Minireview

Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity ☆

John A. Gerlt[a,*], Patricia C. Babbitt[b], Ivan Rayment[c]

[a] *Departments of Biochemistry and Chemistry, University of Illinois, Urbana, IL 61801, USA*
[b] *Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, University of California, San Francisco, CA 94143, USA*
[c] *Department of Biochemistry, University of Wisconsin, Madison, WI 53705, USA*

Abstract

The members of the mechanistically diverse enolase superfamily catalyze different overall reactions. Each shares a partial reaction in which an active site base abstracts the α-proton of the carboxylate substrate to generate an enolate anion intermediate that is stabilized by coordination to the essential Mg^{2+} ion; the intermediates are then directed to different products in the different active sites. In this minireview, our current understanding of structure/function relationships in the divergent members of the superfamily is reviewed, and the use of this knowledge for our future studies is proposed.
© 2004 Published by Elsevier Inc.

Keywords: Enolase superfamily; Divergent evolution

In 1990, the structure-based discovery was made that: (1) the three-dimensional structures of mandelate racemase (MR)[1] from *Pseudomonas putida* and muconate lactonizing enzyme (MLE), also from *P. putida*, are remarkably superimposable (Fig. 1); and (2) the active site carboxylate residues that bind an essential Mg^{2+} ion and mediate proton transfer reactions from the carbon acid substrate and to the resulting enolate ion intermediate are highly conserved [1]. The enzymes share a bidomain structure, in which the active sites are located at the interface between flexible loops in a capping domain formed from segments contributed by the N- and C-terminal regions of the polypeptide and the C-terminal ends of the β-strands of a modified TIM-barrel domain $[(\beta/\alpha)_7\beta$ instead of $(\beta/\alpha)_8]$ where the conserved active site residues are positioned. The reactions catalyzed by MR and MLE were immediately recognized to share a partial reaction in which an active site base abstracts the α-proton of the carboxylate substrate to generate an enolate anion intermediate that is stabilized by coordination to the essential Mg^{2+} ion; the intermediates are directed to different products in the different active sites (Fig. 2). The conservation of the bidomain structure provided convincing evidence that MR and MLE are homologues, i.e., derived from a common progenitor by divergent evolution. At that time, structural conservation was thought to be necessary to "prove" evolution from a common ancestor, because the pair-wise sequence identities were ∼25%. No other

[1] *Abbreviations used:* MR, mandelate racemase; MLE, muconate lactonizing enzyme; 2-PGA, 2-phosphoglycerate; OSBS, *o*-succinylbenzoate synthase; AE Epim, L-Ala-D/L-Glu epimerase, GlucD, D-glucarate/L-idarate dehydratase; AltD/ManD, D-altronate/D-mannonate dehydratase; GalD, D-galactonate dehydratase; GlcD, D-gluconate dehydratase; RhamD, L-rhamnonate dehydratase; MAL, 3-methylaspartate ammonia lyase; NAAAR, *N*-acylamino acid racemase; SHCHC, 2-succinyl-6-hydroxyl-2,4-cyclohexadiene-1-carboxylate.
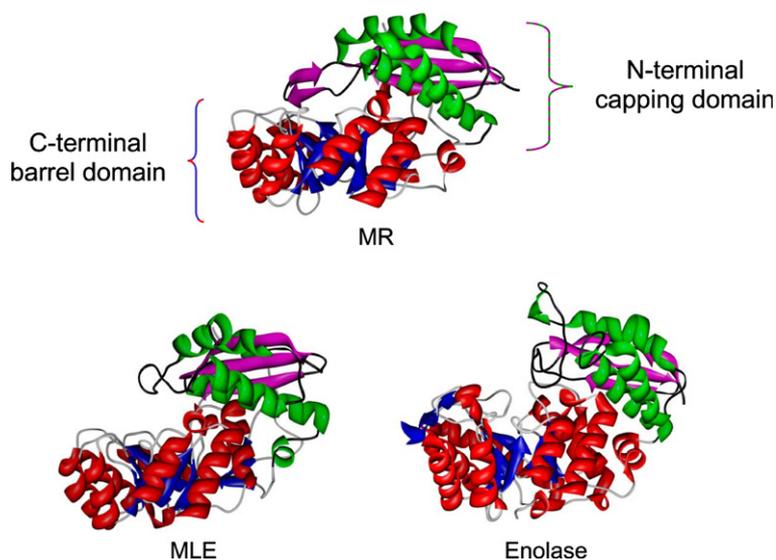
Fig. 1. Comparison of the structures of the polypeptides of mandelate racemase (MR), muconate lactonizing enzyme (MLE), and enolase, showing the two homologous domains that "prove" divergent evolution.
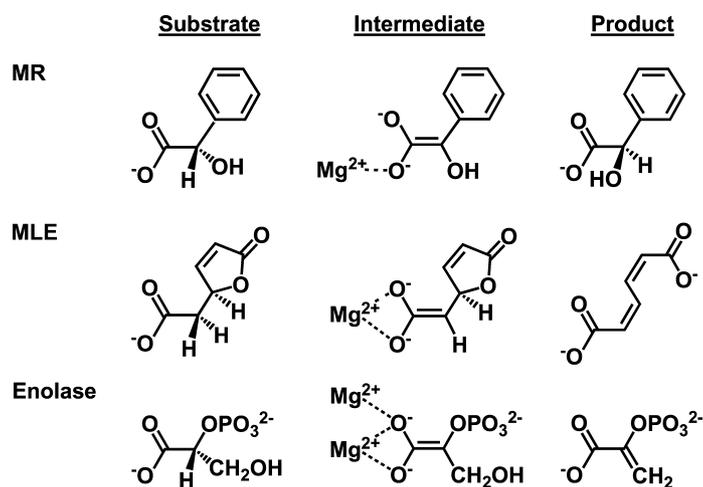

Fig. 2. The substrates, enolate anion intermediates, and products of the MR, MLE, and enolase reactions.

homologues could be recognized in the structure (or sequence) databases, so the importance of this discovery in elucidating relationships between structure and function was unknown.

In 1995, again based on structural evidence, yeast enolase was recognized to share the same bidomain structure as that observed in MR and MLE as well as some of the functionally essential active site residues (Fig. 1), thereby providing a third reaction involving enolization of a carbon acid that can be catalyzed by active sites derived from that of a common ancestor (Fig. 2) [2,3]; in enolases, the substrate carboxylate group is coordinated to two Mg^{2+} ions, one of which is liganded to the three conserved carboxylate residues. Even at that time, the pair-wise sequence identities relating enolases with either MR or MLE were too low to permit the con-

clusion that the sequences and, therefore, the structures were homologous. Eventually, the sequence databases were sufficiently populated that the homologous relationship of enolases with both MR and MLE could be recognized by sequence alignments. But, even the structure-based discovery of three homologous enzymes that catalyze different reactions provided persuasive evidence that the process of divergent evolution could give rise to unexpected and unprecedented functional diversity. The term "mechanistically diverse" was used to describe the functional relationships in the enolase superfamily, because the reactions the members catalyze share a common partial reaction, Mg^{2+}-assisted enolization of a carbon acid, but the enolate anion intermediates are directed to different products by different partial reactions that usually involve general acid catalysis [4].

Since the initial structure-based discovery of the enolase superfamily, we now recognize that the sequence databases contain hundreds of members of the enolase superfamily, many of which catalyze as yet unknown reactions. Structures are now available for members of the superfamily that catalyze eight different overall reactions: enolase [5]; mandelate racemase [6]; muconate lactonizing enzyme I [7,8]; muconate lactonizing enzyme II [9]; D-glucarate dehydratase [10–12]; D-galactonate dehydratase [13]; o-succinylbenzoate synthases [14–16]; L-Ala-D/L-Glu epimerases [17]; and 3-methylaspartate ammonia lyase [18,19]. With these growing sequence and structure databases, we now are in the position to understand how changes in sequence and structure permit changes in substrate specificity and reaction mechanism and the interplay between these in the evolution of new enzymatic reactions. Such understanding has already allowed the (re)design of members of the superfamily to catalyze different reactions [20] and is also expected to allow the development of strategies for discovery of the functions of the functionally unknown members. This minireview summarizes the current state of this knowledge as well as provides directions for future studies.

Active site motifs in the enolase superfamily

By "definition," all members of the superfamily contain ligands (almost always Glu or Asp) for the essential Mg^{2+} ion located at the ends of the third, fourth, and fifth β-strands; the conservation of homologues of these at appropriate locations in the sequences of proteins >300 residues in length is the primary criterion for identifying "new" members of the superfamily in the sequence databases. With the sequence data now available, the members of the enolase superfamily can be partitioned into four subgroups, based on the identities of the acid/base functional groups located at the ends of the second, third, fifth, sixth, and seventh β-strands. These partitions, based only on variation in the active site residues, correlate well with the evolutionary trees generated using the overall sequences, indicating the importance of these motifs during evolutionary divergence.

The orthologous members of the enolase subgroup contain a conserved Lys at the end of the sixth β-strand. This Lys is the general base that abstracts the proton of the carbon acid substrate, 2-phosphoglycerate (2-PGA); the general acid that facilitates departure of the hydroxide leaving group is located on a characteristic loop following the second β-strand. To date, the members of the enolase subgroup are thought to be isofunctional, catalyzing the conversion of 2-PGA to PEP (Fig. 3). The current databases contain sequences for >600 enolases.

The heterofunctional members of the MLE subgroup contain Lys residues at the ends of the second and sixth β-strands; sometimes, one of these is substituted with an Arg. At least one of these Lys residues is the general base that abstracts the proton of the carbon acid substrate. The
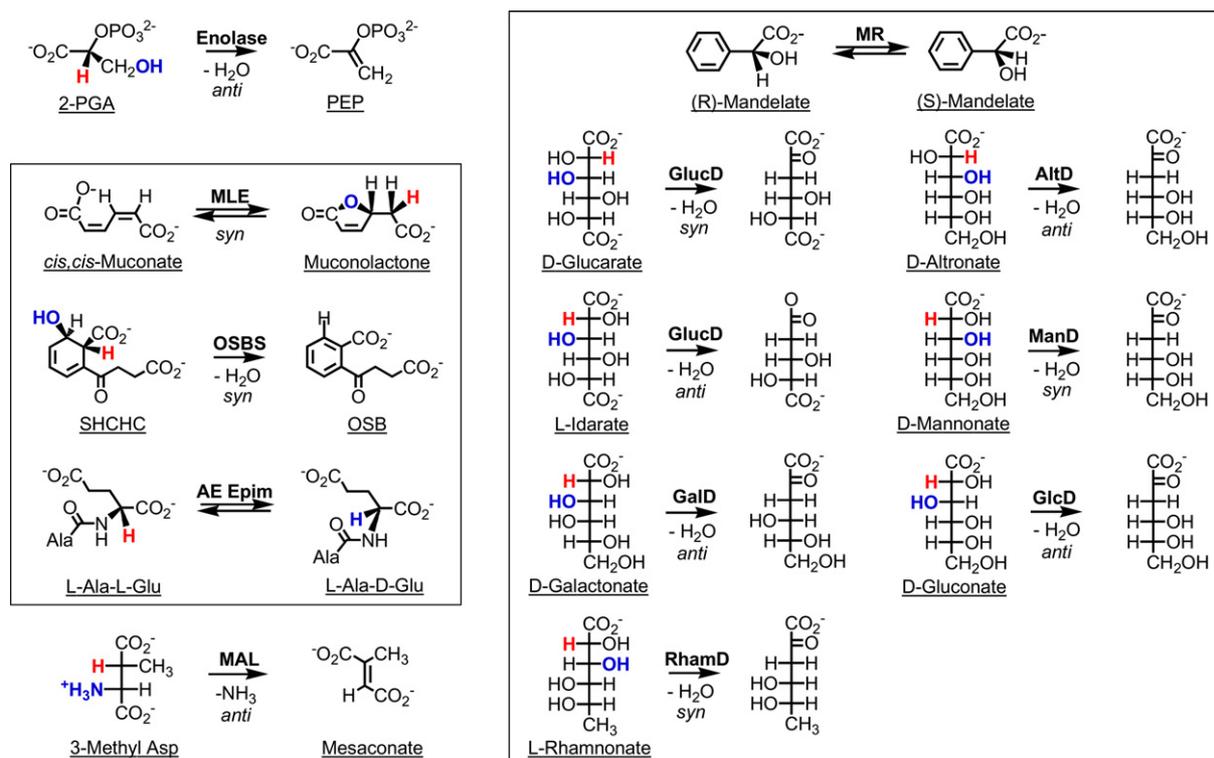

Fig. 3. The reactions catalyzed by the four subgroups of the enolase superfamily.

identities of metal ligands reinforce the identification of a member of the MLE subgroup, i.e., the ligand at the end of the fifth β-strand is always an Asp followed by a Glu. To date, three reactions are known to be catalyzed by members of the MLE subgroup: the "paradigm" MLE reaction, the *o*-succinylbenzoate synthase (OSBS) reaction, and the L-Ala-D/L-Glu epimerase (AE Epim) reaction (Fig. 3). The current databases contain sequences for ~300 members of the MLE subgroup; ~50% of these are functionally assigned with confidence based on their similarities to characterized members of the subgroup.

The heterofunctional members of the MR subgroup contain a His-Asp dyad, in which the His is located at the end of the seventh β-strand and the Asp is located at the end of the sixth β-strand. In those members of the MR group for which structure/function relationships have been established, the His is the general base that abstracts the proton of the carbon acid substrate. The active sites usually contain an general acid catalyst that can be located at the end of the second, third, or fifth β-strand, depending on the identity and stereochemical course of the reaction that is catalyzed. As in the case of the MLE subgroup, the identities of metal ligands reinforce the identification of a member of the MR subgroup: the ligand at the end of the fifth β-strand is almost always a Glu preceded or followed by a residue other than Asx/Glx (the D-glucarate dehydratases are an exception in that the ligand at this position is an Asn). To date, five reactions are known to be catalyzed by members of the MR subgroup: the "paradigm" MR reaction, the D-glucarate/L-idarate dehydratase (GlucD) reaction, the D-altronate/D-mannonate dehydratase (AltD/ManD) reaction, the D-galactonate dehydratase (GalD) reaction, the D-gluconate dehydratase (GlcD) reaction, and the L-rhamnonate dehydratase (RhamD) reaction (Fig. 3). The current databases contain sequences for ~400 members of the MLE subgroup; ~50% of these are functionally assigned.

A fourth, less populated, highly diverged subgroup includes 3-methylaspartate ammonia lyase (MAL). The members of this subgroup contain a Lys at the end of the sixth β-strand that is likely the general base that abstracts the proton of the carbon acid substrate. This subgroup appears to contain members that catalyze at least two other reactions, although the identities of these have not been established (Fig. 3). The current databases contain sequences for ~20 members of the MAL subgroup; five are functionally unassigned.

Applying the superfamily paradigm: prediction of natural catalytic promiscuity by GlucD

Even before the discovery of the enolase superfamily, detailed structure–function relationships had been established for the MR-catalyzed reaction, the "paradigm"

member of the MR subgroup. These disclosed the importance of the His-Asp dyad as the (*R*)-specific base and a Lys (in a characteristic Lys-X-Lys motif) at the end of the second β-strand as the (*S*)-specific base in the 1,1-proton transfer reaction [21–25]. Subsequent work performed by Bearne and co-workers has further investigated the mechanism of the MR-catalyzed reaction [26–29], including quantitation of the reaction coordinate [30].

By both sequence comparisons and high resolution structural analyses, the active site of D-glucarate dehydratase (GlucD) is homologous to that of MR, i.e., the only potential acid/base catalysts are Lys 207 located in a Lys-X-Lys motif at the end of the second β-strand and the "required" His339-Asp319 dyad on the opposite face of the active site [10]. What are the functions of these groups in the GlucD active site?

The structures of liganded complexes of the GlucD from *Escherichia coli*, including both the product and the inert substrate analog 4-deoxy-D-glucarate, revealed the orientation of the substrate with respect to the active site functional groups [11] (Fig. 4). From these structures, together with the phenotypes of substitutions for the active site residues, we concluded that GlucD catalyzes the *syn*-dehydration of D-glucarate, with His339 functioning first as the base that abstracts the proton from carbon-5 to generate a Mg^{2+}-stabilized enolate anion intermediate and then, as its conjugate acid, facilitates the departure of the hydroxide leaving group from carbon-4 to generate an enol precursor to the final product [12] (Fig. 5). We also determined that the stereochemical course of the reaction is retention of relative configuration, i.e, the solvent-derived hydrogen is located in the same position as the hydroxide leaving group; the simplest explanation for this is that His339 also catalyzes the stereospecific tautomerization of the enol derived from dehydration to the 5-keto-4-deoxy product [31].

What is the functional role of Lys 207 at the end of the second β-strand? In analogy with the mechanism established for 1,1-proton transfer reaction catalyzed by MR,
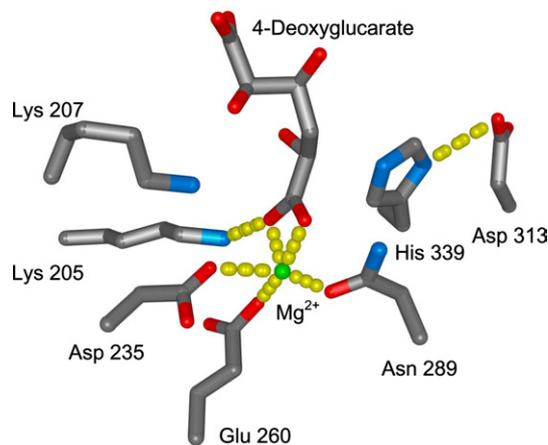

Fig. 4. The active site of D-glucarate dehydratase from *P. putida* in the presence of the inert substrate analog 4-deoxy-D-glucarate.
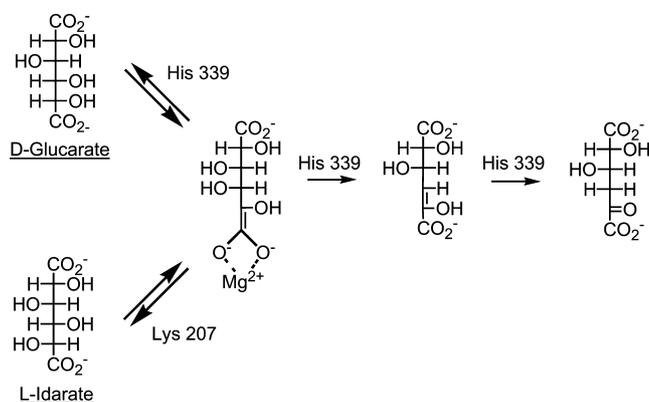
Fig. 5. The mechanisms of the reactions catalyzed by D-glucarate dehydratase.

we reasoned that Lys 207 might be able to function as the general base to abstract the proton from carbon-5 of the L-idarate, the 5-epimer of the D-glucarate, thereby allowing GlucD to generate the same Mg^{2+}-stabilized enolate anion intermediate as that derived from D-glucarate and catalyze the accidental *anti*-dehydration of L-idarate. Furthermore, we also predicted that GlucD might be able to catalyze the epimerization of D-glucarate/L-idarate via the shared Mg^{2+}-stabilized enolate anion intermediate in competition with dehydration. Both predictions were verified; in fact, GlucD is comparably efficient in dehydrating both L-idarate ($k_{cat} = 21 \text{ s}^{-1}$; $k_{cat}/K_m = 2.5 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$) and D-glucarate ($k_{cat} = 17 \text{ s}^{-1}$; $k_{cat}/K_m = 1.2 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$) and catalyzing their epimerization [32,33]. To the best of our knowledge, L-idarate is an unnatural diacid sugar (although the GlucD-catalyzed epimerization of D-glucarate that competes with dehydration does constitute a "formal" biosynthetic route to L-idarate), so we assume that this promiscuity is an in vitro "accident."

Thus, the GlucD active site is able to catalyze dehydration reactions of opposite stereochemical courses with nearly equivalent efficiencies: the Lys and His-Asp catalysts are able to abstract a proton from carbons of opposite configurations, and the His-Asp dyad catalyzes the departure of the hydroxide leaving group from carbons of identical configuration. So, the dehydration of D-glucarate proceeds with a *syn*-stereochemical course and that of L-idarate occurs with an *anti*-stereochemical course [31]. This conclusion invalidated the previous "dogma" that β-elimination reactions of carboxylate anion substrates necessarily proceed via an *anti*-stereochemical course.

Applying the superfamily paradigm: assignment of the D-galactonate dehydratase function to a member of the enolase superfamily

The *E. coli* genome encodes eight members of enolase superfamily, including enolase. Five are mem-

bers of the MR subgroup, and two are members of the MLE subgroup. In 1995, as the first partial sequences of the *E. coli* genome were deposited in the databases, only three members had known functions: enolase; OSBS, a member of the MLE subgroup whose gene had been sequenced by Meganathan and co-workers [34] in their characterization of the enzymes in the biosynthesis of menaquinone; and GlucD, a member of the MR subgroup and an orthologue of an enzyme of characterized function found in *P. putida*. The identities of the remaining five members of the superfamily were unknown.

The first application of our recognition that homologous proteins can catalyze different reactions that share a common partial reaction was the assignment of function to one of the five members of unknown function. Both then and now, independent information about the substrate specificity of an unknown protein is essential for assignment of function to an unknown member of the superfamily. Fortuitously, Cooper and co-workers had determined that the catabolism of D-galactonate is encoded by an operon mapped at 81.7 min on the *E. coli* chromosome [35]; the sequence of one of the unknown members of the MR subgroup of the enolase superfamily was located in approximately the same position. The steps in the catabolic pathway also had been elucidated: dehydration of D-galactonate produces 2-keto-3-deoxy-D-galactonate which, after conversion to the 6-phosphate, is cleaved to pyruvate and D-glyceraldehyde 3-phosphate [36]. The operon that had been sequenced at 81.7 min encoded the member of the enolase superfamily apparently fused to an aldolase (587 residues) as well as a kinase. Based on superfamily paradigm, we hypothesized that the 587 residue protein included D-galactonate dehydratase. As is often typical of protein sequences determined by genome projects, sequencing errors had been made, and the gene encoding the bifunctional protein actually encoded a 382 residue member of the enolase superfamily and a 205 residue aldolase. The purified 382 residue protein was demonstrated to be an efficient catalyst for the dehydration of D-galactonate ($k_{cat} = 3.5 \text{ s}^{-1}$ and $k_{cat}/K_m = 2.3 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$), thereby accomplishing the functional assignment as D-galactonate dehydratase (GalD) [37].

As a member of the MR subgroup, the sequence of GalD necessarily predicted presence of a His-Asp dyad; however, the sequence did not reveal the presence of Lys-X-Lys motif at the end of the second β-strand, even though a dehydration reaction would be expected to require general acid catalysis for departure of the hydroxide leaving group. The structure of a complex of GalD with L-threonohydroxamate, an analog of the enediolate intermediate, suggested that His 185, located at the end of the third β-strand, is the required general acid catalyst to facilitate departure of the 3-OH group in the *anti*-dehydration reaction (Fig. 6); that hypothesis was verified
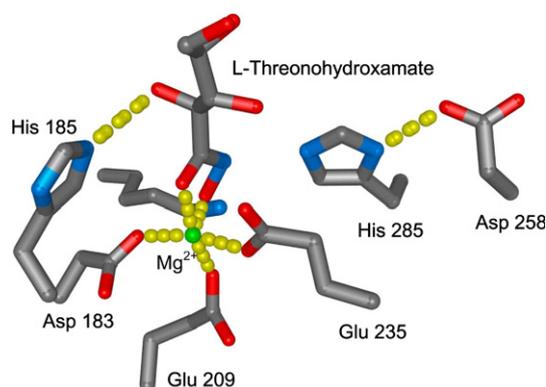
Fig. 6. The active site of D-galactonate dehydratase from *Escherichia coli* in the presence of the enolate anion intermediate analog L-threonohydroxamate.

by chemical rescue of the H185N and H185Q mutants with 3-deoxy-3-fluoro-D-galactonate, a substrate analog that does not require an acid catalyst for the β-elimination (in this case, dehydrofluorination) reaction [13].

Assignment of function to other acid sugar dehydratases

At this time, we know that other members of the MR subgroup can catalyze the dehydration of D-altronate, D-mannonate, D-gluconate, and L-rhamnonate (Fig. 3). As summarized below, the *E. coli* genome encodes L-rhamnonate dehydratase (RhamD) and a bifunctional D-altronate/D-mannonate dehydratase (AltD/ManD); D-gluconate dehydratase is not encoded by the *E. coli* genome but is encoded by other bacteria, including several species of archeae.

Although structure/function relationships for the reactions catalyzed by AltD/ManD, GlcD, and RhamD are still incomplete, we are confident that we understand how the differing stereochemical requirements for proton abstraction from carbon-2 and departure of the hydroxide leaving group from carbon-3 are accomplished (Fig. 3). In particular, examination of the structure of the (β/α)$_7$β-barrel domain of the structurally characterized members reveals that the acid/base catalysts and ligands for the essential Mg^{2+} are located at the ends of the various β-strands so that they surround the bound substrate and enolate anion intermediate that are sequestered from solvent by the capping domain [38]. In these structurally "independent" locations, mutational events in divergent evolution can add or delete functional groups as the demands for efficient catalysis of the "new" reaction are realized. For example, an *anti*-dehydration reaction requires a general base and a general acid on opposite faces of the active site, and a *syn*-dehydration requires a general base and a general acid on the same face of the active site.

In the case of the reaction catalyzed by the RhamD from *E. coli*, originally annotated by the genome project

as a protein of "unknown function" and more recently, as a "putative racemase," we were able to accomplish the assignment of function by screening a panel of mono- and diacid sugars as potential substrates. Two substrates were identified, L-rhamnonate and L-lyxonate [39], that share the same structures from carbon-1 to carbon-4, so it is not surprising that both are substrates. In the case of GlcD that is not encoded by the *E. coli* genome, the encoding genes in other bacteria are often found in operons that encode a D-glucose 1-dehydrogenase; in one case, the enzyme has been isolated and demonstrated to have GlcD activity [40].

Both the GlcD- and RhamD-catalyzed reactions require abstraction of a proton from a 2-carbon with the *R*-absolute configuration, but the leaving groups are positioned on three carbons of opposite absolute configuration. Based on the established structure/function relationships established for GalD, we expect that GlcD catalyzes an *anti*-dehydration and RhamD catalyzes a *syn*-dehydration. Our sequence analyses allow the prediction that for both GlcD and RhamD the dehydration reactions are initiated by abstraction of the proton from carbon-2 by the His-Asp dyad. In the case of GlcD, the sequence analyses reveal a conserved His following an uncommon Glu ligand for Mg^{2+} at the end of the third β-strand, so we expect that the *anti*-dehydration is catalyzed like that established for GalD. In the case of RhamD, the sequence analyses reveal two conserved His residues following the usual Glu ligand for Mg^{2+} at the end of the fifth β-strand; in this case, we expect that the His-Asp dyad is the general base that abstracts the proton from carbon-2 and one of the His residues at the end of the fifth β-strand is the general acid that facilitates departure of the leaving group. Perhaps the other His catalyzes stereospecific tautomerization of the enol generated by dehydration? Studies are in progress to investigate these predictions.

Although we now know that the *E. coli* genome encodes AltD/ManD (previously designated as "starvation sensing protein RspA [41]" with an unknown enzymatic function), the assignment of function was based on the operon context of the gene encoding an orthologue in *Novosphingobium aromaticivorans*. The operon context of the gene in *E. coli*, with a gene encoding a dehydrogenase, was insufficient to assign the identity of the likely substrate. In *N. aromaticivorans*, the gene encoding an orthologue of RspA is located in a cluster of genes encoding enzymes involved in the catabolism of D-glucuronate and D-galacturonate. In *E. coli*, that also catabolizes these hexuronates, parallel degradative pathways have been established that include distinct (and nonhomologous) enzymes to catalyze the dehydration of D-mannonate (from D-glucuronate) and the 3-epimeric D-altronate (from D-galacturonate), with both reactions yielding 2-keto-3-deoxy-D-gluconate. In *N. aromaticivorans*, the various enzymes producing this epi-

meric pair of monoacid sugars apparently have promiscuous substrate specificities, so that the analogous transformations required for catabolism of each hexuronate are catalyzed by the same enzyme. And, the genes encoding the canonical D-mannonate and D-altronate dehydratases are absent, but a gene encoding an orthologue of RspA is present in the hexuronate gene cluster.

We have determined that the orthologues from *E. coli* (RspA) and from *N. aromaticivorans* are, in fact, bifunctional dehydratases, utilizing both D-mannonate and D-altronate as substrates [42]. Again, based on the structure/function relationships we have established for GalD, we expect that the AltD reaction is an *anti*-dehydration, and the ManD reaction is a *syn*-dehydration. Interestingly, the sequence analyses reveal the presence the usual His-Asp dyad as well as two conserved His residues at the end of the third β-strand following the "first" Asp ligand for Mg^{2+}. As the absolute configurations of carbon-2 of the substrates are opposite to those of D-galactonate and D-gluconate, we hypothesize that one of the His residues at the end of the third β-strand is the general base that initiates both dehydration reactions. We also hypothesize that: (1) the other His at the end of the third β-strand is the general acid catalyst that facilitates the *syn* β-elimination of the 3-hydroxide leaving group from D-mannonate; and (2) the His-Asp dyad is the general acid catalyst that the *anti*-β-elimination of the 3-epimeric leaving group from D-altronate. Thus, in contrast to all of the other acid sugar dehydratases, our expectation is that the His-Asp dyad functions as the general acid and *not* the general base that initiates the reaction. Again, studies are in progress to investigate these predictions.

Thus, the stereochemical requirements of the *syn*- and *anti*-dehydration reactions catalyzed by the bifunctional AltD/ManD differ in significant detail from the *syn*- and *anti*-dehydration reactions catalyzed by GlucD: in the former case, the leaving groups depart from carbons of opposite configuration, and, in the latter case, the protons are abstracted from carbons of opposite configuration.

Based on these and the previously summarized studies, the four functionally characterized members of the MR subgroup encoded by the *E. coli* genome are GlucD, GalD, RhamD, and AltD/ManD. The function of the last remaining member of the MR subgroup (D-glucarate dehydratase-related protein or GlucDRP) is unknown and will be discussed in a later section.

The design of "new" reactions by altering reaction mechanism

Our work in this area is still underway, but we have begun to use Nature's design principles for catalyzing *syn*- and *anti*-dehydration reactions, i.e., adding or deleting acid/base catalysts at the ends of the appropriately located β-strands in the $(\beta/\alpha)_7\beta$-barrel domain to change the stereochemical course of a reaction. In particular, we have redesigned the active site of GlucD, that catalyzes the *syn*-dehydration of D-glucarate with the participation of only the His-Asp dyad, so that it can catalyze the *anti*-dehydration of galactarate. D-Glucarate and galactarate are epimers at carbon-4, the location of the leaving group, so the redesign required the introduction of an appropriately position general acid on the opposite face of the active site to allow the departure of the epimeric leaving group.

Taking a lesson from the GalD-catalyzed reaction, we installed a His at the end of the third β-strand in the GlucD from *E. coli* by constructing the N237H substitution (homologous to His 185 in GalD). This point mutant catalyzes the efficient dehydration of both D-glucarate and galactarate [42].

We note that our studies have both identified a natural example (AltD/ManD) and engineered an unnatural example (the N237H mutant of GlucD) of members of the MR subgroup that share the ability to catalyze *syn*- and *anti*-dehydration reactions from an epimeric pair of substrates that differ in the configuration of the carbon from which the leaving group departs.

Natural mechanistic promiscuity: an OSBS is also an *N*-acylamino acid racemase

The dynamic kinetic resolution of achiral *N*-acylamino acids with a coupled-enzyme system has the potential to serve as an environmentally friendly process for the commercial production of chiral amino acids [43]. One approach would utilize a specific acylase together with an *N*-acylamino acid racemase so that an achiral *N*-acylamino acid precursor can be converted quantitatively and irreversibly to a chiral amino acid. Both D- and L-acylases are available; the "problem" is access to an *N*-acylamino acid racemase. A few *N*-acylamino acids are known in metabolism, including *N*-acetyl-L-glutamate and *N*-acetylornithine in arginine biosynthesis and *N*-succinyl-L,L-diaminopimelate in lysine biosynthesis. However, the natural reactions involving these do not involve racemization of the α-carbon. Therefore, a large number of microorganisms have been screened for the presence of an enzymatic activity for racemization of *N*-acylamino acids.

Such an activity has been found in several strains of *Streptomyces*, gram positive microorganisms [44,45]. An "*N*-acylamino acid racemase" (NAAAR) was first purified from a strain of *Amycolatopsis*, and the encoding gene was cloned and sequenced [46]. The NAAAR is a member of the MLE subgroup of the enolase superfamily. Although not clearly disclosed in the original publications, the NAAAR is not a very efficient catalyst

for the racemization of its "best" substrate, N-acetylmethionine; we determined that the values of k_{cat}, K_m, and k_{cat}/K_m are 6.4 s^{-1}, 11 mM, and 590 M^{-1} s^{-1}, respectively [47]. The reason for the low activity is that NAAAR is not the natural reaction catalyzed by this protein; instead, the protein is an o-succinylbenzoate synthase (OSBS) as judged by the value of its "superior" kinetic constants for this reaction, 120 s^{-1}, 0.48 mM, and 2 × 10^5 M^{-1} s^{-1}, respectively, for k_{cat}, K_m, and k_{cat}/K_m.

OSBSs are found in many bacteria, including E. coli, and catalyze a step in the biosynthesis of menaquinone, an essential cofactor for anaerobic growth. The dehydration reaction catalyzed by OSBS is the "easiest" biochemical reaction initiated by abstraction of the α-proton of a carbon acid: the uncatalyzed rate is 1.6 × 10^{-10} s^{-1}, and the rate acceleration associated with the reaction catalyzed by the OSBS from E. coli is 2 × 10^{11} [48]. For comparison, the uncatalyzed rate for the racemization of mandelate is 3 × 10^{-13} s^{-1}, and the rate acceleration associated with the MR-catalyzed reaction is 1.5 × 10^{15} [26]. Although the value of the pK_a of the α-proton is unknown, the enhanced reactivity of the substrate for the OSBS reaction is likely the result of the exergonicity of the reaction and the accompanying early transition state that would be expected for proton abstraction.

In a structure of the wild-type OSBS from E. coli, the OSB product is sandwiched between Lys 133 and Lys 235, located at the ends of the second and sixth β-strands; one carboxylate oxygen of the substrate is coordinated to the essential Mg^{2+} [14]. The syn-stereochemical course of the dehydration reaction was established by the structure of the complex of the 2-succinyl-6-hydroxyl-2,4-cyclohexadiene-1-carboxylate (SHCHC) substrate with the inactive K133R mutant. In this structure, the 1-proton and 6-hydroxide leaving group are syn-substituents of the σ-bond, and the Arg133 group is located on the same face of the bond, implying that Lys133 is the single base/acid catalyst for the stepwise dehydration that involves the transient formation of a Mg^{2+}-coordinated enolate anion intermediate [15]. In both structures, Lys235 is in close proximity to the aromatic (wild type)/cyclohexadienyl (K133R) rings, suggesting that it assists in stabilization of the enolate anion intermediate by a π-cation interaction.

So, the NAAAR reaction catalyzed by the OSBS from Amycolatopsis is an example of functional promiscuity, arising because the hydrophobic substrate for the NAAAR reaction can bind in the same cavity as the substrate for the OSBS reaction and be properly positioned between Lys163 and Lys263, located at the ends of the second and sixth β-strands, respectively, so that the adventitious 1,1-proton transfer reaction can occur (Fig. 7). The mechanisms of both the natural OSBS reaction and the promiscuous NAAAR reaction are ini-
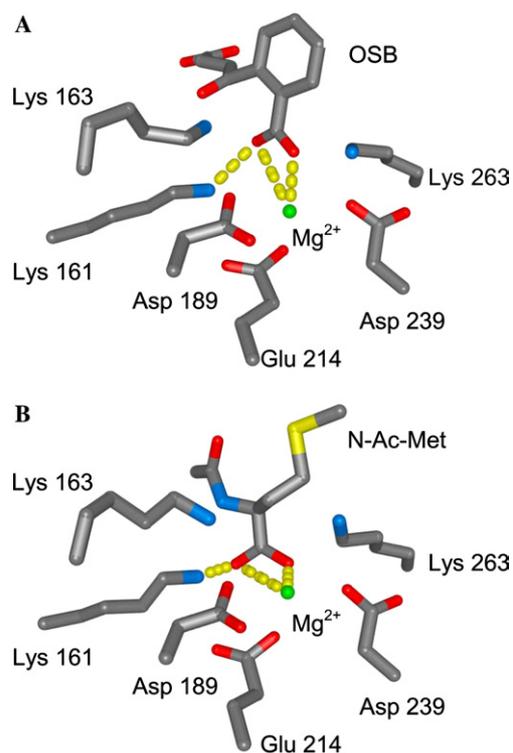

Fig. 7. The active site of the promiscuous OSBS/NAAAR from Amycolatopsis. (A) The presence of the o-succinylbenzoate product of the OSBS reaction; (B) the presence of the N-acetylmethionine substrate for the NAAAR reaction.

tiated by abstraction of the α-proton of the substrate to generate an enolate anion intermediate stabilized by coordination to the essential Mg^{2+} (Fig. 8).

The efficiency of the promiscuous NAAAR reaction could be enhanced by modifying the structures of N-acylamino acid substrate so that it more closely resembles the structure of the SHCHC substrate for the OSBS reaction. Indeed, using N-succinyl-L-phenylglycine as substrate, the values of k_{cat}, K_m, and k_{cat}/K_m are 190 s^{-1}, 0.95 mM, and 2 × 10^5 M^{-1} s^{-1}, respectively; these should be compared to those for the OSBS reaction, 120 s^{-1}, 0.48 mM, and 2.5 × 10^5 M^{-1} s^{-1}, respectively [49].

The structures of the complexes of the NAAAR/OSBS with OSB and the various N-acylamino acid substrates confirmed the expectation that the aromatic ring of the OSB product and the hydrophobic side chains of the N-acylamino acids occupied the same hydrophobic cavity; in addition, the conformations and interactions of the succinyl side chains of the OSB product and N-acylamino acids were similar, thereby providing a comprehensive structural explanation for the promiscuity [16].

The OSBS from E. coli does not catalyze a detectable NAAAR reaction, at least with the N-acylamino acid substrates for the promiscuous OSBS/NAAAR. Comparisons of the structures of the promiscuous OSBS/NAAAR with the monofunctional OSBS from E. coli suggest that the latter protein cannot be promiscuous
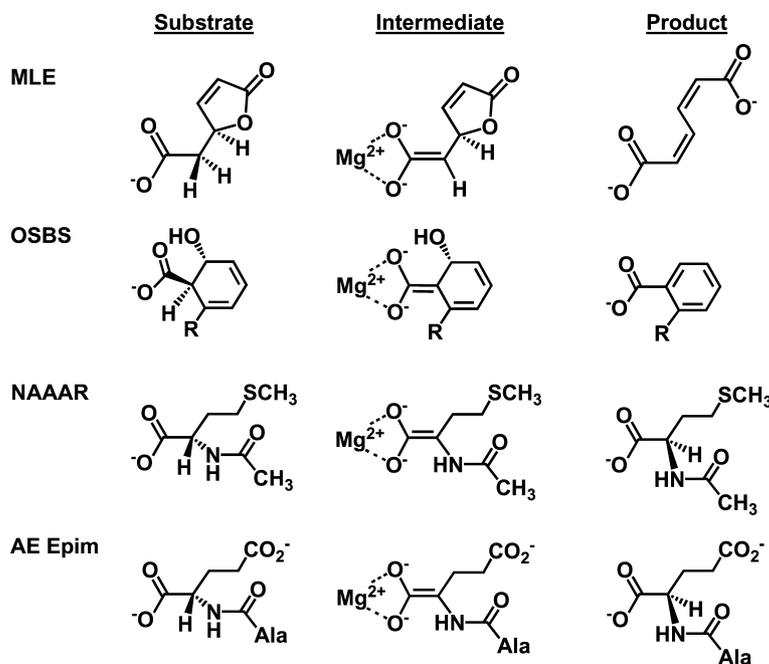
Fig. 8. The substrates, enolate anion intermediates, and products of the MLE, OSBS, NAAAR, and AE Epim reactions.

for the NAAAR reaction because the *N*-acyl group cannot form the same hydrogen bonding interactions with the active site due to steric interactions [16].

Assignment of function to L-Ala-D/L-Glu epimerases

The second member of MLE subgroup encoded by the *E. coli* genome was annotated as a "putative muconate cycloisomerase." Given the ability of the OSBS from *Amycolatopsis* to catalyze the promiscuous racemization of *N*-acylamino acids, we considered that this member of the MLE subgroup might catalyze a 1,1-proton transfer reaction as its natural reaction (Fig. 8). The gene encoding this protein is located in a cluster of genes, one of which had previously been assigned as encoding a periplasmic binding protein for the murein tripeptide, L-Ala-γ-D-Glu-meso-diaminopimelate, and another which encodes a protein distantly homologous to an endopeptidase from *Bacillus subtilis* that catalyzes the cleavage of the murein peptides. Accordingly, in our first experiment, conducted in an NMR tube, we determined that the unknown member of the MLE subgroup is L-Ala-D/L-Glu epimerase (AE Epim) that catalyzes the 1,1-proton transfer reaction that interconverts L-Ala-D-Glu and L-Ala-L-Glu, with the L-Ala-L-Glu presumably the substrate for a dipeptidase that allows either catabolism and/or recycling of the components of the murein peptide [50]. Subsequent studies revealed that an homologous member of the MLE subgroup encoded by the *B. subtilis* also catalyzes this reaction. Despite the identical functions, the enzymes share only 31% sequence identity. The substrate specificity for the epimerase from *E. coli* is much broader than

that from *B. subtilis*, but the kinetic constants determined for the both epimerases using L-Ala-D-Glu, the optimal substrate for both [*E. coli* ($k_{cat} = 10$ s^{-1}; $k_{cat}/K_m = 7.7 \times 10^4$ M^{-1} s^{-1}); *B. subtilis* ($k_{cat} =; 15$ s^{-1}; $k_{cat}/K_m = 4.7 \times 10^4$ M^{-1} s^{-1})], support this biochemical function.

Structures are available for both epimerases in the absence of a substrate or substrate analog, and these confirmed the expected locations of Lys residues at the ends of the second and sixth β-strands where they would catalyze the 1,1-proton transfer reactions [17]. More recently, the structure of the epimerase from *B. subtilis* with L-Ala-L-Glu bound in the active site was solved [51]. In this structure, the carboxylate group of the substrate is a bidentate ligand of the Mg^{2+}, and the α-carbon is located between Lys 162 and Lys 268 at the ends of the second and sixth β-strands. The pronounced specificity of this epimerase for L-Ala-D/L-Glu is determined by a hydrogen bond between the γ-carboxylate group of the Glu moiety of the substrate and the conserved Arg 24 in a flexible loop in the capping domain. The ammonium group of the substrate is hydrogen-bonded to both Asp 321 and Asp 323 at the end of the eighth β-strand.

With this functional assignment, seven of the eight members of the enolase superfamily encoded by the *E. coli* genome have been established: enolase, GlucD, GalD, RhamD, AltD/ManD, OSBS, and AE Epim.

The design of "new: reactions by altering the substrate specificity

To date, three reactions have been assigned to members of the MLE subgroup: MLE, OSBS, and AE Epim.

High-resolution structures are available for two of more examples of enzymes that catalyze these reactions: MLE [7–9]; OSBS [14–16], and AE Epim [17]. Each has the bidomain structure expected for a member of the enolase superfamily. In the $(\beta/\alpha)_7\beta$-barrel domain of each, the three strictly conserved ligands for the essential Mg^{2+} located at the ends of the third, fourth, and fifth β-strands and the Lys residues at the ends of the second and sixth β-strand are superimposable. However, the overall reactions they catalyze are different, albeit they share Lys-catalyzed abstraction of the α-proton of the substrate to yield a Mg^{2+}-stabilized enolate anion intermediate. How can superimposable active site functional groups catalyze different reactions?

Starting with the chloromuconate lactonizing enzyme (MLEII) from *Pseudomonas* sp. P51 that does not catalyze either the OSBS or AE Epim reaction and the AE Epim from *E. coli* that does not catalyze either the MLE or OSBS reaction, we have been able to identify a single active site mutation in each that allows catalysis of the OSBS reaction at the partial expense of the reaction catalyzed by the progenitors [20]. The OSBS activity was introduced into the MLEII scaffold by directed evolution, i.e., random mutagenesis followed by selection for a mutant gene that would allow anaerobic growth; the OSBS activity was introduced into the AE Epim scaffold by design, i.e., superposition of the active sites of the unliganded AE Epim and the product-liganded OSBS from *E. coli* suggested a rational, structure-based substitution.

Both approaches identified a Gly substitution for an acidic residue at the end of the eighth β-strand of the $(\beta/\alpha)_7\beta$-barrel domain: Glu 323 in the MLE II and Asp 297 in the AE Epim. Although a structure is not yet available for either mutant, the design of the OSBS activity in the AE Epim scaffold was based on the expectation that Asp 297 both sterically and electrostatically occludes the binding of the SHCHC substrate for the OSBS reaction to the wild-type enzyme; mutation to a Gly was predicted to remove the repulsive interaction. That the D297G possessed OSBS activity, albeit at a low level ($k_{cat} = 0.013$ s^{-1}; $k_{cat}/K_m = 7.4$ M^{-1} s^{-1}), supports the importance of the design strategy. Superposition of the structure of an unliganded MLE II with the product-liganded OSBS from *E. coli* suggests that the active site of the MLE also prevents binding of the SHCHC substrate. The Gly substitutions in both progenitors decreased the efficiency of the natural reaction, because the acidic residue at the end of the eighth β-strand is important for binding of the natural substrate.

Thus, we conclude that the active sites of the members of the MLE subgroup that share strictly conserved ligands for the essential Mg^{2+} located at the ends of the third, fourth, and fifth β-strands and Lys residues at the ends of the second and sixth β-strand are "hard-wired" for acid/base chemistry, including general base-cata-

lyzed enolization of the substrate and subsequent general acid-catalyzed conversion of the enolate anion intermediate to different products (intramolecular β-elimination of a carboxylate in the MLE reaction, dehydration in the case of the OSBS reaction, and protonation in the AE Epim reaction). The identity of the reaction is determined by the structure of the substrate that is "presented" to the conserved acid/base residues. Presumably, the natural evolution of a new function involves interdependent mutations that enhance specificity (decreasing K_m and, therefore, increasing k_{cat}/K_m) and catalysis (increasing k_{cat}), although a selective advantage would be accorded to promiscuous progenitors that accidentally catalyze the "new" reaction, e.g., the NAAAR reaction catalyzed by the OSBS from *Amycolatopsis*. Based on these results, we hypothesize that a single substitution and, therefore, a statistically probable evolutionary pathway, could have allowed the natural evolution of a 1,1-proton transfer function, e.g., AE Epim, from an OSBS progenitor.

Future directions: prediction of substrate specificity and, therefore, function

The physiological function of one member of MR subgroup encoded by the *E. coli* genome remains unknown. The protein is encoded by the same operon that encodes GlucD and shares 65% sequence identity with GlucD; we refer to this protein as GlucD-related protein or GlucDRP [52]. Given the high level of pair-wise sequence identity, the expectation was that GlucDRP is a redundant GlucD. However, our studies revealed that GlucDRP catalyzes the very inefficient dehydration of both D-glucarate and L-idarate, the comparably efficient and promiscuous substrates for GlucD; furthermore, the D-glucarate/L-idarate epimerase activity is also seriously compromised. So, what is the substrate for GlucDRP and what reaction does it catalyze?

In the context of the superfamily, the high level of sequence identity that relates GlucD and GlucDRP suggests that the natural substrate for GlucDRP is also a diacid sugar. One approach to solving the problem of identifying the substrate for GlucDRP, which is underway, is to prepare and screen a comprehensive library of mono- and diacid sugars, derived from all of the D- and L-hexoses, pentoses, and tetroses. A complementary approach that we also are exploring is *prediction* of the substrate specificity for GlucDRP by in silico docking the structures of the library into both experimentally determined structures of orthologous GlucDRPs as well as homology modeled structures based on the experimentally determined structure of GlucD from *E. coli* [53]. The expected requirement that the substrate bind in the active site with its carboxylate group coordinated to the essential Mg^{2+} and its α-proton juxtaposed to a

general base located at the end of a β-strand in the barrel domain is expected to restrict the identities of the potential substrates that can be productively accommodated in the active site cavity formed by residues in the flexible loops in the capping domain.

Although experimental screening of the library is certainly feasible, the parallel computational studies are expected to allow refinement of the algorithms for high resolution homology modeling and ligand docking so that the more general and important problem of assignment of function to more divergent, unknown members of the superfamily might be facilitated. Of course, the problem of functional assignment of unknowns is not limited to members of the enolase superfamily "discovered" in genome projects—it is a general and, arguably, one of the most important problems in genomic biology.

Conclusions

Our studies of the enolase superfamily started with detailed mechanistic and structural studies of the reaction catalyzed by MR. As the sequence and structure databases have become more populated, our studies have made the transition to simultaneous, synergistic structure/function studies of homologues that have accelerated our understanding of Nature's design principles for evolving new functions in the enolase superfamily. We now expect that our future studies of the enolase superfamily will provide the basis for both predicting of the functions of unknown members and engineering new functions so that novel reactions can be catalyzed.

Acknowledgment

References

[1] D.J. Neidhart, G.L. Kenyon, J.A. Gerlt, G.A. Petsko, Nature 347 (1990) 692–694.

[2] L. Lebioda, B. Stec, Nature 333 (1988) 683–686.

[3] J.E. Wedekind, G.H. Reed, I. Rayment, Biochemistry 34 (1995) 4325–4330.

[4] P.C. Babbitt, M.S. Hasson, J.E. Wedekind, D.R. Palmer, W.C. Barrett, G.H. Reed, I. Rayment, D. Ringe, G.L. Kenyon, J.A. Gerlt, Biochemistry 35 (1996) 16489–16501.

[5] T.M. Larsen, J.E. Wedekind, I. Rayment, G.H. Reed, Biochemistry 35 (1996) 4349–4358.

[6] D.J. Neidhart, P.L. Howell, G.A. Petsko, V.M. Powers, R.S. Li, G.L. Kenyon, J.A. Gerlt, Biochemistry 30 (1991) 9264–9273.

[7] S. Helin, P.C. Kahn, B.L. Guha, D.G. Mallows, A. Goldman, J. Mol. Biol. 254 (1995) 918–941.

[8] M.S. Hasson, I. Schlichting, J. Moulai, K. Taylor, W. Barrett, G.L. Kenyon, P.C. Babbitt, J.A. Gerlt, G.A. Petsko, D. Ringe, Proc. Natl. Acad. Sci. USA 95 (1998) 10396–10401.

[9] T. Kajander, L. Lehtio, M. Schlomann, A. Goldman, Protein Sci. 12 (2003) 1855–1864.

[10] A.M. Gulick, D.R. Palmer, P.C. Babbitt, J.A. Gerlt, I. Rayment, Biochemistry 37 (1998) 14358–14368.

[11] A.M. Gulick, B.K. Hubbard, J.A. Gerlt, I. Rayment, Biochemistry 39 (2000) 4590–4602.

[12] A.M. Gulick, B.K. Hubbard, J.A. Gerlt, I. Rayment, Biochemistry 40 (2001) 10054–10062.

[13] S.W. Wieczorek, K.A. Kalivoda, J.G. Clifton, D. Ringe, G.A. Petsko, J.A. Gerlt, J. Am. Chem. Soc. 121 (1999) 4540–4541.

[14] T.B. Thompson, J.B. Garrett, E.A. Taylor, R. Meganathan, J.A. Gerlt, I. Rayment, Biochemistry 39 (2000) 10662–10676.

[15] V.A. Klenchin, E.A. Taylor Ringia, J.A. Gerlt, I. Rayment, Biochemistry 42 (2003) 14427–14433.

[16] J.B. Thoden, E.A. Taylor Ringia, J.B. Garrett, J.A. Gerlt, H.M. Holden, I. Rayment, Biochemistry 43 (2004) 5716–5727.

[17] A.M. Gulick, D.M. Schmidt, J.A. Gerlt, I. Rayment, Biochemistry 40 (2001) 15716–15724.

[18] M. Asuncion, W. Blankenfeldt, J.N. Barlow, D. Gani, J.H. Naismith, J. Biol. Chem. 277 (2002) 8306–8311.

[19] C.W. Levy, P.A. Buckley, S. Sedelnikova, Y. Kato, Y. Asano, D.W. Rice, P.J. Baker, Structure (Camb.) 10 (2002) 105–113.

[20] D.M. Schmidt, E.C. Mundorff, M. Dojka, E. Bermudez, J.E. Ness, S. Govindarajan, P.C. Babbitt, J. Minshull, J.A. Gerlt, Biochemistry 42 (2003) 8387–8393.

[21] J.A. Landro, A.T. Kallarakal, S.C. Ransom, J.A. Gerlt, J.W. Kozarich, D.J. Neidhart, G.L. Kenyon, Biochemistry 30 (1991) 9274–9281.

[22] J.A. Landro, J.A. Gerlt, J.W. Kozarich, C.W. Koo, V.J. Shah, G.L. Kenyon, D.J. Neidhart, S. Fujita, G.A. Petsko, Biochemistry 33 (1994) 635–643.

[23] A.T. Kallarakal, B. Mitra, J.W. Kozarich, J.A. Gerlt, J.G. Clifton, G.A. Petsko, G.L. Kenyon, Biochemistry 34 (1995) 2788–2797.

[24] B. Mitra, A.T. Kallarakal, J.W. Kozarich, J.A. Gerlt, J.G. Clifton, G.A. Petsko, G.L. Kenyon, Biochemistry 34 (1995) 2777–2787.

[25] S.L. Schafer, W.C. Barrett, A.T. Kallarakal, B. Mitra, J.W. Kozarich, J.A. Gerlt, J.G. Clifton, G.A. Petsko, G.L. Kenyon, Biochemistry 35 (1996) 5662–5669.

[26] S.L. Bearne, R. Wolfenden, Biochemistry 36 (1997) 1646–1656.

[27] M. St Maurice, S.L. Bearne, Biochemistry 39 (2000) 13324–13335.

[28] M. St Maurice, S.L. Bearne, Biochemistry 43 (2004) 2524–2532.

[29] M. St Maurice, S.L. Bearne, W. Lu, S.D. Taylor, Bioorg. Med. Chem. Lett. 13 (2003) 2041–2044.

[30] M. St Maurice, S.L. Bearne, Biochemistry 41 (2002) 4048–4058.

[31] D.R. Palmer, S.W. Wieczorek, B.K. Hubbard, G.T. Mrachko, J.A. Gerlt, J. Am. Chem. Soc. 119 (1997) 9580–9581.

[32] D.R. Palmer, J.A. Gerlt, J. Am. Chem. Soc. 118 (1996) 10323–10324.

[33] D.R. Palmer, B.K. Hubbard, J.A. Gerlt, Biochemistry 37 (1998) 14350–14357.

[34] V. Sharma, R. Meganathan, M.E. Hudspeth, J. Bacteriol. 175 (1993) 4917–4921.

[35] R.A. Cooper, Arch. Microbiol. 118 (1978) 199–206.

[36] J. Deacon, R.A. Cooper, FEBS Lett. 77 (1977) 201–205.

[37] P.C. Babbitt, G.T. Mrachko, M.S. Hasson, G.W. Huisman, R. Kolter, D. Ringe, G.A. Petsko, G.L. Kenyon, J.A. Gerlt, Science 267 (1995) 1159–1161.

[38] P.C. Babbitt, J.A. Gerlt, J. Biol. Chem. 272 (1997) 30591–30594.

[39] B.K. Hubbard, J.A. Gerlt, unpublished observations.

[40] B. Siebers, B. Tjaden, K. Michalke, C. Dorr, H. Ahmed, M. Zaparty, P. Gordon, C.W. Sensen, A. Zibat, H.P. Klenk, S.C. Schuster, R. Hensel, J. Bacteriol. 186 (2004) 2179–2194.

[41] G.W. Huisman, R. Kolter, Science 265 (1994) 537–539.

[42] C. Millikin, J.A. Gerlt, unpublished observations.

[43] O. May, S. Verseck, A. Bommarius, K. Drauz, Org. Process Res. Dev. 6 (2002) 452–457.

[44] S. Tokuyama, K. Hatano, Appl. Microbiol. Biotechnol. 42 (1995) 853–859.

[45] S. Verseck, A. Bommarius, M.R. Kula, Appl. Microbiol. Biotechnol. 55 (2001) 354–361.

[46] S. Tokuyama, K. Hatano, Appl. Microbiol. Biotechnol. 42 (1995) 884–889.

[47] D.R. Palmer, J.B. Garrett, V. Sharma, R. Meganathan, P.C. Babbitt, J.A. Gerlt, Biochemistry 38 (1999) 4252–4258.

[48] E.A. Taylor, D.R. Palmer, J.A. Gerlt, J. Am. Chem. Soc. 123 (2001) 5824–5825.

[49] E.A. Taylor Ringia, J.B. Garrett, J.B. Thoden, H.M. Holden, I. Rayment, J.A. Gerlt, Biochemistry 43 (2004) 224–229.

[50] D.M. Schmidt, B.K. Hubbard, J.A. Gerlt, Biochemistry 40 (2001) 15707–15715.

[51] V.A. Klenchin, D.M. Schmidt, J.A. Gerlt, I. Rayment, Biochemistry 43 (2004) 10370–10378.

[52] B.K. Hubbard, M. Koch, D.R. Palmer, P.C. Babbitt, J.A. Gerlt, Biochemistry 37 (1998) 14369–14375.

[53] P.C. Babbitt, M. Jacobson, A. Sali, B. Schoichet, unpublished observations.