EVOLUTION OF ENZYME SPECIFICITY IN THE OSBS FAMILY

An Undergraduate Research Scholars Thesis

by

CHRISTINE MICHELLE JONES

Submitted to Honors and Undergraduate Research
Texas A&M University
in partial fulfillment of the requirements for the designation as

UNDERGRADUATE RESEARCH SCHOLAR

Approved by
Research Advisor: Dr. Margaret E. Glasner

May 2013

Major: Genetics

TABLE OF CONTENTS

ABSTRACT

Evolution of Enzyme Specificity in the OSBS Family. (May 2013)

Christine Michelle Jones
Department of Biochemistry and Biophysics
Texas A&M University

Research Advisor: Dr. Margaret E. Glasner
Department of Biochemistry and Biophysics

In addition to their primary biological function, many proteins are at least moderately capable of catalyzing secondary, promiscuous activities that may have a major role in enzyme evolution. Mounting evidence supports the idea that new enzymes can evolve when natural selection optimizes these weak promiscuous activities. It is not known what features of promiscuous enzymes enable the evolution of new activities. The *o*-succinylbenzoate synthase (OSBS) family of proteins has a subfamily capable of conducting *N*-succinylamino acid racemase (NSAR) activity in vitro. Analysis of genomic operon context has indicated that many of these NSAR/OSBS proteins conduct the NSAR activity in vivo. A specific asparagine residue in the active site is conserved only in the NSAR subfamily and not the rest of the OSBS family and is suspected to function directly in the NSAR reaction. The residue was replaced in a member of the NSAR subfamily by the corresponding residue from a non-promiscuous OSBS via site-directed mutagenesis. If the asparagine plays a direct role in the chemistry of the racemization, there should be a significantly larger effect on the NSAR activity relative to OSBS activity. Preliminary results, however, indicate that the effect on OSBS activity is much larger than anticipated. We were also interested in the extreme sequence divergence within the OSBS family as a whole. Generally, homologous enzymes that catalyze the same reaction share at least 40% sequence identity. Sequence identity within the OSBS family is as low as 15%. We hypothesize

that while OSBS activity was conserved, functionally important residues diverged within OSBS subfamilies. From the crystal structure of the OSBS from *Thermosynechococcus elongatus*, a member of the Cyanobacteria 1 subfamily, residues were selected for site-directed mutagenesis based on proximity to the active site and ability to orient ligand for binding and catalysis. Kinetics data from mutants indicates that these non-catalytic residues have significant impact on the efficiency of enzyme activity.

CHAPTER I

INTRODUCTION

A number of problems result from our limited understanding of protein evolution and the relationship between protein structure and function. Misannotation levels are extremely high in many public sequence databases, limiting their usefulness as a resource, because reliance on homology for determining function ignores the fact that many closely related or highly sequence identical enzymes have diverged in function[1,2]. This eliminates opportunities for applications in many fields in medicine and industry. Gaining a better understanding of enzyme evolution allows for improvements in both function prediction and in protein engineering.

We use the *o*-succinylbenzoate synthase (OSBS) family as a model system to help us answer questions about the intersections of enzyme structure and function. The OSBS family is part of the enolase superfamily, a very diverse group of related enzymes[3]. Members of the enolase superfamily catalyze a number of distinct reactions, from epimerization to lactonization to dehydration[3,4]. But there is one feature shared by all members of the superfamily: a conserved set of active site residues that conduct a conserved partial reaction[3] (Figure 1). Namely, a base abstracts a carboxylate α-carbon to form an enolate anion intermediate stabilized by a metal ion. While these conserved catalytic residues are responsible for conservation of the shared partial reaction, other residues in these proteins must be the determinants of specificity, by influencing factors like ligand orientation and binding[1]. Members of the OSBS family use this partial reaction to catalyze a dehydration reaction in Vitamin K biosynthesis (Figure 2). There are two primary reasons we are interested in the OSBS family as a model for enzyme evolution: 1)

members of the family are extremely sequence divergent, and 2) one subfamily of the OSBS family is catalytically promiscuous.

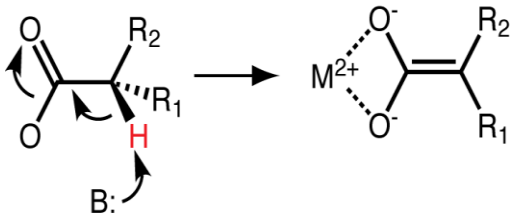

Figure 1: The partial reaction conserved throughout the enolase superfamily. A base abstracts a proton from an α-carbon of a carboxylate to form an enolate anion intermediate that is stabilized by a metal ion.

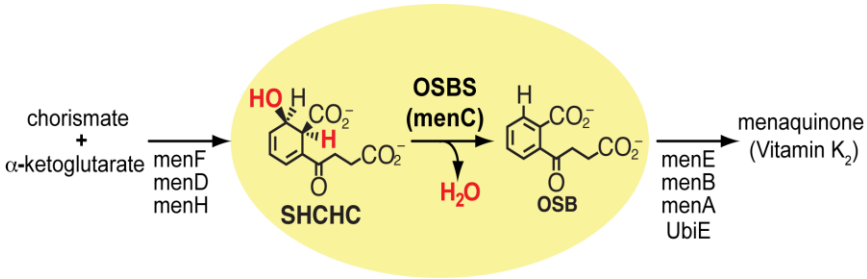

Figure 2: The OSBS family catalyzes a dehydration reaction in the pathway for synthesis of menaquinone, or Vitamin K. The gene responsible for the OSBS function is menC.

Sequence Diversity

Most homologous, isofunctional enzymes share around 40% sequence identity[5]. However, members of the OSBS family are extremely sequence divergent, with some members sharing as little as 15% sequence identity[1]. The only residues conserved throughout the family are those shared with the entirety of the enolase superfamily. One would expect residues that determine reaction specificity to be shared among homologous proteins that catalyze the same reaction. How do proteins that show so little sequence similarity catalyze the same reaction? We hypothesize that while function is conserved, functionally important residues have diverged between OSBS subfamilies. Our hypothesis is that different subfamilies are using different sets of non-catalytic, functionally important residues to bind and orient the ligand for catalysis. To test this hypothesis, we have divided the OSBS family into eight subfamilies in which the members share >40% identity[6] (Figure 3). We are determining which residues affect OSBS activity in representative members of each subfamily.

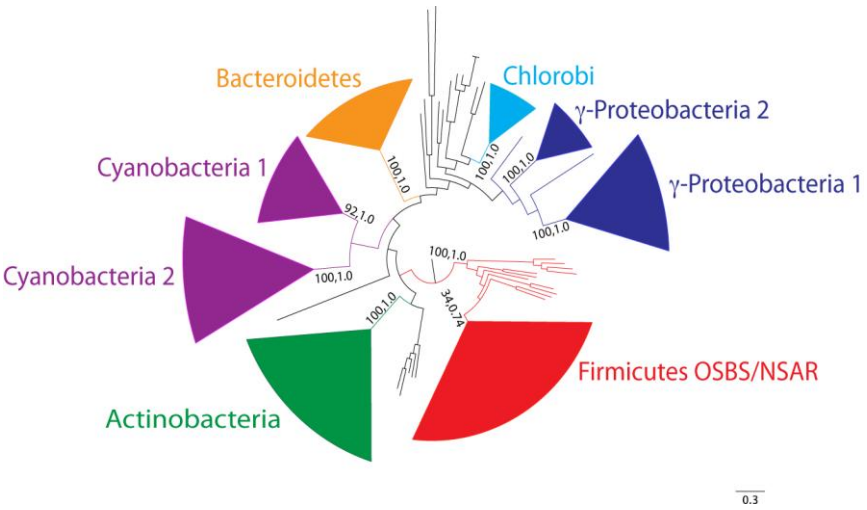

Figure 3[6]: The OSBS family has been divided into eight subfamilies based on a 40% sequence identity cutoff.

To determine which residues may be important for substrate specificity in the Cyanobacteria 1 subfamily, I examined the crystal structure of the OSBS from *Thermosynechococcus elongatus*. Six residues were chosen for mutation based on their proximity to the substrate binding site (Figure 4). Site-directed mutagenesis yielded five of the chosen mutants. The five mutants were subjected to a complementation assay, but results were obscured by a poor expression system. Two of the mutants have since been measured for enzyme activity relative to the wildtype. One of the mutants had no significant effect on activity, while the activity of the other was drastically reduced.

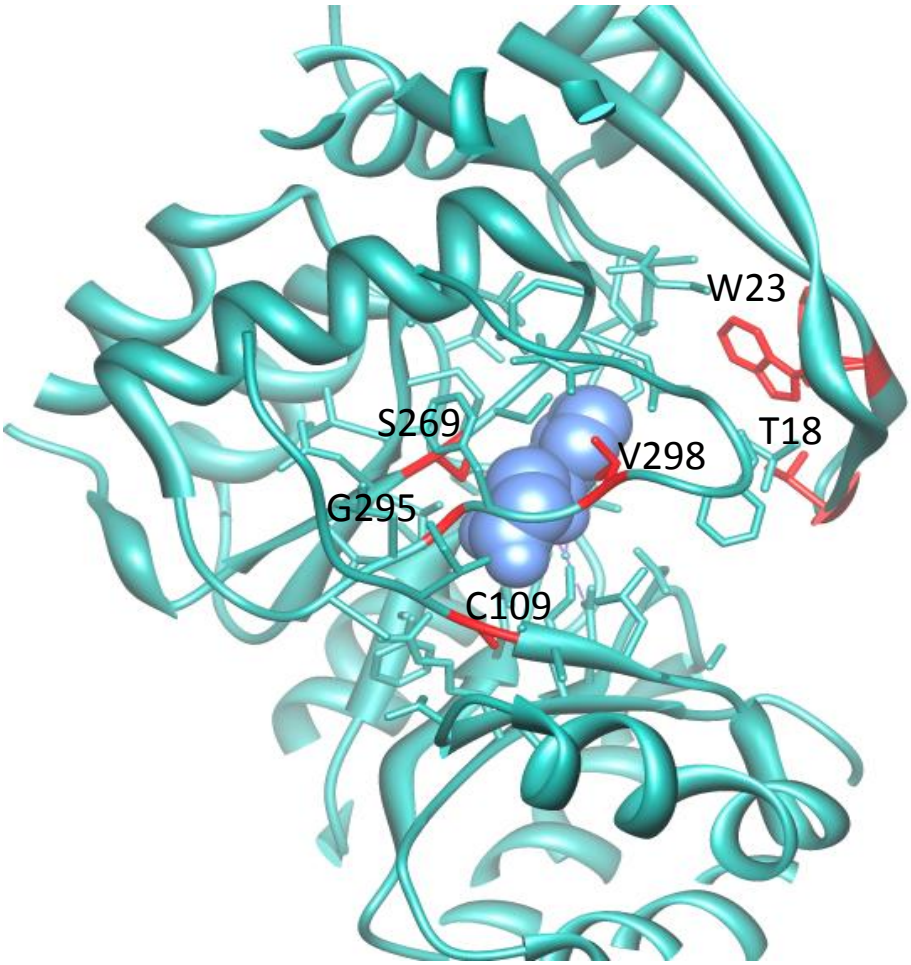

Figure 4: Mutations selected for *T. elongatus* OSBS. Position of OSB in the active site was modeled by aligning *T. elongatus OSBS* (PDB: 2OZT) to *Thermobifida fusca* OSBS bound to OSB (PDB: 2QVH).

Promiscuity

According to the traditional view of protein evolution, a gene duplication or amplification event is followed by the accumulation of mutations in either duplicate gene, and eventually the functions of the two proteins diverge (Figure 5). The problem with this model is that deleterious mutations are much more common than advantageous mutations, so the duplicate gene would more likely lose function altogether. It is also relatively unlikely that this model is the mechanism for the evolution of all proteins, since evolution would have to go through a series of low-likelihood events to create a function *de novo*. There is a more promising alternative model[7] (Figure 6). In this model, promiscuous activity serves as raw material for evolution. Promiscuous proteins are proteins capable of catalyzing at least two distinct chemical reactions in the same active site, only one of which tends to be physiological. If both activities contribute to the fitness of the organism, the enzyme can undergo selection toward both functions and become bifunctional. If a gene duplication event occurs, each activity can be optimized in one of the two paralogs. Recent evidence indicates that many proteins are likely to exhibit promiscuous activities[7], which means a lot of raw material with which to develop new function exists in nature.

Many members of the Firmicutes subfamily (Figure 3) are catalytically promiscuous for *N*-succinylamino acid racemization, or NSAR[8] (Figure 7). NSAR is a step in the pathway for converting D-amino acids to L-amino acids[9]. In the biochemically characterized promiscuous NSAR/OSBS enzymes, the NSAR and OSBS reactions are catalyzed at similar rates *in vitro*[9], so the only way to determine which activity is the biological function of a particular enzyme is to

look at where its gene is located within the organism's genome. If the gene is located near other genes in the Vitamin K biosynthesis pathway, we expect it to function as an OSBS *in vivo*. On the other hand, if it is located near genes that function in the D-amino acid to L-amino acid conversion pathway, its biological activity is most likely NSAR. Examination of the genome context of the genes from the NSAR/OSBS subfamily (data not shown) and other previous work[9] reveals that some of these NSAR/OSBS proteins are likely to be bifunctional *in vivo*, while most are more likely to only be responsible for one of the two activities[8,9]. We are using the Firmicutes NSAR/OSBS subfamily (Figure 3) to help elucidate what features of promiscuous proteins are utilized for the evolution of new activities.

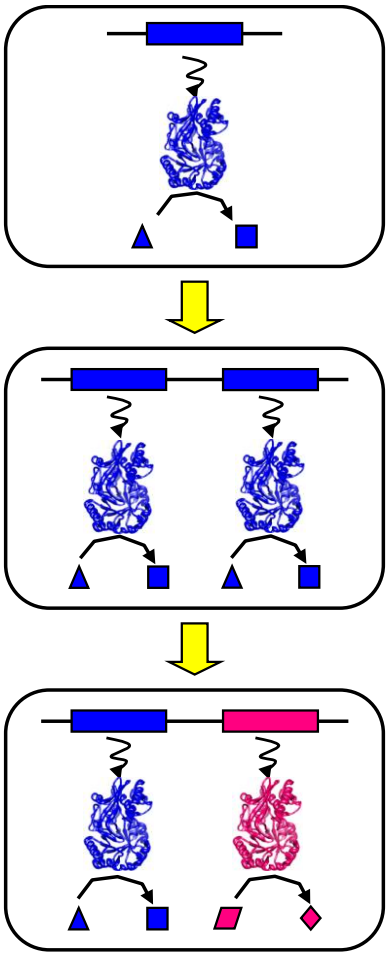

Figure 5: The traditional model of protein evolution states that gene duplication is followed by accumulation of mutations, which eventually leads to functional divergence in the two paralogous proteins.

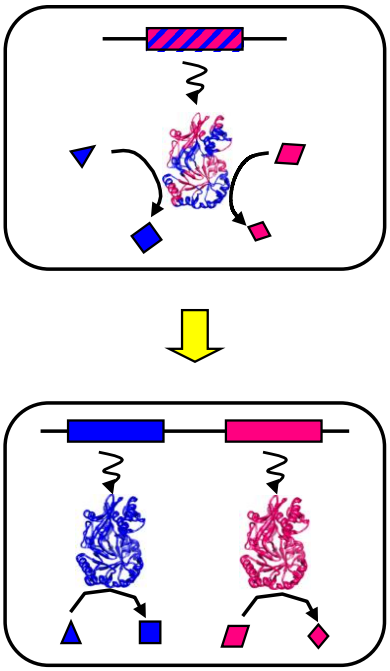

Figure 6: The newer model of evolution states that the gene for a promiscuous protein can undergo duplication, and the existent promiscuous function can be optimized in one or the other paralog. The end result is the same, but the process of developing a new function is higher likelihood.

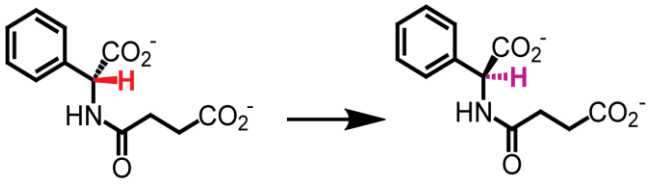

Figure 7: *N*-succinylamino acid racemization, or NSAR, is an additional activity exhibited by many catalytically promiscuous members of the Firmicutes NSAR/OSBS subfamily.

Since catalytic residues are conserved in all subfamilies, but only the NSAR/OSBS proteins also catalyze a chemically distinct reaction, other residues within the enzymes must be responsible for varying reaction specificities. Sequence analysis suggests that an asparagine residue in close proximity to a conserved catalytic lysine may be responsible for these differing activities[3]. The asparagine is conserved within the NSAR/OSBS subfamily, but replaced by other, generally hydrophobic amino acids (such as valine and leucine), in the other subfamilies. The asparagine is also conserved in the Dipeptide Epimerase family (part of the enolase superfamily), which catalyzes a similar racemization reaction. We hypothesize that this asparagine functions in the NSAR reaction by lowering the pK_a of an adjacent catalytic lysine to allow it to participate in the abstraction of a proton from the substrate during racemization (Figure 8). I mutated the

asparagine in an NSAR/OSBS from *Amycolatopsis* sp. T-1-60 to valine, which is the residue at the corresponding position in the non-promiscuous OSBS from *Escherichia coli*. My data indicate that OSBS activity is much lower in the mutant.

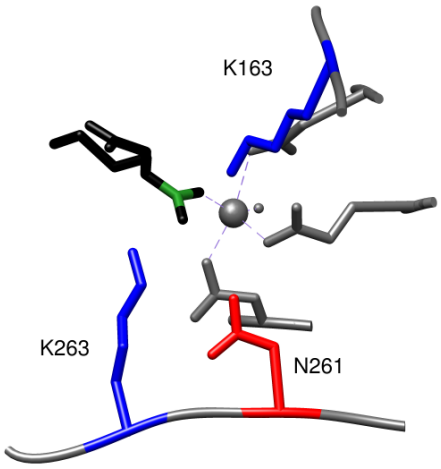

Figure 8: Conserved asparagine at position 261 in *Amycolatopsis* sp. T-1-60 NSAR/OSBS (PDB: 1SJA[10]). The lysines highlighted in blue are involved in abstraction of the proton from the α-carbon (highlighted in green) of *N*-acetylmethionine (shown here in black) during racemization.

CHAPTER II

METHODS

Site-directed mutagenesis and plasmid purification

Site-directed mutagenesis was performed using the QuikChange mutagenesis protocol with a two-stage polymerase chain reaction (PCR). Primer sequences are included in Table 1. The templates used for mutagenesis were the menC genes from *T. elongatus* subcloned into a pET15b vector and *Amycolatopsis* sp. T-1-60 cloned into pET17b (Novagen).

Table 1: Primer sequences for mutagenesis. Reverse primers are reverse complement to the forward.

Target Mutation	Primer Sequence (forward)
T. elongatus W23A	GCGCAGGGCGTGGCGCGCTCTCGTTCTGG
T. elongatus G295A	CCTGCCACGCACTTGCGTTCGGTGTGGACC
T. elongatus C109A	CCGTGGCCAATCGCGGCACTCCTGGGCTCCG
T. elongatus S269A	CGCAACGTCTGGTTTTCGCGAGTGCCCTCGAGGG
T. elongatus V298A	CCACGCACTTGGCTTTGGGGCGGATCGCTGGAGAAGCGC
Amycolatopsis N261V	CCAAATCGTGGTGATCAAACCGGGC

Plasmid obtained from PCR were purified (Qiagen), transformed by electroporation at 1500 V into *E. coli* DH5a cells, and then plated and incubated at 37° C for 16 hours on solid LB media with added carbenicillin. Liquid LB media cultures of single colonies taken from these plates were incubated in a shaker for approximately 12 hours and then centrifuged (4° C, 3000 rpm) and pelleted. From these pellets, the plasmids were purified using the Quick Plasmid Miniprep Kit from Invitrogen. Purified plasmids were sequenced in both directions to confirm the mutated sequence (Eton).

Protein purification

Amycolatopsis mutants were expressed in *E. coli* strain BW25113 (menC::kan) (gift from Dr. John Gerlt, University of Illinois). This strain was converted into a DE3 strain to express T7 RNA polymerase using the λDE3 lysogenization kit from Novagen. Expressing the mutants in the menC$^-$ strain ensured that the purified proteins would not be contaminated with wild-type OSBS. These mutants were plated and incubated at 30° C for 16 hours in liquid LB media with added carbenicillin and kanamycin. Cultures were pelleted, resuspended with 20 mM Tris, pH 8 and 5mM MgCl2, and lysed. Lysed cells were again pelleted, and the supernatant was injected onto a 20 mL HiTrap^{TM} 16/10 DEAE FF (GE) to separate by charge. Fractions were run on an SDS-PAGE gel, and fractions with the most pure protein were combined, then run on 3, 5 mL HiTrap^{TM} Phenyl FF column (GE) connected in series. Fractions containing the most pure protein from this purification were concentrated and used for enzyme kinetics.

Enzyme kinetic assays

Preparations of purified protein were tested for OSBS activity using varying concentrations of the substrate 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate (SHCHC) under different pH conditions, and analyzed with a spectrophotometer. Change in absorbance of SHCHC at 310 nm over time was recorded. From these data, the initial rate of reaction for each concentration of SHCHC was determined. These data were fitted to the Michaelis-Menten equation to determine K_m and k_{cat} using Kaleidograph.

Complementation

Plasmids with correct mutations (as well as the wildtype strain) were transformed into the same *E. coli* strain BW25113 (menC::kan) as used in protein expression and purification and then plated on solid LB media with added carbenicillin and kanamycin and grown overnight. Single colonies from these plates were selected for 1 mL overnight cultures in liquid LB with added carbenicillin and kanamycin. 1:20 dilutions of these cultures were made with liquid menC minimal media (described below) and then 2 μL of the dilution added to anaerobic tubes containing liquid menC minimal media supplemented with carbenicillin and kanamycin. Anaerobic cultures were incubated at 37° C, and measurements of optical density at 600 nm were made approximately every 3 hours.

Liquid menC minimal media was made by adding 10.49 g KH_2PO_4, 5.44 g K_2HPO_4, 2 g $(NH_4)_2SO_4$, 29.6 μL 1M $MnSO_4 \bullet H_2O$, 415.3 μL 1M $MgSO_4$, 3.4 μL 1M $CaCl_2$,

0.125 mg $FeSO_4 \bullet 7\ H_2O$, 1 g casamino acids (acid hydrolysate), 1 mg thiamine HCl, 5.1 g trimethylamine *N*-oxide, 6.64 mL 50% glycerol, and H_2O to a final volume 1 L.

CHAPTER III

RESULTS

Sequence divergence

From the crystal structure of OSBS from *Thermosynechococcus elongatus*, six residues were selected for mutation to alanine (Figure 4). The crystal structure for the *T. elongatus* OSBS does not include the native ligand, so the ligand from a similar protein (the OSB from *Thermobifida fusca*) was modeled in using a sequence/structure alignment with the program Chimera. Residues were chosen based on a loose assessment of proximity to the bound ligand or ability to contact the ligand during binding (entry into the active site). Of these six, five were successfully obtained from site-directed mutagenesis. These five mutants – C109A, G295A, W23A, V298A, and S269A – were subjected to complementation to determine whether a significant difference in growth between any of the mutants and the wildtype could be observed. No significant difference between any of the mutants and the wildtype was seen (data not shown). Unfortunately, the vector used for the expression of the mutant and wildtype proteins is designed for overexpression – meaning that any differences in OSBS activity as a result of mutation would be masked and compensated by the sheer amount of protein produced during growth.

In order to more directly examine the effects of mutation on activity, activity assays were carried out by Denis Odokonyero and summarized here (Table 2). Activity assays measured the change in absorbance of the substrate during initial catalysis. W23A and G295A were the only mutants included in preliminary data. While W23A saw effectively no change, there was an

approximately 80-fold decrease in activity for G295A compared to the wildtype. Mutation of the residue corresponding to G295 in *E. coli* OSBS, G288, is known to have an approximately 500-fold decrease in activity compared to wildtype (Table 2).

Table 2: Activity data from wildtype *T. elongatus* and *E. coli* OSBS, as well as mutants.

OSBS	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
***T. elongatus* WT**	80	78	1.0×10^6
***T. elongatus* G295A**	3.5	304	1.0×10^4
***T. elongatus* W23A**	33	29	1.1×10^6
****E. coli* WT**	24	12	2.0×10^6
****E. coli* G288A**	n.d.	n.d.	3.8×10^3

*Courtesy of Denis Odokonyero. **Courtesy of Wan Wen Zhu[6].

Catalytic promiscuity

To determine whether a specific conserved asparagine in the Firmicutes NSAR/OSBS subfamily was responsible for the chemistry of the NSAR reaction, I mutated the residue to valine, a residue conserved in non-promiscuous members of the OSBS family, and then conducted kinetics experiments in an attempt to determine the effects of mutation on OSBS and NSAR activities. Compared to wildtype *Amycolatopsis* NSAR/OSBS, the mutant N261V appeared to

express less well (data not shown). During protein purification, the peak for the mutant was significantly smaller than for wildtype. After concentration and storage, the protein began to visibly precipitate and aggregated within a few days. Preliminary results indicate that OSBS activity in the mutant is significantly lower than for wildtype (Figure 9), with very little activity detected at the highest protein concentration I was able to use (Figure 10). The corresponding residue in a non-promiscuous OSBS (*E. coli* OSBS), V233, has no significant effect on OSBS activity when mutated to asparagine (Table 3).

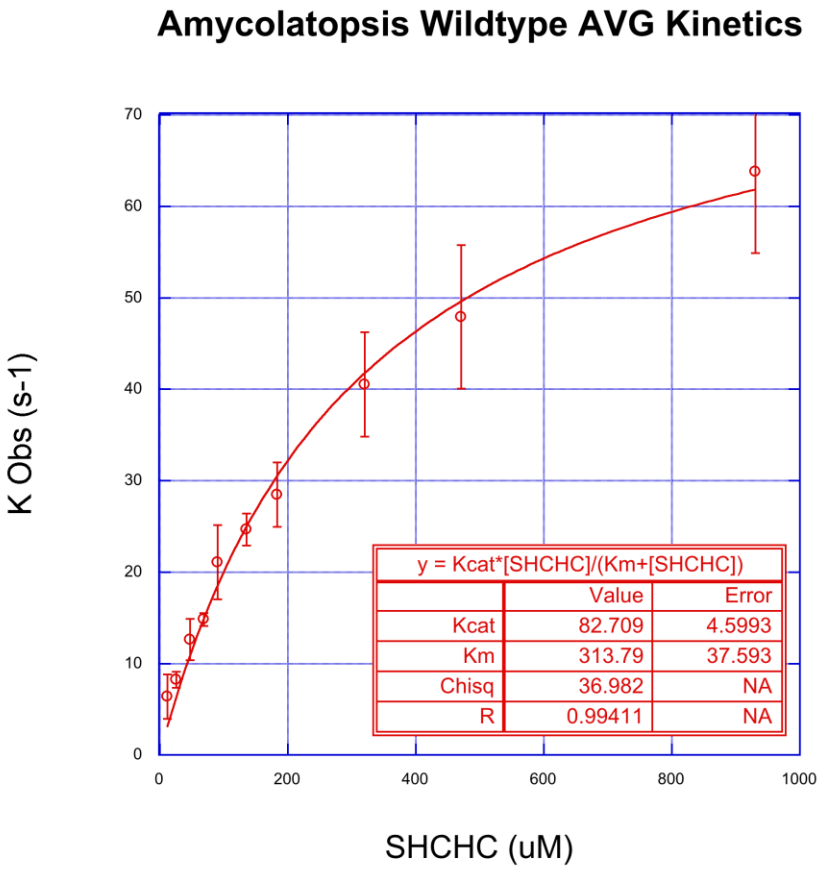

Figure 9: Activity data for wildtype *Amycolatopsis*, courtesy of Andy McMillan.

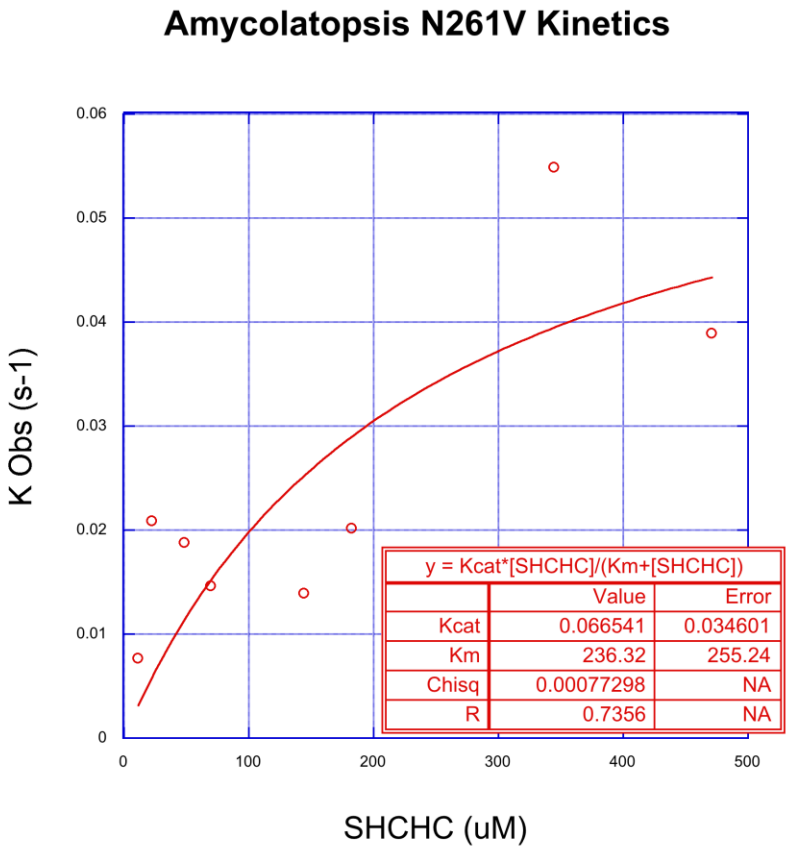

Figure 10: Activity data for N261V in *Amycolatopsis* NSAR/OSBS.

Table 3: Activity data from *Amycolatopsis* and *E. coli* wildtype OSBS, as well as mutants.

OSBS	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
***Amycolatopsis* WT**	83	314	2.6×10^5
***Amycolatopsis* N261V**	0.07	236	3.0×10^2
****E. coli* WT**	24	12	9.6×10^5
****E. coli* V233A**	94	99	2.0×10^6

*Courtesy of Andy McMillan. **Courtesy of Wan Wen Zhu[6].

CHAPTER IV

CONCLUSIONS

Sequence Divergence

Preliminary data includes the kinetics data for only two of the five mutants – G295A and W23A. These two mutants were prioritized over the others for different reasons. G295A has a corresponding residue in *Escherichia coli* (part of the γ-Proteobacteria 1 subfamily) for which we already had data[6]. This corresponding residue – G288 in *E. coli* OSBS– has a drastic effect on activity when mutated to alanine[6] (Table 2). The approximately 500-fold difference in activity may be due to a steric conflict encountered when glycine, a very small amino acid, is replaced with alanine, a slightly larger amino acid (Figure 11). The methyl side chain on alanine is probably displacing the ligand from the proper orientation for binding and preventing efficient catalysis. We have determined that G295A in *T. elongatus* OSBS also has a significant effect on activity relative to the wildtype. In this case, the activity drops by around 80-fold (Table 2). We believe that this drop in activity is also due to steric conflict, although the less drastic difference in activity compared to the effects of G288A in *E. coli* may be due to the differences in orientation of the ligand within the different OSBS's. G295 may be located further from the bound ligand, which would give more flexibility in terms of chemistry and activity when changing the residue to become slightly larger.

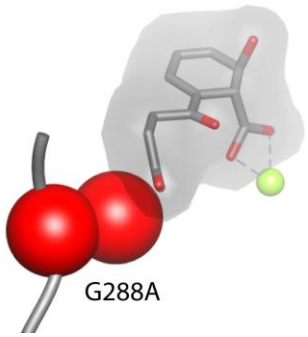

Figure 11[6]: G288A in *E. coli* dramatically lowers OSBS activity, likely due to steric conflict when glycine is changed to alanine.

W23A was prioritized for assay because of its specific position within the active site. Tryptophan has a relatively large side-chain, and in the case of W23 in *T. elongatus* OSBS, this side chain is present in a loop that 'caps' the active site during binding and catalysis. We hypothesized that since the side-chain juts into the active site and toward the ligand, it may play a role in closing or restricting the active site. Data from kinetics (Table 2) indicates that the mutation to alanine has no significant effect on OSBS activity, with perhaps even a slight increase in activity. This could be due to the fact that switching the large amino acid tryptophan with the small amino acid alanine opens the active site and allows for the substrate to more easily move into position for catalysis.

Preliminary results from mutants in *T. elongatus* OSBS have shown interesting results in terms of the importance of non-catalytic residues for function. Further analysis, with data from more mutants, is necessary for any significant conclusions, but it is obvious from these and other experiments that the structure of the active site is intricate and complicated and that some non-catalytic residues are more essential for specificity than others.

Catalytic Promiscuity

To determine whether an asparagine conserved in the Firmicutes NSAR/OSBS subfamily is involved in the chemistry of the promiscuous function, I mutated it to the corresponding valine residue from a non-promiscuous OSBS and then conducted kinetics experiments. We had expected mild to moderate changes in OSBS activity, with significantly higher effects on NSAR activity. My results (Table 3) indicate that OSBS activity in *Amycolatopsis* NSAR/OSBS N261V has been significantly affected by the mutation. The deviation from expectations may be due less to a direct reliance on N261 for OSBS chemistry and more to an issue of folding or stability. This idea is supported by the expression levels and swift, visible precipitation of the mutant upon purification. It is interesting to note that the corresponding mutation in a non-NSAR OSBS from *E. coli*, V233N, has been shown to have very little effect on OSBS activity, with around only a 2-fold decrease in efficiency relative to wildtype[6] (Table 3). This reinforces the idea that different groups within the OSBS family are utilizing different residues in distinct manners.

Further experimentation will be done to determine the exact effects on both OSBS and NSAR activity. At higher protein concentration it should be easier to accurately detect the lower activity

of the mutant. If so, we can vary the pH of the reaction mixture to determine the role of N261 in NSAR chemistry.

REFERENCES

1. Palmer, D. R. J., Garrett, J. B., Sharma, V., Meganathan, R., Babbitt, P. C., & Gerlt, J. A. (1999). Unexpected divergence of enzyme function and sequence: "*N*-acylamino acid racemase" is *o*-succinylbenzoate synthase. *Biochemistry*, **38**, 4252-4258.

2. Devos, D. & Valencia, A. (2000). Practical limits of function prediction. *Proteins*, **41**, 98-107.

3. Babbitt, P. C., Hasson, M. S., Wedekind, J. E., Palmer D. R. J., Barrett, W. C., Reed, G. H., Rayment, I., Dagmar, R., Kenyon, G. L., & Gerlt, J. A. (1996). The enolase superfamily: a general strategy for enzyme-catalyzed abstraction of the α-protons of carboxylic acids. *Biochemistry*, **35**, 16489-16501.

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

5. Wilson, C. A., Kreychman, J., & Gerstein M. (2000). Assessing annotation transfer for genomics: quantifying the relations between protein sequence, structure, and function through traditional and probabilistic scores. *J. Mol. Biol.* **297**, 233-249.

6. Zhu, W. W., Wang, C., Jipp, J., Ferguson, L., Lucas, S. N., Hicks, M. A., & 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.

7. Khersonsky, O. & Tawfik D. S. (2010). Enzyme promiscuity: a mechanistic and evolutionary perspective. *Biochemistry*, **79**, 471-505.

8. Glasner, M. E., Fayazmanesh, N., Chiang, R. A., Sakai, A., Jacobson, M. P., Gerlt, J. A., & 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.

9. Sakai, A., Xiang, D. F., Xu, C., Song, L., Yew, W. S., Raushel, F. M., & 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.

10. Thoden, J. B., Taylor, E. A., Garrett, J. B., Gerlt J. A., Holden, H. M., Rayment, I. (2004). Evolution of enzymatic activity in the enolase superfamily: structural studies of the promiscuous o-succinylbenzoate synthase from Amycolatopsis. *Biochemistry*, **43**, 5716-5727.