BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Residues Required for Activity in *Escherichia coli* *o*-Succinylbenzoate Synthase (OSBS) Are Not Conserved in All OSBS Enzymes

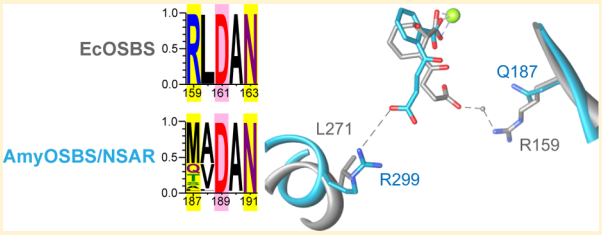
Wan Wen Zhu,[†] Chenxi Wang,[†] Jacob Jipp,[†] Lance Ferguson,[†] Stephanie N. Lucas,[‡] Michael A. Hicks,[‡] and Margaret E. Glasner*[,†]

[†]Department of Biochemistry and Biophysics, Texas A&M University, 2128 TAMU, College Station, Texas 77843-2128, United States

[‡]Department of Biopharmaceutical Sciences, School of Pharmacy and California Institute for Quantitative Biomedical Research, University of California, 1700 4th Street, San Francisco, California 94158, United States

Supporting Information

ABSTRACT: Understanding how enzyme specificity evolves will provide guiding principles for protein engineering and function prediction. The *o*-succinylbenzoate synthase (OSBS) family is an excellent model system for elucidating these principles because it has many highly divergent amino acid sequences that are <20% identical, and some members have evolved a second function. The OSBS family belongs to the enolase superfamily, members of which use a set of conserved residues to catalyze a wide variety of reactions. These residues are the only conserved residues in the OSBS family, so they are not sufficient to determine reaction specificity. Some enzymes in the OSBS family catalyze another reaction, *N*-succinylamino acid racemization (NSAR). NSARs cannot be segregated into a separate family because their sequences are highly similar to those of known OSBSs, and many of them have both OSBS and NSAR activities. To determine how such divergent enzymes can catalyze the same reaction and how NSAR activity evolved, we divided the OSBS family into subfamilies and compared the divergence of their active site residues. Correlating sequence conservation with the effects of mutations in *Escherichia coli* OSBS identified two nonconserved residues (R159 and G288) at which mutations decrease efficiency ≥200-fold. These residues are not conserved in the subfamily that includes NSAR enzymes. The OSBS/NSAR subfamily binds the substrate in a different orientation, eliminating selective pressure to retain arginine and glycine at these positions. This supports the hypothesis that specificity-determining residues have diverged in the OSBS family and provides insight into the sequence changes required for the evolution of NSAR activity.



Understanding how enzyme specificity evolves is a fundamental problem in biochemistry, which has implications for discerning protein structure−function relationships, predicting protein functions, and protein engineering. One basic principle is that homlogous enzymes typically share some aspect of catalysis, such as a step in their mechanism, even if they have very different overall activities.[1] For example, all enzymes in the enolase superfamily catalyze a partial reaction in which a base abstracts a proton α to a carboxylate, forming a metal-stabilized enolate anion intermediate (Figure 1).[2] The enolase superfamily can be divided into more than 20 families that use this mechanistic step to catalyze different overall reactions, including dehydration, racemization, and cycloisomerization of a wide variety of substrates.[2,3]

This shared catalytic feature is mediated by a set of conserved catalytic residues, while specificity is determined by additional catalytic and ligand binding residues, as well as other residues that are outside the active site.[2,4−7] Classifying proteins in enzyme superfamilies according to their specificity is challenging. Indeed, 8−30% of sequences in public databases are misannotated, mainly because of the transfer of annotations from homologous proteins that do not have the same function.[8−11] One likely reason for this problem is the inability of current functional annotation methods to differentiate between residues that are conserved because proteins have the same specificity and residues that are conserved because proteins have some functional similarity, similar structures, and a shared evolutionary history.

Another possibility is that residues that determine specificity are not always conserved among homologous proteins that have the same function. This would help to explain the extraordinary divergence of proteins in the *o*-succinylbenzoate synthase (OSBS) family (EC 4.2.1.113).[12] The OSBS family belongs to the enolase superfamily. OSBS catalyzes the conversion of 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate (SHCHC) to *o*-succinylbenzoate (OSB) (Figure 1C). This reaction is required for menaquinone synthesis in a wide variety of bacteria, a few Archaea, and plants.[12,13] OSBS enzymes can share as little as 15% sequence identity, even

Received: June 6, 2012
Revised: July 6, 2012
Published: July 9, 2012

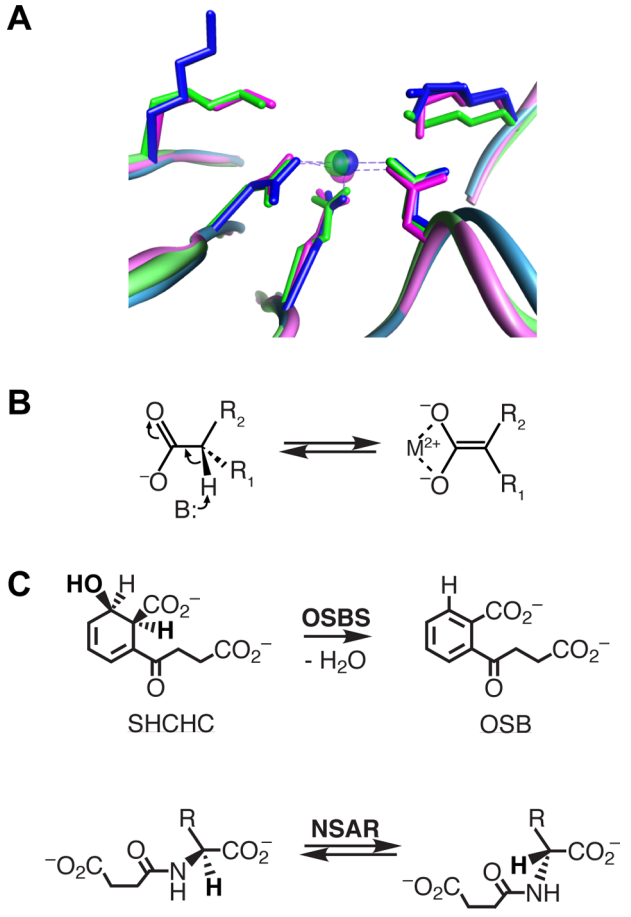


Figure 1. Enolase superfamily. (A) Conserved catalytic residues in the active sites of *Escherichia coli* OSBS (blue, PDB entry 1FHV), *Bacillus subtilis* DE (pink, PDB entry 1TKK), and *Pseudomonas fluorescens* MLE (green, PDB entry 3DGB).[16,23,43] Metal ions are shown as spheres. Starting from the upper left and going counterclockwise, we identify the conserved catalytic residues in EcOSBS as K133, D161, E190, D213, and K235. (B) Conserved partial reaction catalyzed by all enolase superfamily members. (C) *o*-Succinylbenzoate synthase (OSBS) and *N*-succinylamino acid racemase (NSAR) reactions. R is a hydrophobic amino acid side chain.

though they have a single evolutionary origin and a conserved function.[12]

On the basis of the phylogeny, we originally divided the OSBS family into five subfamilies. One subfamily includes proteins that catalyze a second reaction, *N*-succinylamino acid racemization (NSAR). NSAR is utilized in a pathway for converting D-amino acids to L-amino acids.[14] All characterized enzymes with NSAR activity also have OSBS activity (A. Sakai and J. Gerlt, personal communication).[14,15] Operon context indicates that some of these proteins are bifunctional in vivo, while the biological function of others is either OSBS or NSAR. Because of high levels of sequence similarity and the bifunctionality of some enzymes, the NSARs cannot be easily segregated into a family separate from the OSBS family. The discovery that both NSAR and OSBS are biologically relevant activities supports the hypothesis that new enzyme activities evolve through promiscuous intermediates.[12] However, the biochemical basis for evolving this new activity is not well-understood.

Like all members of the enolase superfamily, OSBS enzymes are composed of a C-terminal catalytic $(\beta/\alpha)_7\beta$-barrel domain and an N-terminal capping domain with an $\alpha+\beta$ fold that is unique to the enolase superfamily. Two loops from the capping domain, one around position 20 (the 20s loop) and one around position 50 (the 50s loop), form the top of the active site and help determine specificity in some members of the enolase superfamily.[16,17] In the barrel domain, the second lysine found in a KxK motif at the end of the second β-strand is the catalytic base [K133 in EcOSBS (Figure 1A)]. The first acidic residue in motifs DxN, ExP, and DEx on β-strands 3−5 bind the divalent metal ion, and a lysine or arginine on β-strand 6 helps stabilize the transition state.[18] Mutations of these catalytic residues in *Escherichia coli* OSBS (EcOSBS) abolish its activity.[18,19]

The catalytic motifs on strands 2−5 are the only absolutely conserved residues in the whole OSBS family. The lysine on β-strand 6 is also highly conserved, but it is replaced by arginine in one OSBS subfamily. All of these catalytic amino acids are also conserved in other members of the enolase superfamily, including the muconate lactonizing enzyme (MLE) family, the dipeptide epimerase (DE) family, and a number of uncharacterized proteins. Thus, these conserved residues do not determine specificity for the OSBS reaction.[12] Together, the families that have all of these catalytic motifs are classified as the MLE subgroup, which comprises ∼30% of the sequences in the enolase superfamily.[20]

Two hypotheses could explain the apparent lack of conserved specificity-determining residues in the OSBS family. Both offer plausible evolutionary scenarios that are not mutually exclusive. First, the OSBS reaction might not require additional catalytic or ligand-binding residues because thermodynamically it is a relatively easy reaction compared to other reactions in the enolase superfamily.[21] After the proton abstraction step that is shared by all members of the enolase superfamily, dehydration and formation of the aromatic ring are energetically favorable. Substrate binding via coordination of the carboxylate to the metal ion and the hydrophobic pocket formed by the 50s loop could be sufficient to position the substrate for catalysis. These features are shared by many families in the enolase superfamily.[12] In support of this hypothesis, mutation of a single amino acid in L-Ala-D/L-Glu epimerase from *E. coli* or MLE II from *Pseudomonas* sp. P51 relieves a steric clash that allows them to catalyze the OSBS reaction.[22]

Second, additional residues in the active site contribute to the binding and orientation of the substrate for catalysis, but they have diverged within the OSBS family. As a result, each OSBS lineage would have a different set of residues that contribute to catalysis. These residues are expected to be conserved within the lineage but not in the whole family. Support for this hypothesis comes from a comparison of the crystal structures of *E. coli* OSBS (EcOSBS) and the promiscuous OSBS/NSAR from *Amycolatopsis* sp. T-1-60 (AmyOSBS/NSAR). The relative orientation of the barrel and capping domains differs by 18°, which shifts the position of the 20s loop so that it cannot contact the product or the barrel domain the same way in the two structures (Figure 2).[12,23,24] In addition, the conformation of the ligand is different, with the succinyl tail of OSB bent down in EcOSBS and extended in AmyOSBS/NSAR. We expect that structural differences of this magnitude would be accompanied by the evolution of specific contacts that determine ligand conformation, rather than resulting from the accumulation of neutral mutations.

The goal of this work is to test these hypotheses by assessing the contribution of active site residues to the activity of EcOSBS and comparing their conservation in different OSBS

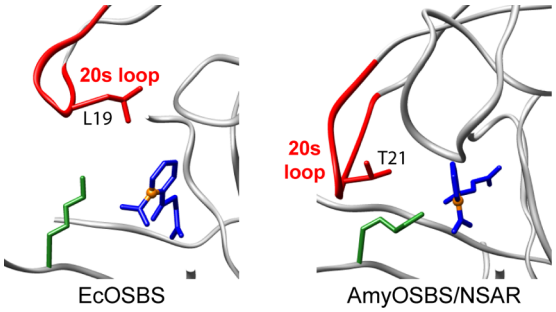


Figure 2. Differences in the position of the 20s loop in EcOSBS (PDB entry 1FHV) and AmyOSBS/NSAR (PDB entry 1SJB).[23,24] The 20s loop (red) in EcOSBS is shifted up relative to that of AmyOSBS/NSAR, leaving its active site open. The catalytic lysine (green) has been modeled in an extended conformation, although its C_δ, C_ε, and N_ζ atoms are disordered in the structure.[23] In contrast, the 20s loop of AmyOSBS/NSAR closes the active site and sits on the catalytic lysine, positioning it for proton abstraction.[24] The 20s loop residue that is closest to the catalytic lysine is colored red (L19 in EcOSBS and T21 in AmyOSBS/NSAR). Both structures are bound to OSB (blue). The carbon from which the proton was abstracted is represented as an orange sphere.

subfamilies. Our results show that most active site mutations have modest effects on activity. This provides some support for the first hypothesis by suggesting that the only active site residues under strong selective pressure are the known catalytic residues, which are conserved in the whole MLE subgroup. However, two mutations (R159M and G288A) decreased enzyme efficiency by 180- and 530-fold, respectively. Comparing six OSBS subfamilies revealed that conservation of these residues is subfamily specific. In particular, differences in sequence between the γ-Proteobacteria subfamily (to which EcOSBS belongs) and the Firmicutes OSBS/NSAR subfamily (to which AmyOSBS/NSAR belongs) showed why these enzymes bind the substrate in different conformations. This result supports the second hypothesis that different OSBS subfamilies require different subsets of amino acids for optimal catalysis. In addition, identification of differences between the γ-Proteobacteria and Firmicutes OSBS/NSAR subfamilies provides insights into the sequence changes required for the evolution of NSAR activity.

EXPERIMENTAL PROCEDURES

Mutagenesis. Site-directed mutagenesis was performed by the QuikChange mutagenesis protocol using a two-stage polymerase chain reaction (PCR) (see the Supporting Information for primers and PCR conditions).[25] The template for mutagenesis was the *menC* gene from *E. coli* (GI 16130196) subcloned into a modified pET15b vector (Novagen) with a 10-histidine N-terminal tag (a gift from J. A. Gerlt, University of Illinois, Urbana, IL). Mutations were confirmed by sequencing in both directions (Eton Bioscience, Inc.).

Protein Purification. Wild-type EcOSBS was expressed in *E. coli* strain BL21(DE3). Mutant EcOSBS enzymes were expressed in *E. coli* strain BW25113 (*menC::kan*) (a gift from J. A. Gerlt). 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. Cultures were grown overnight at 37 °C without induction in 300 mL of Luria-Bertani broth supplemented with carbenicillin and kanamycin. Cells were harvested by centrifugation at 1700*g* for 15 min at 4 °C. They were resuspended in buffer containing 20 mM Tris (pH 8.0), 500 mM NaCl, and 5 mM imidazole. Resuspended pellets were lysed using a Microfluidizer (Microfluidics Corp.) at 1800 psi. After centrifugation at 18000*g* for 30 min at 4 °C, the filtered lysate was applied to a 5 mL HisTrap FF column charged with Ni^{2+} (GE Healthcare). The protein was eluted with a buffer containing 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 apparently homogeneous protein were identified by sodium dodecyl sulfate−polyacrylamide gel electrophoresis and pooled. Amicon Ultra-15 centrifugal filters (30 kDa cutoff) (Millipore) were used to exchange the buffer and concentrate the pooled fractions. Purified proteins were stored in 10 mM Tris (pH 8.0) and 5 mM $MgCl_2$ supplemented with 25% glycerol for storage at −80 °C.

SHCHC Synthesis. SHCHC was synthesized from chorismate and α-ketoglutarate using the enzymes preceding OSBS in the menaquinone synthesis pathway. Chorismate was obtained by isolating it from the culture media of chorismate mutase-deficient *E. coli* strain KA12 (a gift from D. Hilvert, ETH Zürich, Zürich, Switzerland).[26,27] The protocol for synthesizing SHCHC was modified from ref 15 by adding the protein MenH, in addition to EntC and MenD.[28] SHCHC was synthesized by combining 42 mM chorismate, 210 mM α-ketoglutarate, 250 μM thiamine pyrophosphate, 30 μM EntC, 45 μM MenD, 30 μM MenH, 7.5 mM $MgSO_4$, and 50 mM $NaPO_4$ (pH 7.0). SHCHC was purified as described previously.[15] The plasmids encoding EntC and MenD were gifts from J. A. Gerlt.

OSBS Activity Assay. Wild-type and mutant OSBS enzymes were assayed with varying concentrations of SHCHC in 50 mM Tris (pH 8.0) and 0.1 mM $MnCl_2$ at 25 °C. The assays were performed by quantifying the decrease in absorbance at 310 nm ($\Delta\varepsilon$ = −2400 M^{-1} cm^{-1}), as previously described.[15,29] The His tag was not removed prior to assays because control experiments showed that it did not significantly affect activity (data not shown). Initial rates were calculated using VisionPro (Thermo Scientific) and were fit to the Michaelis−Menten equation using Kaleidagraph (Synergy Software).

Data Set. Starting with our manually curated alignment from 2006,[12] we expanded our original data set by downloading all sequences annotated as OSBS from the Structure-Function Linkage Database, which uses Hidden Markov Models (HMMs) to divide superfamilies into families of proteins that are expected to have the same function.[20] Because the extreme divergence of the OSBS family increases the likelihood of misannotation, we retained only proteins that have amino acid sequences >40% identical to those of OSBSs that had been verified based on phylogeny and operon context.[12] Previous results demonstrated that all proteins with sequences >40% identical to those of known OSBSs fall into a monophyletic clade in the phylogeny of the MLE subgroup. This data set includes both OSBS and NSAR enzymes. NSAR enzymes cannot be segregated on the basis of sequence similarity.

The data set was divided into clusters in which proteins share >40% identity with at least one other protein. New sequences in each cluster were aligned to the previously aligned sequences using the profile option in MUSCLE.[30] The resulting alignment was manually adjusted according to a structural alignment of the OSBS enzymes from *E. coli* (PDB entry 1FHV),

dx.doi.org/10.1021/bi300753j | *Biochemistry* 2012, 51, 6171−6181

Table 1. Effect of Active Site Mutations on the Activity of *E . coli* OSBS

structural position[a]	variant	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (M^{-1} s^{-1})	relative k_{cat}/K_M[b]
	wild-type	24 ± 0.8	12 ± 1.8	2.0×10^6	1.0
20s loop	L19A	3.9 ± 0.1	13 ± 2.4	3.0×10^5	0.15
	20s switch[c]	0.4 ± 0.02	9.6 ± 1.6	4.3×10^4	0.022
	20s del[d]	0.2 ± 0.003	29 ± 1.3	6.9×10^3	0.0034
50s loop	L48A	14 ± 0.8	111 ± 17	1.3×10^5	0.063
	F51A	49 ± 2.6	122 ± 21	4.0×10^5	0.20
	L48A/F51A	11 ± 0.8	89 ± 21	1.3×10^5	0.062
	L48M	29 ± 1.7	52 ± 13	5.7×10^5	0.28
	F51Y	18 ± 1	51 ± 14	3.5×10^5	0.18
	L48M/F51Y	27 ± 0.5	73 ± 6	3.7×10^5	0.19
after bar-β1	L109A	2.9 ± 0.2	36 ± 8	8.0×10^4	0.042
bar-β3	R159M	0.9 ± 0.05	80 ± 15	1.1×10^4	0.0056
after bar-β6	V233N	94 ± 4	99 ± 13	9.6×10^5	0.48
after bar-β7	S262G	10 ± 0.5	21 ± 4	4.8×10^5	0.24
	S262T	65 ± 3	343 ± 44	1.9×10^5	0.095
	S263G	73 ± 6	158 ± 33	4.6×10^5	0.23
	S264A	12 ± 0.5	29 ± 4.7	4.1×10^5	0.21
	I265A	6 ± 0.2	10 ± 1.5	5.8×10^5	0.30
	S262A/S263A/S264A/I265A	2.7 ± 0.1	51 ± 10	5.3×10^4	0.027
after bar-β8	G288A	nd[e]	nd[e]	3.8×10^3	0.0019
	T291S	67 ± 3	206 ± 30	3.3×10^5	0.16

[a]Structural position indicates the position of the residues in the context of the fold. The 20s and 50s loops are in the N-terminal capping domain. bar signifies the C-terminal $(\beta/\alpha)_7\beta$-barrel domain, and βn indicates the β-strand number. [b]Ratio of mutant to wild-type k_{cat}/K_M. [c]Residues were changed to the amino acid at the corresponding position in AmyOSBS/NSAR: G16P, V17F, V18R, L19T, R20S, D21F, R22G, R23T, and L24Q. [d]The 20s loop was truncated by deleting residues V17, V18, L19, R20, D21, R22, R23, and L24. [e]Not determined because saturation could not be achieved.

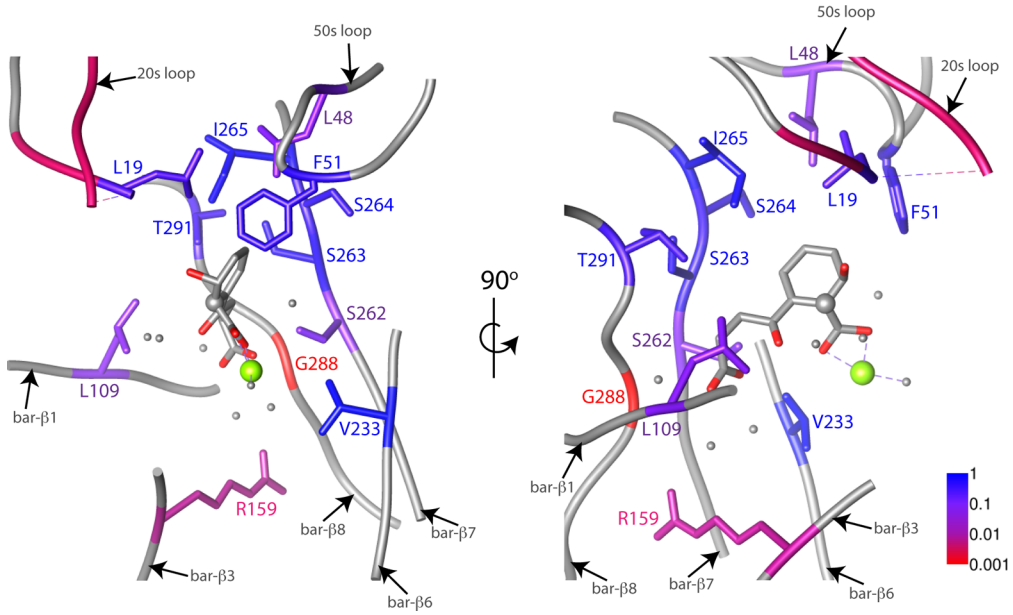


Figure 3. Effect of mutating active site residues of EcOSBS on k_{cat}/K_M (PDB entry 1R6W). The color scale indicates the ratio of each mutant's k_{cat}/K_M to that of the wild type. If more than one mutation was constructed, the color is shown for the more detrimental mutation. Structural elements of the active site (loops and β-strands) are labeled, using the nomenclature described in Table 1. The known catalytic residues on β-strands 2–6 have been removed for the sake of clarity. They would be in the foreground in the first panel and to the right in the second panel. The oxygens on the substrate are colored red, and the α-carbon from which the proton is abstracted is shown as a sphere. Water atoms are shown as small spheres, and magnesium is shown as a lime green sphere. A disordered residue in the 20s loop is represented by the dashed line, as is the coordination of water with the magnesium ion.

Thermosynechococcus elongatus (PDB entry 2OZT), *Desulfotalea psychrophila* (PDB entry 2PGE), *Thermobifida fusca* (PDB entry 2QVH), *Staphylococcus aureus* (PDB entry 2OKT), and *Amycolatopsis* sp. T-1-60 (PDB entry 1SJB). Structural alignments and all structural images were produced using the University of California, San Francisco, Chimera Package from the Resource for Biocomputing, Visualization and Informatics at UCSF (supported by National Institutes of Health Grant

dx.doi.org/10.1021/bi300753j | *Biochemistry* 2012, 51, 6171–6181

2P41RR001081).[31] The final data set consisted of 408 sequences.

Phylogeny. The phylogeny of the whole OSBS family was determined for a representative set of 198 proteins in which no two proteins are >70% identical. This set was selected using CD-HIT.[32,33] Trees were constructed using MrBayes 3.1.2 under the WAG substitution matrix and a gamma distribution to approximate rate variation among sites.[34,35] MrBayes was run on CIPRES-Portal 2.0.[36] The results were analyzed using Tracer to evaluate tree convergence and burn-in.[37] Trees were also constructed by maximum likelihood using the RaxML BlackBox web server (http://phylobench.vital-it.ch/raxml-bb/) with a WAG substitution matrix and a gamma distribution to approximate rate variation among sites.[38,39]

RESULTS

Contribution of Active Site Residues to the Activity of EcOSBS. To test our hypothesis that different OSBS subfamilies rely on different sets of amino acids for optimal catalysis, we examined the effect of mutating residues in the active site of EcOSBS. On the basis of the structure of EcOSBS bound to SHCHC or OSB, we identified amino acids with active site accessible side chains that could interact with the substrate.[18,23] Amino acids were changed to alanine or to an amino acid found at the corresponding position in the Firmicutes OSBS/NSAR subfamily. We chose to make conservative mutations because they reflect the acceptable amino acid substitutions in other members of the OSBS family and enolase superfamily. This is a better test of our hypothesis than introducing larger or charged residues, which are expected to be detrimental because of predictable steric and electrostatic conflicts.

Most mutations decreased k_{cat}/K_M <10-fold, reducing it from ~10^6 to ~10^5 M^{-1} s^{-1}, which is similar to the efficiency of wild-type *Bacillus subtilis* OSBS (BsOSBS) and AmyOSBS/NSAR (Table 1 and Figure 3).[15] Notably, mutations in the 20s loop, which helps determine specificity in other enolase superfamily proteins, did not have a large effect. Changing the only residue in the 20s loop that is within 5 Å of the substrate (L19) decreased k_{cat}/K_M only 7-fold, entirely by affecting k_{cat}. The magnitude of this effect is comparable to that observed in mutations of 20s loop residues in mandelate racemase, another member of the enolase superfamily.[40] Even deleting the 20s loop, changing its sequence to a run of eight alanines, or swapping its sequence for that of AmyOSBS/NSAR did not abolish activity (Table 1 and data not shown).

The only individual mutations that decreased k_{cat}/K_M >100-fold were G288A and R159M. G288A introduces a steric clash with the substrate, reducing k_{cat}/K_M >500-fold (Figure 4A). R159 is 6 Å from the succinyl carboxylate of SHCHC, but they are linked by hydrogen bonds to an intervening water molecule (Figure 4B).[23] K131, which is on β-strand 2, is much closer to this carboxylate, and it is oriented so that it could interact with both carboxylates and the carbonyl of SHCHC or with two acidic residues that bind the metal ion. Perhaps interactions between the succinyl carboxylate and K131 are not sufficient in the R159M mutant because the charge of K131 is effectively neutralized by its other interactions. Loss of R159 might also change how K131 interacts with the succinyl carboxylate, weakening K131's ability to stabilize the substrate and product complexes.[16,41] This could explain the 25-fold decrease in k_{cat} caused by the R159M mutation.

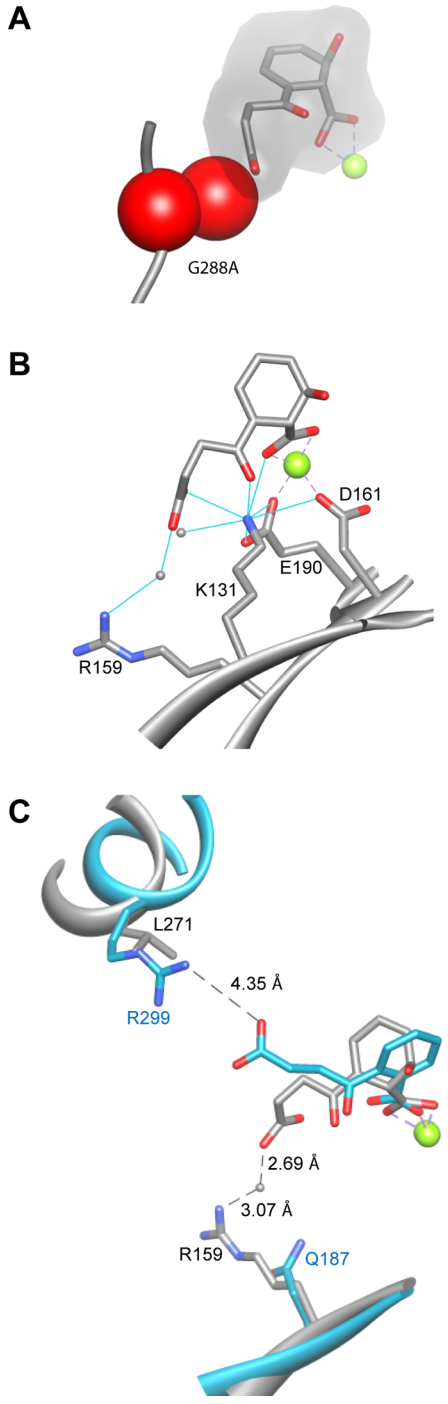


Figure 4. Interactions of residues G288 and R159 with the ligand. (A) Model of the steric clash with SHCHC created by introducing alanine at position 288 of EcOSBS (PDB entry 1R6W).[18] (B) Hydrogen bond network involving SHCHC, R159, and K131 in EcOSBS (PDB entry 1R6W).[18] Predicted hydrogen bonds are colored cyan. (C) Arginines at different positions stabilize ligand conformation in EcOSBS and AmyOSBS/NSAR. EcOSBS bound to SHCHC is colored gray (PDB entry 1R6W),[18] and AmyOSBS/NSAR bound to OSB is colored blue (PDB entry 1SJB).[24] The magnesium ions are shown as lime green spheres.

Several mutations (F51A, V233N, S263G, S262T, and T291S) had an unusual effect: both k_{cat} and K_M increased. For V233N, k_{cat} and K_M increased equally, so that k_{cat}/K_M is approximately the same as that of the wild type. Intriguingly,

dx.doi.org/10.1021/bi300753j | *Biochemistry* 2012, 51, 6171−6181

the k_{cat} and K_M values of these mutants are intermediate between those of wild-type EcOSBS and those of the Firmicutes OSBS/NSAR subfamily members, AmyOSBS/NSAR and BsOSBS (BsOSBS does not have NSAR activity, although sequence similarity places it in the Firmicutes OSBS/NSAR subfamily). Their k_{cat} values are 120 and 150 s^{-1} and their K_M values 480 and 200 μM, respectively.[15] V233, S263, and T291 were mutated to residues found in the Firmicutes OSBS/NSAR subfamily. This increase in k_{cat} suggests that the rate-limiting step changed. The kinetic isotope effect (KIE = 6.1) suggests that proton abstraction is rate-limiting,[21] but it is possible that product release is partially rate-limiting in the wild-type enzyme. The increase in K_M observed for these mutations could reflect weaker ligand binding, making product release faster. Alternatively, introducing larger amino acids at positions 233 and 262 rotates the substrate toward the catalytic lysine (K133) on the opposite side of the barrel. Repositioning the substrate this way would make its position in the active site more similar to the orientation of the product in AmyOSBS/NSAR (Figure 2). We also speculate that the differences in kinetic constants between γ-Proteobacteria subfamily 1 and the Firmicutes OSBS/NSAR subfamily are related to the evolution of NSAR activity.

Defining Subfamilies in the OSBS Family. To determine whether the active site residues examined above are conserved within the γ-Proteobacteria or other OSBS subfamilies, we examined phylogenetic relationships and sequence variation in the OSBS family. Public sequence databases have grown substantially since our initial analysis of the OSBS family.[12] Addition of new OSBS sequences from the Structure-Function Linkage Database increased the size of our data set to 408 sequences, more than twice as many as in our original data set.[20]

We next constructed a phylogenetic tree of the OSBS family to determine if adding new sequences altered the previously defined subfamilies, in which sequences belonging to the same bacterial phylum were grouped together. Because of the large number of sequences, we filtered the OSBS family to a nonredundant set of 198 proteins in which no two proteins are >70% identical. The previously defined subfamilies are still well-supported (Figure 5). The smallest subfamily from our previous analysis, the Bacteroidetes subfamily, grew from 4 to 36 sequences, and expansion of the Chlorobi subfamily from 1 to 11 sequences defined another major subfamily.

However, adding more sequences made deep divisions in the previously defined subfamilies very obvious. There are two deeply branching Cyanobacteria groups, and the γ-Proteobacteria subfamily has several small groups that branch deeply. The original Firmicutes OSBS/NSAR subfamily (including the red wedge and other red branches) is also more diverse than most other subfamilies, but divergence of the basal branches does not correlate with NSAR activity. Thus, we redefined subfamilies by restricting membership to sequences that have sequences >40% identical with that of at least one other subfamily member. Phylogenetic support for the redefined Firmicutes OSBS/NSAR subfamily is weak using this cutoff. However, the sequence diversity is more uniform between the redefined families, so differences in sequence between subfamilies are less likely to be due to differences in evolutionary rate or divergence time.

Correlation between Sequence Conservation and Activity of Mutants. To determine if the active site residues examined above are conserved in all OSBS subfamilies, we

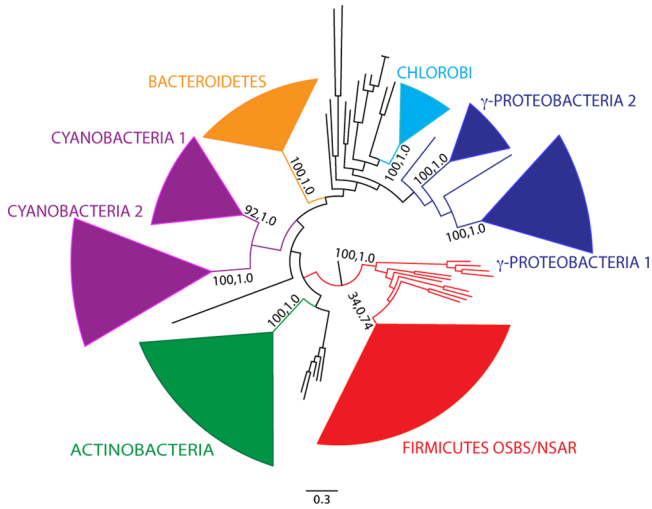


Figure 5. Division of the OSBS family into eight subfamilies. The width of the wedges is proportional to the number of sequences, and the wedge radius corresponds to the longest branch length. Proteins represented by individual branches share too little similarity to be included in the major subfamilies and are therefore left unassigned. Maximum likelihood bootstrap values and Bayesian posterior probabilities are shown for each designated subfamily. The tree is rooted on the basis of the phylogeny of the MLE subgroup.[12] A full tree is shown in Figure 1 of the Supporting Information.

constructed Sequence Logos, in which the height of the letter is proportional to the probability that a sequence in the group will have that amino acid (Figure 6).[42] Residue probabilities were compared among all members of the MLE subgroup, the whole OSBS family, and individual OSBS subfamilies. In addition, we compared the OSBS family to two other protein sets in the enolase superfamily: the *syn*-MLE family and the Firmicutes dipeptide epimerase (DE) subfamily. Like the OSBS subfamilies, these sequence sets were assembled by grouping all sequences that are >40% identical with that of at least one other subfamily member. For the MLE family, this set includes all characterized *syn*-MLE enzymes that utilize muconate or chloromuconate. It excludes the *anti*-MLE family, whose mechanism differs in stereochemistry.[43] The dipeptide epimerases were restricted to the Firmicutes DE subfamily, which includes *B. subtilis* L-Ala-D/L-Glu epimerase (BsDE). Although most enzymes in the Firmicutes DE subfamily prefer Ala-Glu dipeptides, some enzymes in this subfamily are specific for Val-Tyr or other dipeptides.[17] We did not attempt to restrict either the *syn*-MLE family or the Firmicutes DE subfamily to proteins of a single substrate specificity for three reasons. First, specificity has not been experimentally verified for all sequences. Second, the residues that determine specificity in the experimentally characterized MLE and DE enzymes do not correspond to the residues mutated in this study. Finally, the purpose of the comparison was to determine if the active site residues of EcOSBS are subfamily specific, not to identify residues that determine specificity between families.

Strikingly, the 20s loop, which is important for determining specificity in some members of the enolase superfamily, is poorly conserved in the OSBS family (Figure 6, 20s loop). Leucine is conserved at position 19 in γ-Proteobacteria, but it is usually a threonine in other OSBS subfamilies. Leucine and threonine are also the predominant residues at this position in the MLE family and Firmicutes DE subfamily. In most structures from these families, this residue sits on top of the

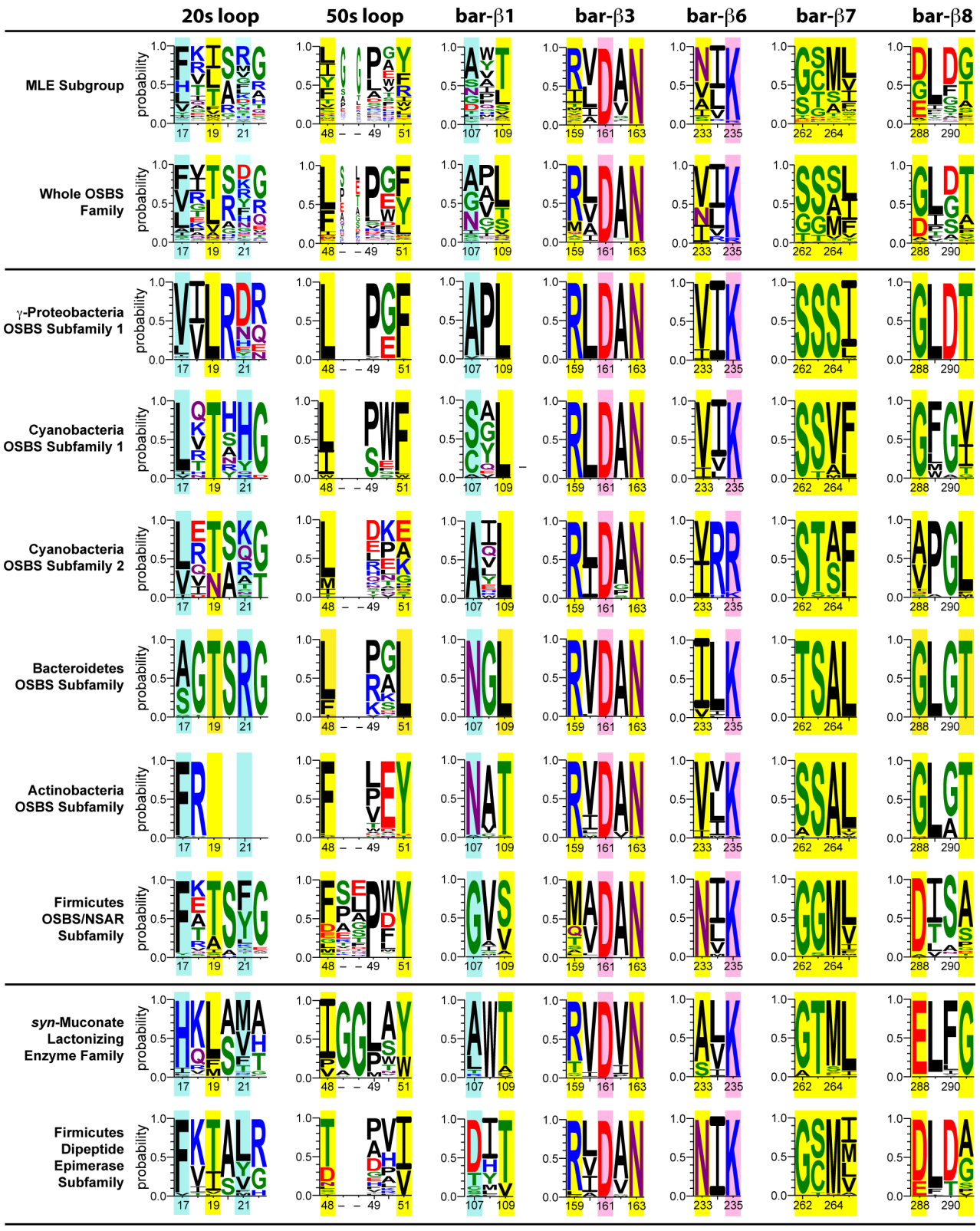


Figure 6. Sequence Logos of active site residues with unknown roles in catalysis.[42] Columns list the logos for each structural position in the active site, using the nomenclature described in Table 1. Each row shows logos that were constructed using the designated protein sequence alignment. The height of each letter represents the probability that a sequence in the indicated sequence alignment has that amino acid. The side chains of residues highlighted in yellow point into the active site of EcOSBS and most other members of the enolase superfamily. Side chains of residues highlighted in blue contact the ligand in some members of the enolase superfamily but are too distant in EcOSBS. The residues highlighted in pink are conserved catalytic residues. The sequence numbering is that of EcOSBS.

dx.doi.org/10.1021/bi300753j | *Biochemistry* 2012, 51, 6171−6181

lysine that abstracts the proton, positioning it for catalysis. In EcOSBS, the capping and barrel domains are twisted, preventing the 20s loop from clamping down on the catalytic lysine (K133). Although the 20s loop in EcOSBS does not contact K133 directly, it forms a ledge that might restrict K133's conformation (Figure 2). In AmyOSBS/NSAR, side chains at positions corresponding to V17 and D21 in EcOSBS also contact the ligand. These residues are too far from the substrate in EcOSBS, and D21 is a surface residue that points out of the active site. In the *syn*-MLE from *Pseudomonas fluorescens*, H24, which corresponds to V17 in EcOSBS, forms a hydrogen bond to muconolactone.[43] In the Firmicutes DE subfamily, the position corresponding to R22 in EcOSBS determines specificity, with arginine in BsDE interacting with the γ-carboxylate of glutamate, and a glycine in *Enterococcus faecalis* DE correlating with its preference for tyrosine instead of glutamate. Thus, the 20s loop determines specificity in some members of the enolase superfamily, but the positions of the specificity determinants are not conserved.

Residues in the 50s loop whose side chains are in the active site are usually large hydrophobic residues in the OSBS family (Figure 6, 50s loop). Mutating these residues (L48 and F51) reduced k_{cat}/K_M 4−16-fold, primarily by increasing K_M. Replacing these amino acids with other large hydrophobic amino acids from corresponding positions in AmyOSBS/NSAR (L48M and F51Y) had a weaker effect than changing them to alanine, but the efficiency was still lower than that of the wild type. This could be because their conformation is not optimal without the two-amino acid insertion found in the 50s loop of AmyOSBS/NSAR. This could not be tested because swapping the sequence of the whole 50s loop for the sequence of AmyOSBS/NSAR's 50s loop yielded an insoluble protein. One exception to the hydrophobic nature of the 50s loop is that Cyanobacteria subfamily 2 has a charged or polar residue at position 51. No structure is available for this subfamily, but we speculate that this substitution is tolerated because this residue is at the surface of the protein, near the edge of the active site. In the *syn*-MLE family and Firmicutes DE subfamily, residues at these positions are also large hydrophobic amino acids. Thus, the hydrophobic pocket formed by the 50s loop is not unique to the OSBS family, although mutations indicate that it does play a moderate role in maintaining protein activity.

The side chains of two residues on the first β-strand of the barrel also point into the active site. In EcOSBS, only one of these (L109) contacts the substrate, sandwiching it against S262. The L109A mutation had a stronger effect than many mutations, decreasing k_{cat}/K_M 24-fold, primarily by affecting k_{cat}. By enlarging the active site, this mutation might affect k_{cat} by altering the orientation of the substrate. Alternatively, L109 is adjacent to the catalytic base K133, so changing it to alanine might also affect the position of the base. Leucine is found at this position in several other OSBS subfamilies (Figure 6, bar-β1). In the crystal structure of *Thermobifida fusca* OSBS, which is a member of the Actinobacteria subfamily, both the asparagine and threonine at the positions corresponding to A107 and L109 in EcOSBS, respectively, form hydrogen bonds to the carboxylate of the succinyl tail. In contrast, a serine at the position corresponding to L109 in AmyOSBS/NSAR forms a hydrogen bond to the carbonyl of the succinyl tail instead of the carboxylate. Threonine is also found at this position in the MLE family and Firmicutes DE subfamily, and it forms a hydrogen bond with the bound ligand. Thus, kinetic and structural data demonstrate that the first β-strand is important for substrate recognition in the OSBS and related families. However, the nature of this interaction varies in the OSBS family, and some OSBS subfamilies have the same residues as MLE or DE enzymes. Because neither residue identities nor the modes of ligand interaction at these positions are conserved, they are not considered to be specificity determinants for the whole OSBS family.

The catalytic motifs on β-strands 2−6 are highly conserved, and we did not expect to discover differences between subfamilies in these regions. However, we made two mutations that had interesting effects. First, replacing R159 of EcOSBS with methionine, the residue found at that position in many members of the Firmicutes OSBS/NSAR subfamily, reduced efficiency ~200-fold. This effect is much larger than those caused most other mutations we made. All OSBS subfamilies have arginine at this position, except the Firmicutes OSBS/NSAR subfamily, in which it is polar or hydrophobic (Figure 6, bar-β3). This correlates with the difference in substrate orientation between EcOSBS and AmyOSBS/NSAR (Figure 4C). In AmyOSBS/NSAR, an arginine enters the active site from a different location, which corresponds to a buried leucine in EcOSBS. Thus, R159 helps confer specificity in OSBS enzymes that bind the substrate in the "bent" conformation. However, arginine is also found at this position in the MLE family and the Firmicutes DE subfamily. In the MLE family, it is >10 Å from the ligand, and its primary role might be to plug the bottom of the barrel. In the Firmicutes DE subfamily, this arginine has no active site accessibility, but it forms a hydrogen bond to an aspartate at the position corresponding to A107, which also contacts the amino terminus of the dipeptide substrate. Because this arginine is conserved in proteins that are not in the OSBS family, R159 would not be detected as a specificity determinant by bioinformatic sequence comparison methods.

The second mutation in a highly conserved region, V233N, had the surprising effect of making k_{cat} and K_M similar to those of characterized members of the Firmicutes OSBS/NSAR subfamily, without changing the efficiency of the enzyme. It is difficult to rationalize this result on the basis of the crystal structures of the proteins, because modeling the mutation does not create any obvious contacts that would adjust the position of the substrate or catalytic residues. Both the Firmicutes OSBS/NSAR subfamily and nearly all dipeptide epimerases have a conserved asparagine at this position (Figure 6, bar-β6). This asparagine is two residues N-terminal of the catalytic lysine on β-strand 6 (K235), which is an acid/base catalyst in the racemization reaction. Future experiments will determine if the asparagine is required for racemization by the Firmicutes OSBS/NSAR subfamily.

The last two β-strands do not have any recognized catalytic residues, but they make some of the closest interactions with the ligand. Thus, we were surprised that mutating residues on β-strand 7 did not have a large effect. However, we introduced conservative mutations to glycine, alanine, and threonine to determine if the amino acids found at those positions in other superfamily members would affect activity. In all members of the MLE subgroup, positions 262 and 263 are small and usually hydrophilic, while position 265 is hydrophobic. Position 264 is a large hydrophobic residue when position 262 is a glycine (Figure 6, bar-β7). The back wall of the active site is formed by both the side chains and backbone atoms of these amino acids. Perhaps backbone accessibility in the active site limits the types and conformations of side chains that are allowed. Bulkier side

chains could distort the shape of the active site, make the active site cleft too shallow to allow the 20s loop to close over the substrate, or prevent the substrate from properly orienting with the catalytic residues. These effects would universally affect members of the MLE subgroup. As a result, the seventh β-strand appears to primarily provide shape complementarity rather than other interactions that determine specificity.

In contrast to β-strand 7, β-strand 8 is very important for determining specificity in EcOSBS. Mutating G288 to alanine reduced activity >500-fold because of a steric clash with the succinyl group of the substrate (Figure 4A). Glycine is conserved at this position in other OSBS subfamilies, except Cyanobacteria subfamily 2 and the Firmicutes OSBS/NSAR subfamily (Figure 6, bar-β8). In the Firmicutes OSBS/NSAR subfamily, the extended conformation of the succinyl tail prevents steric conflict with the aspartate that corresponds to G288. The fact that members of Cyanobacteria OSBS subfamily 2 also lack glycine at this position suggests that the conformation of the succinyl group cannot be bent in this subfamily either. Members of the *syn*-MLE family and Firmicutes DE subfamily have aspartate or glutamate at this position. In the Firmicutes and other DE subfamilies, the conserved DxD motif on β-strand 8 forms a hydrogen bond to the α-ammonium of the dipeptide. Mutating the acidic residue at the position corresponding to G288 in MLE II from *Pseudomonas* sp. P51 and *E. coli* L-Ala-D/L-Glu epimerase was sufficient to alter their specificity to catalyze the OSBS reaction.[22] Thus, G288 is a specificity determinant in some but not all OSBS subfamilies.

DISCUSSION

Subfamily Specific Strategies for Substrate Binding and Catalysis. We proposed two hypotheses to explain the extreme sequence divergence of the OSBS family. The first hypothesis is that features shared by the OSBS family and the rest of the MLE subgroup are sufficient to support catalysis, so there is weak selective pressure to maintain additional specificity determinants. Therefore, mutating residues that are poorly conserved in the OSBS family should have little effect on activity. Because no mutations completely abolished the activity of EcOSBS, we cannot completely reject this hypothesis. However, the evolutionary history of OSBS activity suggests that this hypothesis is insufficient. If the shared active site features of enolase superfamily members were sufficient for OSBS activity, we might expect that OSBS activity could evolve independently from different ancestors within the enolase superfamily. However, our previous study determined that all OSBS enzymes share a common ancestor, despite several instances of lateral transfer of OSBS genes to distantly related organisms and the OSBS family's extreme sequence divergence.[12] In contrast, NSAR activity did evolve twice, once in the OSBS family and once in the DE family, albeit with different amino acid specificities.[14,44] There is also evidence that MLE activity evolved twice, with different stereospecificities, although the evidence is not as strong because the proteins separating the two MLEs in the phylogeny are not experimentally characterized.[43] Thus, even though in vitro experiments show that an MLE and a DE enzyme can be engineered to have OSBS activity by making a single mutation,[22] natural evolution does not appear to have exploited this capability.

Our data provide stronger support for the second hypothesis, which is that different sets of residues contribute to OSBS activity in different subfamilies. Under this hypothesis, active site residues make moderate contributions to ligand binding and catalysis, so that mutating them would be mildly deleterious but not catastrophic. Compensatory mutations to restore optimal activity could occur at different positions in different OSBS lineages. This hypothesis predicts that certain mutations will significantly decrease activity, and the amino acids at these positions will be conserved in only a subset of the OSBS family. Nearly all mutations decreased efficiency at least slightly, and two mutations decreased efficiency by >100-fold. These two residues exhibit subfamily specific conservation. R159 is conserved in all OSBS subfamilies except the Firmicutes OSBS/NSAR subfamily, and G288 is conserved in all OSBS subfamilies except the Firmicutes subfamily and Cyanobacteria subfamily 2. Both residues help determine differences in substrate conformation between EcOSBS and AmyOSBS/NSAR.

A similar phenomenon was recently identified in the DE family. Crystal structures of DE enzymes that are specific for Ala-Glu dipeptides from *B. subtilis* and *Bacteroides thetaiotamicron* show that specificity for glutamate is determined by arginines that are in different places in the structures.[17] These examples in the OSBS and DE families are consistent with "treadmill adaptation", as proposed by Dean and Golding.[45] In treadmill adaptation, sequence variation does not arise solely from the accumulation of neutral mutations. Instead, mutations are continually selected, not because they contribute to a new function but because they are needed to maintain function in changing environments. Although Dean and Golding mainly referred to this idea for amino acids far from the active site, our data indicate that residues in the active site can follow this pattern as well. Similarly, the extent of sequence variation can increase because of coevolution of residues that compensate for deleterious mutations.

Mutating three residues (V233, S263, and T291) to the amino acids found in the members of the Firmicutes OSBS/NSAR subfamily changed the kinetic parameters of EcOSBS to resemble those of proteins in the Firmicutes OSBS/NSAR subfamily. At these positions, the sequences of most OSBS subfamilies are more similar to those of the γ-Proteobacteria subfamily than to those of the Firmicutes OSBS/NSAR subfamily. This raises the possibility that these sequence differences contributed to the evolution of NSAR activity. Thus, the experiments in this work will be the foundation for studies on the biochemical basis for NSAR activity.

Role of the 20s Loop. Although certain residues in the 20s loop appear to determine substrate specificity in the *syn*-MLE family and some dipeptide epimerases, our data indicate that this is not universal. Consistent with our data, mutagenesis of several residues in the 20s loop of mandelate racemase, a member of the enolase superfamily that is not in the MLE subgroup, reduced efficiency by only 6–40-fold.[40] The authors of this study proposed that mutations that change amino acid size but not polarity are tolerated because the 20s loop is a "flexible flap" that closes after substrate binding. Because the 20s loop is solvent-exposed, bulkier residues can distort the flap without severely disrupting function. Such flexible flaps can have a variety of roles in addition to substrate recognition, including exclusion of water and preventing release of reactive intermediates.[46] Because the 20s loop is disordered in most apo structures of the enolase superfamily, its role as a flexible flap appears to be more universal than its potential roles in determining specificity.[16,23,41,47−50] In enolase superfamily

members that use the lysine on the second β-strand for proton abstraction, we further suggest that the main role of the 20s loop is to orient the lysine for catalysis. Indeed, T24 and A25 in mandelate racemase (corresponding to L19 and R20 in EcOSBS, respectively) are the amino acids closest to the catalytic lysine and were particularly sensitive to mutations.

Implications for Predicting Protein Functions. Our previous study found that OSBS enzymes were frequently misannotated because there is as much sequence diversity within the OSBS family as within the whole MLE subgroup.[12] On top of that, divergence of biological function in the Firmicutes OSBS/NSAR subfamily makes accurate annotation impossible without combining sequence similarity, operon context, and possibly other information. By comparing the functional contribution of amino acids to their conservation in OSBS subfamilies, this study takes a step toward determining the relevant sequence differences among OSBSs, NSARs, and other members of the enolase superfamily.

Given the extreme sequence divergence in the OSBS family, we expected to discover major differences among all OSBS lineages. However, the active sites of most OSBS proteins are fairly similar, except those of the Firmicutes OSBS/NSAR subfamily. This is very reassuring, because operon context and experimental data suggest that the Firmicutes OSBS/NSAR subfamily is the only OSBS subfamily that includes enzymes with different biological functions. Further experiments to determine which residues are necessary for NSAR function will determine if sequence information alone can subdivide the Firmicutes OSBS/NSAR subfamily according to the biological functions of its members. If so, this family will be an excellent model system for developing automated functional annotation tools based on identification of specificity determinants.

■ ASSOCIATED CONTENT

Supporting Information

Sequences of primers, PCR conditions for site-directed mutagenesis, and the full phylogenetic tree of the OSBS family. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Department of Biochemistry and Biophysics, Texas A&M University, 2128 TAMU, College Station, TX 77843-2128. E-mail: margy.glasner@tamu.edu. Phone: (979) 458-0123.

Funding

This work was supported by a grant from the Robert A. Welch Foundation (A-1758).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Prof. D. Hilvert and Prof. J. A. Gerlt for sharing reagents.

■ ABBREVIATIONS

OSBS, *o*-succinylbenzoate synthase; NSAR, *N*-succinylamino acid racemase; SHCHC, 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate; MLE, muconate lactonizing enzyme; DE, dipeptide epimerase; EcOSBS, *E. coli* OSBS; AmyOSBS/NSAR, *Amycolatopsis* sp. T-1-60 OSBS/NSAR; BsOSBS, *B. subtilis* OSBS; BsDE, *B. subtilis* L-Ala-D/L-Glu epimerase; PDB, Protein Data Bank.

■ REFERENCES

(1) Gerlt, J. A., and Babbitt, P. C. (2001) Divergent evolution of enzymatic function: Mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu. Rev. Biochem. 70*, 209−246.

(2) 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.

(3) Rakus, J. F., Kalyanaraman, C., Fedorov, A. A., Fedorov, E. V., Mills-Groninger, F. P., Toro, R., Bonanno, J., Bain, K., Sauder, J. M., Burley, S. K., Almo, S. C., Jacobson, M. P., and Gerlt, J. A. (2009) Computation-facilitated assignment of the function in the enolase superfamily: A regiochemically distinct galactarate dehydratase from *Oceanobacillus iheyensis*. *Biochemistry 48*, 11546−11558.

(4) Glasner, M. E., Gerlt, J. A., and Babbitt, P. C. (2006) Evolution of enzyme superfamilies. *Curr. Opin. Chem. Biol. 10*, 492−497.

(5) Mildvan, A. S., Xia, Z., Azurmendi, H. F., Saraswat, V., Legler, P. M., Massiah, M. A., Gabelli, S. B., Bianchet, M. A., Kang, L. W., and Amzel, L. M. (2005) Structures and mechanisms of Nudix hydrolases. *Arch. Biochem. Biophys. 433*, 129−143.

(6) Seibert, C. M., and Raushel, F. M. (2005) Structural and catalytic diversity within the amidohydrolase superfamily. *Biochemistry 44*, 6383−6391.

(7) Allen, K. N., and Dunaway-Mariano, D. (2004) Phosphoryl group transfer: Evolution of a catalytic scaffold. *Trends Biochem. Sci. 29*, 495−503.

(8) Brenner, S. E. (1999) Errors in genome annotation. *Trends Genet. 15*, 132−133.

(9) Devos, D., and Valencia, A. (2001) Intrinsic errors in genome annotation. *Trends Genet. 17*, 429−431.

(10) Jones, C. E., Brown, A. L., and Baumann, U. (2007) Estimating the annotation error rate of curated GO database sequence annotations. *BMC Bioinf. 8*, 170.

(11) Schnoes, A. M., Brown, S. D., Dodevski, I., and Babbitt, P. C. (2009) Annotation error in public databases: Misannotation of molecular function in enzyme superfamilies. *PLoS Comput. Biol. 5*, e1000605.

(12) 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.

(13) Meganathan, R. (2001) Biosynthesis of menaquinone (vitamin K_2) and ubiquinone (coenzyme Q): A perspective on enzymatic mechanisms. *Vitam. Horm. 61*, 173−218.

(14) 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.

(15) 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.

(16) Klenchin, V. A., Schmidt, D. M., Gerlt, J. A., and Rayment, I. (2004) Evolution of enzymatic activities in the enolase superfamily: Structure of a substrate-liganded complex of the L-Ala-D/L-Glu epimerase from *Bacillus subtilis*. *Biochemistry 43*, 10370−10378.

(17) 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.

(18) Klenchin, V. A., Taylor Ringia, E. A., Gerlt, J. A., and Rayment, I. (2003) Evolution of enzymatic activity in the enolase superfamily:

Structural and mutagenic studies of the mechanism of the reaction catalyzed by *o*-succinylbenzoate synthase from *Escherichia coli*. *Biochemistry 42*, 14427−14433.
(19) Nagatani, R. A., Gonzalez, A., Shoichet, B. K., Brinen, L. S., and Babbitt, P. C. (2007) Stability for function trade-offs in the enolase superfamily "catalytic module". *Biochemistry 46*, 6688−6695.
(20) Pegg, S. C., Brown, S. D., Ojha, S., Seffernick, J., Meng, E. C., Morris, J. H., Chang, P. J., Huang, C. C., Ferrin, T. E., and Babbitt, P. C. (2006) Leveraging enzyme structure-function relationships for functional inference and experimental design: The Structure-Function Linkage Database. *Biochemistry 45*, 2545−2555.
(21) 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.
(22) Schmidt, D. M. Z., 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.
(23) Thompson, T. B., Garrett, J. B., Taylor, E. A., Meganathan, R., Gerlt, J. A., and Rayment, I. (2000) Evolution of enzymatic activity in the enolase superfamily: Structure of *o*-succinylbenzoate synthase from *Escherichia coli* in complex with Mg^{2+} and *o*-succinylbenzoate. *Biochemistry 39*, 10662−10676.
(24) Thoden, J. B., Taylor Ringia, E. A., Garrett, J. B., Gerlt, J. A., Holden, H. M., and 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.
(25) Wang, W., and Malcolm, B. A. (1999) Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange site-directed mutagenesis. *BioTechniques 26*, 680−682.
(26) Grisostomi, C., Kast, P., Pulido, R., Huynh, J., and Hilvert, D. (1997) Efficient *in vivo* synthesis and rapid purification of chorismic acid using an engineered *Escherichia coli* strain. *Bioorg. Chem. 25*, 297−305.
(27) Rieger, C. E., and Turnbull, J. L. (1996) Small scale biosynthesis and purification of gram quantities of chorismic acid. *Prep. Biochem. Biotechnol. 26*, 67−76.
(28) Jiang, M., Chen, X., Guo, Z. F., Cao, Y., Chen, M., and Guo, Z. (2008) Identification and characterization of (1*R*,6*R*)-2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase in the menaquinone biosynthesis of *Escherichia coli*. *Biochemistry 47*, 3426−3434.
(29) 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.
(30) Edgar, R. C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res. 32*, 1792−1797.
(31) Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera: A visualization system for exploratory research and analysis. *J. Comput. Chem. 25*, 1605−1612.
(32) Li, W., and Godzik, A. (2006) CD-HIT: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics 22*, 1658−1659.
(33) Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010) CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics 26*, 680−682.
(34) Ronquist, F., and Huelsenbeck, J. P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics 19*, 1572−1574.
(35) Whelan, S., and Goldman, N. (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol. 18*, 691−699.
(36) Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *Proceedings of the Gateway Computing Environments Workshop (GCE)*, pp 1−8, IEEE, Piscataway, NJ.
(37) Rambaut, A., and Drummond, A. J. (2007) *Tracer*, version 1.4 (http://beast.bio.ed.ac.uk/Tracer).
(38) Stamatakis, A. (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics 22*, 2688−2690.
(39) Stamatakis, A., Hoover, P., and Rougemont, J. (2008) A rapid bootstrap algorithm for the RAxML Web servers. *Syst. Biol. 57*, 758−771.
(40) Bourque, J. R., and Bearne, S. L. (2008) Mutational analysis of the active site flap (20s loop) of mandelate racemase. *Biochemistry 47*, 566−578.
(41) Landro, J. A., Gerlt, J. A., Kozarich, J. W., Koo, C. W., Shah, V. J., Kenyon, G. L., Neidhart, D. J., Fujita, S., and Petsko, G. A. (1994) The role of lysine 166 in the mechanism of mandelate racemase from *Pseudomonas putida*: Mechanistic and crystallographic evidence for stereospecific alkylation by (*R*)-α-phenylglycidate. *Biochemistry 33*, 635−643.
(42) Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: A sequence logo generator. *Genome Res. 14*, 1188−1190.
(43) Sakai, A., Fedorov, A. A., Fedorov, E. V., Schnoes, A. M., Glasner, M. E., Brown, S., Rutter, M. E., Bain, K., Chang, S., Gheyi, T., Sauder, J. M., Burley, S. K., Babbitt, P. C., Almo, S. C., and Gerlt, J. A. (2009) Evolution of enzymatic activities in the enolase superfamily: Stereochemically distinct mechanisms in two families of *cis,cis*-muconate lactonizing enzymes. *Biochemistry 48*, 1445−1453.
(44) Song, L., Kalyanaraman, C., Fedorov, A. A., Fedorov, E. V., Glasner, M. E., Brown, S., Imker, H. J., Babbitt, P. C., Almo, S. C., Jacobson, M. P., and Gerlt, J. A. (2007) Prediction and assignment of function for a divergent *N*-succinyl amino acid racemase. *Nat. Chem. Biol. 3*, 486−491.
(45) Dean, A. M., and Golding, G. B. (1997) Protein engineering reveals ancient adaptive replacements in isocitrate dehydrogenase. *Proc. Natl. Acad. Sci. U.S.A. 94*, 3104−3109.
(46) Knowles, J. R. (1991) Enzyme catalysis: Not different, just better. *Nature 350*, 121−124.
(47) Lebioda, L., and Stec, B. (1988) Crystal structure of enolase indicates that enolase and pyruvate kinase evolved from a common ancestor. *Nature 333*, 683−686.
(48) Neidhart, D. J., Howell, P. L., Petsko, G. A., Powers, V. M., Li, R. S., Kenyon, G. L., and Gerlt, J. A. (1991) Mechanism of the reaction catalyzed by mandelate racemase. 2. Crystal structure of mandelate racemase at 2.5-Å resolution: Identification of the active site and possible catalytic residues. *Biochemistry 30*, 9264−9273.
(49) Wedekind, J. E., Poyner, R. R., Reed, G. H., and Rayment, I. (1994) Chelation of serine 39 to Mg^{2+} latches a gate at the active site of enolase: Structure of the bis(Mg^{2+}) complex of yeast enolase and the intermediate analog phosphonoacetohydroxamate at 2.1-Å resolution. *Biochemistry 33*, 9333−9342.
(50) Gulick, A. M., Hubbard, B. K., Gerlt, J. A., and Rayment, I. (2000) Evolution of enzymatic activities in the enolase superfamily: Crystallographic and mutagenesis studies of the reaction catalyzed by D-glucarate dehydratase from *Escherichia coli*. *Biochemistry 39*, 4590−4602.