Understanding the Importance of Protein Structure to Nature's Routes for Divergent Evolution in TIM Barrel Enzymes

ERIC L. WISE AND IVAN RAYMENT*

Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706

Received October 14, 2003

ABSTRACT

It is widely agreed that new enzymes evolve from existing ones through the duplication of genes encoding existing enzymes followed by sequence divergence. While evolution is an inherently random process, studies of divergently related enzymes have shown that the evolution of new enzymes follows one of three general routes in which the substrate specificity, reaction mechanism, or active site architecture of the progenitor enzyme is reused in the new enzyme. Recent developments in structural biology relating to divergently related (β/α)8 enzymes have brought new insight into these processes and have revealed that conserved structural elements play an important role in divergent evolution. These studies have shown that, although evolution occurs as a series of random mutations, stable folds such as the (β/α)8 barrel and structural features of the active sites of enzymes are frequently reused in evolution and adapted for new catalytic purposes.

Introduction

The existence of proteins in Nature with sequence and structural similarity suggests that new enzyme functions arise from existing ones through divergent evolution.[1] This process is thought to require the duplication of the gene encoding the progenitor enzyme in order to produce a second copy of the progenitor enzyme so that the original metabolic function is not lost.[2,3] Through random mutations and genetic drift, the new enzyme evolves so that it is capable of catalyzing a reaction that is distinct from that catalyzed by the progenitor enzyme. While evolution is an entirely random process and the exact mechanisms by which it occurs in Nature are likely a product of the unique evolutionary demands of the environment, divergent enzyme evolution has been shown to follow one of three general routes in which the substrate specificity, reaction mechanism, or active site architecture of the progenitor enzyme is conserved and serves as a template for the evolution of new enzyme function.[1]

Eric Wise received a B.S. in chemistry from the University of Michigan in 1997, and is currently a graduate student in Dr. Rayment's group at the University of Wisconsin. His research focuses on using structural biology as a tool to better understand divergent enzyme evolution.

Ivan Rayment is currently a professor in the Biochemistry Department at the University of Wisconsin−Madison. He received his B.Sc and Ph.D. from the University of Durham, England, in 1972 and 1975, respectively. After postdoctoral study at Purdue University with Michael Rossmann and at Brandeis University with Don Caspar, he joined the faculty at the University of Arizona in 1984. He moved to the University of Wisconsin in 1988.

In the first of these three evolutionary routes, substrate specificity serves as the template for divergent evolution so that the progenitor enzyme and the newly evolved enzyme catalyze different reactions using chemically unrelated mechanisms but bind a common substrate. Evidence for this evolutionary route is provided by studies of homologous enzymes that bind a common substrate but catalyze unrelated reactions, such as the homologous enzymes that catalyze successive steps in the tryptophan and histidine biosynthetic pathways described in this Account.[4]

In the second route for divergent enzyme evolution, a conserved reaction mechanism serves as the template for divergent evolution. Evidence for this evolutionary route comes from studies of groups of homologous enzymes, known as enzyme superfamilies, which share a conserved reaction mechanism but catalyze different reactions on different substrates.[5] Over 20 enzyme superfamilies, including the enolase superfamily described herein, have been identified and are each comprised of homologous enzymes that share a common reaction mechanism.[6]

In the third route for divergent enzyme evolution, neither substrate specificity nor reaction mechanism is conserved. Instead, a conserved active site architecture is reused and serves as the template for divergent evolution. A group of enzymes that are homologous to orotidine monophosphate decarboxylase (OMPDC) form the only definite example of enzymes related in this way.[7] Although these enzymes share a conserved active site structure, each uses it to catalyze a reaction that is distinct in terms of substrate specificity and reaction mechanism.

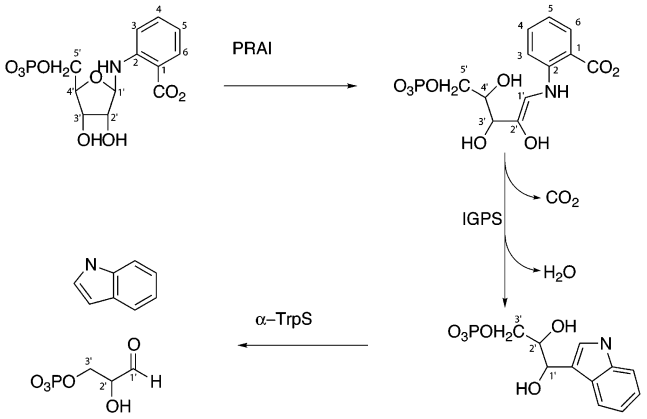
Because evolution is a slow and irregular process, the mechanisms of divergent evolution are difficult to study directly. From the study of divergently related enzymes, however, the routes through which enzymes have evolved in Nature may be inferred on the basis of common features that have been conserved in evolution. Structural studies, in particular, have been important for understanding enzyme evolution because structural similarity is often found in divergently related enzymes where there is little similarity in primary sequence. In highly divergent enzymes, this structural similarity may amount to only a few structurally conserved catalytic residues.[8] Nevertheless, these conserved structural elements may be critical for enzyme function, and understanding their roles in divergently related enzymes may reveal their evolutionary importance.

Recent crystallographic studies of divergently related (β/α)8 barrel enzymes that will be described in this Account have revealed the evolutionary importance of conserved structural elements to each of the three strategies for divergent evolution. These studies have focused on enzymes that adopt the (β/α)8 barrel fold because it appears to have been reused repeatedly in divergent

* Address correspondence to this author at Department of Biochemistry, 433 Babcock Dr., Madison, WI 53706. Phone: (608) 262-0437. Fax: (608) 262-1319. E-mail: Ivan_Rayment@biochem.wisc.edu.

10.1021/ar030250v CCC: $27.50
Published on Web 02/20/2004

Scheme 1. Reactions Catalyzed by PRAI, IGPD, and TrpS



evolution.[9] These investigations, however, may serve as models that further our understanding of the functional and structural requirements for the evolution of new enzymes throughout Nature.

The Tryptophan Biosynthetic Pathway: Divergent Evolution with a Conserved Substrate Specificity

One of the earliest attempts to identify the origins of new enzymes was made by Nathaniel Horowitz, who in 1945 proposed that biosynthetic pathways evolve in a backwards fashion such that each enzyme in a pathway evolves from a progenitor enzyme that catalyzes the next step in the pathway.[10] Accordingly, substrate specificity is conserved and has been proposed to serve as a template for divergent evolution because consecutive steps in a biosynthetic pathway must bind a common substrate, even if they catalyze mechanistically unrelated reactions.[11] While Horowitz's initial model has since been amended by many others, including Horowitz himself in 1965, examples of homologous enzymes that share a conserved substrate specificity have been identified in several biosynthetic pathways.[3,12−15]

One of the clearest examples of divergently related enzymes that share conserved substrate specificity involves three enzymes in the tryptophan biosynthetic pathway: phosphoribosyl anthrilate isomerase (PRAI), indole glycerol phosphate synthase (IGPS), and the α-subunit of the bifunctional tryptophan synthase (α-TrpS), Scheme 1. These enzymes, which catalyze subsequent reactions that constitute three of the last four steps in tryptophan biosynthesis, are homologous and share between 15% and 20% sequence identity.[16] Because they catalyze subsequent biosynthetic steps, each enzyme shares a common substrate specificity with the previous and subsequent enzymes in the pathway, although they catalyze mechanistically unrelated reactions, Scheme 1. Recent structural investigations of PRAI, IGPS, and α-TrpS have shed light on the role that conserved substrate specificity has played in the evolution of these three enzymes.[4,17,18] The X-ray crystal structures of PRAI and α-TrpS were each determined with their respective substrates, while the crystal structure of IGPS, which catalyzes

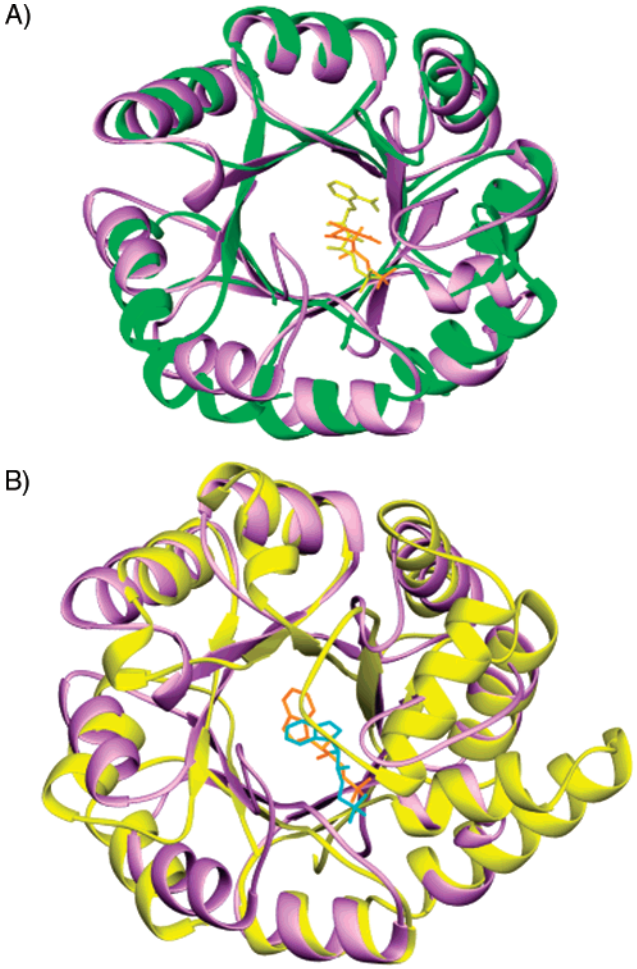


FIGURE 1. Structural alignments of (a) PRAI (green) and IGPS (purple) and (b) IGPS (purple) and α-TrpS (yellow). The ligand for the PRAI reaction is shown in yellow, the ligand for the IGPS reaction is shown in orange, and the ligand for the α-TrpS reaction is shown in blue.

the conversion of the PRAI reaction product into the substrate for α-TrpS, was determined with both a substrate analogue and the product bound.[18] These studies have revealed that all three enzymes share a similar $(\beta/\alpha)_8$-barrel fold, and the structures of each can be overlaid on top of one another with a root-mean-square deviation between 2.2 and 2.8 Å for approximately 140 paired Cα atoms, supporting the belief that all three enzymes share a common ancestor.[16]

The active sites of PRAI, IGPS, and α-TrpS are equally similar and are found, as in most $(\beta/\alpha)_8$ barrel enzymes, near the C-terminal ends of the barrel.[16] When cocrystal structures of PRAI and IGPS with their common substrate, 1-(*o*-carboxyphenylamino)-1-deoxyribulose-5-phosphate, are aligned, the substrate binds in essentially an identical manner in each active site (Figure 1). The phosphoribosyl group in particular is found in an almost identical position, and conserved interactions are formed between the phosphoribosyl group and backbone amide nitrogen atoms from structurally conserved loop regions of PRAI and IGPS that create a conserved substrate-binding site (Figure 2A). Similarly, when cocrystal structures of IGPS and α-TrpS with their common substrate,

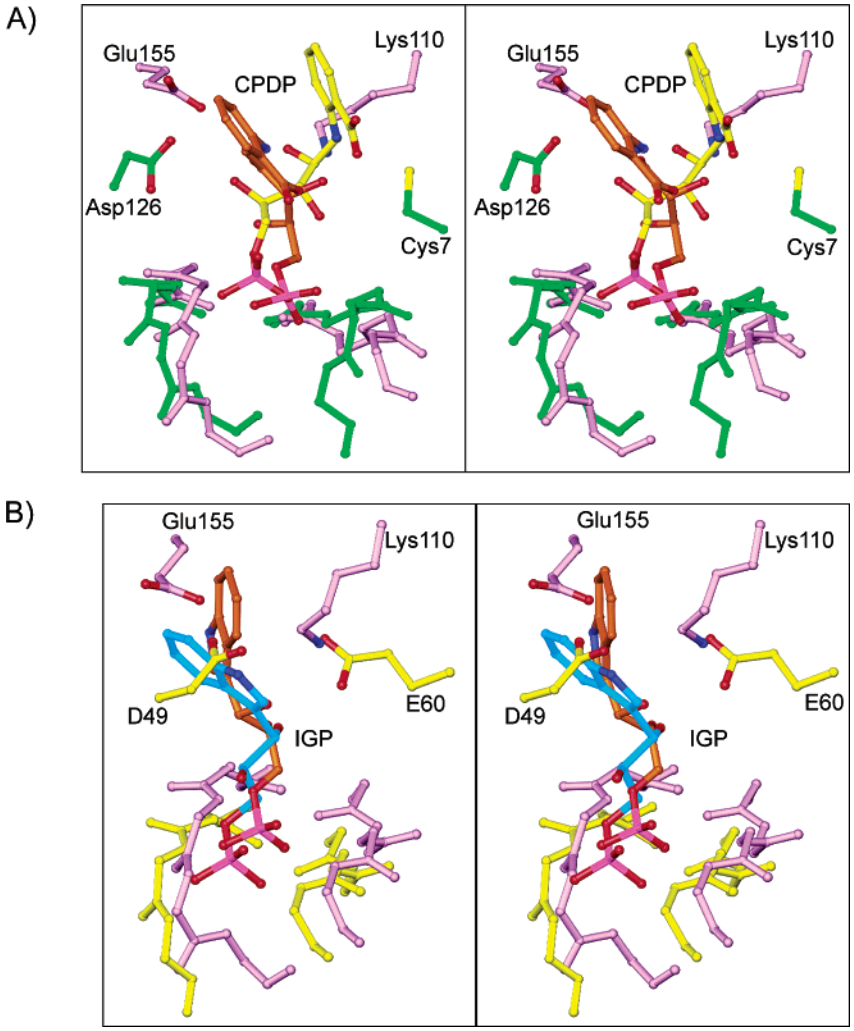


FIGURE 2. Stereo closeup views of the active sites of (a) PRAI (green) and IGPS (purple) overlayed and (b) IGPS (purple) and α-TrpS (yellow) overlayed. A conserved substrate binding site allows each enzyme to bind as a substrate the product of the previous reaction in the pathway. The catalytic residues, which are labeled for each enzyme, are not conserved, so each enzyme may catalyze different reactions.

indole glycerol 3-phosphate, are aligned, this substrate binds in an almost identical manner in the structurally conserved substrate-binding site (Figure 2B). In all three enzymes, carbons C3 through C5 and the phosphate group of each substrate, which are not involved in catalysis, interact with structurally conserved loops in the active sites of each enzyme. These conserved interactions allow each enzyme to bind the substrate of the next enzyme in the pathway and serves as the structural basis for the conserved substrate specificity.

Although PRAI, IGPS, and α-TrpS share common substrate specificities, the reaction mechanisms and the identities of catalytic active site residues are unique to each enzyme. PRAI catalyzes an Amadori rearrangement that results in the opening of the ribose sugar ring.[19] The mechanism of this reaction involves either the formation of a Schiff base or simple acid−base catalysis. While this ambiguity has yet to be resolved, Asp126 and Cys7 have been identified as likely candidates for general acid and general base.[4]

The IGPS-catalyzed reaction, which results in indole formation, proceeds via an entirely different mechanism from those of PRAI or α-TrpS and makes use of a completely different set of catalytic residues. The most plausible mechanism for the IGPS reaction involves a condensation reaction between carbon C1 of the phenyl ring and carbon C2′ of the ribulose moiety, followed by successive decarboxylation and dehydration reactions to yield the aromatic product. All available structural and biochemical evidence suggests that Glu159 acts as the general base and Lys110 acts as the general acid to catalyze the dehydration reaction.[18,20]

The α-TrpS-catalyzed reaction results in the cleavage of IGP to yield D-glyceraldehyde 3-phosphate and indole. Indole is channeled to the β subunit of TrpS for the final step in tryptophan biosynthesis. The reaction mechanism for the α-TrpS reaction is unrelated to those of the PRAI- and IGPS-catalyzed reactions and is thought to involve tautomerization to form an indolenine, followed by carbon−carbon bond cleavage in which Glu49 and Asp60 are thought to be involved in acid−base catalysis.[21]

The variability in the identities and functions of the catalytic residues in PRAI, IGPS, and α-TrpS exemplifies the flexibility of the $(\beta/\alpha)_8$ fold for enzyme evolution. While

these enzymes share common structural features that confer substrate specificity, each has evolved to use a unique set of catalytic residues to catalyze completely different reactions. Other examples of divergently related enzymes that share substrate specificity include, most notably, two enzymes in the histidine biosynthetic pathway, HisA and HisF.[22] While the lack of structural information for either of these enzymes has made it difficult to understand their evolutionary relationships, it is known that, like PRAI, IGPS, and α-TrpS, these two enzymes share sequence identity, structural similarity, and a common substrate.[22,23] In addition, HisA and HisF also share internal sequence identity and structural similarity between the N- and C-terminal halves of the protein, which has led to speculation that the $(\beta/\alpha)_8$ barrel fold initially evolved from the fusion of two $(\beta/\alpha)_4$ units.[23,24] Further investigations will be required to decipher the evolutionary relationships between these two enzymes.

The Enolase Superfamily: Divergent Evolution with a Conserved Chemical Mechanism

The examples of the homologous $(\beta/\alpha)_8$ barrel enzymes in the tryptophan and histidine biosynthetic pathways demonstrate how substrate specificity may serve as a template for divergent evolution and be reused in the evolution of new enzyme functions. A second route for divergent enzyme evolution, evidence for which is found in several enzyme superfamilies, sharply contrasts this route in that the progenitor enzyme and the newly evolved enzyme do not necessarily share any common substrate specificity. Instead, the progenitor enzyme and the newly evolved enzyme share a conserved mechanistic feature that is used in each reaction mechanism, even though the overall reactions may be unrelated.[25] Examples of enzyme superfamilies that are formed from enzymes that share a conserved mechanistic feature include the amidohydrolase/phosphotriesterase superfamily, the vicinal-oxygen-chelate superfamily, the crotonase superfamily, and the enolase superfamily.[5,26−28]

The enolase superfamily is of particular interest because its members have been extensively characterized both structurally and functionally and because it was the first example of an enzyme superfamily to be described.[29] Like other enzyme superfamilies, the overall reactions catalyzed by members of the enolase superfamily vary greatly, but each involves the abstraction of a proton α to a carboxylate group of the substrate to generate a Mg^{2+} ion-stabilized enolate anion intermediate.[5] The enolase superfamily was initially identified when it was discovered that mandelate racemase (MR) and muconate lactonizing enzyme (MLE) were homologous and used this common reaction step, even though they catalyzed unrelated reactions on different substrates, Scheme 2.[29] Both MR and MLE, which share approximately 25% sequence identity, adopt similar two-domain folds that consist of a central $(\beta/\alpha)_8$ barrel domain and an $\alpha+\beta$ domain that is formed from the N- and C-termini of the protein (Figure 3).[5] MR catalyzes the racemization of (*R*)-mandelate to (*S*)-mandelate; MLE catalyzes the cycloisomerization of *cis,cis*-muconate.[29,30] This common mechanistic step was conserved through divergent evolution and served as the template for the evolution of new enzyme functions within the enolase superfamily.[5] Other members of the enolase superfamily that catalyze reactions that involve an enolate

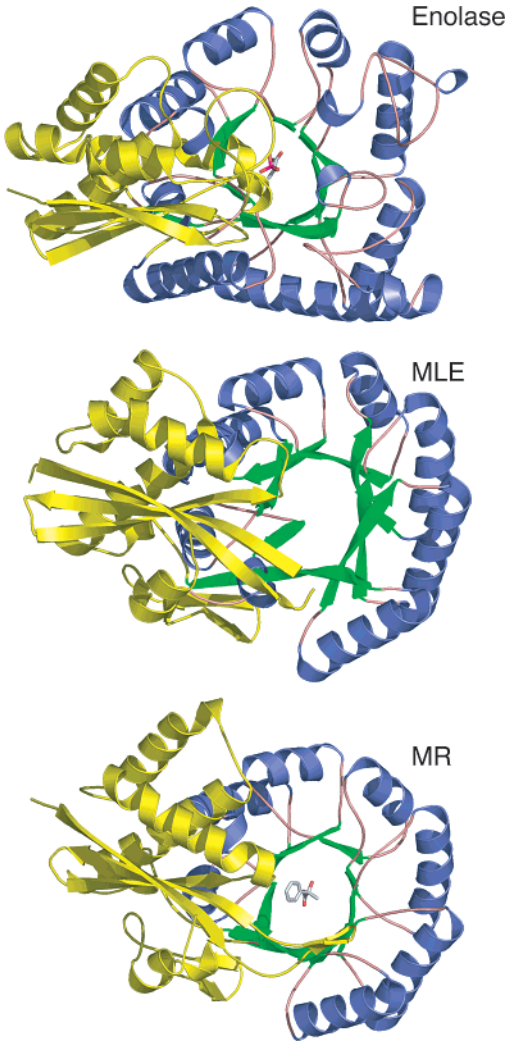
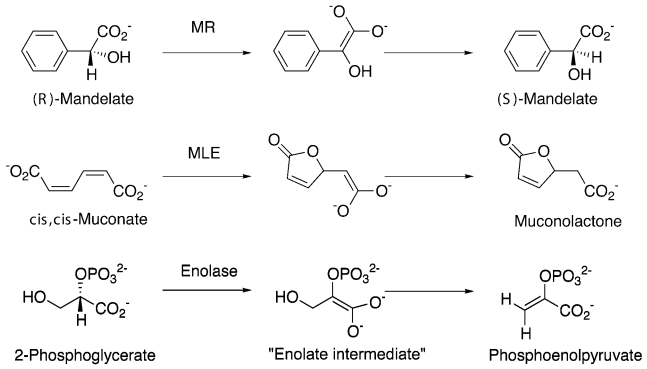


FIGURE 3. Structures of MLE, MR, and enolase. The enolase superfamily fold, which is found in all known members, consists of a central $(\beta/\alpha)_8$ domain (green and blue) and a separate $\alpha+\beta$ domain (yellow) that is constructed from both N- and C-teminal regions of the protein.

Scheme 2. Reactions Catalyzed by Mandelate Racemase (MR), Muconate Lactonizing Enzyme (MLE), and Enolase



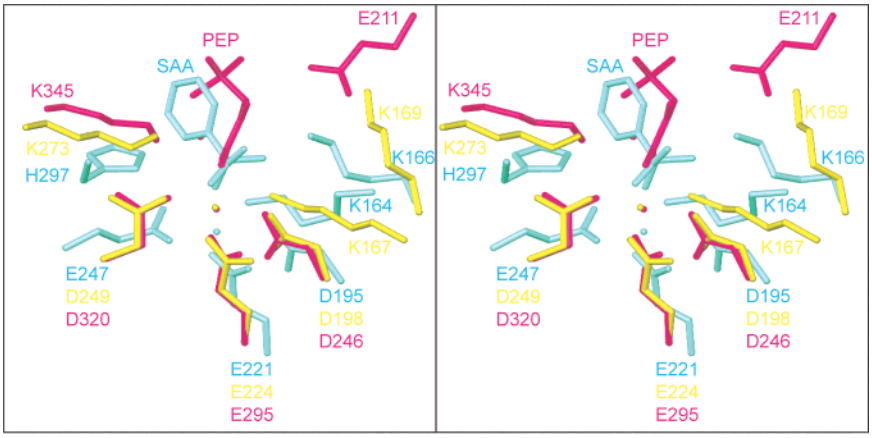


FIGURE 4. Overlay of the active sites of MLE (yellow), MR (blue), and enolase (magenta). Three conserved carboxylate side chains function as ligands for the catalytic Mg^{2+} ion. The identities of the catalytic acids/bases as well as other important active site residues vary within each subgroup.

intermediate have since been identified and shown to possess structural and sequence similarity to MR and MLE, including enolase, from which the enolase superfamily derives its name.[31,32] Structural alignments of the active sites of enolase superfamily members have revealed that a trio of conserved aspartate and glutamate residues is found in a structurally conserved arrangement (Figure 4).[30,32,33] These residues, located on the C-terminal ends of the third, fourth, and fifth β strands of the $(\beta/\alpha)_8$ barrel domain, function as ligands for a catalytic metal ion.[5] This metal ion, which is typically a Mg^{2+} ion, is critical for enzyme function and acts to stabilize the negatively charged enolate intermediate through electrostatic interactions. The three-dimensional arrangement of the carboxylate groups of these conserved residues is highly conserved so as to place the metal ion in a position that allows it to interact with the oxygen atoms of the carboxylate group of the enolate intermediate in each reaction.

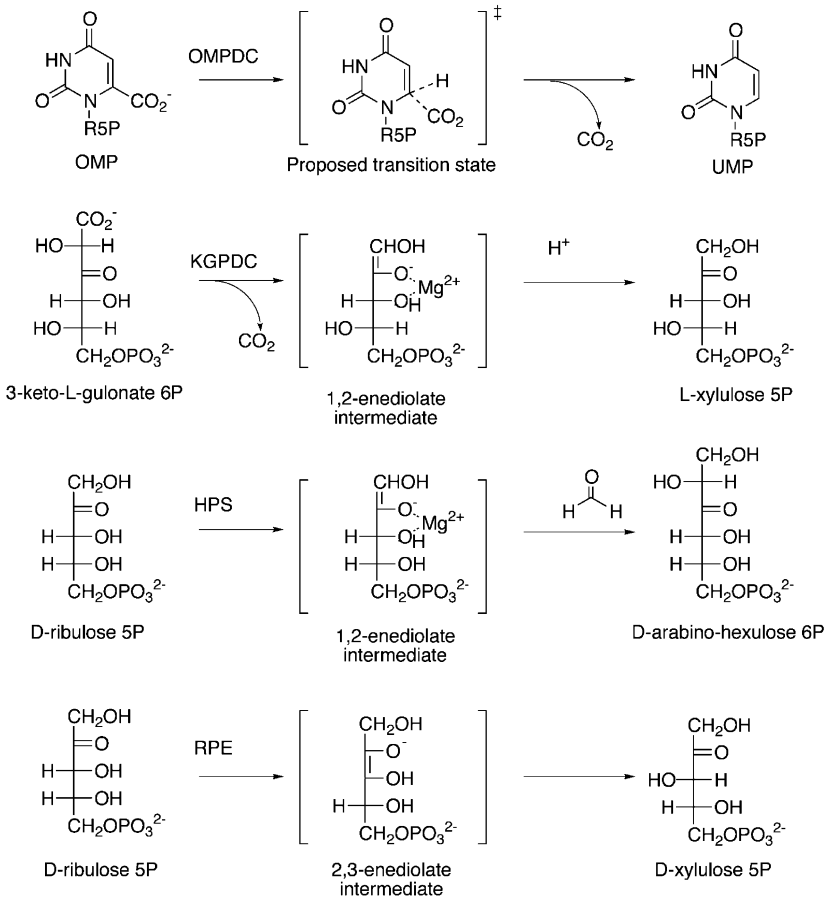
Conserved residues that function as the acid/base catalysts are also found in the active sites of members of the enolase superfamily.[5] The identities of these residues are not strictly conserved, however, but rather fall into one of three groups. As a result, the enolase superfamily has been subdivided into three subgroups on the basis of similarity to MR, MLE, or enolase, depending on which residues function as the catalytic acids/bases.[34] In the MR subgroup, conserved lysine and histidine residues, homologous to Lys166 and His297 in MR, assume this function. In the MLE subgroup, two lysine residues, homologous to Lys167 and Lys273 in MLE, function as the general base(s). In the enolase subgroup, a single conserved lysine, homologous to Lys345 in enolase, functions as the general base.[5] While there are considerable variations in the identities of the acid/base catalysts, their locations in the active site are roughly conserved so that acid/base catalysis may occur from either face of the planar enolate intermediate formed in the active site of each enzyme.

Like the homologous enzymes in the tryptophan biosynthetic pathway, the enolase superfamily has made use of the versatility of the $(\beta/\alpha)_8$ barrel fold to catalyze a wide variety of reactions. Eleven reactions have been identified in the enolase superfamily, including the cycloisomerization reaction catalyzed by MLE, the 1,1-proton transfer reactions catalyzed by MR, OSBS, and L-Ala-D/L-Glu epimerases, and dehydration and deamination reactions catalyzed by RhamD and MAL.[1] Unlike the homologous enzymes in the tryptophan biosynthetic pathway, there is no common mode of substrate binding or conserved residues involved in substrate specificity, which is not surprising since substrate specificity is not conserved within the enolase superfamily. Substrate specificity is thought to be conferred upon individual members of the enolase superfamily by residues in the $\alpha + \beta$ domain, in which few conserved residues are found, that "caps" the C-terminal end of the barrel.[1] Thus, new reactions have evolved that do not share any common substrate specificity but make use of the enolase active site to catalyze reactions that make use of an enolate intermediate.

The OMPDC Suprafamily: Divergent Evolution with a Conserved Active Site Architecture

In the examples of the homologous tryptophan biosynthetic enzymes and the enolase superfamily, a common feature of the reaction catalyzed by divergently related enzymes—either substrate specificity or reaction mechanism—is reused as a template for divergent evolution. In both cases, structural elements in the active site that confer one of these features upon divergently related enzymes are conserved. Recent investigations into a group of enzymes with sequence similarity to orotidine monophosphate decarboxylase (OMPDC) have provided evidence for a third route for the evolution of new enzymes in Nature in which, unlike the previous two examples, neither substrate specificity nor reaction mechanism is conserved.[7] Instead, the overall structure of the active site itself is conserved, which suggests that this overall struc-

Scheme 3. Reactions Catalyzed by OMPDC, KGPDC, HPS, and RPE[a]



[a] The proposed transition state for OMPDC and the proposed intermediates for KGPDC, HPS, and RPE are shown.

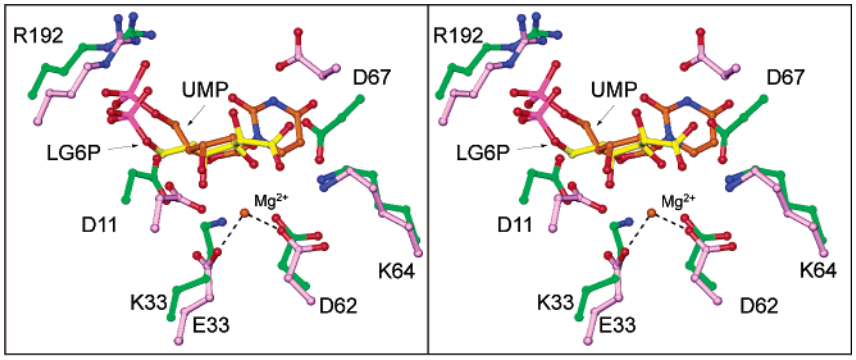


FIGURE 5. Overlay of the active sites of KGPDC (purple) and OMPDC (green). The active sites of KGPDC and OMPDC are highly similar, despite the fact that they catalyze unrelated reactions and bind different substrates. Conserved active site residues found in each enzyme are used for different functions.

ture, rather than a particular reaction feature, was reused as a template for divergent evolution in these enzymes.[1]

Structural and biochemical investigations into OMPDC and 3-keto-L-gulonate 6-phosphate decarboxylase (KGPDC) have revealed that OMPDC and KGPDC are homologous enzymes that arose from a common ancestral enzyme through divergent evolution but share neither a common substrate specificity nor a common reaction mechanism. OMPDC catalyzes the metal ion-independent decarboxylation of orotidine 5′-monophosphate (OMP) to uridine 5′-monophosphate (UMP); KGPDC catalyzes the metal ion-dependent decarboxylation of 3-keto-L-gulonate 6-phosphate to L-xylulose 5-phosphate (Scheme 3).[35,36] Structural alignments of recently reported X-ray crystal structures of four OMPDCs and the KGPDC encoded by the UlaD gene in *Escherichia coli* have revealed that the active sites of KGPDC and OMPDC are remarkably similar, even though KGPDC and OMPDC share only around 20% sequence identity and lack any similarity in substrate identity or reaction mechanism (Figure 5).[7,37−40] In both enzymes, two identical active sites are created at a conserved interface of two $(\beta/\alpha)_8$ barrels in a homo-dimer, which are related to one another by a conserved two-fold rotational axis (Figure 6). Despite the relatively low

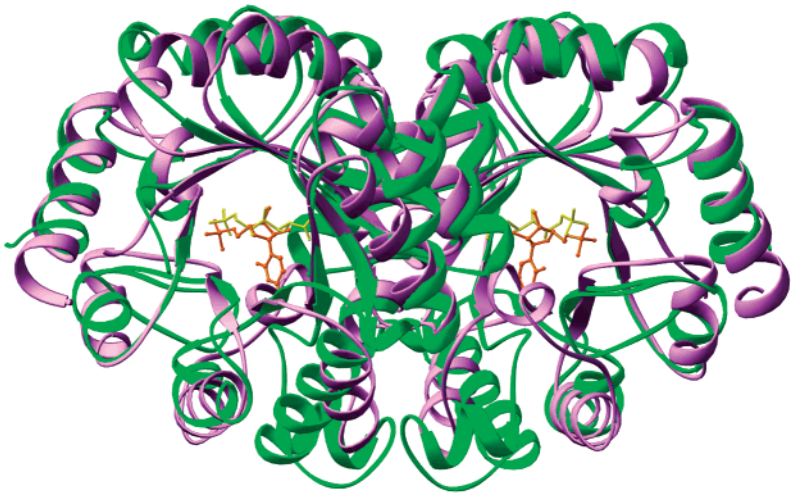


FIGURE 6. Structural alignment of KGPDC (purple) with bound L-gulonate 6-phosphate (yellow) and OMPDC (green) with bound UMP (orange). Although they share limited sequence identity, both enzymes adopt a conserved $(\beta/\alpha)_8$ barrel fold. The quaternary relationship between the two individual subunits in the dimer is highly conserved as well.

sequence similarity shared between KGPDC and OMPDC, the relative three-dimensional positions of conserved active site residues, particularly those in an Asp-X-Lys-X-X-Asp motif, are nearly identical in each enzyme. This motif spans both active sites in the homo-dimer so that the first aspartate and the lysine are found in one active site in the dimer and the second aspartate is found in the other active site in the dimer.[7] In addition, while the substrates for each enzyme are chemically unrelated except for a common phosphate group, they also are found in nearly identical positions in each active site.

Much controversy still exists about the roles of these conserved residues in the OMPDC-catalyzed reaction, due in part to the incredible rate enhancement achieved by the enzyme so that the ratio of k_{cat}/k_{uncat} exceeds 10^{17}.[41] Although the exact mechanism for the enzyme-catalyzed reaction is unclear, the reaction has been shown to proceed without the use of a metal ion or cofactor.[42] Several competing mechanisms for the OMPDC reaction have been proposed, for which there is little consensus on the details of the reaction. A common but not universally held belief is that the driving force behind the reaction is ground-state destabilization created by repulsive interactions between the negatively charged carboxyl leaving group of OMP and the negatively charged side chain of Asp60.[40,43] The high energy of binding created by these energetically unfavorable interactions has been proposed to be offset by a large number of energetically favorable interactions between the polar functional groups of the substrate and a subset of polar functional groups in the enzyme active site.[44] It is generally agreed that the side chain of the conserved Lys64 functions as the proton source and transfers a proton to carbon C5 of OMP.[37] The mechanism of the KGPDC-catalyzed reaction is better understood and is completely unrelated to the mechanism of the OMPDC-catalyzed reaction. The KGPDC-catalyzed reaction almost certainly proceeds through an enediolate intermediate and, unlike the OMPDC reaction, requires a metal ion, which acts to stabilize a 1,2-enediolate intermediate through electrostatic interactions with the negatively charged oxygen on carbon C2.[45]

The same conserved active site residues in KGPDC and OMPDC are used by each enzyme to catalyze their respective reactions, although the two reactions bear no resemblance in terms of substrate identity or mechanism and the roles of these residues in each reaction are completely different.[1,7] In OMPDC, the side chain Asp60 functions to provide ground-state destabilization through interactions with OMP.[41] In KGPDC this residue instead functions as a ligand for the catalytic Mg^{2+} ion. In OMPDC, Asp67 contributes to the extensive network of hydrogen bonds between the enzyme and the substrate and interacts with the ribose ring, away from the site of catalysis.[44] The side chain of Lys64 acts as the proton source in the OMPDC-catalyzed reaction to add a proton to the C5 position of OMP following decarboxylation.[37] In the KGPDC reaction, Lys64 and Asp67 act to orient and stabilize the intermediate through interactions with the oxygen on carbon C1. The proton source the KGPDC-catalyzed reaction appears to be unrelated to that of the OMPDC reaction and likely falls upon two activated water molecules in the active site. These two water molecules are located on the *re* and *si* faces of the enediolate intermediate, so either could act as the general acid and act as proton shuttles to transfer protons from the side chains of an arginine and a histidine in the active site.[46−47]

Given the vast differences in reaction mechanism and substrate specificity between KGPDC and OMPDC, the conservation of structural features between these two enzymes, in particular those of the enzyme active sites, is quite remarkable. Rather than a conserved substrate specificity or reaction mechanism, OMPDC and KGPDC share a conserved overall active site architecture that has been reused by Nature to perform unrelated functions. Enzymes related in this way are said to form a mechanistically diverse enzyme *supra*family. The OMPDC suprafamily is the first and only definite example to date of an enzyme suprafamily.[1]

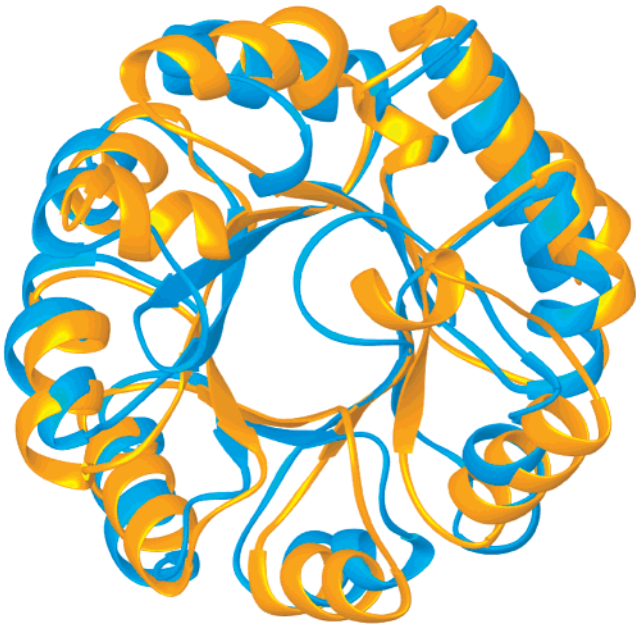


FIGURE 7. Structural alignment of KGPDC (blue) and RPE (gold). Although they share only limited sequence identity, both enzymes adopt similar $(\beta/\alpha)_8$ barrel folds.

Two additional potential members of the OMPDC suprafamily have been identified on the basis of sequence similarity to KGPDC and OMPDC. Hexulose phosphate synthase (HPS), which shares several conserved active site residues with OMPDC and KGPDC, including the conserved Asp-X-Lys-X-X-Asp motif, appears to be homologous to OMPDC and KGPDC.[48] HPS catalyzes the metal ion-dependent aldol condensation of D-ribulose 5-phosphate and formaldehyde to form D-*arabino*-3-hexulose 6-phosphate. The mechanism of HPS reaction is unrelated to that of OMPDC but is similar to that of KGPDC in that it likely involves the formation of an enediolate intermediate.[49] This similarity to the KGPDC reaction suggests that HPS and KGPDC may form an enzyme superfamily within the OMPDC suprafamily, which would imply that membership in an enzyme superfamily and membership in an enzyme *supra*family are not mutually exclusive.[1]

Further blurring the definition of an enzyme suprafamily is another, more distantly related homologue of OMPDC and KGPDC, D-ribulose 5-phosphate 3-epimerase (RPE). RPE catalyzes the interconversion of D-xylulose 5-phosphate and D-ribulose 5-phosphate and may also be a member of the OMPDC suprfamily.[50] RPE has slight sequence and structural similarity to KGPDC and OMPDC, although the identities of most active site residues, including those in the Asp-X-Lys-X-X-Asp motif, are not conserved.[1,50] While X-ray crystal structures of two RPEs have been reported, neither was solved with a bound ligand, and the exact mechanism of the RPE reaction, as well as the relationship between RPE and other members of the OMPDC suprafamily, is not clear.[51,52]

An overlay of the structures of RPE from rice and KGPDC, however, still clearly demonstrates that structural features are conserved between the two enzymes (Figure 7). The locations of the proposed active sites of both enzymes overlay almost exactly, and a bound sulfate ion in the RPE active site, which is proposed to occupy the same location as the phosphate group of the substrate, is found in nearly the exact same position as the phosphate group of L-gulonate 6-phosphate in KGPDC. The identities of several active site residues are conserved among RPE, OMPDC, and KGPDC, although the side chains in the RPE active site are not conserved. Even so, the overall active site geometry is largely conserved from RPE to other members of the OMPDC suprafamily (Figure 8).

The example of the OMPDC suprafamily, like the tryptophan biosynthetic enzymes and the enolase superfamily, highlights the versatility of the $(\beta/\alpha)_8$ barrel fold for the evolution of new enzyme functions. Clearly a deeper knowledge of the mechanisms of the HPS- and RPE-catalyzed reactions will be required to better understand the relationships among enzymes in the OMPDC suprafamily and to fully appreciate the role that active site architecture may play in divergent enzyme evolution.

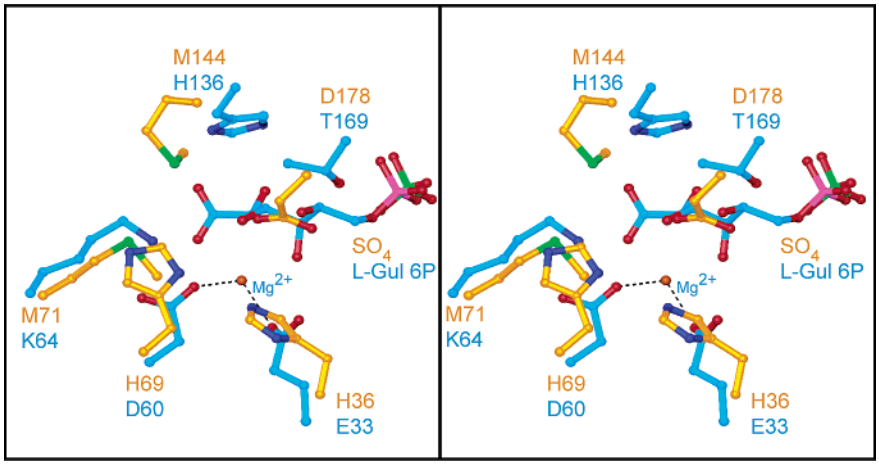


FIGURE 8. Overlay of the active sites of KGPDC (blue) and RPE (yellow). Although the side chains of catalytic residues are not conserved, their positions in the active site are.

Conclusions

The presence of structural similarity in the overall folds and active sites of divergently related enzymes suggests that new enzymes evolve by reusing structural features of enzymes and adapting them for new purposes. The examples of the homologous tryptophan biosynthetic enzymes, the enolase superfamily enzymes, and the OMPDC suprafamily enzymes demonstrate the versatility of conserved protein structural features in divergent evolution. These and other examples of divergently related enzymes may serve as paradigms for understanding the structural and functional requirements for the evolution of new enzymes, and may ultimately lead to a greater comprehension of the extensive diversity in protein structure and function in Nature.

This work was supported in part by the NIH Grants GM-52594 (to J. A. Gerlt and I.R.) and GM-65155 (to J. A. Gerlt and I.R.) from the National Institutes of Health. E.L.W. was supported by the NIH Biophysics Training Grant GM08293.

References

(1) Gerlt, J. A.; Babbitt, P. C. Divergent Evolution of Enzymatic Function: Mechanistically Diverse Superfamilies and Functionally Distinct Suprafamilies. *Annu. Rev. Biochem.* **2001**, *70*, 209−246.

(2) O'Brien, P. J.; Herschlag, D. Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* **1999**, *6*, R91−R105.

(3) Jensen, R. A. Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **1976**, *30*, 409−425.

(4) Henn-Sax, M.; Thoma, R.; Schmidt, S.; Hennig, M.; Kirschner, K.; Sterner, R. Two (betaalpha)(8)-barrel enzymes of histidine and tryptophan biosynthesis have similar reaction mechanisms and common strategies for protecting their labile substrates. *Biochemistry* **2002**, *41*, 12032−12042.

(5) 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.; Gerlt, J. A. The enolase superfamily: a general strategy for enzyme-catalyzed abstraction of the alpha-protons of carboxylic acids. *Biochemistry* **1996**, *35*, 16489−16501.

(6) Todd, A. E.; Orengo, C. A.; Thornton, J. M. Evolution of protein function, from a structural perspective. *Curr. Opin. Chem. Biol.* **1999**, *3*, 548−556.

(7) Wise, E.; Yew, W. S.; Babbitt, P. C.; Gerlt, J. A.; Rayment, I. Homologous $(\beta/\alpha)_8$-barrel enzymes that catalyze unrelated reactions: orotidine 5′-monophosphate decarboxylase and 3-keto-L-gulonate 6-phosphate decarboxylase. *Biochemistry* **2002**, *41*, 3861−3869.

(8) Bartlett, G. J.; Borkakoti, N.; Thornton, J. M. Catalysing new reactions during evolution: economy of residues and mechanism. *J. Mol. Biol.* **2003**, *331*, 829−860.

(9) Nagano, N.; Orengo, C. A.; Thornton, J. M. One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J. Mol. Biol.* **2002**, *321*, 741−765.

(10) Horowitz, N. H. On the evolution of biochemical syntheses. *Proc. Natl. Acad. Sci. U.S.A.* **1945**, *31*, 153−157.

(11) Gerlt, J. A.; Raushel, F. M. Evolution of function in (beta/alpha)-(8)-barrel enzymes. *Curr. Opin. Chem. Biol.* **2003**, *7*, 252−264.

(12) Gerlt, J. A. New wine from old barrels. *Nat. Struct. Biol.* **2000**, *7*, 171−173.

(13) Lazcano, A.; Miller, S. L. On the origin of metabolic pathways. *J. Mol. Evol.* **1999**, *49*, 424−431.

(14) Ycas, M. On earlier states of the biochemical system. *J. Theor. Biol.* **1974**, *44*, 145−160.

(15) Horowitz, N. H. The evolution of biochemical syntheses-retrospect and prospect. In *Evolving Genes and Proteins*; Bryson, V., Vogel, H. J., Eds.; Academic Press: New York, 1965; pp 15−23.

(16) Wilmanns, M.; Hyde, C. C.; Davies, D. R.; Kirschner, K.; Jansonius, J. N. Structural conservation in parallel beta/alpha-barrel enzymes that catalyze three sequential reactions in the pathway of tryptophan biosynthesis. *Biochemistry* **1991**, *30*, 9161−9169.

(17) Weyand, M.; Schlichting, I. Crystal structure of wild-type tryptophan synthase complexed with the natural substrate indole-3-glycerol phosphate. *Biochemistry* **1999**, *38*, 16469−16480.

(18) Hennig, M.; Darimont, B. D.; Jansonius, J. N.; Kirschner, K. The catalytic mechanism of indole-3-glycerol phosphate synthase: crystal structures of complexes of the enzyme from Sulfolobus solfataricus with substrate analogue, substrate, and product. *J. Mol. Biol.* **2002**, *319*, 757−766.

(19) Hommel, U.; Eberhard, M.; Kirschner, K. Phosphoribosyl anthranilate isomerase catalyzes a reversible amadori reaction. *Biochemistry* **1995**, *34*, 5429−5439.

(20) Darimont, B.; Stehlin, C.; Szadkowski, H.; Kirschner, K. Mutational analysis of the active site of indoleglycerol phosphate synthase from Escherichia coli. *Protein Sci* **1998**, *7*, 1221−1232.

(21) Rhee, S.; Miles, E. W.; Davies, D. R. Cryo-crystallography of a true substrate, indole-3-glycerol phosphate, bound to a mutant (alphaD60N) tryptophan synthase alpha2beta2 complex reveals the correct orientation of active site alphaGlu49. *J. Biol. Chem.* **1998**, *273*, 8553−8555.

(22) Fani, R.; Lio, P.; Chiarelli, I.; Bazzicalupo, M. The evolution of the histidine biosynthetic genes in prokaryotes: a common ancestor for the hisA and hisF genes. *J. Mol. Evol.* **1994**, *38*, 489−495.

(23) Lang, D.; Thoma, R.; Henn-Sax, M.; Sterner, R.; Wilmanns, M. Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion. *Science* **2000**, *289*, 1546−1550.

(24) Hocker, B.; Beismann-Driemeyer, S.; Hettwer, S.; Lustig, A.; Sterner, R. Dissection of a (betaalpha)8-barrel enzyme into 2-folded halves. *Nat. Struct. Biol.* **2001**, *8*, 32−36.

(25) Babbitt, P. C.; Gerlt, J. A. Understanding enzyme superfamilies. Chemistry As the fundamental determinant in the evolution of new catalytic activities. *J. Biol. Chem.* **1997**, *272*, 30591−30594.

(26) Holm, L.; Sander, C. An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. *Proteins* **1997**, *28*, 72−82.

(27) Armstrong, R. N. Mechanistic diversity in a metalloenzyme superfamily. *Biochemistry* **2000**, *39*, 13625−13632.

(28) Holden, H. M.; Benning, M. M.; Haller, T.; Gerlt, J. A. The crotonase superfamily: divergently related enzymes that catalyze different reactions involving acyl coenzyme a thioesters. *Acc. Chem. Res.* **2001**, *34*, 145−157.

(29) Neidhart, D. J.; Kenyon, G. L.; Gerlt, J. A.; Petsko, G. A. Mandelate racemase and muconate lactonizing enzyme are mechanistically distinct and structurally homologous. *Nature* **1990**, *347*, 692−694.

(30) Helin, S.; Kahn, P. C.; Guha, B. L.; Mallows, D. G.; Goldman, A. The refined X-ray structure of muconate lactonizing enzyme from Pseudomonas putida PRS2000 at 1.85 A resolution. *J. Mol. Biol.* **1995**, *254*, 918−941.

(31) Schmidt, D. M.; Hubbard, B. K.; Gerlt, J. A. 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* **2001**, *40*, 15707−15715.

(32) Larsen, T. M.; Wedekind, J. E.; Rayment, I.; Reed, G. H. A carboxylate oxygen of the substrate bridges the magnesium ions at the active site of enolase: structure of the yeast enzyme complexed with the equilibrium mixture of 2-phosphoglycerate and phosphoenolpyruvate at 1.8 A resolution. *Biochemistry* **1996**, *35*, 4349−4358.

(33) Landro, J. A.; Gerlt, J. A.; Kozarich, J. W.; Koo, C. W.; Shah, V. J.; Kenyon, G. L.; Neidhart, D. J.; Fujita, S.; Petsko, G. A. The role of lysine 166 in the mechanism of mandelate racemase from Pseudomonas putida: mechanistic and crystallographic evidence for stereospecific alkylation by (R)-alpha-phenylglycidate. *Biochemistry* **1994**, *33*, 635−643.

(34) Gulick, A. M.; Hubbard, B. K.; Gerlt, J. A.; Rayment, I. 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* **2001**, *40*, 10054−10062.

(35) Miller, B. G.; Smiley, J. A.; Short, S. A.; Wolfenden, R. Activity of yeast orotidine-5′-phosphate decarboxylase in the absence of metals. *J. Biol. Chem.* **1999**, *274*, 23841−23843.

(36) Yew, W. S.; Gerlt, J. A. Utilization of L-ascorbate by Escherichia coli K-12: assignments of functions to products of the yjf-sga and yia-sgb operons. *J. Bacteriol.* **2002**, *184*, 302−306.

(37) Appleby, T. C.; Kinsland, C.; Begley, T. P.; Ealick, S. E. The crystal structure and mechanism of orotidine 5′-monophosphate decarboxylase. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2005−2010.
(38) Harris, P.; Navarro Poulsen, J. C.; Jensen, K. F.; Larsen, S. Structural basis for the catalytic mechanism of a proficient enzyme: orotidine 5′-monophosphate decarboxylase. *Biochemistry* **2000**, *39*, 4217−4224.
(39) Miller, B. G.; Hassell, A. M.; Wolfenden, R.; Milburn, M. V.; Short, S. A. Anatomy of a proficient enzyme: the structure of orotidine 5′- monophosphate decarboxylase in the presence and absence of a potential transition state analog. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2011−2016.
(40) Wu, N.; Mo, Y.; Gao, J.; Pai, E. F. Electrostatic stress in catalysis: structure and mechanism of the enzyme orotidine monophosphate decarboxylase. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2017−2022.
(41) Miller, B. G.; Wolfenden, R. Catalytic proficiency: the unusual case of OMP decarboxylase. *Annu. Rev. Biochem.* **2002**, *71*, 847−885.
(42) Begley, T. P.; Appleby, T. C.; Ealick, S. E. The structural basis for the remarkable catalytic proficiency of orotidine 5′-monophosphate decarboxylase. *Curr. Opin. Struct. Biol.* **2000**, *10*, 711−718.
(43) Miller, B. G.; Snider, M. J.; Short, S. A.; Wolfenden, R. Contribution of enzyme-phosphoribosyl contacts to catalysis by orotidine 5′-phosphate decarboxylase. *Biochemistry* **2000**, *39*, 8113−8118.
(44) Miller, B. G.; Snider, M. J.; Wolfenden, R.; Short, S. A. Dissecting a charged network at the active site of orotidine-5′-phosphate decarboxylase. *J. Biol. Chem.* **2001**, *276*, 15174−15176.
(45) Wise, E.; Yew, W. S.; Gerlt, J. A.; Rayment, I. Structural Evidence for an Enediolate Intermediate in the Reaction Catalyzed by 3-ket-L-gulonate 6-phosphate decarboxylase, a member of the OMPDC Suprafamily. *Biochemistry* 2003, *42*, 12133−12142.
(46) Wise, E.; Yew, W. S.; Gerlt, J. A.; Rayment, I. Results not yet published.
(47) Yew, W. S.; Wise, E.; Rayment, I.; Gerlt, J. A. Results not yet published, 2003.
(48) Traut, T. W.; Temple, B. R. The chemistry of the reaction determines the invariant amino acids during the evolution and divergence of orotidine 5′-monophosphate decarboxylase. *J. Biol. Chem.* **2000**, *275*, 28675−28681.
(49) Sahm, H.; Schutte, H.; Kula, M. R. Purification and properties of 3-hexulosephosphate synthase from Methylomonas M 15. *Eur. J. Biochem.* **1976**, *66*, 591−596.
(50) Chen, Y. R.; Hartman, F. C.; Lu, T. Y.; Larimer, F. W. D-Ribulose-5-phosphate 3-epimerase: cloning and heterologous expression of the spinach gene, and purification and characterization of the recombinant enzyme. *Plant Physiol.* **1998**, *118*, 199−207.
(51) Jelakovic, S.; Kopriva, S.; Suss, K. H.; Schulz, G. E. Structure and catalytic mechanism of the cytosolic D-ribulose-5-phosphate 3-epimerase from rice. *J. Mol. Biol.* **2003**, *326*, 127−135.
(52) Kopp, J.; Kopriva, S.; Suss, K. H.; Schulz, G. E. Structure and mechanism of the amphibolic enzyme D-ribulose-5-phosphate 3-epimerase from potato chloroplasts. *J. Mol. Biol.* **1999**, *287*, 761−771.

AR030250V