Evolutionary Potential of $(\beta/\alpha)_8$-Barrels: Stepwise Evolution of a "New" Reaction in the Enolase Superfamily†

Jacob E. Vick and John A. Gerlt*

Departments of Biochemistry and Chemistry, University of Illinois at Urbana-Champaign, 600 S. Mathews Avenue, Urbana, Illinois 61801

Received September 18, 2007

ABSTRACT: The molecular details of the processes involved in divergent evolution of "new" enzymatic functions are ill-defined. Likely starting points are either a progenitor promiscuous for the new reaction or a progenitor capable of catalyzing the new reaction following a single substitution that results from a single base change. However, the molecular (sequence) pathway by which the selective advantage provided by this protein can be improved and ultimately optimized is unclear. In the mechanistically diverse enolase superfamily, we discovered that a monofunctional progenitor could acquire the ability to catalyze a "new" reaction by a single base change: the D297G mutant of the monofunctional L-Ala-D/L-Glu epimerase (AEE) from *Escherichia coli* catalyzed a low level of the *o*-succinylbenzoate synthase (OSBS) reaction as well as a reduced level of the AEE reaction [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) *Biochemistry 42*, 8387−8393]. We then discovered that the selective advantage and OSBS activity of the D297G mutant are both enhanced by the I19F substitution [Vick, J. E., Schmidt, D. M. Z., and Gerlt, J. A. (2005) *Biochemistry 44*, 11722−11729]. Both the D297G and I19F substitutions are positioned to alter the substrate specificity so that the substrate for the OSBS reaction is more productively positioned *vis a vis* the active site catalytic groups. We now report that both the selective advantage and OSBS activity of the D297G/I19F double mutant are enhanced by the R24C (one base change from the wild type Arg codon), R24W (two base changes from the wild type Arg codon and one base change from the R24C codon), and L277W (one base change from the wild type Leu codon) substitutions. The effects of the R24C and L277W mutants are "additive" in the D297G/I19F/R24C/L277W mutant. The greatest selective advantage and OSBS activity are associated with the D297G/I19F/R24W mutant. These "new" substitutions that enhance both the selective advantage and kinetic constants are positioned in the active site where they can alter the specificity, highlighting that the evolution of the "new" OSBS function can be accomplished by changes in substrate specificity.

Homologous enzymes often catalyze different reactions. In many examples, the substrate specificities for the homologues differ but the identity of the reaction is conserved, as in the serine protease family. In contrast, the members of mechanistically diverse superfamilies catalyze different reactions using different substrates that share a conserved partial reaction, as in the enolase, amidohydrolase, enoyl-CoA hydratase, and vicinal oxygen chelate superfamilies (*1*−*3*). The members of functionally distinct suprafamilies catalyze different reactions that share no discernible relationships in either substrate specificity or reaction mechanism (*4*). Because the sequence identities that relate homologues with different functions are usually <40%, comparisons of sequences provide little information about the location and identity of the mutational events that provide both initial and subsequent additional selective advantage in the divergent evolution of new functions. However, such information is important for both understanding Nature's strategies for divergent evolution as well as devising structural solutions for the invention of new catalysts.

We are interested in discovering the sequence and structural bases for the evolution of new functions in the mechanistically diverse enolase superfamily. The members of the superfamily share an N-terminal $\alpha+\beta$ capping domain and a C-terminal $(\beta/\alpha)_7\beta$-barrel (modified TIM-barrel) domain (*5*, *6*). The surfaces of the barrel-domain formed by the loops at the ends of the β-strands and those of two flexible loops in the capping domain form the active site cavity whose structure determines the substrate specificity. As expected for TIM-barrel-containing proteins, the active site functional groups, including ligands for an essential Mg^{2+} ion as well as the acid/base catalysts for generation of an enediolate intermediate and its further processing to a specific product, are located at the C-terminal ends of various β-strands in the barrel domain (Mg^{2+} ligands at the ends of the third, fourth, and fifth β-strands and acid/base catalysts at the ends of the second, third, fifth, sixth, and/or seventh β-strands). The locations of the active site residues in the barrel domain of members of the muconate lactonizing

† This research was supported by Grant GM-52594 (to J.A.G.) from the National Institutes of Health.

* To whom correspondence should be addressed. Phone: (217) 244-7414. Fax: (217) 265-0385. E-mail: j-gerlt@uiuc.edu.

10.1021/bi7019063 CCC: $37.00
Published on Web 11/20/2007

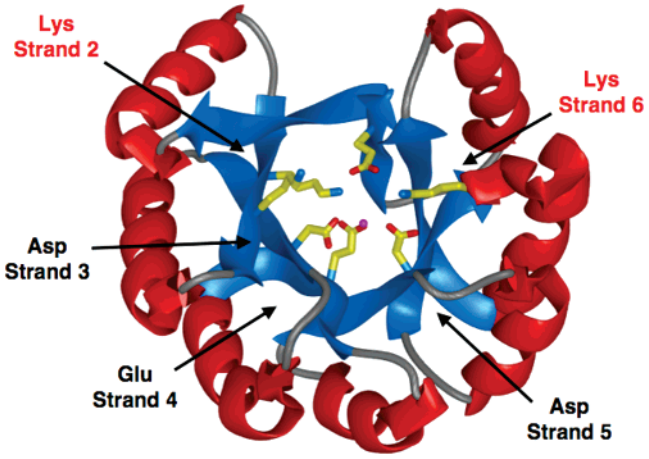

FIGURE 1: Location of the active site residues in the TIM-barrel domain of members of the MLE subgroup of the enolase superfamily.

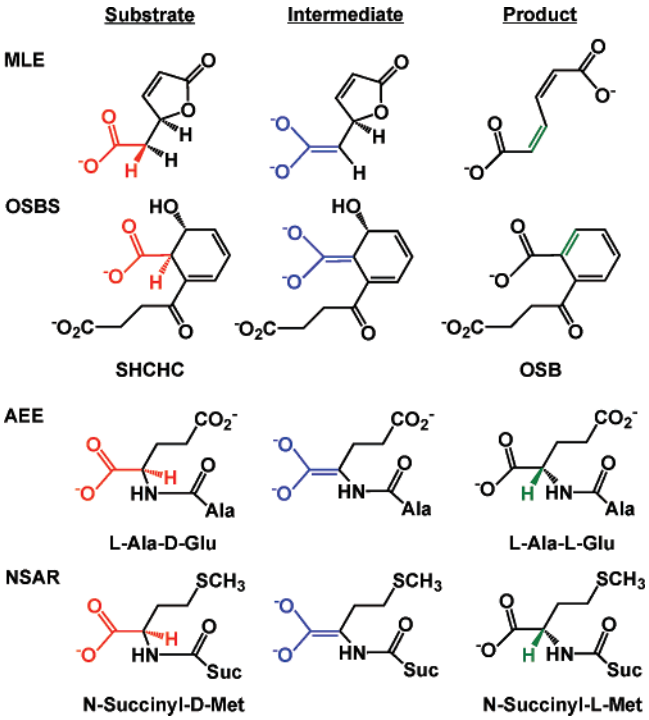

FIGURE 2: Reactions catalyzed by members of the MLE subgroup of the enolase superfamily.

enzyme (MLE[1]) subgroup of the superfamily are shown in Figure 1.

The members of the MLE subgroup of the enolase superfamily catalyze three different chemical transformations in four different physiological reactions (Figure 2): (1) cycloisomerization (intramolecular addition/elimination) in the MLE-catalyzed reaction; (2) β-elimination (dehydration) in the reaction catalyzed by *o*-succinylbenzoate synthase (OSBS); (3) 1,1-proton transfer in the reactions catalyzed by L-Ala-D/L-Glu epimerase (AEE) and *N*-succinylamino acid racemase (NSAR)(*7*).

On the basis of sequence alignments and the available crystal structures, the members of the MLE subgroup share conserved Lys residues at the ends of the second and sixth β-strands (Figure 1). Despite the different reactions, each is initiated by abstraction of the α-proton of the carboxylate anion substrate by the Lys at the end of the second β-strand to generate an enediolate anion stabilized by coordination to the Mg^{2+}. In the MLE-catalyzed reaction, subsequent vinylogous intramolecular elimination of the carboxylate group does not require acid catalysis so the Lys at the end of the sixth β-strand may assist in stabilization of the enediolate intermediate. In the "two-base mechanisms" catalyzed by AEE and NSAR, the Lys residue at the end of the second β-strand is the (*R*)-specific acid/base catalyst and the Lys at the end of the sixth β-strand is the (*S*)-specific acid/base catalyst (*8*, *9*). And, in the OSBS-catalyzed reaction, the Lys at the end of the second β-strand catalyzes both abstraction of the substrate proton and subsequent departure of the hydroxide leaving group from the 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate (SHCHC) substrate in a syn-dehydration reaction in the biosynthesis of menaquinone, an essential cofactor for anaerobic growth by many bacteria; the Lys at the end of the sixth β-strand likely stabilizes the enediolate intermediate (*10*). However, the various enzymes have different residues at the ends of the eighth β-strand in their barrel domains and in the loops of their capping domains; these differences confer distinct substrate specificities and, consequently, the identities of the different reactions.

Divergent evolution of function in the enolase superfamily presumably begins with duplication of the gene encoding a progenitor so that the original function can be retained as the mutations required for a new function accumulate in the copy (*11*−*14*). The progenitor would catalyze a different chemical reaction, although it would employ the conserved catalytic strategy of using the Mg^{2+} to stabilize an enediolate intermediate. Important considerations include (1) whether the progenitor is promiscuous and already catalyzes the new function, (2) the number and location of mutations required for the new function to provide a selective advantage, and (3) the number and location of additional mutations required to optimize catalysis.

To address these questions, we first demonstrated that a single mutation is sufficient to allow two functionally distinct members of the MLE subgroup to catalyze a "new" reaction (*15*). The designed Asp 297 to Gly substitution [D297G, a single-base change (GAC → GGC) and, therefore, naturally accessible] in the AEE from *E. coli* and a selected Glu 323 to Gly substitution [E323G, also a single-base change, (GAG → GGG)] in MLE II from *Pseudomonas sp.* P51 are each sufficient to generate a level of OSBS activity that allows growth of an OSBS-deficient strain of *E. coli* under anaerobic conditions (*15*). Neither progenitor catalyzes detectable levels of the OSBS reaction nor complements the OSBS-deficient strain. For the designed D297G substitution of AEE, the rate acceleration for the OSBS reaction produced by the single substitution (single-base change) is a factor of 10^8; for the E323G substitution of MLE II, the rate acceleration is an even more remarkable 10^{10}. The rate acceleration associated with the reaction catalyzed by the natural OSBS from *E. coli* is 2×10^{11} (*16*).

The D297G and E323G substitutions are structurally homologous, each located at the end of the eighth β-strand of the barrel domain. In the AEE, Asp 297, the second Asp

[1] Abbreviations: AEE, L-Ala-D/L-Glu epimerase; MLE, muconate lactonizing enzyme; NSAR, *N*-succinylamino acid racemase; OSB, *o*-succinylbenzoate; OSBS, *o*-succinylbenzoate synthase; SHCHC, 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate.

in an Asp-x-Asp motif that is conserved in all orthologues, is expected to participate in an electrostatic interaction with the α-ammonium group of the dipeptide substrate based on the structure of the homologous (31% sequence identity) substrate-liganded AEE from *Bacillus subtilis* (*8*). In the MLE II, the role of Glu 323 is unknown because no liganded structure is available for any MLE. But many natural OSBSs have a Gly residue at this position, thereby contributing to an electrostatically neutral pocket in which the succinyl moiety of the SHCHC substrate binds (*17*). In the case of the designed D297G substitution in AEE, the replacement of the carboxymethyl side chain of Asp with a proton in Gly was predicted to eliminate unfavorable electrostatic and steric interactions that would exclude SHCHC from the active site. Presumably, the same reasoning would explain the effect of the selected E323G substitution in MLE II.

Using the gene encoding the D297G mutant of the AEE as the template for error-prone PCR and metabolic selection under anaerobic conditions, we discovered that the I19F substitution, a one-base change (ATT → TTT), enhances the catalytic efficiency of the designed OSBS reaction (*via* a 2.5-fold increase in the value of k_{cat} and a 12-fold increase in the value k_{cat}/K_m) (*18*). Although structures are not available for the wild type AEE in the presence of substrate or for either the D297G mutant or the catalytically enhanced D297G/I19F double mutant either in the absence or presence of substrate, we hypothesized that the substitutions alter the structure of the substrate binding site, thereby allowing introduction (D297G) and then enhancement (I19F) of the OSBS-catalyzed reaction. We concluded that changes in physiological function result from changes in substrate specificity, with the Lys catalysts located at the ends of the second and sixth β-strands structurally invariant, thereby "hardwired" for acid/base chemistry on carboxylate anions that can be positioned productively between the catalysts.

We now have extended these studies and identified substitutions of both Arg 24 and Leu 277 that further enhance the OSBS activity of the D297G/I19F mutant *via* single-base changes. These residues are located in the substrate binding pocket, providing additional support for our earlier conclusion that enhancements in the catalytic efficiency of a "new" reaction result from changes in the residues that form the active site cavity and determine substrate specificity. The selective advantages provided by these substitutions require the presence of the D297G and I19F substitutions. Thus, we have defined an *in vitro* stepwise pathway for evolution of a "new" function that could mimic a natural pathway for divergent evolution of the OSBS function.

MATERIALS AND METHODS

Construction of the Error-Prone Library. The pDMS1a vector containing the gene encoding the D297G/I19F mutant of the AEE from *E. coli* (*18*) was used as the template for the "GeneMorph II" kit (Stratagene, La Jolla, CA) to create an error-prone library following the procedures previously described to achieve one to three mutations (*18*). The gel-purified PCR product was digested with *Bam*HI and *Nde*I and ligated into the pDMS1a vector. Ligation reactions (20 μL) contained 150 ng of the vector, 50 ng of the digested PCR product, 4 μL of 5x ligase buffer, and 2 μL of T4 ligase (Invitrogen, Carlsbad, CA) and were carried out for 11 h at 16 °C. The ligated products were purified by addition of 50 μL of H_2O and 500 μL of 1-butanol, vortexing for 5 min, and centrifugation for 10 min. The supernatant was decanted, and the pellet was resuspended in 10 μL of H_2O. Plasmids from three ligations (in 1 μL aliquots) were transformed *via* electroporation into XL1Blue cells, and the cells were allowed to recover for 30 min in 200 μL of SOC medium. An aliquot (10 μL) from each transformation was plated on an LB-Amp plate to quantitate the number of transformants, and the remainder was plated on an LB-AMP plate. Thirty transformations were required to obtain a library size of 10^6 transformants. The colonies were recovered, and the plasmids were isolated using the Maxi-prep procedure (Qiagen, Valencia, CA). The diversity of the library (expected number of base changes) was estimated using the program PEDEL (*19*) from the error rate assessed by DNA sequencing.

Construction of Site-Specific Random Libraries. Site-specific random libraries at codons 17, 21, 24, 48, 50, and 277 were constructed by the overlap extension method. The primers used for construction of the libraries are detailed in the Supporting Information. The PCR products containing the randomized codons were gel-purified and used as the template for the second PCR reaction which afforded the full length genes containing the desired randomized codon. The PCR products were digested with *Bam*HI and *Nde*I and ligated into the pDMS1a vector. The libraries were obtained as described in the previous section. The randomized codons were verified by DNA sequence analysis. The distribution of mutants in the library was estimated by using the program GLUE (*19*).

Construction of Site-Specific Mutants. Site-specific mutants were constructed using the same overlap extension method as the site-specific randomized libraries. The primers used for construction of the libraries are detailed in the Supporting Information.

Anaerobic Selection for OSBS Activity. Minimal medium for anaeroboic growth was prepared as previously described (*15*, *20*). Libraries were transformed into the *menC*::kan strain of *E. coli* (*15*) and grown aerobically overnight at 37 °C in the minimal medium containing 100 μg/mL ampicillin and 50 μg/mL kanamycin. These cultures were used to inoculate anaerobic growths in completely filled 8.5 mL screw-top vials. The cultures were incubated at 37 °C in an anaerobic chamber; growth was monitored at 600 nm using a Spectronic 20 spectrophotometer. The cultures were allowed to grow to midlog phase (OD at 600 nm = 0.6), and an aliquot was used as an inoculum for the next round of selection. In the case of the site-specific libraries, the convergence of the randomized codon(s) on the selected sequence was monitored by DNA sequence analysis.

Purification of Proteins. All enzymes were expressed in and purified from the menC::kan/ycjG::cam strain of *E. coli* (*18*) to eliminate contaminating OSBS and AEE in enzymatic assays. The purification procedure previously described was used for all variants (*15*).

Enzymatic Activity Assays. OSBS and AEE activities were quantitated as previously described (*18*). The program ENZFIT (Elsevier-Biosoft) was used to determine the values of k_{cat} and K_M. If an enzyme could not be saturated with substrate, the slope of a plot of velocity vs substrate concentration was divided by the concentration of the enzyme to obtain the value of k_{cat}/K_M.

RESULTS AND DISCUSSION

As summarized in the Introduction, we first designed the AEE from *E. coli* to catalyze the OSBS reaction, with the strategy based on the hypothesis that Asp 297 at the end of the eighth β-strand in the barrel domain prevented binding of the SHCHC substrate for the OSBS reaction by both unfavorable steric and electrostatic interactions (*15*). We then subjected the active D297G mutant to random mutagenesis followed by the selection under anaerobic conditions and identified the sequential I19F mutant that enhanced the OSBS activity (*18*). The D297G substitution is located at the "back" of the active site where the succinyl moiety is predicted to bind, and the I19F substitution is located in the 20s loop of the capping domain that forms the "roof" of the active site. In both cases, we proposed that the substitutions altered the geometry of the active site, so that the SHCHC substrate could be located appropriately for the syn-dehydration reaction facilitated by coordination of the carboxylate group to the essential Mg^{2+} and presentation of the α-proton of the substrate to Lys 151 at the end of the second β-strand. Both the D297G and I19F substitutions result from single-base changes, mutations expected to occur naturally, thereby allowing incremental divergent evolution of a "new" function from a homologous, but inactive, progenitor that would mimic natural processes.

Random Mutagenesis of D297G/I19F. We sought to determine whether additional single-base changes could be identified that would allow further enhancement(s) in the ability of the D297G/I19F mutant to complement an OSBS auxotroph as the result of increases in the kinetic constants, thereby further defining additional steps in an *in vitro* pathway for divergent evolution of the OSBS reaction from the AEE reaction. The rate acceleration for the OSBS reaction catalyzed by D297G/I19F is 10^9, remarkable considering that the wild type AEE progenitor had no detectable OSBS activity. The "gold standard" for evolution of the OSBS reaction is the rate acceleration of 2×10^{11} measured for the OSBS from *E. coli*.

An error-prone library was constructed containing from one to three base changes in the gene encoding D297G/I19F (321 codons) and subjected to metabolic selection by complementation of the host in which the gene encoding OSBS, *men*C, had been insertionally disrupted. After several rounds of selection by anaerobic growth in liquid culture (aliquots for the next round taken at midlog phase), colonies were isolated from plates grown anaerobically, and the genes contained in the plasmids were sequenced. Five different mutants were obtained, in which each of the substitutions resulted from a single base change:

(1) D297G/I19F/**R188H**/**L277W** (two substitutions);
(2) D297G/I19F/**F7S**/**K71E**/**L277W** (three substitutions);
(3) D/297G/I19F/**L115I**/**L277W** (two substitutions);
(4) D297G/I19F/**R24C** (one substitution); and
(5) D297G/I19F/**R24C**/**S136G** (two substitutions).

As shown in the growth curves presented in Figure 3, panel A, each of these mutants grew faster than the D297G/I19F progenitor.

Both Arg 24 and Leu 277 are located within the active site (Figure 4; at the C-terminal end of the disordered 20s loop in the capping domain and at the end of the seventh β-strand in the barrel domain, respectively), suggesting that substitutions for these can have a direct effect on enhancement of the OSBS reaction. However, Phe 7, Leu 115, Ser 136, Lys 177, and Arg 188 are located on the surface of the protein and remote from the active site, suggesting that substitutions for these substitutions do not directly influence substrate specificity. The D297G/I19F/**R24C** (without S136G) was isolated from the selection, and the D297G/I19F/**L277W** mutant (without F7S, K71E, L115I, or R188H) was constructed to establish the relationships of the selected phenotypes on the presence of the remote substitutions. As shown in the growth curves presented in Figure 3a, both retain the enhanced growth phenotypes, although the additional mutants (all single base changes) may confer marginally improved growth phenotypes. Our subsequent studies focused on the "simplified" D297G/I19F/**R24C** and D297G/I19F/**L277W** mutants.

We also constructed the I19F/**R24C** and I19F/**L277W** mutants (without D297G), the D297G/**R24C** and D297G/**L277W** mutants (without I19F), and **R24C** and **L277W** (without both D297G and I19F) to determine whether the selective advantage conferred by the R24C and L277W substitutions is dependent on the presence of these previously identified substitutions. None of the mutants lacking the D297G substitution complemented the OSBS auxotroph, establishing that this is substitution is a required first step in the evolution of OSBS activity (data not shown). The I19F substitution also is required for L277W, although D297G/R24C mutant complemented the auxotroph with an efficiency similar to that of D297G alone (data not shown).

We constructed the D297G/I19F/**R24C**/**L277W** mutant; this combination mutant more efficiently complemented the OSBS auxotroph than either the D297G/I19F/**R24C** or D297G/I19F/**L277W** mutants identified in the selections (Figure 3, panel B). This growth phenotype of this mutant also required the presence of both D297G and I19F (data not shown).

Thus, the R24C (CGC → TGC) and L277W (TTG → TGG) mutations represent additional incremental steps along a possible pathway for divergent evolution of the OSBS function from the AEE function, with their presence together further enhancing the growth advantage.

Effects of R24C and L277W on OSBS Activity. The D297G/I19F/**R24C**, D297G/I19F/**L277W**, and D297G/I19F/**R24C**/**L277W** mutants were purified from a strain of *E. coli* in which the chromosomal genes encoding both AEE and OSBS were insertionally disrupted. The proteins were assayed for OSBS activity; the kinetic constants are displayed in Table 1 along with the values for the D297G and D29G/I19F progenitors.

Although values of k_{cat} and k_{cat}/K_m for the OSBS reaction could be obtained for the D297G/I19F/**R24C** mutant, the D297G/I19F/**L277W** mutant could not be saturated with the SHCHC substrate. For the former mutant, the values for both kinetic constants were increased. For the latter mutant, the value of k_{cat}/K_m was similar to that of the D297G/I19F progenitor, although the substitutions provided a growth advantage. The expression levels of the various mutants were similar, based on SDS-PAGE of cell extracts (data not shown), but the enhanced growth rate could reflect a small increase in the level of enzyme.

The value of k_{cat} for the D297G/I19F/**R24C**/**L277W** mutant was compromised 8-fold relative to the D297G/I19F/

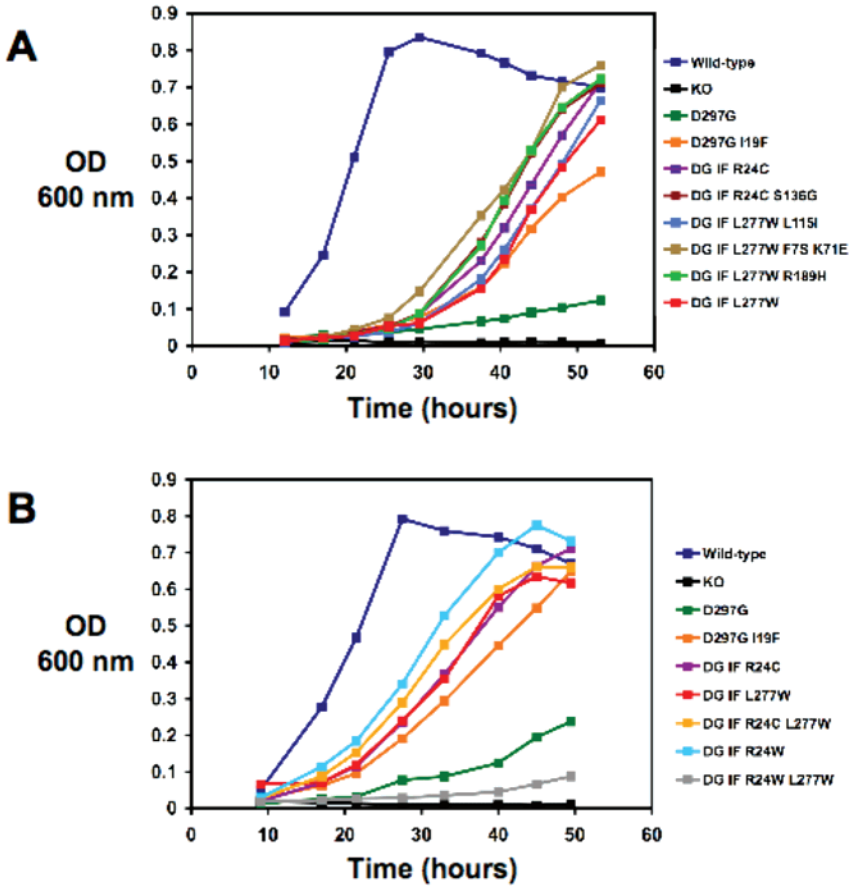

FIGURE 3: Anaerobic growth of various strains of *E. coli*: WT, wild type BW25113; KO, the *menC::kan* mutant of BW25113; D297G, the engineered mutant of AEE (*15*); D297G I19F, the selected double mutant that enhances OSBS activity (*18*); additional mutants identified as described in the text. Panel A, the results of selections following error-prone mutagenesis of the gene encoding the D297G/I19F mutant; panel B, additional mutants constructed as described in the text.

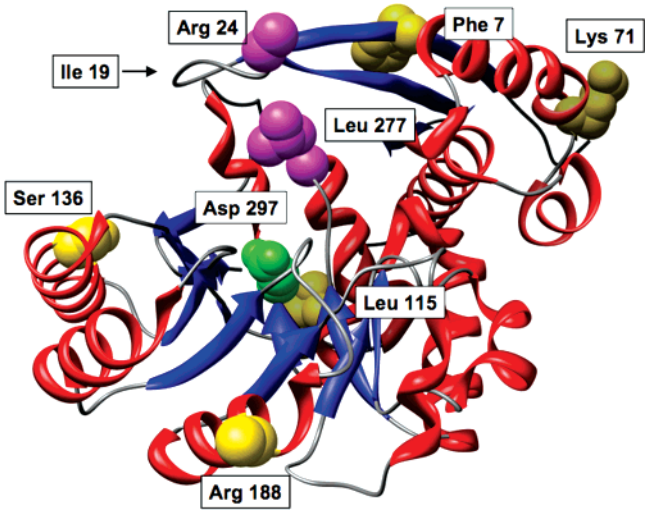

FIGURE 4: Locations of the substitutions described in this article in the three-dimensional structure of the unliganded AEE from *E. coli* (*22*). In this structure (1JPD), the 20s loop is disordered. Ile 19 is presumed to be located at the tip of the loop, pointing into the active site cavity.

R24C, but the value of k_{cat}/K_m was improved 2-fold relative to D297G/I19F/**R24C** and 4-fold relative to D297G/I19F. Thus, selective advantage is correlated with the value of k_{cat}/K_m, as expected if the concentration of the SHCHC substrate for the OSBS-catalyzed reaction is subsaturating *in vivo*.

Site-Specific Random Mutagenesis of Arg 24 and Leu 277. Because we identified both Arg 24 and Leu 277 at potential "hot spots" for improving the efficiency of the OSBS reaction, we generated site-specific random libraries for each residue as well as a library in which both residues were simultaneously randomized. In contrast to the error-prone PCR experiments, saturation mutagenesis coupled with metabolic selection allows identification of substitutions also resulting from two- or three-base pair substitutions in the same codon. The metabolic selections were performed in anaerobic liquid cultures, also with aliquots taken from midlog phase growth to inoculate each successive round of selection. The identities of the selected residues at each position were determined by DNA sequencing of plasmids isolated from the successive inocula.

In the library for Arg 24, we identified a mixture of two substitutions that provided selective advantage: R24C (CGC → TGC) and R24W (CGC → TGG). The former substitution was identified in the error-prone library; the latter substitution was not expected in the error-prone library because the substitution results from a two base pair change in the codon. In the Leu 277 library, we identified only L277W (TTG → TGG), the single base change identified in the error prone-library.

We also prepared a library in which the codons for both Arg 24 and Leu 277 were simultaneously randomized. We again identified the R24W mutant as the fastest growing mutant, with no other combination of substitutions for these residues providing an additional selective advantage relative to that provided by the R24W substitution identified in the

Table 1: Kinetic Parameters for OSBS and AEE Activities

enzyme	k_{cat} (s^{-1})	K_M (M)	k_{cat}/K_M ($M^{-1}s^{-1}$)
	OSBS Activity[a]		
OSBS (*E. coli*)[b]	28	1.6×10^{-5}	1.8×10^{6}
AEE (wild-type)[c]	−	−	$\leq 5.2 \times 10^{-3}$ [b]
single base pair change relative to wild-type			
D297G[e]	$(1.3 \pm 0.6) \times 10^{-2}$	$(1.8 \pm 0.2) \times 10^{-3}$	7.4
two base pair changes			
D297G/I19F[e]	$(3.1 \pm 0.9) \times 10^{-2}$	$(3.4 \pm 0.27) \times 10^{-4}$	90
three base pair changes			
D297G/I19F/**R24C**	$(1.1 \pm 0.1) \times 10^{-1}$	$(5.3 \pm 0.5) \times 10^{-4}$	206
D297G/I19F/**L277W**	−	−	81[d]
four base pair changes			
D297G /I19F/**R24C**/**L277W**	$(1.3 \pm 0.45) \times 10^{-2}$	$(3.6 \pm 0.9) \times 10^{-5}$	365
D297G /I19F/**R24W**/**L277W**	−	−	−
D297G I19F/**R24W**	$(1.6 \pm 0.05) \times 10^{-1}$	$(7.6 \pm 0.6) \times 10^{-5}$	2105
D297G/**R21Y**	$(1.0 \pm 0.3) \times 10^{-2}$	$(4.6 \pm 0.5) \times 10^{-4}$	20
five base pair changes			
D297G/I19F/**R21Y**	$(2.4 \pm 0.1) \times 10^{-1}$	$(8.9 \pm 0.8) \times 10^{-4}$	274
	AEE Activity[a]		
AEE (*E. coli*)[c]	10 ± 0.4	$(1.3 \pm 0.3) \times 10^{-4}$	7.7×10^{4}
D297G[c]	$(4.3 \pm 0.05) \times 10^{-2}$	$(4.4 \pm 0.3) \times 10^{-3}$	9.8
D297G/I19F[e]	−	−	9.7×10^{-2} [d]
D297G/I19F/**R24C**	5.6×10^{-4}		
D297G/I19F/**L277W**	2.6×10^{-3}		
D297G/I19F/**R24C**/**L277W**	4.2×10^{-3}		
D297G/**I19F**/**R24W**	1.8×10^{-3}		

[a] Assay conditions described in Materials and Methods. [b] From ref *16*. [c] From ref *15*. [d] Does not saturate. [e] From (*18*).

site-specific random library for Arg 24. We constructed the D297G/I19F/**R24W**/**L277W** mutant and observed that it displayed weak growth under anaerobic conditions (Figure 3, panel B).

We also constructed the D297G/**R24W**, I19F/**R24W**, and **R24W** mutants as well as the D297G/I19F/**R24W**/**L277W**, D297G/**R24W**/**L277W**, and **R24W**/**L277W** mutants to assess their abilities to complement the auxotroph. Only the D297G/I19F/**R24W**/**L277W** mutant showed detectable growth (slower than D297G alone; Figure 3, panel B); the other mutants did not grow under anaerobic conditions (data not shown).

The D297G/I19F/**R24W** mutant was purified from the host in which the chromosomal genes encoding both AEE and OSBS were insertionally disrupted; the values of the kinetic constants are included in Table 1. The value of k_{cat} for the OSBS reaction was increased 5-fold and the value of k_{cat}/K_m was increased 23-fold relative to the kinetic constants for the D297G/I19F progenitor. The D297G/I19F/**R24W**/**L277W** mutant also was isolated and had no detectable OSBS activity; perhaps the presence of two sterically demanding Trp side chains in the active site prevents effective binding of the SHCHC substrate.

Site-Specific Random Mutagenesis of Other Residues in the 20s and 50s Loops. The R24C and R24W substitutions that confer selective advantage and improved kinetic constants are located at the C-terminal end of the mobile 20s loop that presumably closes to sequester the substrate from the solvent and, also, define the shape and polarity of the "top" of the cavity in which the substrate binds. Based on the structure of the homologous active site of the structurally characterized liganded complex of the AEE from *B. subtilis* with its L-Ala-L-Glu substrate (*8*) as well as the structure of the OSBS from *E. coli* liganded with the OSB product (*21*) (Figure 5), we hypothesized that in addition to Ile 19 and Arg 24, both Phe 17 and Arg 21 in the 20s loop likely participate in forming the active site cavity (Figure 5). In addition, in the structure of the unliganded AEE from *E. coli* both Tyr 48 and Arg 50 in the 50s loop also participate in the active site cavity (*22*). Thus, we used saturation mutagenesis to probe whether substitutions for these residues would enhance both the selective advantage and kinetic constants for the OSBS reaction relative to the D297G/I19F progenitor.

Separate site-specific random libraries were constructed for Phe 17, Tyr 48, and Arg 50. As before, the selections were performed in anaerobic liquid cultures, with aliquots taken from midlog phase growth to innoculate each successive round of selection. The libraries for Phe 17 and Arg 50 identified no substitutions better than the wild type residues (data not shown). The library for Tyr 48 resulted in a mixture of wild type as well Y48F (TAT → TTT), although no selective advantage was associated with the one-base pair substitution.

A randomized library for spatially proximal Ile 19 and Arg 21 residues allowed identification of the R21Y substitution (CGG → TAC, a three-base pair change) that provided selective advantage. This protein (D297G/I19F/**R21Y**) was isolated from the AEE/OSBS deficient strain of *E. coli* and assayed for OSBS activity (Table 1). Despite a slight reduction in the value of k_{cat}, the value of k_{cat}/K_m was improved 3-fold. However, because the improvements in selective advantage and kinetic constants associated with the Arg to Tyr substitution require three base changes, this protein could not occur (at least in a single step) on a natural pathway for divergent evolution.

Effects of R24C, R24W, and L277W on AEE Activity. Our previous studies demonstrated that as the level of the OSBS activity is first introduced by the D297G substitution and then enhanced by the I19F substitution, the level of the

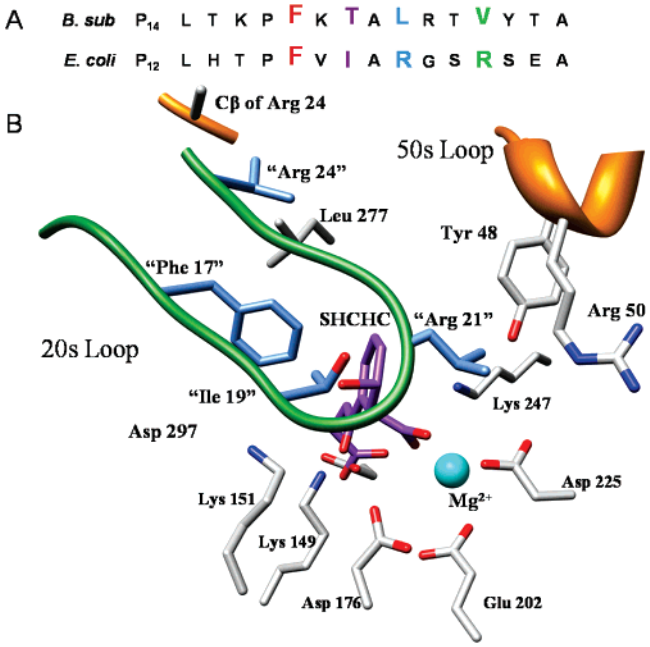

FIGURE 5: Panel A: Alignment of the sequences of the 20s loops in the AEEs from *E. coli* and *B. subtilis*. Panel B: The active site of the AEE from *B. subtilis* highlighting the positions of the residues in the 20s and 50s loop relative to the Ala-Glu substrate (shaded in yellow). The side chains of Phe 19, Thr 21, Ile 23, Arg 24, and Ile 54 (shaded in gray) are pointed into the active site cavity as partial determinants of shape and polarity. The Lys residues at the ends of the second (Lys 162) and sixth (Lys 268) β-strands, the Mg^{2+} ligands at ends of the third (Asp 191), fourth (Glu 219), and fifth (Asp 244) β-strands, and the α-ammonium group binding residues at the end of the eighth β-strand (Asp 321 and Asp 323) of the barrel domain are shaded in cyan.

progenitor's AEE activity is decreased (*18*). Presumably, because the substrates for both the AEE and OSBS reactions are large and fill the volume of the active site cavity, changes that allow the SHCHC substrate for the OSBS reaction to bind also will interfere with binding of the dipeptide substrate for the AEE reaction.

The AEE activities associated with the four "new" mutants that provided a selective advantage relative to the D297G/I19F progenitor are also included in Table 1. These were obtained at 10 mM L-Ala-D-Glu using the coupled enzyme assay. For each mutant, the AEE activity was barely distinguishable from background. These low levels of activity support our earlier conclusion that the progenitor's AEE reaction necessarily diminishes as the evolved OSBS reaction is enhanced by changes in specificity determining residues (*18*, *23*).

Structural Basis for Selective Advantage and Enhanced Kinetic Parameters. Unfortunately, we have been unable to obtain structures for any mutants of the AEE, either previously described or discovered in this study. So, an exact structure-based explanation for the enhanced OSBS activity is not possible. However, the backbone atoms and Cβ of the side chain of Arg 24 and all of Leu 277 are visible in the unliganded structure of the AEE (Figures 4 and 5). The side chain of Arg 24 is expected to be solvent exposed, suggesting that the hydrophobic R24C and R24W substitutions likely alter the conformation of the 20s loop that closes over and sequesters the substrate from the active site rather than directly interacting with the SHCHC substrate and directly changing its position in the active site. The previously identified I19F substitution that is required for the growth and kinetic phenotypes of the R24C and R24W substitutions is located at the end of the 20s loop, with its side chain likely pointing into the active site and, therefore, forming part of the cavity in which the SHCHC substrate for the OSBS reaction necessarily binds. Changes in the conformation of the 20s loop may change the position of the I19F side chain relative to the bound substrate, thereby allowing more efficient catalysis.

In contrast, the side chain of Leu 277, located at the end of the seventh β-strand in the barrel-domain, is both pointed into the active site and, also, located beneath the base of the 20s loop. Thus, the L277W substitution could alter the structure of the cavity for the SHCHC substrate, indirectly by altering the conformation of the 20s loop and/or directly by altering the shape of the surface formed by this and neighboring residues.

Implications for Divergent Evolution. Nature has many potential progenitors and much time for the divergent evolution of "new" enzymatic functions. Even so, in the enolase superfamily as well as many others, families of homologous proteins that catalyze different reactions typically share <40% sequence identity, suggesting a complex interplay of structure, function, and stability as new functions evolve. In the case of the OSBS function, several homologous families have been identified that share <15% sequence identity but catalyze the OSBS reaction with comparable catalytic efficiencies; these families likely arose by independent lineages of divergent evolution, in which different progenitors were selected and subjected to adaptive selection (although the OSBSs and the progenitors apparently have diverged significantly in sequence so that the progenitors cannot be identified on the basis of sequence comparisons). In each family, the catalytic machinery (ligands for the essential Mg^{2+} and two Lys acid/base catalysts; Figure 1) is strictly conserved, so natural evolution of the OSBS function evidently involved changes in the specificity determining residues, as we describe in this manuscript for laboratory evolution.

We do not expect that attempts to generate a "new" function within a laboratory time scale can imitate the entire process used by Nature to evolve and optimize a new function, but our studies illustrate that the acquisition of sequential mutations can produce a continuous pathway for the evolution of a "new" function. Using rounds of either gene-wide or site-specific random mutagenesis coupled with a metabolic selection, we will not identify those substitutions that may be neutral but, with further mutations, provide unique access to structural changes that allow enhancement and eventual optimization of the "new" function. Presumably, such processes occur in Nature and are essential for optimization of the "new" function and explain the large amount of sequence divergence that routinely accompanies the acquisition of the "new" function.

Figure 6 summarizes the enhancements in the OSBS function that we have discovered starting with the AEE from *E. coli*, a progenitor that has no detectable OSBS activity. The designed D297G mutant is a prerequisite to further enhancements in OSBS activity and was responsible for a remarkable 10^8 increase in the rate acceleration (as measured by the value of k_{cat} (*15*)).

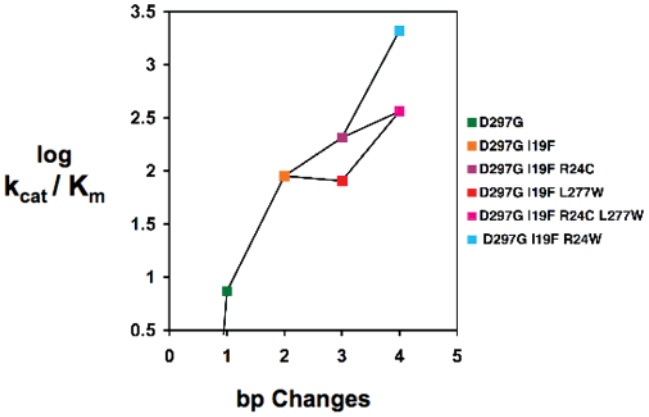

FIGURE 6: Plot of the value of log k_{cat}/K_M for the evolved OSBS reaction as a function of the number of base pair changes relative to the wild type AEE progenitor.

Using D297G as the initial intermediate in the evolution of the OSBS function, we identified Ile 19 as a "hot spot" with several one-base substitutions providing growth advantage (*18*). Of these, the I19F substitution provided the superior selective advantage, and the isolated protein also had the greatest values for both k_{cat} and k_{cat}/K_m. We demonstrated that the growth advantage provided by I19F requires the presence of the D297G substitution.

We then used the D297G/I19F double mutant as the next intermediate in the evolution of OSBS function and, as described in this manuscript, identified two additional codons that, with a single substitution, allow additional selective advantage, Arg 24 and Leu 277. The R24C substitution, a single base change (from CGC to TGC), also enhances the values of k_{cat} and k_{cat}/K_m. The L277W substitution, also a single substitution (from TTG to TGG), does not enhance the value of k_{cat}/K_m (k_{cat} could not be measured). However, when these substitutions are combined, the resulting quadruple mutant provides an additional growth advantage, and the values of the kinetic constants are further enhanced. The selective advantage provided by these substitutions requires the presence of both the D297G and I19F substitutions. Thus, the sequence D297G/I19F → D297G/I19F/**R24C** → D297G/I19F/**R24C**/**L277W** describes an extension of the *in vitro* pathway for divergent evolution of the OSBS function, with each step providing a selective advantage.

Also, when the codon for Arg 24 was randomized, we identified the R24W substitution, a two-base change (from CGC to TGG), that provides the greatest selective advantage yet observed in our studies. Again, this selective advantage requires the presence of both the D297G and I19F substitutions. Because the codon for Trp (TGG) is a one-base change relative to the codon for Cys (TGC), the sequence D297G/I19F → D297G/I19F/**R24C** → D297G/I19F/**R24W** describes a second *in vitro* pathway for divergent evolution of the OSBS function, again with each incremental step providing a selective advantage.

Perhaps, if we had selected another of the several substitutions for Ile 19 that provide selective advantage as the progenitor for our studies, we would have identified other distinct pathways for the evolution of the OSBS function. However, in contrast to natural evolution, laboratory evolution does not have the capabilities to explore all possibilities.

The observations parallel those reported recently for the acquisition of resistance of a β-lactamase to an antibiotic involving five substitutions (*24*). Although 120 possible pathways ("trajectories) for the sequential acquisition of these substitutions are statistically possible, only 10 were found to provide sequential, selectable enhancements in activity. Thus, despite the considerable sequence divergence that relates homologous proteins with different functions, the molecular pathway for divergent evolution of "new" functions is likely well-defined. However, comparisons of the sequences of the homologues are unlikely to allow that pathway to be described.

CONCLUSIONS

The values of k_{cat} and k_{cat}/K_m for D297G/I19F/**R24W** are 0.16 s^{-1} and 2.1 × 10^3 M^{-1} s^{-1}, respectively; the values of k_{cat} and k_{cat}/K_m for the wild type OSBS from *E. coli* are 28 s^{-1} and 1.8 × 10^6 M^{-1} s^{-1}, respectively. With four substitutions, our *in vitro* evolution of the OSBS function produced a rate acceleration of 10^9, 2 orders of magnitude less than that achieved by natural evolution; the value of k_{cat}/K_m we achieved is 3 orders of magnitude less than that achieved by natural evolution. Despite these shortcomings, our discovery of two pathways that involve successive single-base changes that provide continuous improvement in both selective advantage and the values of the kinetic constants (as reflected by k_{cat}/K_m) is noteworthy. Indeed, our experiments illustrate that divergent evolution of a new function can occur by a very limited number of substitutions, with the altered residues positioned to influence catalysis *via* changes in substrate specificity and not changes in the identities and positions of the catalytic residues.

SUPPORTING INFORMATION AVAILABLE

Sequences of the primers used for construction of the random libraries and site-directed substitutions. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Babbitt, P. C., and Gerlt, J. A. (1997) Understanding enzyme superfamilies. Chemistry As the fundamental determinant in the evolution of new catalytic activities, *J. Biol. Chem. 272*, 30591−30594.
2. 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.
3. Gerlt, J. A., and Raushel, F. M. (2003) Evolution of function in $(\beta/\alpha)_8$-barrel enzymes, *Curr. Opin. Chem. Biol. 7*, 252−264.
4. Wise, E., Yew, W. S., Babbitt, P. C., Gerlt, J. A., and Rayment, I. (2002) Homologous $(\beta/\alpha)_8$-barrel enzymes that catalyze unrelated reactions: orotidine 5′-monophosphate decarboxylase and 3-keto-L-gulonate 6-phosphate decarboxylase, *Biochemistry 41*, 3861−3869.
5. Babbitt, P. C., Mrachko, G. T., Hasson, M. S., Huisman, G. W., Kolter, R., Ringe, D., Petsko, G. A., Kenyon, G. L., and Gerlt, J. A. (1995) A functionally diverse enzyme superfamily that abstracts the alpha protons of carboxylic acids, *Science 267*, 1159−1161.
6. 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.
7. 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.
8. Klenchin, V. A., Schmidt, D. M., Gerlt, J. A., and Rayment, I. (2004) Evolution of enzymatic activities in the enolase superfam-

ily: structure of a substrate-liganded complex of the L-Ala-D/L-Glu epimerase from *Bacillus subtilis*, *Biochemistry 43*, 10370−10378.
9. 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.
10. 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.
11. Jensen, R. A. (1976) Enzyme recruitment in evolution of new function, *Annu. Rev. Microbiol. 30*, 409−425.
12. Ohno, S. (1970) *Evolution by Gene Duplication*, Springer-Verlag, New York.
13. Hughes, A. L. (1994) The evolution of functionally novel proteins after gene duplication, *Proc. Biol. Sci. 256*, 119−124.
14. Lynch, M., and Katju, V. (2004) The altered evolutionary trajectories of gene duplicates, *Trends Genet. 20*, 544−549.
15. Schmidt, D. M., Mundorff, E. C., Dojka, M., Bermudez, E., Ness, J. E., Govindarajan, S., Babbitt, P. C., Minshull, J., and Gerlt, J. A. (2003) Evolutionary potential of $(\beta/\alpha)_8$-barrels: functional promiscuity produced by single substitutions in the enolase superfamily, *Biochemistry 42*, 8387−8393.
16. 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.
17. 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(II) and *o*-succinylbenzoate, *Biochemistry 39*, 10662−10676.
18. Vick, J. E., Schmidt, D. M., and Gerlt, J. A. (2005) Evolutionary potential of $(\beta/\alpha)_8$-barrels: in vitro enhancement of a "new" reaction in the enolase superfamily, *Biochemistry 44*, 11722−11729.
19. Patrick, W. M., Firth, A. E., and Blackburn, J. M. (2003) User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries, *Protein Eng. 16*, 451−457.
20. Spencer, M. E., and Guest, J. R. (1973) Isolation and properties of fumarate reductase mutants of Escherichia coli, *J. Bacteriol. 114*, 563−570.
21. 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.
22. Gulick, A. M., Schmidt, D. M., Gerlt, J. A., and Rayment, I. (2001) Evolution of enzymatic activities in the enolase superfamily: crystal structures of the L-Ala-D/L-Glu epimerases from *Escherichia coli* and *Bacillus subtilis*, *Biochemistry 40*, 15716−15724.
23. Khersonsky, O., Roodveldt, C., and Tawfik, D. S. (2006) Enzyme promiscuity: evolutionary and mechanistic aspects, *Curr. Opin. Chem. Biol. 10*, 498−508.
24. Weinreich, D. M., Delaney, N. F., Depristo, M. A., and Hartl, D. L. (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins, *Science 312*, 111−114.

BI7019063