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*[†,‡]

Vadim A. Klenchin,[§] Dawn M. Schmidt,[||] John A. Gerlt,*[,||] and Ivan Rayment*[,§]

Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706, and Departments of Biochemistry and Chemistry, University of Illinois, Urbana, Illinois 61801

Received April 21, 2004; Revised Manuscript Received June 10, 2004

ABSTRACT: The members of the mechanistically diverse enolase superfamily share a bidomain structure formed from a $(\beta/\alpha)_7\beta$-barrel domain [a modified $(\beta/\alpha)_8$- or TIM-barrel] and a capping domain formed from N- and C-terminal segments of the polypeptide. The active sites are located at the interface between the C-terminal ends of the β-strands in the barrel domain and two flexible loops in the capping domain. Within this structure, the acid/base chemistry responsible for formation and stabilization of an enediolate intermediate derived from a carboxylate anion substrate and the processing of it to product is "hard-wired" by functional groups at the C-terminal ends of the β-strands in the barrel domain; the identity of the substrate is determined in part by the identities of residues located at the end of the eighth β-strand in the barrel domain and two mobile loops in the capping domain. On the basis of the identities of the acid/base functional groups at the ends of the β-strands, the currently available structure−function relationships derived from functionally characterized members are often sufficient for "deciphering" the identity of the chemical reaction catalyzed by sequence-divergent members discovered in genome projects. However, insufficient structural information for liganded complexes for specifying the identity of the substrate is available. In this paper, the structure of the complex of L-Ala-L-Glu with the L-Ala-D/L-Glu epimerase from *Bacillus subtilis* is reported. As expected for the 1,1-proton transfer reaction catalyzed by this enzyme, the α-carbon of the substrate is located between Lys 162 and Lys 268 at the ends of the second and sixth β-strands in the barrel domain. The α-ammonium group of the L-Ala moiety is hydrogen bonded to both Asp 321 and Asp 323 at the end of the eighth β-strand, revealing a novel strategy for substrate recognition in the superfamily. The δ-carboxylate group of the Glu moiety is hydrogen bonded to Arg 24 in one of the flexible loops in the capping domain, thereby providing a structural explanation for the restricted substrate specificity of this epimerase [Schmidt, D. M., Hubbard, B. K., and Gerlt, J. A. (2001) *Biochemistry 40*, 15707−15715]. These studies provide important new information about the structural bases for substrate specificity in the enolase superfamily.

The members of the enolase superfamily share a conserved bidomain structure: a $(\beta/\alpha)_7\beta$-barrel domain [a modified $(\beta/\alpha)_8$- or TIM-barrel] that positions the ligands for an essential Mg^{2+} and one or more acid/base catalysts at the C-terminal ends of the various β-strands, and a capping domain formed from segments at the N- and C-termini of the polypeptide that interact with the substrate and sequester the active site from bulk solvent (*1*, *2*). In the reaction catalyzed by each member of the superfamily, an active site base abstracts a proton from the carbon adjacent to a carboxylate group (α-proton) to generate an enediolate anion intermediate. The intermediate is rendered kinetically competent by electrostatic interactions with a coordinated Mg^{2+} ion. An active site acid converts the intermediate to product by protonation either of the α-carbon of the intermediate (in 1,1-proton transfer reactions) or of a leaving group at the β-carbon (in β-elimination reactions, including dehydration and cycloisomerization). The carboxylate ligands for the Mg^{2+} are always located at the ends of the third, fourth, and fifth β-strands of the barrel domain. In contrast to the ligands for the metal ion, the identities and locations of the acid/base catalysts depend on the specific reaction and can be located at the ends of the second, third, fifth, sixth, or seventh β-strand. The identity of the substrate is determined in part by residues located at the end of the eighth β-strand in the barrel domain and two mobile loops in the capping domain.

[†] This research was supported by Grants GM-52594 (to J.A.G. and I.R.) and GM-65155 (to J.A.G. and I.R.) from the National Institutes of Health. Use of the Argonne National Laboratory Structural Biology Center beamline at the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Energy Research, under Contract W-31-109-ENG-38.

[‡] The X-ray coordinates and structure factors for the complex with the L-Ala-D/L-Glu epimerase from *B. subtilis* have been deposited in the Protein Data Bank as entry 1TKK.

* To whom correspondence should be addressed. I.R.: Department of Biochemistry, 433 Babcock Dr., Madison, WI 53706; phone, (608) 262-0437; fax, (608) 262-1319; e-mail, Ivan_Rayment@biochem.wisc.edu. J.A.G.: Department of Biochemistry, University of Illinois, 600 S. Mathews Ave., Urbana, IL 61801; phone, (217) 244-7414; fax, (217) 244-6538; e-mail, j-gerlt@uiuc.edu.

[§] University of Wisconsin.

[||] University of Illinois.

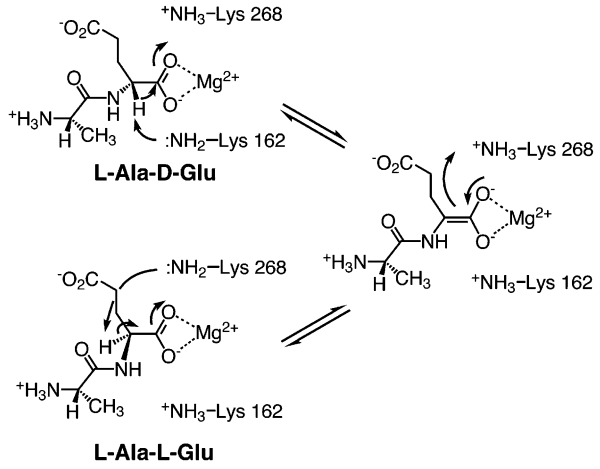
10.1021/bi049197o CCC: $27.50
Published on Web 07/24/2004

Three subgroups for the enolase superfamily whose names are based on their first identified member have been defined. These are the enolase, mandelate racemase (MR),[1] and muconate lactonizing enzyme (MLE) subgroups. This classification is based on the nature and disposition of the ligands that coordinate the metal ion and catalytic residues that initiate the abstraction of the α-proton of the substrate. From an extensive series of studies, it has become clear that the reactions catalyzed by members of the superfamily result from "hard-wiring" of the acid/base chemistry, with the identity of the reaction determined by the structure of the carboxylate anion that is allowed to coordinate to the essential Mg^{2+} and be positioned proximal to the acid/base catalysts (*3*).

In contrast to the ability to "decipher" the identity of the reaction, including its stereochemical course, from the positions of the acid/base catalysts, an understanding of the structural features that determine substrate specificities is more poorly developed. As a result, it is difficult to assign functions to many members that have been discovered in genome projects. If the ubiquitous enolases are excluded, more than 700 nonredundant members can be identified in the sequence databases (J. A. Gerlt, unpublished observations). Of these, at least one-half have unknown functions, because the genome (operon) context is insufficient for identifying the substrate, so it is impossible to predict the identities of the substrates from either sequence or structure.

In addition to the large pool of unknown genomic proteins, the problem of assignment of function is also confounded by the apparent plasticity and ease of alteration of function within previously defined members of the superfamily. For example, the *o*-succinylbenzoate synthase (OSBS) from *Amycolatopsis* catalyzes an *N*-acylamino acid racemase reaction in addition to its required activity (*4*−*6*). Introduction of new activities into this group of enzymes is also apparently quite facile as exemplified by the demonstration that single substitutions in two members of the MLE subgroup are sufficient to introduce functional promiscuity (*7*). In the case of the L-Ala-D/L-Glu epimerase from *Escherichia coli*, substitution of Asp 297 with Gly at the end of the eighth β-strand in the barrel domain allows the catalysis of the *o*-succinylbenzoate synthase reaction as well as a reduced level of the epimerase reaction. In the case of MLE II from *Pseudomonas* sp. P51, the analogous substitution of Glu 323 with Gly at the end of the eighth β-strand in the barrel domain allows catalysis of the OSBS reaction as well as a reduced level of the cycloisomerase reaction. In both cases, a single substitution is able to relax the substrate specificity so that the Lys residues located on opposite faces of the active site at the ends of the second and sixth β-strands can catalyze the proton transfer reactions necessary for both the original and introduced reactions. High-resolution structures are available for only a small number of liganded complexes (*8*−*12*). In these, one or more loops in the capping domain, which often are disordered in the absence of ligand, form direct interactions with the bound ligand. In addition, the residue at the end of the eighth β-strand in the barrel domain sometimes contacts the bound ligand. In MR, the phenyl group of the substrate is located in a hydrophobic cavity formed by the capping domain, and Glu 317 at the end of the eighth β-strand participates in a hydrogen bond with one carboxylate oxygen of the substrate to assist in stabilization of the enediolate anion intermediate (*8*). In D-glucarate dehydratase, the distal portion of the substrate contacts several residues in the capping domain, and Asp 366 at the end of the eighth β-strand interacts via a water molecule with a carboxylate oxygen of the substrate (*10*). In addition, in two OSBS's, the cyclohexadienyl moiety of the substrate is bound in a hydrophobic cavity formed by the capping domain, but the end of the eighth β-strand does not contact the substrate (*11*−*13*). Unfortunately, this structural database of liganded structures is much too limited to allow predictions, or even to place geometric restrictions on, the identities of the substrates for functionally unassigned members.

Scheme 1



In an effort to improve our ability to assign function to unassigned proteins, the structure of the complex of L-Ala-D-Glu and the L-Ala-D/L-Glu epimerase from *Bacillus subtilis*, a member of the MLE subgroup (*14*, *15*), has been determined (Scheme 1). This study provides insight into the disposition of the catalytic residues together with a better understanding of those factors that control the specificity and function within the enolase superfamily.

MATERIALS AND METHODS

Protein Purification. The L-Ala-D/L-Glu epimerase from *B. subtilis* (YkfB or AE epimerase) was expressed and purified as described previously (*14*). Briefly, the *ykfB* gene from *B. subtilis* strain 168 was PCR-amplified and cloned into the pET17b vector to produce a protein without an N-terminal His tag. The recombinant plasmid was transformed into *E. coli* BL21(DE3) cells, and expression of soluble protein occurred without induction at room temperature. The protein was purified by ion exchange chromatography with DEAE-Sepharose where the final protein was estimated to be greater than 95% pure as judged by SDS−PAGE.

Crystallization and X-ray Data Collection. The protein was concentrated to 20 mg/mL and dialyzed against 5 mM

[1] Abbreviations: APS, Advanced Photon Source at Argonne National Laboratory; MePEG, methyl ether polyethylene glycol; MLE, muconate lactonizing enzyme; rms, root-mean-square; MOPS, 3-(*N*-morpholino)-propanesulfonic acid; HEPES, *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid; OSBS, *o*-succinylbenzoate synthase; TCEP, tris(2-carboxyethyl)phosphine hydrochloride; MR, mandelate racemase; YkfB, L-Ala-D/L-Glu epimerase from *B. subtilis*; AE epimerase, L-Ala-D/L-Glu epimerase from *B. subtilis*.

HEPES, 2 mM $MgCl_2$, 50 mM NaCl, 0.25 mM TCEP, and 1 mM NaN_3 (pH 7.0), drop frozen as small pellets in liquid nitrogen, and stored at −80 °C. The frozen concentrated protein was thawed and diluted to 10 mg/mL with 5 mM HEPES, 2 mM $MgCl_2$, and 1 mM NaN_3 (pH 7.5). A stock solution of the peptide substrate L-Ala-L-Glu (Bachem) was neutralized to pH 7.5 with NaOH and brought to a final concentration of 1.6 M. This solution was added to the protein as 1 part per 50 by volume and the resultant mixture utilized for crystallization experiments. The initial crystals were grown by the hanging drop method by mixing 5 μL of the protein solution and 5 μL of a solution containing 9−11% dimethyl-PEG 5000, 50 mM MOPS, 1% MPD, and 1 mM NaN_3 (pH 7.0) and suspending the droplets over 600 μL of precipitant at 20 °C. Large rods (~0.2 mm × ~0.3 mm × ~1.0 mm) grew spontaneously within 1 week. They belong to space group *P*212121 with an octamer in the asymmetric unit and diffracted to at least 2.0 Å resolution with the following cell dimensions: $a = 117.4$ Å, $b = 134.6$ Å, and $c = 194.9$ Å.

For cryopreservation, crystals were first transferred into a precipitation solution containing 2 mM $MgCl_2$ and 16 mM L-Ala-D/L-Glu. Thereafter, the crystals were transferred in four equal steps at 5 min intervals to a solution composed of 35% PEG 8000, 400 mM KCl, 50 mM MOPS, 1% ethanol, 2 mM $MgCl_2$, and 16 mM Ala-Glu (pH 7.0) and rapidly frozen in a stream of nitrogen at −160 °C.

Data were collected to 2.1 Å resolution at beamline 19ID of the Structural Biology Center at the Advanced Photon Source (APS, Argonne National Laboratory, Argonne, IL) with a single 180° scan with 0.5° oscillations. The data were processed and integrated with d*Trek (*16*) which was more capable of resolving the high mosaicity of the data than Denzo (*17*). The number of integrated intensities and their R_{merge}, together with the subsequent quality of the structural determination, were far superior with the data processed with d*Trek than the data processed with Denzo. The latter program gave only 50% completeness to 2.1 Å resolution, because of its inability to resolve the overlapping reflections caused by the high mosaicity.

Structure Determination and Refinement. The structure was determined by molecular replacement with Molrep (*18*) starting from the coordinates for the substrate free enzyme [PDB entry 1JPM (*15*)]. Thereafter, the structure was refined with Refmac (*19*, *20*). Water molecules were added to the coordinate set with ARP/wARP and subsequent manual verification (*19*, *21*). Iterative cycles of maximum likelihood refinement and manual model building reduced the R_{work} to 21.9% for all measured X-ray data from 30.0 to 2.1 Å resolution. The R_{free} was 28.5% for 5% of the data that was excluded from the refinement. Refinement statistics are presented in Table 1. Analysis of the coordinates with PROCHECK (*22*) revealed that 92.1% of the residues lie in the most favored regions of the Ramachandran plot, whereas the remaining 7.9% of the residues lie in additionally allowed areas. No residues are located in the disallowed regions.

Table 1: Data Collection and Refinement Statistics

space group	*P*212121
unit cell (Å)	$a = 117.4$, $b = 134.6$, $c = 194.9$
wavelength (Å)	0.99987
resolution (Å)	50−2.1 (2.18−2.10)
total no. of reflections	1081194
no. of unique reflections	173946
completeness (%)[a]	96.7 (96.7)
redundancy	6.15 (6.4)
average I/σ (raw/scaled)	8.6 (2.6)
R_{merge} (%)[b]	0.159 (0.508)
refinement and model statistics	
no. of protein atoms	21626
no. of heteroatoms[c]	128
no. of waters	1521
R_{work} (%)	21.9 (32.4)
R_{free}[d] (%)	28.5 (36.2)
Wilson *B* value (Å^2)	33
average *B* factor (Å^2)	
protein	27.9
Mg peptide	26.3
waters	27.5
Ramachandran plot (%)	
most favored regions	92.1
additionally allowed regions	7.3
generously allowed regions	0.6
disallowed regions	0.0
rmsd for bond lengths (Å)	0.01
rmsd for bond angles (deg)	1.2

[a] The values in parentheses give the statistics for the highest-resolution shell that extends from 2.18 to 2.10 Å. [b] $R_{merge} = (\sum|I_{hkl} - I|)/(\sum I_{hkl}) \times 100$, where the average intensity I is taken over all symmetry equivalent measurements and I_{hkl} is the measured intensity for a given reflection. [c] This includes eight Mg^{2+} ions and eight molecules of L-Ala-L-Glu. [d] R_{free} is the R_{factor} for which 5% of the data was excluded from the refinement and phase calculation.

RESULTS AND DISCUSSION

The AE epimerase−L-Ala-D/L-Glu complex crystallizes under conditions different from those of the substrate free protein such that there is now an octamer in the asymmetric unit. The electron density is continuous from Met 1 to Ala 358 in subunits B and F, and continuous from Met 1 to Leu 359 in the other six subunits. The remaining eight or seven residues anticipated from the sequence were not visible in the electron density. As described before, the AE epimerase contains the same ensemble of secondary structural elements characteristic of the enolase superfamily which consists of a central $(\beta/\alpha)_7\beta$-barrel domain that extends from Gly 125 to Leu 328 that is capped by a smaller domain built from the preceding and following segments of the sequence (*15*). The active site is located at the juxtaposition of the loops at the C-terminal ends of β-strands that make up the barrel and is closed off by the capping domain.

The tertiary structure of the AE epimerase substrate complex is essentially identical to that of the apoenzyme except for the loop (Leu 15−Glu 30) that connects the first and second β-strands of the capping domain (Figure 1). When the apoenzyme and the L-Ala-D/L-Glu complex are superimposed, the overall rms difference between this single set of subunits is 0.3 Å for 294 target α-carbon pairs. Likewise, the quaternary structure of the entire complex is very similar to that of the substrate free enzyme. Superposition of the complete octamers shows that the rms difference is 0.87 Å for 2442 (84%) target α-carbon pairs.

Movement of the capping loop is primarily accomplished by a rigid body rotation via changes in the conformational angles associated with Thr 16 and Ala 29 that lie at the beginning and end of the loop. The residues in the middle of the loop move by approximately 12 Å with the result that a section that is solvent-exposed in the apo structure becomes

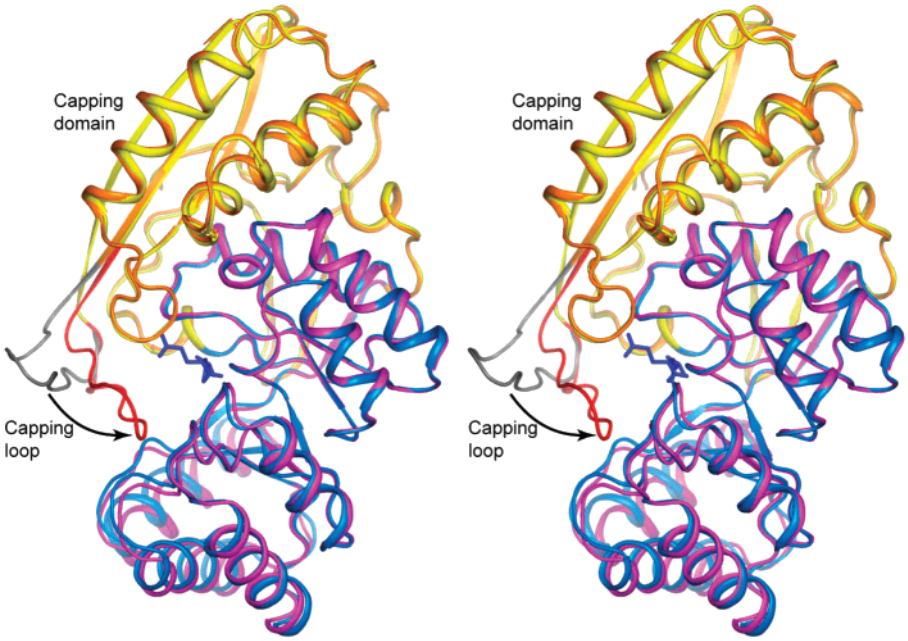


FIGURE 1: Stereo comparison of the apo form and the L-Ala-D/L-Glu complex of AE epimerase that reveals the movement of the capping loop. The substrate complex is depicted in blue for the barrel domain and yellow and red for the capping domain, whereas the structure of the apo form of the AE epimerase is depicted in orange and black for the capping domain and magenta for the barrel domain. The capping loop that extends from Leu 15 to Glu 30 and provides specificity in the liganded structure is depicted in black and red in the apo form and the L-Ala-D/L-Glu complex, respectively. This figure was prepared with PyMOL (*26*). The coordinates for apo AE epimerase from *B. subtilis* were taken from the RCSB Protein Data Bank (entry 1JPM) and superimposed with Align (*27*).

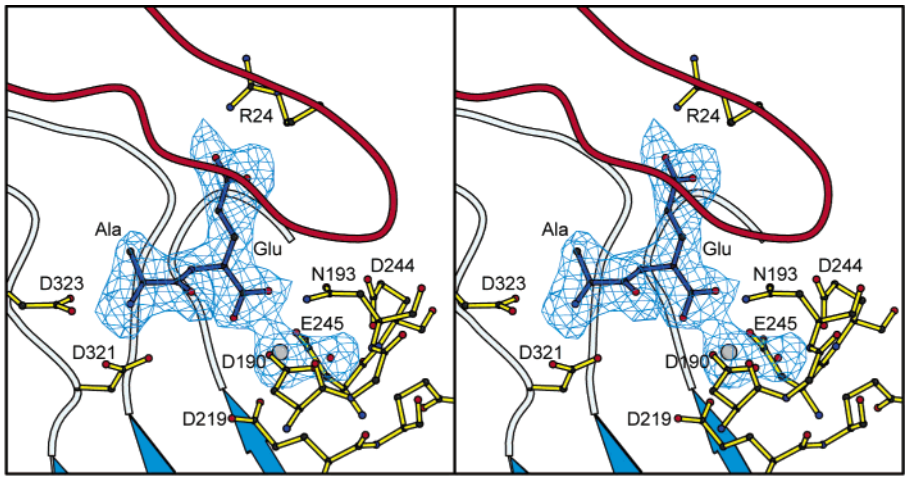


FIGURE 2: Stereoview of the electron density for L-Ala-D/L-Glu, the magnesium ion, and its associated water molecule. This figure is oriented so that the ligand sits above the cavity formed at the C-terminal end of the $(\beta/\alpha)_7$ β-barrel, where the capping loop is colored red. The map was calculated with $F_o - F_c$ coefficients and was contoured at 3σ, where the ligands were omitted from the phase calculation. This figure was prepared with Bobscript (*28*).

packed tightly against the residues that lie at the end of the first β-strand in the L-Ala-D/L-Glu complex. Numerous new specific interactions, both polar and nonpolar, are formed between the translocated capping loop and the residues at the C-terminal end of the $(\beta/\alpha)_7\beta$-barrel when the loop closes over the active site cleft. This movement serves to entrap the substrate within the active site. It is noteworthy that this loop is disordered in the apo structures of MLE (*23*) and the L-Ala-D/L-Glu epimerase from *E. coli* (*15*).

Substrate Coordination Geometry. All eight subunits of the AE epimerase show unequivocal electron density for L-Ala-L-Glu and an associated magnesium ion in the active site (Figure 2) which serves to define the manner in which this enzyme coordinates its substrates. Although a mixture of L-Ala-L-Glu and L-Ala-D-Glu is anticipated to be in the active site, examination of the electron density suggested that the primary constituent is the L-isoform; consequently, the model was built with this stereochemistry. Given that these dipeptides are diastereomers that coordinate an asymmetric active site, they are not expected to bind the enzyme with equal affinity, and thus, one form will predominate, even though the energetic difference between the ground-state complexes might be quite small. As expected, the metal ion forms a bidentate interaction with the α-carboxyl moiety of the glutamate residue in the substrate. The remainder of the metal coordination is completed by monodentate ionic interactions with carboxylate oxygens on Asp 191, Glu 219, and Asp 244, together with a water molecule (Figure 3). There is an important second sphere of ligands that coordi-

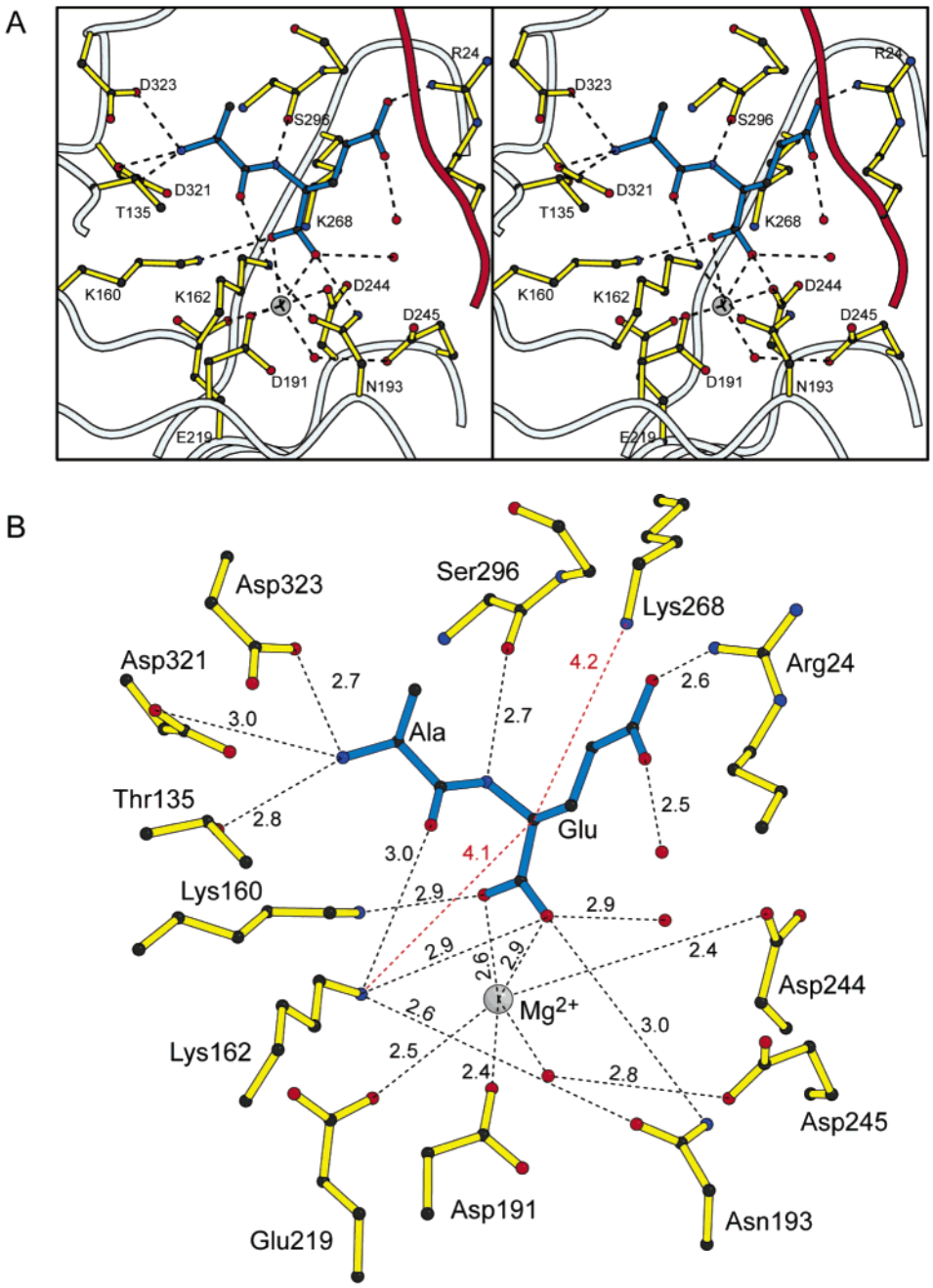


FIGURE 3: Ligand interactions between L-Ala-D/L-Glu and the AE epimerase from *B. subtilis*. (A) Stereo representation of the important polar interactions between L-Ala-D/L-Glu and the active site. (B) Expanded schematic view of the active site revealing the lengths (in angstroms) of the polar contacts between the substrate and metal ion. This figure was prepared with Bobscript (*28*).

nate the metal binding ligands which include Asn 193 and Asp 245.

Our previously published kinetic analysis revealed that the values of k_{cat}, K_m, and k_{cat}/K_m for epimerization of L-Ala-D-Glu catalyzed this epimerase are 15 ± 2.7 s^{-1}, $(3.2 \pm 1.1) \times 10^{-4}$ M, and 4.7×10^4 M^{-1} s^{-1}, respectively (*14*). The reaction is accompanied by incorporation of a solvent-derived hydrogen into the product, as expected for a "two-base mechanism" for the 1,1-proton transfer reaction.

Following the structure−function paradigms previously established for members of the enolase superfamily, including that for the orthologous epimerase from *E. coli*, Lys 162 and Lys 268 are assigned roles as the *R*- and *S*-specific catalytic bases, respectively, in this active site. As expected, these are positioned on opposite sides of the α-carbon of the glutamyl moiety of the substrate. Persuasively supporting this assignment is the fact that mutation of the homologues of either of these residues in the orthologous epimerase from *E. coli* (Lys 151 and Lys 247) reduces the values of k_{cat} and k_{cat}/K_m for the epimerization of L-Ala-D-Glu by factors of $\sim 10^4$ and $\sim 10^3$, respectively (J. A. Gerlt and D. M. Schmidt, unpublished results, and ref *24*).

In all subunits, the Nζ atoms of both Lys 162 and Lys 268 are located approximately 4 Å from the stereochemical center which is racemized, though the exact rotational position of Nζ of Lys 268 varies somewhat between subunits. This is consistent with the premise that a ground-state complex has been crystallized and that as the transition state for enediolate formation is approached these catalytic groups will move closer either to remove or to donate a proton from

the substrate or intermediate. Further evidence that these structures represent ground-state complexes is seen in the ionic interactions between Nζ of Lys 162 and the side chains of Asn 193, the carbonyl oxygen of the alanine residue of the substrate, and the α-carboxyl moiety of the glutamate residue of the substrate (Figure 3). These interactions serve to keep Nζ of Lys 162 opposite the α-carbon, but would need to be disrupted for protonation of the enediolate intermediate to occur. Similar interactions for keeping Nζ of Lys 268 in a position opposite from the α-carbon are not present. Instead, Nζ of Lys 268 is rotated into a polar pocket formed by the side chains of Asp 244 and Asn 266 where it forms a salt bridge with Asp 244. Again for the reaction to occur, the position of Nζ of Lys 268 must change to allow it to abstract or donate a proton. This is also consistent with a substrate ground-state complex.

All of the polar atoms of the substrate are involved in ionic or hydrogen bonding interactions with components of the active site (Figure 3). The α-amino group of the substrate is coordinated to Oγ of Thr 135, and carboxylate oxygens on Asp 321 and Asp 323. These interactions serve to neutralize the positive charge on this end of the Ala-Glu substrate. The amide hydrogen and carbonyl oxygen of the peptide linkage within the substrate form hydrogen bonds with groups on opposite sides of the active site. The main chain oxygen of Ser 296 forms a hydrogen bond with the amide hydrogen, whereas Nζ of Lys 162 is coordinated to the carbonyl oxygen of the substrate. Finally, the glutamate side chain carboxyl forms a salt bridge with the side chain of Arg 24. This latter interaction is the only polar contact between the capping loop and the substrate and would appear to contribute to the specificity of this enzyme as discussed later. Apart from the direct interactions between the substrate and the protein, a substantial number of water-mediated interactions within the active site aid in the neutralization of the charges on the substrate.

In addition to the polar interactions, a significant hydrophobic pocket is associated with the alanine side chain. This is built from the side chains of Phe 19, Ile 298, and Met 327 and would appear to have sufficient room to accept a side chain larger than the methyl group of alanine which is consistent with the observed specificity of the enzyme (*14*).

Previous kinetic studies have shown that the enzyme will epimerize Ala-Glu, Ala-Asp, Ala-Met, Ala-Ser, Pro-Glu, and Ser-Glu where the k_{cat}/K_m is at least 10-fold greater for L-Ala-D-Glu than for L-Ala-D-Asp or L-Ala-D-Met (*14*). The enzyme will not epimerize any of the other normal Ala-amino acid dipeptides or dipeptides such as Lys-Glu or Phe-Glu. The observed specificity is consistent with the structure observed here. Although Arg 24 is well situated for interaction with negatively charged side chains such as glutamate or aspartate, there is a hydrophobic component to this pocket that would readily accept an extended flexible side chain such as methionine. Likewise, epimerization of Pro-Glu or Ser-Glu can readily be explained by the size of the pocket associated with the N-terminus of the peptide, whereas larger or charged side chains would be expected to be inconsistent with productive substrate binding.

Comparison of the Active Sites of Epimerase and a Promiscuous o-Succinylbenzoate Synthase That Also Is an N-Acylamino Racemase. The L-Ala-D/L-Glu epimerase from *B. subtilis* catalyzes a 1,1-proton transfer reaction analogous to the promiscuous *N*-acylamino acid racemase (NAAAR) reaction catalyzed by the *o*-succinylbenzoate synthase (OSBS) from *Amycolatopsis* (*4*−*6*). This enzyme (OSBS/NAAAR) is also a member of the muconate lactonizing enzyme (MLE) subgroup of the enolase superfamily (*1*). The structures of OSBS/NAAAR complexed with the product of the OSBS reaction and three *N*-acylamino acid substrates for the NAAAR reaction, *N*-acetylmethionine, *N*-succinylmethionine, and *N*-succinylphenylglycine, have recently been reported (*12*). Although the degree of sequence identity between the AE epimerase from *B. subtilis* and OSBS/NAAAR is low (28%), their structures are very similar (both enzymes assemble as octamers). Alignment of the L-Ala-D/L-Glu complex of the AE epimerase and the *N*-succinylmethionine complex with OSBS/NAAAR reveals that the rms difference for 300 structurally equivalent target pairs is only 1.5 Å.

The high level of similarity of the tertiary structures of these enzymes is carried forward into the manner in which they bind their substrates. As can be seen in Figure 4, the catalytic lysine residues and the constellation of ligands that bind the metal ions superimpose remarkably well. This correspondence might be expected on the basis of the equivalency of the chemistry that is supported by these enzymes and the structural requirements for catalysis. Even so, this similarity is noteworthy since the physiological function of OSBS/NAAAR is to catalyze the OSBS reaction rather than a 1,1-proton transfer reaction which raises interesting evolutionary questions that are discussed later.

The similarity in the ligand binding to this AE epimerase and OSBS/NAAAR extends beyond the α-carbon and its associated carboxyl group. The overall dispositions of the ligands are very similar. In both cases, the amide hydrogen of the acyl linkage or peptide bond is hydrogen-bonded to a main chain carbonyl oxygen (Ser 296 and Gly 291 in the epimerase and NAAAR, respectively). This interaction in OSBS/NAAAR was identified as a key component in its active site that allows this enzyme to catalyze both the NAAAR and OSBS reactions (*12*). Conversely, the coordination of the carbonyl oxygen of the substrate amide linkage appears to differ between the epimerase and OSBS/NAAAR. In the former, the carbonyl oxygen is coordinated primarily by Nζ of Lys 162, whereas in OSBS/NAAAR, its carbonyl oxygen is hydrogen-bonded to both Oγ of Ser 135 and Nζ of Lys 163. Oγ of Thr 135 is not oriented appropriately for formation of a hydrogen bond to the carbonyl oxygen; however, as noted earlier, the interaction between Nζ of Lys 162 and the carbonyl oxygen is most likely disrupted as the transition state for proton donation is approached. Thus, it is conceivable that an interaction with Thr 135 does exist at some point in the reaction pathway.

The amino acids that coordinate the remainder of the substrates are different in the AE epimerase and OSBS/NAAAR, as would be expected. The glutamate side chain is coordinated by Arg 24 in the epimerase, whereas the side chain for methionine lies in a hydrophobic pocket, although this is provided in part by residues from the capping domain (Phe 23 and Met 50). Greater tertiary structural differences are seen in the manner in which the α-amino group of the alanine is coordinated in the epimerase compared to binding of the succinyl carboxyl group of *N*-succinylmethionine in OSBS/NAAAR. In the latter, the loop at the end of the eighth β-strand is displaced from the barrel axis more than that in

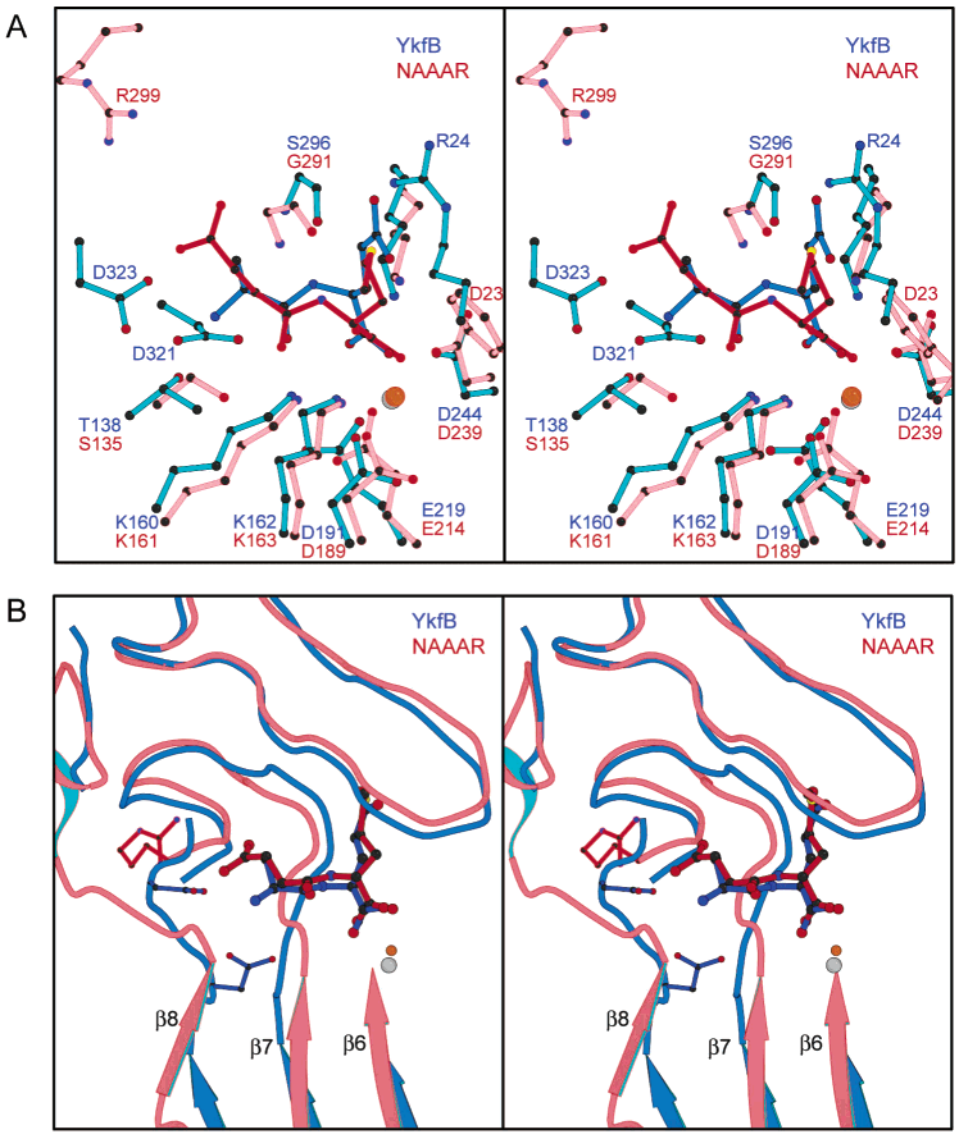


FIGURE 4: Stereoviews of the active site of the L-Ala-D/L-Glu complex of the AE epimerase from *B. subtilis* compared with OSBS/NAAAR from *Amycolatopsis* complexed with *N*-succinylmethionine. This latter enzyme exhibits both OSBS activity and an adventitious ability to racemize *N*-acylamino acids. In panel A, the residues that are directly coordinated to the polar components of the substrates are depicted in cyan and red for the AE epimerase from *B. subtilis* and NAAAR, respectively. In panel B, the sixth, seventh, and eighth β-strands are depicted to demonstrate that the succinyl moiety of *N*-succinylmethionine is accommodated in NAAAR via the movement of the loop that extends from the eighth β-strand into the capping domain. This figure also shows that the capping loop that extends from Leu 15 to Glu 30 in the epimerase is very similar to that seen in NAAAR. The coordinates for the *Amycolatopsis* complex were taken from the RCSB Protein Data Bank [entry 1SJC (*12*)].

the epimerase to allow coordination of the longer succinyl moiety (Figure 4).

In addition to the polar components of the active site of the epimerase, a substantial number of hydrophobic residues interact with the nonpolar components of the substrate. These residues include Phe 19, Ile 54, Met 297, Ile 298, and Met 327. This explains why Ala-Met is a substrate for this enzyme. Interestingly, the equivalent residues in OSBS/NAAAR (Phe 19, Tyr 55, Met 292, Ile 293, and Phe 323) are also hydrophobic and fulfill similar roles. This similarity also explains why a single mutation is capable of introducing OSBS activity in the L-Ala-D/L-Glu epimerase from *E. coli* (*7*), although the exact structural consequence of this mutation or how 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate binds in the active site of the mutated protein is unknown.

Evolutionary Implications. As noted earlier, the degree of sequence identity between this AE epimerase and OSBS/NAAAR is only ~28%. Although this value is low, the level of sequence identity between AE epimerase and the OSBS from the same organism (*B. subtilis*) is even lower (23%). The obvious structural similarity between all of these enzymes is consistent with a common ancestor; however, the pathway along which these proteins evolved is clearly unknown. The distant relationship between the AE epimerase and the OSBS from the same organism (*B. subtilis*) does not support the supposition that they evolved from a gene duplication event from a common ancestor within the same organism, given that the structural frameworks necessary for facilitation of the reactions they catalyze are so similar.

Structural Determinants of Substrate Specificity in the Enolase Superfamily. As noted in the introductory section,

a carboxylate group at the end of the eighth β-strand interacts, either directly or via an intervening water molecule, with one oxygen of the α-carboxylate of the substrate in the active site of MR (Glu 317) and D-glucarate dehydratase from *E. coli* (Asp 366) (*10*). In the case of the as-of-yet structurally uncharacterized acid sugar dehydratases that are members of the MR subgroup and share a conserved Glu residue at the end of the eighth β-strand (L-rhamnonate dehydratase, D-gluconate dehydratase, and a bifunctional D-altronate/D-mannonate dehydratase), it is expected that this residue will play a role in substrate coordination; however, as shown here, the function of these conserved residues and how they influence substrate specificity are difficult to predict within the enolase superfamily.

The structure of the complex of L-Ala-D/L-Glu with the AE epimerase from *B. subtilis* illustrates a different interaction between acidic residues at the end of the eighth β-strand and the substrate: the α-ammonium group of the dipeptide substrate is hydrogen-bonded to Asp 321 and Asp 323, thereby directing the binding of a dipeptide in the active site so that the conserved Lys 162 and Lys 268 located at the ends of the second and sixth β-strands, respectively, catalyze the physiological 1,1-proton transfer reaction, resulting in epimerization of the dipeptide substrate. Other functionally unassigned members of the superfamily also are predicted to contain acidic groups at the ends of the eighth β-strands in the barrel domain, thereby raising the possibility that these may also be dipeptide epimerases, although their sequences are <20% identical with that of the AE epimerase from *B. subtilis*. Alternatively, some of these may be amino acid racemases in which the α-ammonium group is coordinated to the acidic group at the end of the eighth β-strand.

The importance of the loops contributed by the capping domain to the substrate binding pocket is reinforced by the structure presented here. The interactions between the substrate and residues in the capping domain provide insight into substrate specificity. Although they differ in detail between members of the enolase superfamily, they are consistent with the polarity of the substrate. For example, in D-glucarate dehydratase, the distal carboxylate group of its substrate interacts primarily with His 32 in the capping domain and His 368 at the end of the eighth β-strand (*25*). Many of the functionally unassigned members of the superfamily are predicted to contain polar and charged groups in the mobile loops in the capping domain, so even this limited information is expected to provide clues for predicting ligand specificity.

Clearly, "structural rules" cannot yet be formulated that will allow prediction of the substrate specificity of members of the enolase superfamily for which a function has not yet been discovered. Indeed, the number of distinct active site motifs, i.e., identities of the functional groups located at the ends of the β-strands in the barrel domain, is considerably smaller than the expected number of orthologous groups in the superfamily (J. A. Gerlt, unpublished observations). Thus, structural studies of many additional functionally assigned members of the superfamily likely will be required before a general strategy for prediction of ligand specificity and, therefore, functional assignment based on sequence and structure can be developed and implemented.

ACKNOWLEDGMENT

We thank Drs. Randy Alkire and Norma Duke of the Structural Biology Center, Argonne National Laboratory, for outstanding assistance in recording the data reported in this study. We also thank Dr. James Pflugrath for help in processing the data with d*Trek.

REFERENCES

1. Babbitt, P. C., Hasson, M. S., Wedekind, J. E., Palmer, D. R., Barrett, W. C., Reed, G. H., Rayment, I., Ringe, D., Kenyon, G. L., and Gerlt, J. A. (1996) The enolase superfamily: a general strategy for enzyme-catalyzed abstraction of the α-protons of carboxylic acids, *Biochemistry 35*, 16489−16501.
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. Tokuyama, S., and Hatano, K. (1995) Purification and properties of thermostable N-acylamino acid racemase from *Amycolatopsis* sp. TS-1-60, *Appl. Microbiol. Biotechnol. 42*, 853−859.
5. 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.
6. 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.
7. 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.
8. 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.
9. Wieczorek, S. W., Kalivoda, K. A., Clifton, J. G., Ringe, D., Petsko, G. A., and Gerlt, J. A. (1999) Evolution of Enzymatic Activities in the Enolase Superfamily: Identification of a "New" General Acid Catalyst in the Active Site of D-Galactonate Dehydratase from *Escherichia coli*, *J. Am. Chem. Soc. 121*, 4540−4541.
10. 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.
11. 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*, 10662−1076.
12. 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* (in press).
13. 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.
14. Schmidt, D. M., Hubbard, B. K., and Gerlt, J. A. (2001) Evolution of enzymatic activities in the enolase superfamily: functional assignment of unknown proteins in *Bacillus subtilis* and *Escherichia coli* as L-Ala-D/L-Glu epimerases, *Biochemistry 40*, 15707−15715.
15. 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.
16. Pflugrath, J. W. (1999) The finer things in X-ray diffraction data collection, *Acta Crystallogr. D55* (Part 10), 1718−1725.
17. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307−326.
18. Vagin, A., and Teplyakov, A. (2000) An approach to multi-copy search in molecular replacement, *Acta Crystallogr. D56*, 1622−1624.
19. Collaborative Computational Project Number 4 (1994) The CCP4 Suite: Programs for protein crystallography, *Acta Crystallogr. D50*, 760−763.
20. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method, *Acta Crystallogr. D53*, 240−255.
21. Perrakis, A., Harkiolaki, M., Wilson, K. S., and Lamzin, V. S. (2001) ARP/wARP and molecular replacement, *Acta Crystallogr. D57*, 1445−1450.
22. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
23. Helin, S., Kahn, P. C., Guha, B. L., Mallows, D. G., and Goldman, A. (1995) The refined X-ray structure of muconate lactonizing enzyme from *Pseudomonas putida* PRS2000 at 1.85 Å resolution, *J. Mol. Biol. 254*, 918−941.
24. Schmidt, D. M. Z. (2003) Divergent evolution of enzymatic activity: functional assignment and in vitro evolution in the MLE subgroup of the enolase superfamily, Ph.D. Thesis, University of Illinois, Urbana, IL.
25. Gulick, A. M., Hubbard, B. K., Gerlt, J. A., and Rayment, I. (2001) Evolution of enzymatic activities in the enolase superfamily: identification of the general acid catalyst in the active site of D-glucarate dehydratase from *Escherichia coli*, *Biochemistry 40*, 10054−10062.
26. DeLano, W. L. (2002) *PyMOL*, DeLano Scientific, San Carlos, CA.
27. Cohen, G. H. (1997) ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions, *J. Appl. Crystallogr. 30*, 1160−1161.
28. Esnouf, R. M. (1999) Further additions to MolScript version 1.4, including reading and contouring of electron-density maps, *Acta Crystallogr. D55*, 938−940.

BI049197O