in Fig. 3), while arginine is not present at this position in other OSBS subfamilies (black in Fig. 3). This

difference in sequence conservation suggested that this site may be important for the ability to carry out the NSAR reaction. To test this hypothesis, R266 in the *Amycolatopsis* sp. T-1-60 NSAR/OSBS enzyme was mutated to a glutamine. NSAR activity was greatly decreased by mutating this residue to glutamine, while OSBS reactivity was only moderately decreased[35]. To better understand how the NSAR reaction was impacted by the mutation, isotopic exchange experiments were performed on the wildtype and R266Q *Amycolatopsis* sp. T-1-60 variants. Isotopic exchange using ^{1}H NMR allows for the observation of the hydrogen exchange rate of the substrate with the catalytic K163 and K263 residues (Fig. 2). The fold change in the exchange rate for K163 was 3.6-fold lower than wildtype, while the fold change in the exchange rate for K263 was 1340-fold lower than wildtype[35]. This data indicates that the R266Q mutation reduces NSAR reactivity mostly by affecting the proton exchange with K263. Examining the crystal structures of both the wildtype and mutant enzymes reveals that K263 is not positioned correctly in the mutant for proton abstraction or donation (Fig. 5) [35]. This data suggests that gaining R266 was a pre-adaptive mutation that allowed the emergence of NSAR activity and the evolution of the NSAR/OSBS subfamily. If R266 is generally required to modulate K263 reactivity in the NSAR/OSBS subfamily, then the R266Q mutation should have the same effects in other members of the NSAR/OSBS subfamily.

To determine if the impact of the R266Q mutation observed in the *Amycolatopsis* sp. T-1-60 NSAR/OSBS enzyme is general to other NSAR/OSBS enzymes, the equivalent mutation was made in several other NSAR/OSBS enzymes[35]. Chapter 3

describes the effects of this mutation in one of these NSAR/OSBS enzymes, *Lysinibacillus varians* NSAR/OSBS. By examining the observed phenotype of the R266Q in other sequence backgrounds, possible epistatic interactions modulating the impact of the mutation can be elucidated. More details about these previous findings are found in Chapter 3.

A positively charged residue occurs at the same position in the dipeptide epimerase family as R266 in the NSAR/OSBS subfamily, suggesting that the same site could impact the reactivity of the catalytic acid/base that corresponds to K263 in other racemase enzymes within the enolase superfamily (Fig. 5). The dipeptide epimerase subfamily of the enolase superfamily carries out a similar reaction to the NSAR reaction (Fig. 2). In the dipeptide epimerase family, there is a highly conserved lysine residue at the position homologous to R266 in the NSAR/OSBS enzymes. Both of these amino acids are positively charged, suggesting that the charge of the amino acid at that site might be important for racemase and epimerase activity. The residues homologous to R266 in the *Bacillus subtilis* dipeptide epimerase (K271) and the *E. coli* dipeptide epimerase (K250) were mutated to glutamine to examine if the conserved residue plays a similar role in the NSAR/OSBS subfamily and the dipeptide epimerase family. The K271Q mutation decreased k_{cat} 93-fold for the *B. subtilis* dipeptide epimerase[35]. In the *E. coli* dipeptide epimerase, the K250Q mutation decreased k_{cat} by 44-fold [35]. Neither of the enzymes had a change in the K_M value between the wildtype enzyme and the mutant enzyme[35]. Isotopic exchange experiments to examine the reactivity of the catalytic lysines, as described above, were also conducted on both enzymes. The K271Q mutation

in the *B. subtilis* dipeptide epimerase decreased the reactivity of the catalytic acid/base K268 6 times more than the other catalytic acid/base, K162. This data agrees with the hypothesis that the K271Q mutation would preferentially impact the reactivity of the K268 residue, as that was what was observed in *Amycolatopsis* sp. T-1-60 NSAR/OSBS[35]. However, the K250Q mutation in the *E. coli* dipeptide epimerase caused only a 2-fold decrease in the reactivity of K247 and no decrease of the reactivity of K151, which was not significant enough to support the hypothesis[35]. The *E. coli* K250Q dipeptide epimerase was purified from *E. coli* cells that still contained the native *E. coli* dipeptide epimerase gene, which may have caused contamination issues that masked the effect of the K250Q mutation. In Chapter 3 of this thesis, I further evaluate the role of the conserved position, K250, in *E. coli* dipeptide while expressing the enzyme in a strain in which the native dipeptide epimerase gene was deleted.

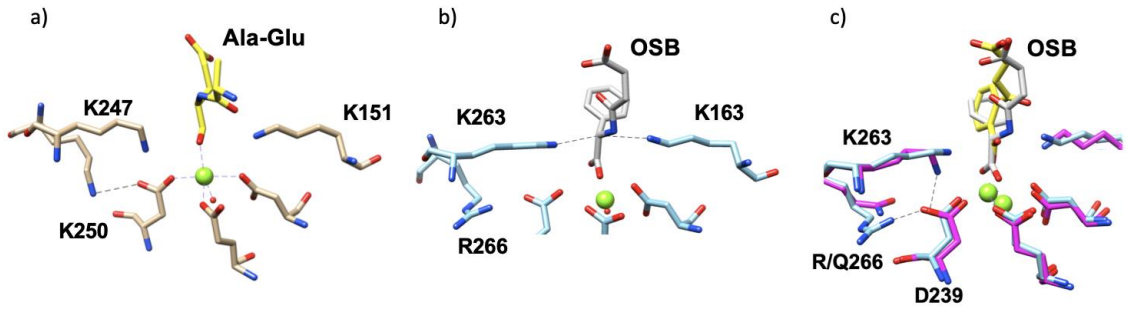

Figure 5: Comparison of the active sites of a) *E. coli* dipeptide epimerase(PDB 1JPD)[36] and b) *Amycolatopsis* sp. T-1-60 NSAR/OSBS (PDB 1SJB)[26] c) the superimposed structures of *Amycolatopsis* sp. T-1-60 NSAR/OSBS WT and R266Q. WT is shown in cyan (PDB ID 1SJB) and R266Q is shown in magenta. OSB, the substrate for the OSBS reaction, is shown in yellow. NSPG, the substrate for the NSAR reaction, is shown in white. Salt bridges are shown as dotted lines[35] Reproduced from Truong, D. P. (2020) Catalytic Promiscuity and the evolutionary mechanism of NSAR reaction specificity in the NSAR/OSBS subfamily. *PhD Dissertation.* Texas A & M University College Station, TX. Copyright 2020 Dat Truong.

Goal of Thesis Research

The goal of the work presented in this thesis is to gain a better understanding about the role and nature of epistatic interactions in the evolution of new enzyme functions, using the OSBS and dipeptide epimerase enzyme families as model enzymes. Chapter 2 addresses how the role of two conserved residues in the OSBS family, R128 and G254 (numbering from *T. fusca* OSBS), differs in enzymes within the family. Chapter 3 explores the role that epistatic interactions play in the evolution of different activities within the enolase superfamily by examining the role of the conserved second-shell amino acid, R266 in the NSAR/OSBS family and the homologous K250 in the dipeptide epimerase family. Together, these studies shed light on the nature of epistatic interactions of residues in or near the active that have a direct effect on ligand binding and catalysis.

CHAPTER II

EPISTATIC INTERACTIONS ALTER THE EFFECTS OF MUTATIONS IN OSBS ENZYMES

Differing Roles of Conserved Amino Acids in *T. fusca* and *E. coli* OSBS enzymes

Amino acids may be conserved in related enzymes due to several reasons. Conserved amino acids may be directly involved in ligand binding, catalysis, or protein-protein interactions. They also may be indirectly involved in enzyme function by impacting folding, stability or dynamics. One might expect that these conserved amino acids play similar roles in homologous enzymes; however, the unique sequence and structure contexts of each enzyme may alter the role that these conserved amino acids play. The *o*-succinylbenzoate synthase (OSBS) family is an excellent model system for examining how different structural and sequence contexts alter the role of conserved residues, as enzymes within the family are highly divergent and contain few conserved residues.

Previously, it has been reported that two amino acids conserved in the majority of enzymes in the OSBS family, corresponding to G254 and R128 in *Thermobifida fusca*, play different roles in different family members[32]. G254 is only 3.8 Å from the succinyl carboxylate of 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate (SHCHC)[32]. R128 is further away, 5.9 Å, but is linked to the succinyl carboxylate of SHCHC by hydrogen bonds to an intervening water molecule (Fig. 4)[32]. Mutations at the sites homologous to G254 and R128 were deleterious to OSBS activity in

Escherichia coli OSBS, while only having a small impact on OSBS activity in *T. fusca* OSBS[32, 34]. *T. fusca* OSBS and *E. coli* OSBS share only 22% sequence identity and have large insertions and deletions relative to each other[32]. Due to these differences in sequences of the two enzymes, it is likely that epistatic interactions are present in *T. fusca* OSBS that are not present in *E. coli* OSBS that have allowed *T. fusca* OSBS to better tolerate mutations at these sites.

As a component of a broader project to understand how epistatic interactions may change the role of conserved amino acids in homologous enzymes, I conducted several experiments. First, I revisited the activity of the G254A mutation, because the plasmid stock was found to be contaminated with a plasmid encoding the wildtype *T. fusca* OSBS. I also examined the role that the R128 site plays in the *T. fusca* OSBS by examining how other mutations within the enzyme, N73A and T75L, modulate effects of the R128M mutation. Lastly, I examined how temperature affects the activity of the wildtype *T. fusca* OSBS and the G254A *T. fusca* OSBS, to determine if elevated temperatures, reflecting the environment of the thermophile *T. fusca*, would make the G254 mutation more deleterious. These experiments, combined with additional experiments carried out by other lab members, help to reveal the effects of epistasis on the structure, thermostability, and activity of enzymes.

Materials and Methods

Mutagenesis

Site-directed mutagenesis was performed using the Q5 mutagenesis protocol (New England BioLabs). The template used was *T. fusca* OSBS (gi number 158430464, PDB code 2QVH) that was subcloned into pet15b. All mutations were confirmed by sequencing in both directions (Eurofins Genomics LLC). The primers used for Q5 mutagenesis were designed using NEDBaseChanger, NEB's online design software (NEBasechanger.com). I performed mutagenesis for the G254A and N73A/T75L/R128M variants using the primers found in Table 1. The template used to generate the N73A/T75L/R128M variant was the previously made N73A/T75L variant (Denis Odokonyero).

Table 1: Primers used to generate the G254A and N73A/T75L/R128M *T. fusca* OSBS variants. All DNA primers are shown in the 5' to 3' direction. The lower-case bases designate the codons where mutations were introduced.

Mutation	**Primers Used**
G254A	Forward: ATACGCTTGTgcgCTGGCAACTCTGCG Reverse: GGCAGTTCCGGCAGTGCA
N73A/T75L/R128M	Forward: TGAGTCTATCatgAGAGCGGAAGACCCTTTG Reverse: TCGGCCGCTATTGGTACG

Protein Purification

All proteins were expressed in the *menC*$^{-}$ strain BW25113 (*menC::kan*, DE3), so that purified proteins were not contaminated with native *E. coli* OSBS[34]. Cell cultures were grown without induction in Luria-Bertani broth supplemented with carbenicillin and kanamycin at a final concentrations of 50 μg/mL overnight at 37 °C. Cells were harvested by centrifugation and resuspended in a buffer containing 20 mM Tris, pH 8.0, 500 mM NaCl, 5 mM imidazole, 0.02 mg/ml DNAse (Worthington), and 2 μM phenylmethylsulfonyl fluoride (PMSF; Thermo Scientific). The resuspended cell pellets were lysed using sonication. The cell lysate was centrifuged, and the supernatant was loaded onto a 5 mL HisTrap FF column charged with Ni^{2+} (GE Healthcare). The protein

was eluted from the column with a buffer that contained 20 mM Tris, pH 8.0, 500 mM NaCl and 500 mM imidazole using a step to 15% elution buffer followed by a linear gradient to 100% elution buffer. Fractions containing homogenous protein were identified using SDS-PAGE and then pooled. Amicon Ultra-15 centrifugal filters with a 10 kD cutoff (Millipore) were used to exchange the buffer and concentrate the pooled fractions. Concentrated purified proteins were stored at -80 °C in 10 mM Tris, pH 8.0, 5 mM $MgCl_2$, and 25% glycerol.

OSBS Activity Assay

2-succinyl-6hydroxy-2,4-cyclohexadiene-1-carboxylate (SHCHC) was prepared using methods as previously described[34]. Enzyme activity was measured in 50 mM Tris pH 8.0 and 0.1 mM $MnCl_2$. Activity was determined by measuring the disappearance of SHCHC at 310 nm ($\Delta\varepsilon$=-2400 $M^{-1}cm^{-1}$) over time[22, 26]. Assays performed at 25 °C were done using a SpectraMax Plus^{384} UV-VIS microplate spectrophotometer (Molecular Devices). Assays at temperatures other than 25 °C were performed using a UV-VIS spectrometer with temperature control. The Tris buffer was modified for each experiment at differing temperatures, so that the pH was 8.0 at the assay temperature. Initial rates were determined by fitting a line to the linear portion of the curve using Excel (Microsoft). Initial rates at varying substrate concentrations were then fit to the Michaelis-Menten equation using Kalediagraph (Synergy Software).

Results and Discussion

G254A Mutation not as well Tolerated as Previously Thought in T. fusca OSBS

We previously published kinetic data for the G254A variant of *T. fusca* OSBS[32]. However, when using this variant for new experiments, I discovered that the plasmid stock that had been used previously for this variant was contaminated with plasmid encoding wildtype *T. fusca* OSBS, which skewed the results of the previous kinetic experiments. Therefore, I remade the mutation and verified there was no wildtype enzyme contamination by sequencing. I then repeated the OSBS assay that had been previously performed to get accurate kinetic constants for the variant (Table 2). Previously, the relative k_{cat}/K_M ($\frac{{k_{cat}}^{\mathrm{G254A}}/{K_M}^{\mathrm{G254A}}}{{k_{cat}}^{\mathrm{WT}}/{K_M}^{\mathrm{WT}}}$) of the G254A *T. fusca* OSBS variant was reported to be 0.28[32]. The relative k_{cat}/K_M using the new correct values for the kinetic constants is 0.04. This value is similar to the relative k_{cat}/K_M that was observed in the G254T mutant, which was 0.03[32]. While the *T. fusca* OSBS is not able to tolerate a mutation at this site as well as previously thought, it is still better able to tolerate the mutation than *E. coli* OSBS. The homologous mutation in *E. coli* OSBS, G288A, dropped the relative k_{cat}/K_M to 0.0019[34]. This data indicates that while previously it was thought that mutations at both the conserved residues, G254 and R128, had little effect on activity, of the *T. fusca* OSBS, mutation at the R128 site is better handled by the enzyme.

Table 2: Kinetic constants of *T. fusca* G254A

Variant	k_{cat} **(s^{-1})**	K_M **(μM^{-1})**	k_{cat}/K_M **(M^{-1} s^{-1})**	**Relative** k_{cat}/K_M
Wildtype*	122 ± 5 *	182 ± 22*	(6.7± 0.9) x 10^5*	1.0
G254A*	16 ± 1*	86 ± 6*	(1.9 ± 0.2) x 10^5*	0.28
G254A	1.1 ± 0.1	43 ± 9	(2.5 ± 0.6) x 10^4	0.04

*Previously published kinetic constants[32]

Effects of N73A, T75L, and R128M mutations in T. fusca OSBS are not additive

Previously, the roles of the residues N73, T75, and R128 were all examined in the *T. fusca* OSBS enzyme. The residue R128 is conserved within most members of the OSBS family, but mutating the position to methionine in *T. fusca* OSBS had little impact on activity[32]. However, mutating the homologous position in *E. coli* OSBS was deleterious to activity[32, 34]. The residues N73 and T75 form hydrogen bonds with the succinyl carboxylate of SHCHC in *T. fusca* OSBS (Fig. 6) [32]. We hypothesized that this interaction could supplant the water-mediated hydrogen bond to R128, so that mutations at this conserved arginine could be tolerated. In *E. coli* OSBS, the residues at the corresponding positions are A107 and L109, which cannot make the same type of interaction with the succinyl group of the substrate (Fig. 6). Changing the N73 residue to alanine in *T. fusca* OSBS dropped the relative k_{cat}/K_M to 0.05[32]. The T75L *T. fusca* OSBS variant reduced the relative k_{cat}/K_M to 0.47[32]. A double mutant variant of *T. fusca* OSBS, N73A/T75L, behaved similarly to the N73A variant. The R128M *T. fusca* OSBS variant

also had a drop in relative k_{cat}/K_M (0. 33)[32]. To test if the role of R128 depends on N73 and T75, I made a triple mutant variant, N73A/T75L/R128M. If the two residues N73 and T75 are responsible for the *T. fusca* OSBS enzyme's ability to tolerate the R128M, I would expect the N73A/T75L/R128M variant to have a drop in relative k_{cat}/K_M similar to that observed in the R159M variant of the *E. coli* OSBS (0.0056)[32]. However, I found that the N73A/T75L/R128M *T. fusca* OSBS variant behaved similarly to the N73A variant (Table 3). This result indicates that the disruption of the hydrogen bonds of N73 and T75L do not have a significant impact the enzyme's dependence on R128.

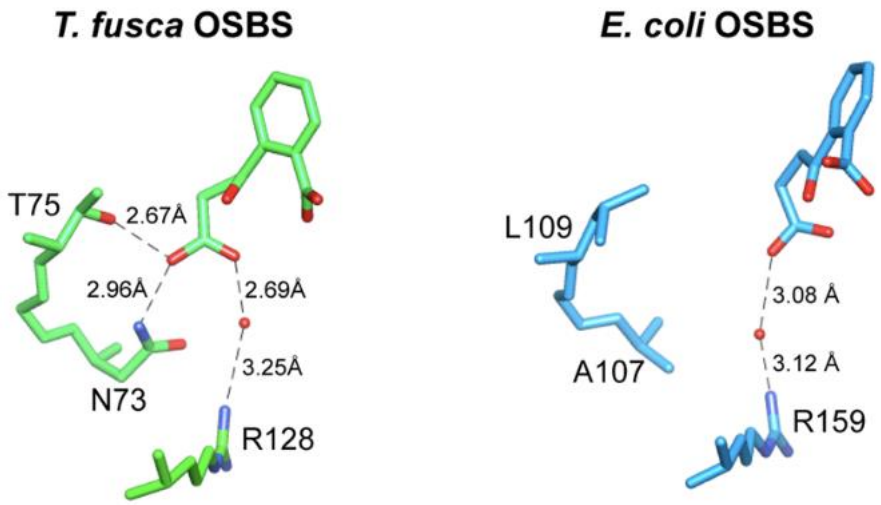

Figure 6: Comparison of the position of the ligand relative to L109 and A107 in the *E. coli* OSBS enzyme and T75 and N73 in *T. fusca* OSBS. Reprinted with permission from Odokonyero, D., Ragumani, S., Lopez, M. S., Bonanno, J. B., Ozerova, N. D., Woodard, D. R., Machala, B. W., Swaminathan, S., Burley, S. K., Almo, S. C., and Glasner, M. E. (2013) Divergent evolution of ligand binding in the o-succinylbenzoate synthase family, *Biochemistry 52*, 7512-7521. Copyright 2013 American Chemical Society.

Table 3: Kinetic constants of *T. fusca* N73A/T75L/R128M variants

Variant	k_{cat} **(s^{-1})**	K_M **(μM^{-1})**	k_{cat}/K_M **(M^{-1} s^{-1})**	**Relative** k_{cat}/K_M
N73A*	10 ± 1*	322 ± 56*	(3.1 ± 0.6) x 10^4*	0.05
T75L*	18 ± 1*	57 ± 5*	(3.2 ± 0.3) x 10^5*	0.47
R128M*	104 ± 4*	470 ± 42*	(2.2 ± 0.2) x 10^5*	0.33
N73A/T75L*	5.1 ± 0.4*	256 ± 51*	(2.0 ± 0.4) x 10^4*	0.03
N73A/T75L/R128M	11.4 ± 0.9	274 ± 55	(4.2 ± 0.9) x 10^4	0.06

*Previously published kinetic constants[32]

T. fusca OSBS Activity is Optimal at a Higher Temperature

Because *T. fusca*'s optimal growth temperature is 55°C, it is likely that the optimal temperature for OSBS activity of the *T. fusca* OSBS is higher than the temperature that the temperature at which the assay is normally performed, 25°C [37]. To test how temperature affects OSBS activity, the k_{obs} was calculated at temperatures ranging from 5 to 63°C for the *T. fusca* wildtype OSBS using saturating substrate concentration of 500 μM. The activity of *T. fusca* OSBS is optimal at 48°C (Fig. 7). The k_{obs} increases at each temperature up to 48°C, reaching a high of 1011 s^{-1}, before beginning to fall. These results are not surprising, as the optimal growth of temperature of the *T. fusca* organism is 55°C[37]. Some other variants of *T. fusca* OSBS should be

assayed at the same temperatures with Michaelis-Menten assays to see if the mutations that display little impact on activity in a room temperature assay, such as R128M, have a different impact at higher temperatures.

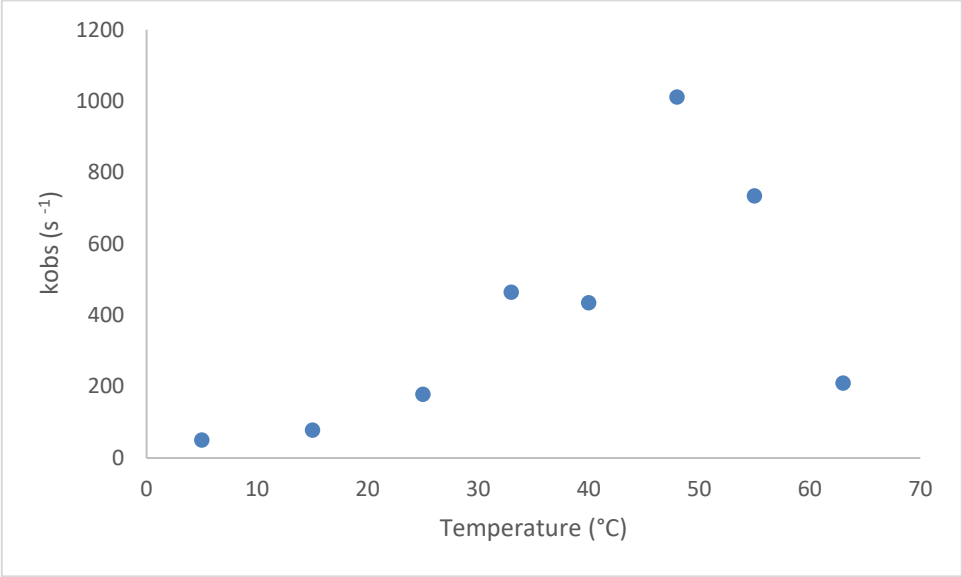

Figure 7: Calculated $k_{obs}s^{-1}$ for the wildtype *T. fusca* OSBS enzyme at temperatures ranging from 5 to 63 °C using saturating substrate

G254A Mutation's effects on OSBS Activity Changes with Temperature

Adaptations that occurred to retain activity in a thermophilic environment may be the reason that the *T. fusca* OSBS is able to tolerate mutations to conserved residues. Measuring the OSBS activity of the G254A *T. fusca* OSBS variant at different temperatures will help elucidate if these temperature adaptations have an effect on the ability of the enzyme to tolerate this change. The OSBS assay was performed at three temperatures, 10, 25, and 40 °C, for both the wildtype and G254A *T. fusca* OSBS variants. The G254A *T. fusca* OSBS variant, while having reduced activity at every

temperature when compared to the wildtype variant, seems to display the same increase in activity as temperature increases (Tables 4). In fact, it appears that the activity of the *T. fusca* G254A variant may be increasing at a faster rate with temperature than the *T. fusca* wildtype variant, based on the way that the relative k_{cat}/K_M increases as temperature increases. There are plans to continue these assays at higher temperatures, including at 48°C, which is the temperature that activity was the highest for the wildtype *T. fusca* OSBS enzyme. They were unable to be completed by the writing of this document, as the instrument used for these measurements needed to be repaired.

Table 4: *T. fusca* wildtype and G254A OSBS kinetic constants at 10, 25 and 40 °C. Only a single replicate was performed.

Temp (°C)	k_{cat} **(s^{-1})**	K_m **(μM^{-1})**	k_{cat}/K_M **(M^{-1} s^{-1})**	**Relative k_{cat}/K_M**
Wildtype				
10	108	198	5.4 x 10^5	1.0
25	178	207	8.5 x 10^5	1.0
40	138	168	8.2 x 10^5	1.0
G254A				
10	0.8	65.3	1.3 x 10^4	0.024
25	1.1	34.3	3.3 x 10^4	0.038
40	1.7	30.4	5.7 x 10^4	0.069

Conclusions and Future Directions

Previously published kinetic values for the G254A *T. fusca* OSBS variant were shown to be incorrect due to a contaminated plasmid stock. The kinetic constants found for the uncontaminated G254A *T. fusca* variant point to mutations at this site being moderately deleterious, but still not nearly as deleterious as the corresponding mutation in *E. coli* OSBS. Therefore, it appears that the conserved residue, R128, may be more tolerant than G254 to mutation in the *T. fusca* OSBS instead of the two sites being similarly tolerant to mutation. Hydrogen bonds between N73 and T75 and the succinyl carboxylate in *T. fusca* OSBS do not appear to be responsible for the enzyme's ability to tolerate the R128M mutation. Optimal OSBS activity for the wildtype *T. fusca* OSBS was found to occur at 48°C. In the future, other *T. fusca* OSBS variants should be assayed at the higher temperature, to see if the mutations impact the way the enzyme is able tolerate increases in temperature. Studying the way temperature impacts OSBS activity will allow us to understand how adaptations in the OSBS enzyme from the thermophilic *T. fusca* impact its ability to tolerate mutations which are not tolerated by the mesophilic *E. coli* OSBS enzyme.

CHAPTER III

ROLE OF A CONSERVED SECOND SHELL AMINO ACID IN THE ENOLASE SUPERFAMILY

Role of a Conserved Second Shell Amino Acid in *N*-succinylamino Acid Racemase and Dipeptide Epimerase Activity

How enzymes evolve new functions and specificities is currently poorly understood. Two factors that are at play in the evolution of new functions and specificities are enzyme promiscuity and epistasis. Enzymes are promiscuous if the enzyme's active site catalyzes a reaction that is biologically relevant and other reactions that are not biologically relevant[4, 5]. Enzyme epistasis occurs when amino acid substitutions have different consequences depending on the sequence background[13-15]. Enzymes in the enolase superfamily serve as excellent model enzymes to observe the interplay between protein promiscuity and epistasis, as it has enzyme families, like the *o*-succinylbenzoate synthase (OSBS) family, with promiscuous members that display high sequence divergence[20, 22]. Studying members of the enolase superfamily enables the observation of how differing sequence contexts changes the way amino acid substitutions are tolerated and how those substitutions may have different impacts on different activities.

The OSBS family is highly divergent in sequence, but all the enzymes in the family retain the ability to carry out the OSBS reaction[20, 22]. Some enzymes in one subfamily, the *N*-succinylamino acid racemase/*o*-succinylbenzoate synthase

(NSAR/OSBS) subfamily, have the ability to carry out both the NSAR reaction and the OSBS reaction (Fig. 2)[20, 22]. Understanding what enables the enzymes within the NSAR/OSBS subfamily to carry out both reactions may allow us to better understand how reaction specificity evolves. To determine if any sites are conserved in the NSAR/OSBS subfamily that are not conserved in other OSBS subfamilies, the sequence conservation was examined for all of the subfamilies[35]. It was found that a second shell amino acid, corresponding to R266 in *Amycolatopsis* sp. T-1-60 NSAR/OSBS, was conserved in the NSAR/OSBS subfamily while not being conserved in the other OSBS subfamilies[35]. Other subfamilies typically have a nonpolar residue at this site, although a small subset has a glutamine at this site.

To test if the R266 site is responsible for the enzymes' ability to carry out the NSAR reaction, the R266 site in *Amycolatopsis* sp. T-1-60 NSAR/OSBS was mutated to glutamine. The relative specificity ($(k_{cat}/K_M)^{OSBS}/(k_{cat}/K_M)^{NSAR}$) of the OSBS and NSAR reactions changed from 1.8 for the wildtype enzyme to 99 for the R266Q enzyme[35]. To evaluate how the R266Q mutation decreased NSAR activity, ^{1}H NMR spectroscopy was used to measure the exchange rate (k_{ex}) between the alpha proton of the NSAR substrate and the catalytic lysine residues[35]. The drop in k_{ex} between K163 and the NSAR substrate in the R266Q mutant was similar to the drop in k_{cat}^{OSBS}, while the drop in k_{ex} between K263 and the NSAR substrate was almost 400-fold higher than the drop in k_{cat}^{OSBS} [35]. This data indicates that R266 primarily changes the reaction specificity of the OSBS and NSAR reactions by affecting the reactivity of K263, which plays a direct role in the NSAR reaction by acting as a general acid or base but only plays an indirect role

in the OSBS reaction[26, 35]. These results indicated that the R266 site was important for NSAR activity in *Amycolatopsis* sp. T-1-60 NSAR/OSBS. To determine if the observed impacts of R266 on activity were a general property of other members of the NSAR/OSBS subfamily, the mutation was introduced into the homologous site in several NSAR/OSBS enzymes[35]. I made the homologous mutation in the *Lysinibacillus varians* NSAR/OSBS, and the impacts of this mutation will be discussed in the results section of this chapter.

Other families from the enolase superfamily, such as the dipeptide epimerase family, carry out activities similar to the NSAR reaction. The OSBS and dipeptide epimerase families utilize the same conserved catalytic lysine residues (Fig. 2). Examining the sequence conservation within the dipeptide epimerase family revealed that the site homologous to the R266 site of the NSAR/OSBS family has a conserved lysine residue[35]. Figure 5 compares the position of this conserved site in *E. coli* dipeptide epimerase with the site in in *Amycolatopsis* sp. T-1-60 NSAR/OSBS. Both of these conserved amino acids, arginine and lysine, are positively charged. It is possible that a positive charge at this site is important for racemase and epimerase evolution in the enolase superfamily. To test if this positively charged residue plays a similar role in the dipeptide epimerase family and NSAR/OSBS subfamily, the homologous site was mutated to glutamine in the *E. coli* dipeptide epimerase and the *B. subtilis* dipeptide epimerase[35]. In the *E. coli* dipeptide epimerase, the mutation (K250Q) decreased k_{cat} by 44-fold, and in the *B. subtilis* dipeptide epimerase, K271Q decreased k_{cat} 93-fold[35]. Neither of the enzymes had a change in the K_M value between the wildtype enzyme and

the mutant enzyme[35]. Isotopic exchange experiments were also conducted on both enzymes. The mutation in *B. subtilis* dipeptide epimerase decreased the reactivity of the catalytic acid/base K268 more than the other catalytic acid/base, K162, agreeing with the hypothesis that the K271Q mutation would preferentially impact the reactivity of the K268 residue[35]. The K250Q mutation in the *E. coli* dipeptide epimerase caused only a 2-fold decrease in the reactivity of K247 and no decrease of the reactivity of K151, which did not match the hypothesis[35]. The protein was expressed in an *E. coli* that still had its native dipeptide epimerase gene, however, which may have impacted the results by co-purifying with the mutant enzyme. In this chapter, I used a dipeptide epimerase knockout strain for protein expression to solve this problem. In addition, I mutated K250 to all other possible amino acids to test the hypothesis that the site is important for epimerase activity. Mutating the site to all of the possible residues allows us to examine how different amino acids behave in the same environment, which can give insights about how epistasis may affect this site.

Materials and Methods

Mutagenesis

Site-directed mutagenesis was performed using the Q5 mutagenesis protocol (New England BioLabs). For the *L. varians* NSAR/OSBS R266Q variant, the template used was the gene encoding *L. varians* GY32 NSAR/OSBS (uniProt entry: X2GR01) that was subcloned into a modified pET21a vector pMCSG7, with a *N*-terminal 6-Histidine tag. For the *E. coli* dipeptide epimerase mutants, the template used was the

gene *ycjG* from *E. coli*, which encodes dipeptide epimerase (UniProt entry: P51981), that was subcloned into a pET15b vector modified with a *N*-terminal 10-histdine tag (Novagen) (gift from J. A Gerlt, University of Illinois, Urbana, IL). All mutations were confirmed by sequencing in both directions (Eurofins Genomics LLC). The primers used for Q5 mutagenesis were designed using NEBaseChanger, NEB's online design software (NEBasechanger.com). The primers used to generate the R266Q mutation in *L. varians* NSAR/OSBS are shown in Table 5. The same reverse primer was used to generate each *E. coli* dipeptide epimerase mutant, while the forward primer varied at the K250 codon position to produce each *E. coli* dipeptide epimerase mutants shown in Table 6. The codon most commonly utilized by *E. coli* was used for each amino acid (Genscript codon frequency table).

Table 5: Primers used for mutagenesis at position 266 in *L. varians* NSAR/OSBS. Primers are shown in the 5' to 3' direction. The lower-case bases designate the codon where mutations were introduced.

Mutation	**Primers Used**
R266Q	Forward: TAAAATTGGAcagGTAGGCGGCATAAC Reverse: ATATTAATTACACCGCAGC

Table 6: Primers used for mutagenesis at position 250 in *E. coli* dipeptide epimerase. All DNA primers are shown in the 5' to 3' direction. The lower-case bases designate the codons where mutations were introduced.

Mutation	**Primers used**
K250A	Forward: TAAGCTCGATgcgACCGGGGGTC Reverse: ATATTAACCATCTCATAGCGC
K250C	Forward: TAAGCTCGATtgcACCGGGGGTC Reverse: ATATTAACCATCTCATAGCGC
K250D	Forward: TAAGCTCGATgatACCGGGGGTC Reverse: ATATTAACCATCTCATAGCGC
K250E	Forward: TAAGCTCGATgaaACCGGGGGTC Reverse: ATATTAACCATCTCATAGCGC
K250F	Forward: TAAGCTCGATtttACCGGGGGTC Reverse: ATATTAACCATCTCATAGCGC
K250G	Forward: TAAGCTCGATggtACCGGGGGTC Reverse: ATATTAACCATCTCATAGCGC
K250H	Forward: TAAGCTCGATcatACCGGGGGTC Reverse: ATATTAACCATCTCATAGCGC
K250I	Forward: TAAGCTCGATattACCGGGGGTC Reverse: ATATTAACCATCTCATAGCGC
K250L	Forward: TAAGCTCGATctgACCGGGGGTC

	Reverse: ATATTAACCATCTCATAGCGC
K250M	Forward: TAAGCTCGATatgACCGGGGGTC Reverse: ATATTAACCATCTCATAGCGC
K250N	Forward: TAAGCTCGATaacACCGGGGGTC Reverse: ATATTAACCATCTCATAGCGC
K250P	Forward: TAAGCTCGATccgACCGGGGGTC Reverse: ATATTAACCATCTCATAGCGC
K250Q	Forward: TAAGCTCGATcagACCGGGGGTC Reverse: ATATTAACCATCTCATAGCGC
K250R	Forward: TAAGCTCGATcgtACCGGGGGTC Reverse: ATATTAACCATCTCATAGCGC
K250S	Forward: TAAGCTCGATagcACCGGGGGTC Reverse: ATATTAACCATCTCATAGCGC
K250T	Forward: TAAGCTCGATaccACCGGGGGTC Reverse: ATATTAACCATCTCATAGCGC
K250V	Forward: TAAGCTCGATgtgACCGGGGGTC Reverse: ATATTAACCATCTCATAGCGC
K250W	Forward: TAAGCTCGATtggACCGGGGGTC

	Reverse: ATATTAACCATCTCATAGCGC
K250Y	Forward: TAAGCTCGATtatACCGGGGGTC Reverse: ATATTAACCATCTCATAGCGC

Protein Purification

L. varians NSAR/OSBS variants were expressed in the *menC*$^{-}$ strain BW25113 (*menC::kan,* DE3), so that purified proteins were not contaminated with native *E. coli* OSBS[34]. *E. coli* dipeptide epimerase variants were expressed in the *ycjG*$^{-}$ strain BW25113 (*ycjG::kan,* DE3) (The Coli Genetic Stock Center, Yale University), so that purified proteins were not contaminated with native *E. coli* dipeptide epimerase. Cell cultures were grown without induction in Luria-Bertani broth supplemented with carbenicillin and kanamycin at a final concentration of 50 μg/mL overnight at 37 °C. Cells were harvested by centrifugation and resuspended in a buffer containing 20 mM Tris, pH 8.0, 500 mM NaCl, 5 mM imidazole, 0.02 mg/ml DNAse (Worthington), and 2 μM phenylmethylsulfonyl fluoride (PMSF; Thermo Scientific). The resuspended cell pellets were lysed using sonication. The cell lysate was centrifuged, and the supernatant was loaded onto a 5 mL HisTrap FF column charged with Ni^{2+} (GE Healthcare). The protein was eluted from the column with a buffer that contained 20 mM Tris, pH 8.0, 500 mM NaCl and 500 mM imidazole using a step to 15% elution buffer followed by a linear gradient to 100% elution buffer. Fractions containing homogenous protein were identified using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-

PAGE) and then pooled. Amicon Ultra-15 centrifugal filters with a 10 kD cutoff (Millipore) were used to exchange the buffer and concentrate the pooled fractions. Concentrated purified proteins were stored at -80 °C in 10 mM Tris, pH 8.0, 5 mM $MgCl_2$, and 25% glycerol.

Western Blot

All proteins were expressed in the *ycjG*$^-$ strain BW25113 (*ycjG::kan*, DE3) (The Coli Genetic Stock Center, Yale University), so that purified proteins were not contaminated with native *E. coli* dipeptide epimerase. Cell cultures were grown without induction in Luria-Bertani broth supplemented with carbenicillin and kanamycin at a final concentration of 50 μg/mL overnight at 37 °C. Cells were harvested by centrifugation and resuspended in a buffer containing 20 mM Tris, pH 8.0, 500 mM NaCl, 5 mM imidazole, 0.02 mg/ml DNAse (Worthington), and 2 μM phenylmethylsulfonyl fluoride (PMSF; Thermo Scientific). The resuspended cell pellets were lysed using sonication. The cell lysate was centrifuged, and the cell pellet and supernatant were separated. The pellet was resuspended in a volume of 8 M urea equivalent to the starting volume of cell lysate. Dithiothreitol (DTT; Sigma) was added to both the supernatant and resuspended pellet achieve a final concentration of 60 mM DTT. Laemmli buffer (Sigma) was added to each sample to achieve a final concentration of 1x buffer. The samples were denatured for 20 minutes at 85 °C in a Bio-Rad thermocycler. The samples and a blue protein standard (Biolabs) were then run on an SDS-PAGE gel in SDS buffer (25 mM Tris pH 8.3, 192 mM glycine, 30 mM SDS) at

100 V for 60 minutes. The gel was transferred onto a Western blot membrane overnight at 200 mA at 4 °C in transfer buffer (25 mM Tris pH 8.3, 192 mM glycine, and 20% methanol). The Western blot was then blocked for one hour in a solution of 5% powdered milk in ddH_2O. The Western blot was then incubated in a 5% powdered milk solution containing the primary antibody (Anti-His tag mouse monoclonal antibody, Sigma-Aldrich) for 30 minutes, followed by incubation with the secondary antibody (anti-mouse IgG-peroxidase antibody, Sigma-Aldrich) for 30 minutes. Enhanced Chemiluminescence substrate (ThermoFisher) was added to the Western blot prior to being imaged using a gel document system (Invitrogen).

OSBS Activity Assay

2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate (SHCHC) was prepared using methods as previously described[34]. Enzyme activity was measured in 50 mM Tris pH 8.0 and 0.1 mM $MnCl_2$. Activity was determined by measuring the disappearance of SHCHC at 310 nm ($\Delta\varepsilon$=-2400 $M^{-1}cm^{-1}$) over time[22, 26]. Assays were performed at 25 °C using a SpectraMax Plus[384] UV-VIS microplate spectrophotometer (Molecular Devices). Initial rates were determined by fitting a line to the linear portion of the curve (Microsoft Excel). Initial rates at varying substrate concentrations were then fit to the Michaelis-Menten equation using Kalediagraph (Synergy Software).

NSAR Activity Assay

L- and D-*N*-succinylphenylglycine (L- and D-NSPG) were prepared as described previously[22]. Each enzyme was assayed in 200 mM Tris (pH 8.0) and 0.1 mM $MnCl_2$ with varying L- or D-*N*-succinylphenylglycine concentrations. The reactions were carried out in a 5 cm path length sample cell. A Jasco P-2000 polarimeter monitored the change in the substrate's optical rotation at 405 nm and 25 °C. Measurements were taken every thirty seconds using a 10 s integration time. 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} [22]. Initial rates were determined by fitting a line to the linear portion of the curve (Microsoft Excel). Initial rates at varying substrate concentrations were then fit to the Michaelis-Menten equation using Kalediagraph (Synergy Software).

Dipeptide Epimerase Activity Assay

The substrate, L-Ala-L-Glu was purchased from Chem Impex Int'l, Inc. and L-Ala-D-Glu was synthesized by Dakota Brock (TAMU) as described previously[21]. The dipeptide epimerase enzyme was assayed in 50 mM Tris (pH 8.0) and 10 mM $MgCl_2$ at varying dipeptide concentrations with a total reaction volume of 1.4 mL. The assays were performed in a cell with a 5 cm path length. The change in the optical rotation of the substrate was by monitored using a Jasco P-2000 polarimeter at 365 nm and 25 °C. The specific rotation of L-Ala-L-Glu at 365 nm is 7.07 deg M^{-1} cm^{-1}. Initial rates were determined by fitting a line to the linear portion of the curve (Microsoft Excel). Initial

rates at varying substrate concentrations were then fit to the Michaelis-Menten equation using Kalediagraph (Synergy Software).

Isotopic Exchange Experiments

The *L. varians* WT and R266Q variants were exchanged three times into 20 mM Tris buffer (pD 8.0 with NaOD) using a Vivaspin Turbo 15 centrifuge filter (Sartorius). A 10 mL aliquot of enzyme was concentrated to 1 mL. After 9 mL of 20 mM Tris buffer (pD 8.0) was added, the solution was again concentrated to 1 mL. The reaction contained either 20 mM L- or D-NSPG (pD 8.0 with NaOD), 50 mM Tris (pD 8.0), 0.1 mM MnCl2, and enzyme in D_2O. The intensity of the α-proton peak was monitored 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 75 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

The T_m of each enzyme was determined. Each sample contained a 1:1250 dilution of SYPRO Orange (Sigma-Aldrich), 0.5 mg/mL enzyme, 25 mM HEPES (pH 7.5), 300 mM NaCl, 10 mM DTT, and 10 mM EDTA[38]. Assays were performed in quadruplicate in sealed 96 well qRT-PCR plates (VWR). The temperature of the samples was increased in sixty second intervals from 25 °C to 99 °C using a BioRad CFX-96

qPCR machine. Protein unfolding was measured using an excitation wavelength of 470 nm and an emission wavelength of 570 nm. The data was exported and graphed using Excel (Microsoft). Each curve was fit to a fourth-degree polynomial. The temperature at the curve's inflection point was determined using the quadratic equation to solve the equation for the second derivative. The solution that matched the visual inflection of the curve is reported at the T_m.

Results and Discussion

Effect of L. varians NSAR/OSBS R266Q on NSAR and OSBS Activity

To test if the conserved R266 residue plays a similar role in *Amycolatopsis* sp. T-1-60 NSAR/OSBS and in other NSAR/OSBS enzymes, I mutated the homologous position in *L. varians* NSAR/OSBS. The *L. varians* NSAR/OSBS enzyme carries out both the OSBS and NSAR reactions, with a strong preference for the OSBS reaction (Table 3). The R266Q mutation reduced NSAR activity more than OSBS activity, which was consistent with the way the mutation behaved in *Amycolatopsis* sp. T-1-60 NSAR/OSBS (Table 3). The relative specificity $(k_{cat}/K_M)^{OSBS}/(k_{cat}/K_M)^{NSAR}$ changed from 166 in the wildtype variant to 382 in the R266Q variant of the *L. varians* NSAR/OSBS, compared to a change from 1.8 to 99 in the *Amycolatopsis* sp. T-1-60 NSAR/OSBS[35]. However, K_M^{OSBS} unexpectedly increased, and k_{cat}^{OSBS} unexpectedly decreased (Table 7). ^{1}H NMR spectroscopy was used to measure the exchange rate (k_{ex}) between the alpha proton of D- and L-NSPG and the two catalytic residues, K163 and K263, to determine if the R266Q mutation greatly decreases the reactivity of K263

without decreasing the reactivity of K163 as was observed in the *Amycolatopsis* sp. T-1-60 NSAR/OSBS. In the R266Q *L.varians* NSAR/OSBS variant, it was found that the reactivity of both catalytic residues was significantly decreased, by 78-fold for K163 and 108-fold for K263 (Table 8). The equivalent change in the reactivity for both of the catalytic lysine further supports that active site distortion is occurring in the R266Q variant. This data points to there being some epistatic interactions at play within the *L. varians* NSAR/OSBS that cause the mutation to have a different impact than it did in the *Amycolatopsis* sp. T-1-60 NSAR/OSBS. The R266Q mutation in *L. varians* reduced NSAR activity more than OSBS activity, which points to the R266 site being important for the evolution of NSAR activity. However, epistatic interactions occurring in different sequence contexts may mask a universal, general role of R266 in modulating reactivity of a catalytic lysine to enable evolution of NSAR activity. It is possible that other substitutions at the R266 site in *L. varians* will have different effects than the glutamine substitution. Examining other substitutions at the same site can help elucidate some of these epistatic interactions that may be occurring.

Table 7: Kinetic constants of *L. varians* WT and R266Q NSAR/OSBS variants

	OSBS			**NSAR**			
	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	71 ± 10	186 ± 15	(2.6 ± 0.4) x10^6	1000 ± 70	16 ± 1	(1.6 ± 0.1) x10^4	166
R266Q	719 ± 130	21 ± 2	(2.9 ± 0.6) x10^4	1600 ±250	0.12 0.01	(1.6 ± 0.1) x10^2	382

Table 8: Hydrogen-Deuterium exchange rate (k_{ex}) of *L. varians* NSAR/OSBS

	WT	**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

Effects of Mutations at K250 in E. coli Dipeptide Epimerase on Protein Expression

To determine how mutations at the K250 site in *E. coli* dipeptide epimerase site affect activity and structure, a saturation mutagenesis library of the site was constructed. Of the 19 possible variants at the K250 site in *E. coli* dipeptide epimerase, 17 of the variants were made using Q5 mutagenesis. Attempts to make the K250L and K250R

mutations were unsuccessful. The expression of each of the variants, with the exception of K250Q, as this variant was unavailable at the time of the expression test, was compared to identify candidates for purification (Fig. 8). The K250Q mutation likely expresses highly as it was previously purified and characterized[35]. Most of the variants had at least some expression with the exception of the K250D and K250E variants. These mutations were likely destabilizing, as aspartic acid and glutamic acid are both negatively charged amino acids, which is the opposite charge of the amino acid that naturally occurs at the site, lysine. The K250F variant expressed poorly, likely due to its bulky aromatic functional group. The K250C, K250N, and K250T variants all had moderate decreases in expression. The K250H and K250S variants had even smaller reductions in expression compared to the wildtype variant. One variant, K250A, expressed at levels very similar to wildtype, which fits expectations as alanine's functional group is non-bulky and chemically inert. It is difficult to determine how some of the mutants (K250G, K250I, K250M, K250P, K250V, K250W, and K250Y) expressed compared to the wildtype enzyme, as they are not on a blot with the wildtype enzyme. However, comparing these variants to one another shows that the K250G, K250P, and K250W mutations had the largest reductions in expression. The K250G variant likely had lower expression levels due to the high conformational flexibility of glycine. The K250P likely had lower expression because of the inability of proline to take certain geometries. The K250W likely had lower expression levels due to the size of the functional group of tryptophan. To better determine how these mutants impacted expression, the expression tests need to be repeated at the same time as a wildtype

expression test. Three of the highly expressing mutants (K250A, K250H, and K250S), in addition to the wildtype enzyme, were purified for additional analysis.

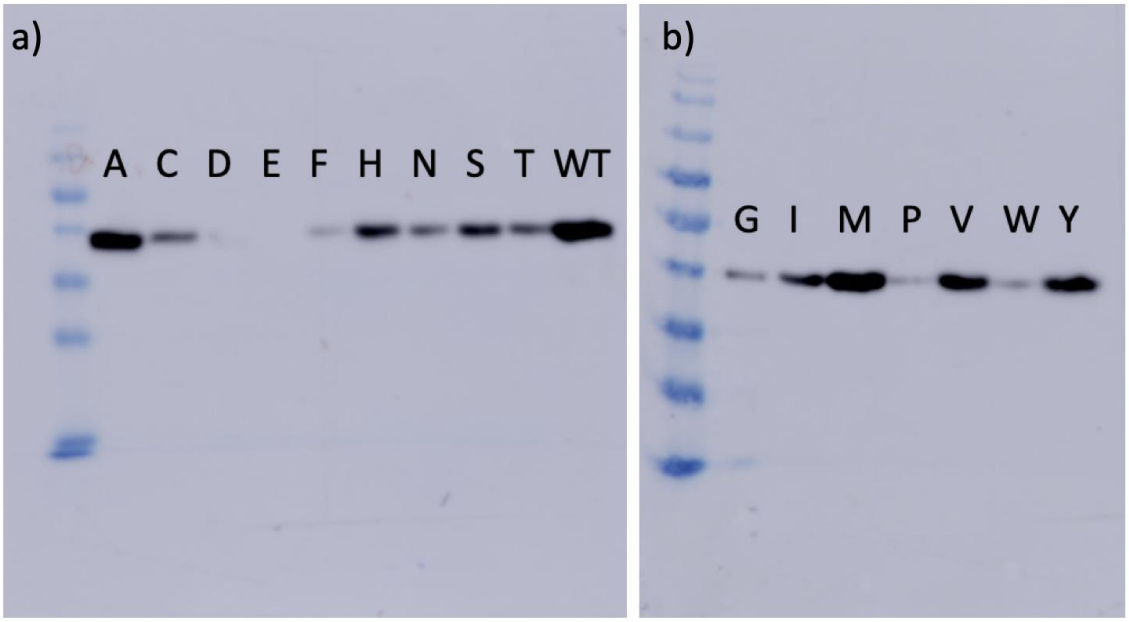

Figure 8: Western blots of *E. coli* dipeptide epimerase variants. The protein was detected using an antibody specific to the His_{10} tag. The first lane of each blot contains a blue protein standard. An equal volume of cell lysate for each protein variant is present on the blot. The letters above each lane correspond to the amino acid identity at position 250. a) The lane labeled A corresponds to K250A, C to K250C, D to K250D, E to K250E, F to K250F, H to K250H, N to K250N, S to K250S, T to K250T, and WT to the wildtype protein. b) G corresponds to K250G, I to K250I, M to K250M, P to K250P, V to K250V, W to K250W, and Y to K250Y.

Effects of Mutations at K250 in E. coli Dipeptide Epimerase on Activity

To examine how each mutation impacted the enzyme's ability to carry out the epimerase reaction, each variant needed to be assayed for dipeptide epimerase activity. Each of the purified variants was assayed; however, it was difficult to decipher a clear trend in the data due to the low signal to noise ratio (a representative data set is shown for the wildtype enzyme in Fig. 9). To determine if the trend in the data was an actual

observation of activity and not just an artifact, the k_{obs} for the wildtype enzyme was calculated for the data collected. The k_{obs} for the data shown in Figure 9 was calculated to be 3.0 s^{-1}. The measurement was taken using saturating substrate, as determined from the published K_M (1.3 x 10^{-4} M) [21]. The previously reported k_{cat} of the wildtype dipeptide epimerase was 10 s^{-1} [21]. Because the measured k_{obs} approaches the reported k_{cat}, the enzyme appears to be active. Due to the large signal to noise ratio at even the highest substrate concentration, fitting any data from assays of the variants would be very difficult. Instead, a different method of assaying the dipeptide epimerase should be used. A coupled enzyme assay for the dipeptide epimerase has been reported, and there are plans to move forward with that assay[21].

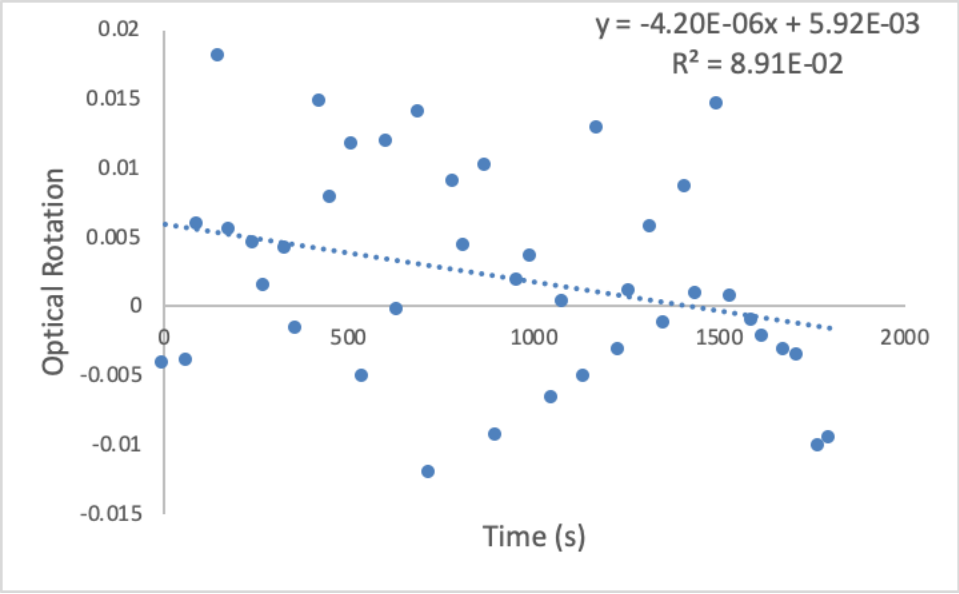

Figure 9: A representative of the data collected from the dipeptide epimerase assay for the *E. coli* dipeptide epimerase variants. The data shown is from a reaction containing 0.02 μM of wildtype enzyme and 5 mM of the substrate, L-Ala-L-Glu. The optical rotation of the substrate was measured over time. The fit line for the data, the equation of the fit line, and the R^2 value of the fit line are all displayed on the graph.

Thermostability of E. coli Dipeptide Epimerase Variants

To understand each mutation's impact on the enzyme's thermostability, Differential Scanning Fluorimetry (DSF) was performed on three of the purified variants (WT, K250A, and K250S). The experiment could not be performed on the other purified variant, K250H, as all of the protein precipitated out of solution during storage. It was anticipated that the K250A variant would have a similar melting temperature to the wildtype enzyme, as the variant had a similar expression level to the wildtype enzyme. K250S, while it still expressed highly, expressed less than the wildtype enzyme, which may indicate that the variant is less stable. The K250A mutant exhibited a small decrease in T_m compared to the wildtype enzyme, while the K250S mutant had a larger decrease (Table 9). These results are consistent with the observed expression levels from the Western blots, as lower enzyme stability may have led to lower observed expression. K250H had similar expression levels to K250S; however, K250H precipitated out of the storage solution. This suggests that the K250H variant may be less stable than the K250A or K250S variant.

Table 9: T_m values and the change in T_m from wildtype for *E. coli* dipeptide epimerase variants calculated using DSF

Variant	**T_m (°C)**	**ΔT_m (°C)**
Wildtype	49.3 ± 1.1	n/a
K250A	47.8 ± 0.7	-1.5
K250S	43.0 ± 0.6	-6.3

Conclusions and Future Directions

In this chapter, I have made advancements towards understanding the role that a conserved second-shell amino acid plays in the dipeptide epimerase family and NSAR/OSBS subfamily. The R266Q mutation in the *L. varians* NSAR/OSBS shifts the relative specificity of the enzyme by reducing NSAR activity more than OSBS activity; however, it does not shift the relative specificity as dramatically as was seen in the R266Q variant of *Amycolatopsis* sp. T-1-60 NSAR/OSBS. Epistatic interactions limit the ability to observe how the R266 site affects the reactivity of the catalytic lysine residues. Developing a saturation mutagenesis library of the K250 residue in *E. coli* dipeptide epimerase allowed for observations about what types of different amino acids are tolerated at that position. Expression and thermostability tests revealed that the impact of mutating the K250 residue in the *E. coli* dipeptide epimerase is different depending on the identity of the amino acid at the site. Due to issues with the dipeptide epimerase assay, how each mutation impacts the dipeptide epimerase function is to be determined.

In the future, the expression tests of the dipeptide epimerase variants should be repeated, so that more quantitative analysis of the expression levels can be conducted. Measuring the total protein in the cell lysate before loading onto the SDS-gel would enable a more quantitative evaluation of protein expression. Other mutants that express well should also be candidates for further observation through assays and isotope exchange experiments. Two variants that were not present in the expression tests presented in this chapter would be of particular interest—K250Q and K250R. The

K250Q variant was previously studied; however, in that study the enzyme may have been contaminated with native *E. coli* dipeptide epimerase due to being purified in cells with the native dipeptide epimerase gene present. Evaluating the K250R variant would help to further test the hypothesis that the positive charge of the amino acid at position 250 is important for modulating reactivity of the adjacent catalytic lysine. The dipeptide epimerase variants should be assayed using another method as the method used here proved to be difficult to interpret. A good candidate for an alternative assay would be the coupled assay that has been previously reported[21]. To understand how the mutation impacts the reactivity of the catalytic lysine of each variant, variants that show a large deficit in enzyme function should be further evaluated through isotopic exchange experiments. The isotopic exchange experiments will reveal if the mutation preferentially impacts the exchange rate of one of the catalytic lysine residues over the other. Based on the data collected about the NSAR/OSBS subfamily, it is expected that the K247 residue will be more impacted by the mutation than the K151 residue. Completing these experiments will lead to a better understanding of how racemase activity evolved in the enolase superfamily, and in turn, will lead to a better understanding of how reaction specificity evolves.

CHAPTER IV

CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions

The work in this thesis advances the current understanding of the roles that enzyme promiscuity and enzyme epistasis play in the evolution of new enzyme functions. Chapter 2 examined the role that two conserved residues, G254 and R128, play in OSBS enzymes. Chapter 3 examined the role that R266/K250 played in the evolution of the NSAR and dipeptide epimerase functions.

Previously published kinetic values for the G254A *T. fusca* OSBS variant were shown to be incorrect due to the plasmid stock containing both the wildtype and mutant enzyme variants. The kinetic constants found for the new verified G254A *T. fusca* variant point to mutations at this site being moderately deleterious. However, this site is still not as deleterious as the corresponding mutation in *E. coli* OSBS. R128 may be more tolerant than G254 to mutation in the *T. fusca* OSBS instead of the two sites being similarly tolerant to mutation. Hydrogen bonds between N73 and T75 and the succinyl carboxylate in *T. fusca* OSBS do not appear to be responsible for the enzyme's ability to tolerate the R128M mutation. Optimal OSBS activity for the wildtype *T. fusca* OSBS was found to occur close to 48°C.

The R266Q mutation in the *L. varians* NSAR/OSBS shifts the relative specificity of the enzyme by reducing the NSAR activity more than the OSBS activity. The mutation does not shift the relative specificity as dramatically as was seen in the R266Q

variant of *Amycolatopsis* sp. T-1-60 NSAR/OSBS. Epistatic interactions may limit the ability to observe how the R266 site affects the reactivity of the catalytic lysine residues. Developing a saturation mutagenesis library of the K250 residue in *E. coli* dipeptide epimerase allowed for observations about what types of amino acids are tolerated at that position. Expression and thermostability tests revealed that the impact of mutating the K250 residue in the *E. coli* dipeptide epimerase is different depending on the identity of the amino acid at the site. Due to issues with the dipeptide epimerase assay, how each mutation impacts the dipeptide epimerase function is to be determined.

Future Directions

The work presented in this thesis serves as a starting point for future work that can further advance the understanding of enzyme evolution. In the future, other *T. fusca* OSBS variants should be assayed at the higher temperature, to see if the mutations impact the way the enzyme is able tolerate increases in temperature. Studying the way temperature impacts OSBS activity will allow us to understand how adaptations in the OSBS enzyme from the thermophilic *T. fusca* impact its ability to tolerate mutations which are not tolerated by the mesophilic *E. coli* OSBS enzymes.

In the future, the expression tests of the dipeptide epimerase variants should be repeated, so that more quantitative analysis of the expression levels can be conducted. Measuring the total protein in the cell lysate before loading onto the SDS-gel would enable a more quantitative evaluation of protein expression. Other mutants that express well should also be candidates for further observation through assays and isotope

exchange experiments. Two variants that were not present in the expression tests presented in this chapter would be of particular interest—K250Q and K250R. The K250Q variant was previously studied; however, in that study the enzyme may have been contaminated with native *E. coli* dipeptide epimerase due to being purified in cells with the native dipeptide epimerase gene present. Evaluating the K250R variant would help to further test the hypothesis that the positive charge of the amino acid at position 250 is important for modulating reactivity of the adjacent catalytic lysine. The dipeptide epimerase variants should be assayed using another method as the method used here proved to be difficult to interpret. A good candidate for an alternative assay would be the coupled assay that has been previously reported[21]. To understand how the mutation impacts the reactivity of the catalytic lysine of each variant, variants that show a large deficit in enzyme function should be further evaluated through isotopic exchange experiments. The isotopic exchange experiments will reveal if the mutation preferentially impacts the exchange rate of one of the catalytic lysine residues over the other. Based on the data collected about the NSAR/OSBS subfamily, it is expected that the K247 residue will be more impacted by the mutation than the K151 residue. Completing these experiments will lead to a better understanding of how racemase activity evolved in the enolase superfamily, and in turn, will lead to a better understanding of how reaction specificity evolves. Knowing the role that the R266/K250 site plays in racemase and epimerase activity in the enolase superfamily may enable function predictions to be made for other uncharacterized members of the enolase superfamily.

REFERENCES

[1] Lutz, S. (2010) Beyond directed evolution--semi-rational protein engineering and design, *Curr Opin Biotechnol 21*, 734-743.

[2] Bloom, J. D., Meyer, M. M., Meinhold, P., Otey, C. R., MacMillan, D., and Arnold, F. H. (2005) Evolving strategies for enzyme engineering, *Curr Opin Struct Biol 15*, 447-452.

[3] Lanio, T., Jeltsch, A., and Pingoud, A. (2000) On the possibilities and limitations of rational protein design to expand the specificity of restriction enzymes: a case study employing EcoRV as the target, *Protein Eng 13*, 275-281.

[4] Glasner, M. E., Truong, D. P., and Morse, B. C. (2019) How enzyme promiscuity and horizontal gene transfer contribute to metabolic innovation, *FEBS J*.

[5] Hult, K., and Berglund, P. (2007) Enzyme promiscuity: mechanism and applications, *Trends Biotechnol 25*, 231-238.

[6] Babtie, A., Tokuriki, N., and Hollfelder, F. (2010) What makes an enzyme promiscuous?, *Curr Opin Chem Biol 14*, 200-207.

[7] Aharoni, A., Gaidukov, L., Khersonsky, O., Mc, Q. G. S., Roodveldt, C., and Tawfik, D. S. (2005) The 'evolvability' of promiscuous protein functions, *Nat Genet 37*, 73-76.

[8] Jensen, R. A. (1976) Enzyme recruitment in evolution of new function, *Annu Rev Microbiol 30*, 409-425.

[9] Soo, V. W., Hanson-Manful, P., and Patrick, W. M. (2011) Artificial gene amplification reveals an abundance of promiscuous resistance determinants in Escherichia coli, *Proc Natl Acad Sci U S A 108*, 1484-1489.

[10] Gould, S. M., and Tawfik, D. S. (2005) Directed evolution of the promiscuous esterase activity of carbonic anhydrase II, *Biochemistry 44*, 5444-5452.

[11] Yoshikuni, Y., Ferrin, T. E., and Keasling, J. D. (2006) Designed divergent evolution of enzyme function, *Nature 440*, 1078-1082.

[12] Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N., and Tawfik, D. S. (2006) Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein, *Nature 444*, 929-932.

[13] Starr, T. N., and Thornton, J. W. (2016) Epistasis in protein evolution, *Protein Sci 25*, 1204-1218.

[14] Breen, M. S., Kemena, C., Vlasov, P. K., Notredame, C., and Kondrashov, F. A. (2012) Epistasis as the primary factor in molecular evolution, *Nature 490*, 535-538.

[15] Harms, M. J., and Thornton, J. W. (2013) Evolutionary biochemistry: revealing the historical and physical causes of protein properties, *Nat Rev Genet 14*, 559-571.

[16] Storz, J. F. (2018) Compensatory mutations and epistasis for protein function, *Curr Opin Struct Biol 50*, 18-25.

[17] Gerlt, J. A., Babbitt, P. C., and Rayment, I. (2005) Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity, *Arch Biochem Biophys 433*, 59-70.
[18] Babbitt, P. C., Hasson, M. S., Wedekind, J. E., Palmer, D. R., Barrett, W. C., Reed, G. H., Rayment, I., Ringe, D., Kenyon, G. L., and Gerlt, J. A. (1996) The enolase superfamily: a general strategy for enzyme-catalyzed abstraction of the alpha-protons of carboxylic acids, *Biochemistry 35*, 16489-16501.
[19] Zhao, S., Kumar, R., Sakai, A., Vetting, M. W., Wood, B. M., Brown, S., Bonanno, J. B., Hillerich, B. S., Seidel, R. D., Babbitt, P. C., Almo, S. C., Sweedler, J. V., Gerlt, J. A., Cronan, J. E., and Jacobson, M. P. (2013) Discovery of new enzymes and metabolic pathways by using structure and genome context, *Nature 502*, 698-702.
[20] Glasner, M. E., Fayazmanesh, N., Chiang, R. A., Sakai, A., Jacobson, M. P., Gerlt, J. A., and Babbitt, P. C. (2006) Evolution of structure and function in the o-succinylbenzoate synthase/N-acylamino acid racemase family of the enolase superfamily, *J Mol Biol 360*, 228-250.
[21] Schmidt, D. M., Hubbard, B. K., and Gerlt, J. A. (2001) 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 40*, 15707-15715.
[22] Palmer, D. R., Garrett, J. B., Sharma, V., Meganathan, R., Babbitt, P. C., and Gerlt, J. A. (1999) Unexpected divergence of enzyme function and sequence: "N-acylamino acid racemase" is o-succinylbenzoate synthase, *Biochemistry 38*, 4252-4258.
[23] Odokonyero, D., Sakai, A., Patskovsky, Y., Malashkevich, V. N., Fedorov, A. A., Bonanno, J. B., Fedorov, E. V., Toro, R., Agarwal, R., Wang, C., Ozerova, N. D., Yew, W. S., Sauder, J. M., Swaminathan, S., Burley, S. K., Almo, S. C., and Glasner, M. E. (2014) Loss of quaternary structure is associated with rapid sequence divergence in the OSBS family, *Proc Natl Acad Sci U S A 111*, 8535-8540.
[24] Brizendine, A. M., Odokonyero, D., McMillan, A. W., Zhu, M., Hull, K., Romo, D., and Glasner, M. E. (2014) Promiscuity of Exiguobacterium sp. AT1b o-succinylbenzoate synthase illustrates evolutionary transitions in the OSBS family, *Biochem Biophys Res Commun 450*, 679-684.
[25] Sakai, A., Xiang, D. F., Xu, C., Song, L., Yew, W. S., Raushel, F. M., and Gerlt, J. A. (2006) 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 45*, 4455-4462.
[26] Taylor Ringia, E. A., Garrett, J. B., Thoden, J. B., Holden, H. M., Rayment, I., and Gerlt, J. A. (2004) Evolution of enzymatic activity in the enolase superfamily: functional studies of the promiscuous o-succinylbenzoate synthase from Amycolatopsis, *Biochemistry 43*, 224-229.
[27] Odokonyero, D., McMillan, A. W., Ramagopal, U. A., Toro, R., Truong, D. P., Zhu, M., Lopez, M. S., Somiari, B., Herman, M., Aziz, A., Bonanno, J. B., Hull,

K. G., Burley, S. K., Romo, D., Almo, S. C., and Glasner, M. E. (2018) Comparison of Alicyclobacillus acidocaldarius o-Succinylbenzoate Synthase to Its Promiscuous N-Succinylamino Acid Racemase/ o-Succinylbenzoate Synthase Relatives, *Biochemistry 57*, 3676-3689.

[28] Kalyanaraman, C., Imker, H. J., Fedorov, A. A., Fedorov, E. V., Glasner, M. E., Babbitt, P. C., Almo, S. C., Gerlt, J. A., and Jacobson, M. P. (2008) Discovery of a dipeptide epimerase enzymatic function guided by homology modeling and virtual screening, *Structure 16*, 1668-1677.

[29] Lukk, T., Sakai, A., Kalyanaraman, C., Brown, S. D., Imker, H. J., Song, L., Fedorov, A. A., Fedorov, E. V., Toro, R., Hillerich, B., Seidel, R., Patskovsky, Y., Vetting, M. W., Nair, S. K., Babbitt, P. C., Almo, S. C., Gerlt, J. A., and Jacobson, M. P. (2012) Homology models guide discovery of diverse enzyme specificities among dipeptide epimerases in the enolase superfamily, *Proc Natl Acad Sci U S A 109*, 4122-4127.

[30] Vick, J. E., Schmidt, D. M., and Gerlt, J. A. (2005) Evolutionary potential of (beta/alpha)8-barrels: in vitro enhancement of a "new" reaction in the enolase superfamily, *Biochemistry 44*, 11722-11729.

[31] Schmidt, D. M., Mundorff, E. C., Dojka, M., Bermudez, E., Ness, J. E., Govindarajan, S., Babbitt, P. C., Minshull, J., and Gerlt, J. A. (2003) Evolutionary potential of (beta/alpha)8-barrels: functional promiscuity produced by single substitutions in the enolase superfamily, *Biochemistry 42*, 8387-8393.

[32] Odokonyero, D., Ragumani, S., Lopez, M. S., Bonanno, J. B., Ozerova, N. D., Woodard, D. R., Machala, B. W., Swaminathan, S., Burley, S. K., Almo, S. C., and Glasner, M. E. (2013) Divergent evolution of ligand binding in the o-succinylbenzoate synthase family, *Biochemistry 52*, 7512-7521.

[33] Taylor, E. A., Palmer, D. R., and Gerlt, J. A. (2001) The lesser "burden borne" by o-succinylbenzoate synthase: an "easy" reaction involving a carboxylate carbon acid, *J Am Chem Soc 123*, 5824-5825.

[34] Zhu, W. W., Wang, C., Jipp, J., Ferguson, L., Lucas, S. N., Hicks, M. A., and Glasner, M. E. (2012) Residues required for activity in Escherichia coli o-succinylbenzoate synthase (OSBS) are not conserved in all OSBS enzymes, *Biochemistry 51*, 6171-6181.

[35] Truong, D. P. (2020) Catalytic Promiscuity and the evolutionary mechanism of NSAR reaction specificity in the NSAR/OSBS subfamily. PhD dissertation. Texas A & M University College Station, TX.

[36] Gulick, A. M., Schmidt, D. M., Gerlt, J. A., and Rayment, I. (2001) 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 40*, 15716-15724.

[37] Lykidis, A., Mavromatis, K., Ivanova, N., Anderson, I., Land, M., DiBartolo, G., Martinez, M., Lapidus, A., Lucas, S., Copeland, A., Richardson, P., Wilson, D. B., and Kyrpides, N. (2007) Genome sequence and analysis of the soil cellulolytic actinomycete Thermobifida fusca YX, *J Bacteriol 189*, 2477-2486.

[38] Niesen, F. H., Berglund, H., and Vedadi, M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability, *Nat Protoc 2*, 2212-2221.