Annu. Rev. Biochem. 2001. 70:209–46

DIVERGENT EVOLUTION OF ENZYMATIC FUNCTION: Mechanistically Diverse Superfamilies and Functionally Distinct Suprafamilies

John A. Gerlt[1] and Patricia C. Babbitt[2]
[1]*Departments of Biochemistry and Chemistry, University of Illinois, Urbana, Illinois 61801; e-mail: j-gerlt@uiuc.edu*
[2]*Department of Biopharmaceutical Sciences, University of California, San Francisco, California 94143; e-mail: babbitt@cgl.ucsf.edu*

Key Words enolase, amidohydrolase, thiyl radical, crotonase, (β/α)-barrels

■ **Abstract** The protein sequence and structure databases are now sufficiently representative that strategies nature uses to evolve new catalytic functions can be identified. Groups of divergently related enzymes whose members catalyze different reactions but share a common partial reaction, intermediate, or transition state (mechanistically diverse superfamilies) have been discovered, including the enolase, amidohydrolase, thiyl radical, crotonase, vicinal-oxygen-chelate, and Fe-dependent oxidase superfamilies. Other groups of divergently related enzymes whose members catalyze different overall reactions that do not share a common mechanistic strategy (functionally distinct *supra*families) have also been identified: (*a*) functionally distinct suprafamilies whose members catalyze successive transformations in the tryptophan and histidine biosynthetic pathways and (*b*) functionally distinct suprafamilies whose members catalyze different reactions in different metabolic pathways. An understanding of the structural bases for the catalytic diversity observed in super- and suprafamilies may provide the basis for discovering the functions of proteins and enzymes in new genomes as well as provide guidance for in vitro evolution/engineering of new enzymes.

CONTENTS

INTRODUCTION

As the numbers of both genome sequences and high-resolution protein structures increase, enzymologists can better decipher and understand nature's strategies for both (*a*) achieving the remarkable catalytic proficiencies of enzymes and (*b*) evolving new catalysts from the relatively small number of protein folds available. Until the early 1990s, enzymologists had little choice but to focus their studies on single examples of specific enzymes. Now, a much larger informational context is available, allowing enzymologists to include the genomic context (sequence families, structures, and functions) relevant to study of their favorite enzyme, rather than describing single-enzyme phenomenology. We suggest the term *genomic enzymology* to describe this expansive strategy for understanding the structural bases for catalysis as well as the design principles that can be used to develop catalysts for new reactions. Unlike approaches that focus primarily on structural comparisons without regard to function or on functional analysis independent of the structural context, the practice of genomic enzymology requires understanding the interplay of structure and function.

A seminal development in genomic enzymology was the realization that enzymes that are related by divergent evolution can catalyze different overall reactions. In 1990, the sequences and structures of muconate lactonizing enzyme and mandelate racemase were discovered to be homologous despite their distinct reactions in the degradation of mandelate by *Pseudomonads* (1). This discovery has focused attention on both elucidating the strategies for divergent evolution of enzyme function and ascertaining the structural bases for functional plasticity, with the expectation that a description of nature's approaches might allow the directed redesign/evolution of existing enzymes to catalyze different, and even unnatural, reactions. In this review, we discuss several groups of functionally diverse/distinct enzymes that are related by divergent evolution.

GLOSSARY FOR DIVERGENT EVOLUTION OF FUNCTION

We begin by examining some of the definitions that are used to describe relationships in sequence, structure, and function.

Homologs, Orthologs, Paralogs, and Analogs

Homologous enzymes (*homologs*) derive from a common ancestor and are therefore structurally related. Frequently, such enzymes show a high degree of sequence similarity easily identifiable from simple database searches and sufficient for inference of function by analogy. Homologous enzymes can also be highly divergent, requiring more sophisticated search strategies for reliable identification of highly diverged sequences (2–4). In other cases, homologous enzymes cannot be identified from sequence information alone but require comparisons of three-dimensional structures. At high levels of divergence, homologous enzymes often do not catalyze the same chemical reaction. Furthermore, conserved sequence elements need not be located in the active site but may be responsible for maintenance or stabilization of common folds that support different reaction mechanisms, e.g. the triose phosphate isomerase (TIM)-barrels. Some insights that can be obtained from investigation of superfamilies of highly divergent homologous enzymes are discussed in this review.

Other terms that are frequently used to described homologous enzymes include *orthologs*, which are homologs in different species that catalyze the same reaction (5), and *paralogs*, homologous enzymes in the same species that likely diverged from one another by gene duplication after speciation.

Analogous enzymes (*analogs*) catalyze the same reaction but are not structurally related. Recently, a systematic study of the Enzyme Commission (EC) nomenclature has shown that many such structurally unrelated analogs are classified with the same EC numbers, leading to substantial confusion in interpreting structure-function relationships in these cases (6). Analogous enzymes are not relevant to genomic enzymology and are not discussed in this review.

Structure-Based Definition of Superfamily

The term *superfamily* is commonly used in both structural and functional contexts; unfortunately, it can have a different meaning in each. These distinctions must be clarified, and we suggest that the correlated structure-function-based definition we describe here be used in both contexts, or that the term be explicitly described as based on structure-only or on function-only information.

Several strategies have been devised for structure-only-based classification of protein structures, notably Structural Classification of Proteins (SCOP; http://scop.mrc-lmb.cam.ac.uk/scop/) (7) and Class-Architecture-Topology-Homologous superfamily (CATH; http://www.biochem.ucl.ac.uk/bsm/cath/) (8, 9). These

provide tools for clustering protein structures that have evolved from a common progenitor, with that evolutionary process retaining discernible structural similarities even if the sequence identity/similarity is insignificant. Although efforts have been described for explicitly incorporating functional properties into these schemes by identifying, for example, consensus residues from sequence alignments as well as structural properties, these semiautomatic methods do not always organize the protein universe into groups of proteins that define the scope of divergent evolution or that describe a structural strategy explicitly associated with defined functional outcomes. The problem is that algorithms for comparing function are difficult to devise and implement, in part because no structurally contextual definitions of enzyme function exist. The only large-scale computationally accessible definition scheme for enzyme function that is available, the EC system, is not reliable, because divergently related enzymes can catalyze different overall reactions (the subject of this review). Functional relationships based on consensus sequences or even motifs are also unreliable, in the absence of structure-function correlations that describe their functional roles.

The definition of superfamily used in SCOP is "proteins [that] have [the same fold and] low sequence identities but whose structures and, in many cases, functional features suggest that a common evolutionary origin is probable." With this definition, enzymes often are grouped according to substrate specificity but without regard to conservation of active site functional groups that mediate the chemical transformation; as a result, this definition of superfamily may partition groups of divergently related enzymes into separate superfamilies. Classifications that fail to include such chemistry-based relationships obscure nature's capabilities and prevent us from fully exploiting divergent evolution for discerning nature's strategies for evolving new catalysts.

Structurally Contextual Function-Based Definitions

We propose and use the following definitions to describe the divergent evolution of enzyme function:

Family: Group of homologous (frequently orthologous) enzymes that catalyze the same reaction (mechanism and substrate specificity). Often, the members of families share greater than 30% sequence identity; however, orthologs may show sequence identities well below 30% and therefore can be difficult to discern in the absence of structural information for at least one member of the family.

Superfamily: Groups of homologous enzymes that catalyze either (*a*) the same chemical reaction with differing substrate specificities or (*b*) different overall reactions that share a common mechanistic attribute (partial reaction, intermediate, or transition state) enabled by conserved active site residues that perform the same functions in all members of the superfamily. Typically, the members of superfamilies share less than 50% sequence identity and often share less than 20% sequence identity (randomized sequences of a typical protein share 6–10% sequence identity). For the first type of superfamily listed here, the adjective

"specificity diverse" should be used to describe the nature of the reactions catalyzed by the members; for the second type of superfamily, the adjective "mechanistically diverse" should be used. Examples of the first type include the serine proteases; examples of the second are the subject of much of this review.

Suprafamily: Groups of homologous enzymes that catalyze different overall reactions but whose reactions do not share any common mechanistic attribute. Although active site residues may be conserved, these perform different functions in the members of the superfamily.

Membership in families, superfamilies, or suprafamilies may not be assigned from sequence data alone but often requires correlated functional and structural characterization. A consequence is that annotations of function for members of mechanistically diverse superfamilies and functionally distinct suprafamilies identified in genome sequencing projects are not possible from sequence data alone.

NATURE'S STRATEGIES FOR DIVERGENT EVOLUTION OF ENZYME FUNCTION

The number of unique protein folds is estimated at no more than a few thousand, many less than the number of proteins encoded by the human genome (10–12). As a result, folds must be reused by divergent evolution or independently invented multiple times (convergent evolution) to provide the functional range represented across the entire protein universe.

Based on the available evidence, three distinct strategies, listed below, can be envisaged for the divergent evolution of enzyme function. Each involves initial duplication of the gene encoding the protein to be evolved so that the original enzyme can be retained in cellular metabolism. The functional bases of the subsequent evolutionary process differ in significant details.

Substrate Specificity Is Dominant

Horowitz proposed that metabolic pathways evolve backwards (13, 14): When the substrate for an enzyme in a biosynthetic pathway is depleted, a new enzyme evolves to supply that substrate from an available precursor by evolution of the enzyme that uses the substrate. Accordingly, evolution is constrained to retain binding specificity, because the original enzyme and the newly evolved enzyme must share the ability to bind the same substrate/product molecule. The mechanisms of the reactions catalyzed by the original and evolved enzymes need not be related, so the old and new enzymes likely would be members of functionally distinct suprafamilies.

Chemical Mechanism Is Dominant

The conceptually opposed hypothesis is that nature selects the protein for divergent evolution from a pool of enzymes whose mechanisms provide the needed partial reaction or strategy for stabilization of energetically unfavorable intermediates or transition states (15, 16). Evolution then alters the other mechanistic characteristics

and/or substrate specificity so that the new reaction is catalyzed with enhanced proficiency; presumably, this process decreases the proficiency of the reaction catalyzed by the progenitor. The old and new enzymes would be members of mechanistically diverse superfamilies.

Perhaps nature selects a progenitor that has the adventitious capability of catalyzing the desired new reaction, albeit at a very low rate but sufficient to provide a selective advantage to the organism. Subsequent evolution then provides enhanced proficiency as the new metabolic pathway is optimized. Thus, experimental support for this strategy is provided by enzymes that are catalytically promiscuous; that is, in addition to their normal reaction, they catalyze a low level of another reaction that involves a shared partial reaction (17).

Active Site Architecture Is Dominant

A third hypothesis does not require that either substrate specificity or chemical mechanism dominates; instead, an active site is able to support an alternate reaction that may use shared functional groups of the active site in different mechanistic and metabolic contexts. The old and new enzymes would be members of functionally distinct suprafamilies.

MECHANISTICALLY DIVERSE SUPERFAMILIES

Enolase Superfamily

The discovery of mechanistically diverse superfamilies occurred in 1990 with the recognition that mandelate racemase (MR) and muconate lactonizing enzyme (MLE) have homologous structures that include a $(\beta/\alpha)_8$ (TIM) barrel domain, share conserved functional groups at the active site, and catalyze different overall reactions that are initiated by abstraction of the α-proton of a carboxylate anion substrate (1). In 1995, (D)-galactonate dehydratase from *Escherichia coli* was recognized as another member of the superfamily, with the assignment of this function to an unknown open reading frame (ORF) in the *E. coli* genome (18). This functional prediction and verification was based on the expectation that the reaction catalyzed by the protein encoded by the ORF would be initiated by abstraction of the α-proton of its substrate. Finally in 1996, after the recognition that the structure of enolase, including the positions of three carboxylate ligands for an essential Mg(II) in the TIM-barrel domain, could be superimposed on those of MR and MLE, the scope of the diversity of the enolase superfamily was fully appreciated (19). Now, we recognize that at least 12 different overall reactions are catalyzed by members of the superfamily, with the reaction types including 1,1-proton transfer (racemization/epimerization), β-elimination of OH^- or NH_3, and cycloisomerization (intramolecular addition/elimination). The stereochemical courses of the β-elimination reactions are either *syn* or *anti* and without conservation of the absolute configurations of the α-carbon from which the proton is

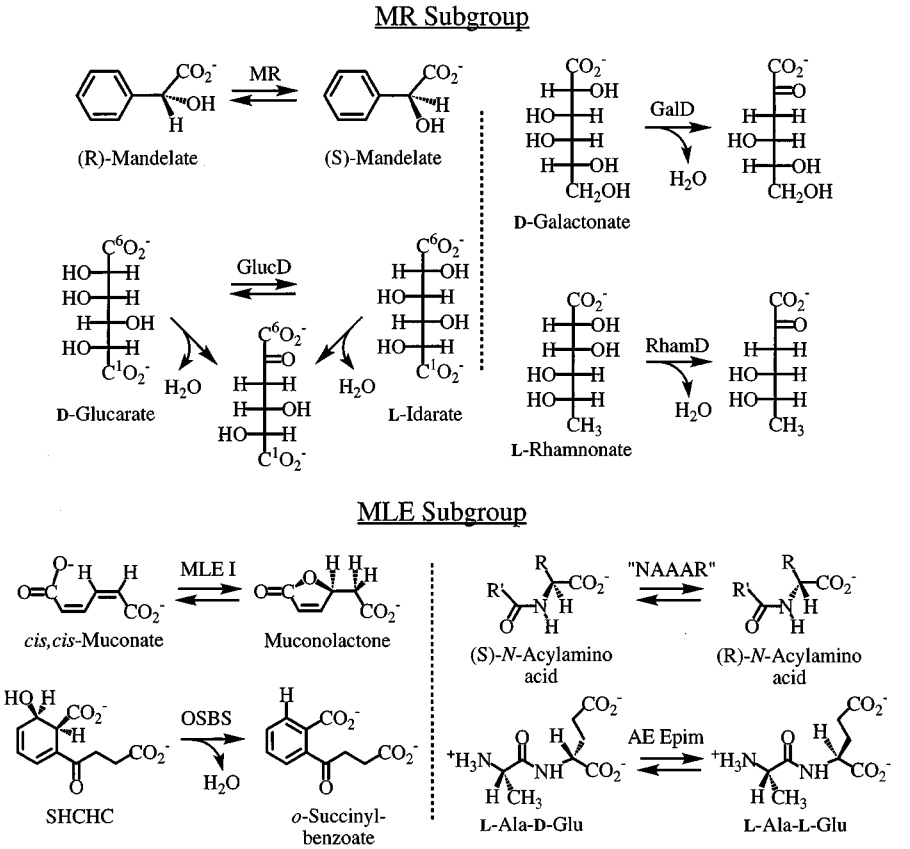

Figure 1 Reactions catalyzed by members of the enolase superfamily.

abstracted or of the β-carbon from which the leaving groups depart. Reactions catalyzed by members of the superfamily are shown in Figure 1, and the members are discussed in more detail below. The known reactions require a divalent metal ion, usually Mg(II), but Mn(II), Co(II), and Ni(II) will frequently support catalysis.

Excluding the ubiquitous enolases, we have identified more than 60 members of the enolase superfamily in sequence databases. Including the enolases, alignments of the sequences reveal the presence of three conserved ligands for the required divalent metal ion, usually Asp. . .Glu. . .Asp or Asp. . .Glu. . .Glu, although the ligands are Asp. . .Glu. . .Asn in (D)-glucarate dehydratases. So, our operational definition of the enolase superfamily includes the required conservation of three metal ion ligands. The conservation of ligands for a divalent metal ion reveals that the underlying catalytic strategy in the superfamily is divalent metal ion–assisted enolization of an enolic intermediate generated by abstraction of the α-proton: In the absence of such stabilization, the pK_as of the α-protons range from 29

to 32, values that otherwise preclude kinetically competent formation of enolic intermediates (20, 21).

Considering what is known about structure-function relationships for the reactions catalyzed by enolase, MR, and MLE, we further expect that members of the superfamily possess at least one general acid/base catalyst that mediates abstraction of the α-proton of the substrate: In enolase, Lys 345 is the general base; in MR, Lys 166 in a Lys 164-X-Lys 166 motif is the (S)-specific base and His 297 hydrogen-bonded to Asp 270 is the (R)-specific base; and in MLE, Lys 169 in a Lys 167-X-Lys 169 motif and Lys 273 are positioned on opposite faces of the active site, although which of these is the expected general base has not been unequivocally assigned. Using these criteria, we partition the members of the superfamily into three subgroups, MR, MLE, and enolase, based on the identities of the putative general acid/base catalysts. Most of the members of the MR and MLE subgroups are distantly related, sharing 25% sequence identity, and we expect that most if not all of these divergent proteins catalyze different reactions. Thus, divergence of enzymatic function is accompanied by significant changes in sequence, although three metal ion ligands and at least one general acid/base catalyst are always conserved.

The structures of MR (22), MLE (23, 24), and enolase (25) reveal that each is organized into two domains: (*a*) a catalytic TIM-barrel domain that contains the metal ion ligands and acid/base catalysts at the C-terminal ends of separate β-sheets and (*b*) a substrate specificity–determining capping domain composed of a large N-terminal segment and usually a shorter C-terminal segment of the polypeptide. The barrel domains in enolase, MR, and MLE are shown in Figure 2. The ligands for the required divalent metal ions are located at the ends of the third, fourth, and fifth β-sheets; the functional groups involved in proton abstraction are located at the ends of the second, sixth, and seventh β-sheets. We expect that this active site design, with the various functional groups located separately at the ends of the β-sheets in a barrel domain, is particularly well suited for divergent evolution of function (15): (*a*) Owing to the pseudo-eight-fold symmetry of the barrel, functional groups can be placed in virtually any position relative to the bound substrates and enolic intermediates; and (*b*) the placement of the functional groups at the ends of distinct β-sheets allows them to evolve independently, thereby permitting new catalytic activities to be generated that retain metal ion binding and the ability to abstract the α-proton of a carboxylate anion substrate.

Mandelate Racemase Subgroup Our continuing studies of the enolase superfamily are focused on understanding how the active site structure can be modified to catalyze different reactions. For example, members of the MR subgroup catalyze dehydration of acid sugars, including (D)-glucarate dehydratase (GlucD), (D)-galactonate dehydratase (GalD), and (L)-rhamnonate dehydratase (RhamD) from *E. coli* (paralogous enzymes in *E. coli*). Given the sophistication of our knowledge about structure-function relationships in the MR-catalyzed reaction (26), we

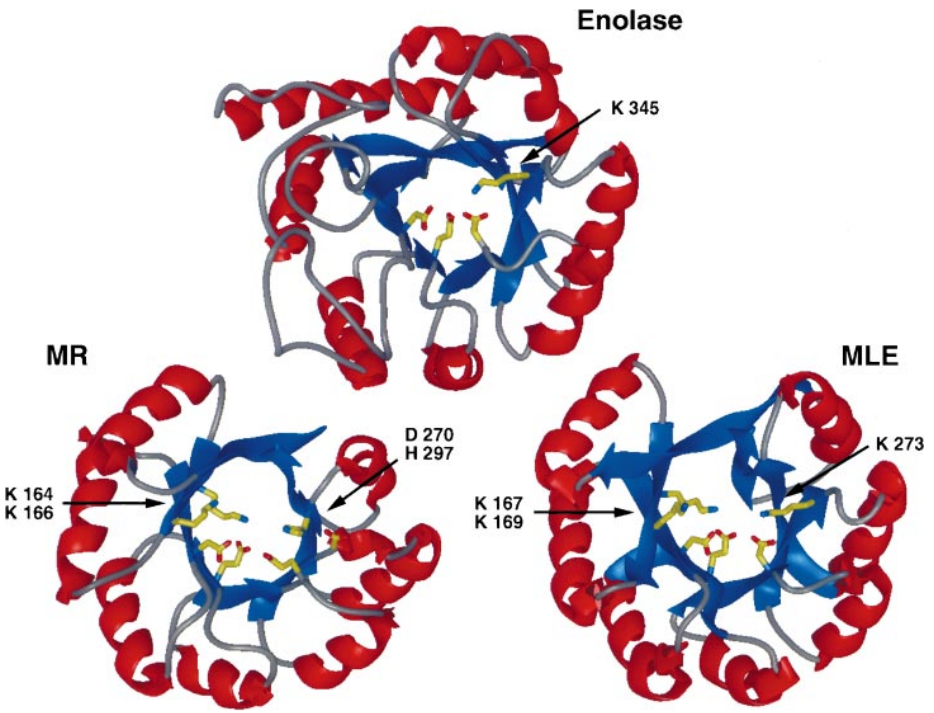

Figure 2 The $(\beta/\alpha)_8$ (TIM) barrel domains of enolase, MR, and MLE, showing the positions of the functional groups in the active site; the identities of the acid/base catalysts are given. β-Sheets are colored blue, α-helices are red.

would like to understand how the vinylogous elimination of the β–leaving group is additionally catalyzed, because we expect that it will require the participation of a general acid catalyst.

In the case of (D)-glucarate dehydratase, the sequence alignment with MR revealed homologs of both sets of functional groups involved in proton abstraction (19). Given that GlucD was known to catalyze the formation of 4-keto-5-deoxy-(D)-glucarate, a reaction that would require only a single general base to initiate the reaction by abstraction of the 5S-proton, the presence of both (S)- and (R)-specific bases was unexpected. We reasoned that GlucD might be promiscuous and catalyze the dehydration of (L)-idarate, the 5-epimer of (D)-glucarate, because it might present its 5R-proton to the base on the opposite face of the active site, and that it might catalyze the epimerization of (D)-glucarate and (L)-idarate (a 1,1-proton transfer reaction analogous to that catalyzed by MR). The kinetic constants (k_{cat} and k_{cat}/K_m) for dehydration of (L)-idarate were found to be similar to those for (D)-glucarate, thereby establishing a new function for GlucD (27). As expected, GlucD catalyzes the epimerization of (D)-glucarate and (L)-idarate in competition with their dehydration (28).

Structural studies of GlucD in the presence of the reaction product and substrate analogs revealed not only the expected placement of the active site residues shared with MR but also a surprise with respect to the details of how the substrate likely binds in the active site (29, 30). In the MR-catalyzed reaction, the substrate is a bidentate ligand of the essential Mg(II), with one carboxylate oxygen and the α-hydroxyl group ligands of the metal ion. The α-proton of (S)-mandelate is presented to Lys 166; the α-proton of (R)-mandelate is presented to His 297. In the GlucD-catalyzed reaction, the substrate is also a bidentate ligand of the Mg(II), but in this case both carboxylate oxygens are the ligands. Furthermore, the geometry of the 4-keto-5-deoxy-(D)-glucarate product and the inert 4-deoxy-(D)-glucarate substrate analog revealed that the 5S-proton of (D)-glucarate likely is presented to His 339 instead of Lys 207. Thus, despite the considerable insights available for structure-function relationships in MR, those were insufficient to predict the details of the GlucD-catalyzed reaction.

In the case of GalD, the sequence alignment with MR revealed a homolog of the His 297-Asp 270 dyad (His 285-Asp 258) but no homolog for Lys 166 (18). The absolute configuration of carbon-2 of (D)-galactonate is R, so His 285 was expected to participate by abstraction of the α-proton. The structure of GalD has been determined in the presence of lyxonohydroxamate, an analog of the enolic intermediate; this structure confirms the expected metal ion ligands and placement of His 285 adjacent to carbon-2 (31). Interestingly, this structure also suggests the close spatial proximity of the 3-OH leaving group of (D)-galactonate with His 185 that is located at the end of the third β-sheet in the barrel domain. Both the H285N and H185N substitutions significantly decrease k_{cat}, as expected for their roles as general basic and general acidic catalysts, respectively, in the dehydration reaction. We reasoned that if His 185 is the general acid, 3-fluoro-(D)-galactonate might serve as a substrate for the H185N mutant, since fluoride anion (F-) departure does not require the assistance of a general acid. This expectation was realized, confirming the identity of His 185 as an evolutionarily new general acid in the enolase superfamily. This suggests that the active sites of other acid sugar dehydratases might have such new general acids located at the ends of this or other β-sheets in the barrel domain.

Muconate Lactonizing Enzyme Subgroup We are also studying the reactions catalyzed by members of the MLE subgroup so that we might both understand and exploit the structural bases for another example of catalytic promiscuity.

The identification of an *N*-acylamino acid racemase (NAAAR) from a species of *Amycolaptosis* as a member of the MLE subgroup (32, 33) attracted our attention because it is predicted to catalyze a 1,1-proton transfer reaction using two Lys acid/base catalysts instead of the Lys and His acid/base catalysts in the MR-catalyzed reaction. A protein with this activity was sought for use in a commercial process in which racemic *N*-acylamino acids would be converted to L-amino acids in high yield using the coupled actions of an L-amino

acylase and NAAAR. Surprisingly, NAAAR is not a remarkable catalyst: Using *N*-acetyl-methionines, the best substrates, the value of k_{cat}/K_m is 400 M^{-1} s^{-1}.

Insight into this "problem" was provided by the identification of a homolog, YtfD in the *Bacillus subtilis* genome, that shares 43% sequence identity with the NAAAR (34). The gene encoding YtfD is the sixth and last in an operon that encodes five enzymes in the menaquinone biosynthetic pathway; the *E. coli* genome also has a six-gene operon encoding the first six enzymes in this pathway. Pairwise matching based on sequence identity/similarity suggested that the function of YtfD was catalysis of the *o*-succinylbenzoate synthase (OSBS) reaction, a highly exergonic dehydration reaction. We verified that both YtfD and the NAAAR were efficient catalysts of the OSBS reaction, with the value of k_{cat}/K_m for the NAAAR being 2.5×10^5 M^{-1} s^{-1}. Thus, we conclude that the metabolic function of the MLE homolog from *Amycolaptosis* is catalysis of the OSBS reaction rather than apparently metabolically silent racemization of *N*-acylamino acids.

The fact that YtfD was not assigned the OSBS function when the *B. subtilis* genome was annotated can be attributed to the low (15%) sequence identity that relates these proteins. We subsequently identified 20 eubacterial OSBSs and found that these are also characterized by low sequence identities, although all contain homologs of the metal ion ligands and the three active site lysines in MLE. Our hypothesis is that the extreme divergence of sequences may be related to the fact that the substrate for the OSBS reaction is chemically unstable at modestly elevated temperatures or pHs that deviate from neutrality (EA Taylor, DR Palmer & JA Gerlt, unpublished information). Perhaps the reactivity of the substrate does not demand precise placement of the acid/base catalysts that are expected to catalyze the dehydration reaction.

A high-resolution structure is available for the OSBS from *E. coli* in the presence of the *o*-succinylbenzoate (OSB) product (36). This structure reveals the expected placement of functional groups shared with MLE: Lys 235 and Lys 133 are appropriately positioned to catalyze the *anti* dehydration reaction.

That one OSBS can catalyze a slow 1,1-proton transfer reaction suggested that other (unidentified) members of the MLE subgroup might catalyze racemization or epimerization reactions as their metabolic functions. In fact, we have discovered that YkfB in the *B. subtilis* genome and YcjG in the *E. coli* genome catalyze the epimerization of certain dipeptides (DZ Schmidt, BK Hubbard & JA Gerlt, unpublished data): YkfB appears to be specific for the epimerization of L-Ala-L-Glu and L-Ala-L-Asp; YcjG catalyzes the epimerization of all L-Ala- L-X dipeptides except when X = Arg, Lys, or Pro. Although our studies are still in progress, our working hypothesis is that both YkfB and YcjG are involved in pathways for recycling and/or catabolism of murein peptides.

Amidohydrolase/Phosphotriesterase Superfamily

Given its frequent occurrence in structurally characterized enzymes, it is not surprising that the TIM-barrel fold provides the conserved architecture of the active site in other mechanistically diverse superfamilies and functionally distinct

suprafamilies; in fact, the frequent occurrence of the TIM-barrel fold suggests and illustrates its versatility in the divergence of enzyme function. Of these families, the amidohydrolase/phosphotriesterase superfamily is among the best characterized: Structures are available for the phosphotriesterase from *Pseudomonas diminuta* (38–40), the urease from *Klebsiella aerogenes* (41–44), and the murine adenosine deaminase (45–48). Each of the functionally characterized members for which structures are available (and many for which structures are not available) requires divalent metal ions for catalysis: two Ni(II) ions in the case of urease, two Zn(II) ions in the case of phosphotriesterase, and one Zn(II) ion in the case of adenosine deaminase.

Each enzyme catalyzes a hydrolysis reaction, although the chemical nature of the scissile bond is variable: a P-O bond in the phosphotriesterase and C-N bonds in differing molecular environments in urease and the deaminase. Representative reactions catalyzed by members of the superfamily (discussed below) are shown in Figure 3. Other members of the superfamily, identified by sequence alignments and for which detailed structural and mechanistic data are missing, extend this repertoire of scissile bonds.

A detailed analysis of sequence/structure relationships in this superfamily was performed by Holm & Sander (49). Their analysis of superpositions of the structures for the three characterized members of the superfamily revealed that the active sites are located at the C-terminal ends of homologous TIM-barrel domains, with the ligands for the essential divalent metal ions located at the ends of β-sheets in these domains; the TIM-barrel domains of urease, phosphotriesterase, and adenosine deaminase are shown in Figure 4. In adenosine deaminase, the single Zn(II) is coordinated by the His-X-His motif at the end of the first β-sheet, a His at the end of the fifth β-sheet, and an Asp at the end of the eighth β-sheet; a His at the end of the sixth β-sheet is thought to activate the nucleophilic water that is coordinated to the Zn(II). In both urease and phosphotriesterase, the homologous His-X-His motif at the end of the first β-sheet and the homologous Asp at the end of the eighth β-sheet coordinate one metal ion, and the homologous His residues at the end of the fifth and sixth β-sheets coordinate the second divalent metal ion. In both enzymes, a carbamylated Lys at the end of the fourth β-sheet provides bridging oxygen ligands for the metal. The conservation of the unusual carbamyl group in urease and phosphotriesterase but not in adenosine deaminase is a striking statement of the nature and extent of divergent evolution in this superfamily. Yet the unrelated ribulose bisphosphate carboxylase also uses a carbamylated Lys as a metal ion ligand to an essential Mg(II) (50).

The divalent metal ion(s) play essential roles in catalysis. In the case of the deaminase, the Zn(II) is thought to decrease the pK_a of the directly coordinated water, thereby enhancing its nucleophilicity at neutral pH (48). In the case of urease, the C=O bond of the urea substrate is thought to be polarized by the binding of its oxygen to one Ni(II) while the second Ni(II), as in the case of the deaminase, enhances the nucleophilicity of the directly coordinated water (51). In the case of phosphotriesterase, the available structural information suggests that in the

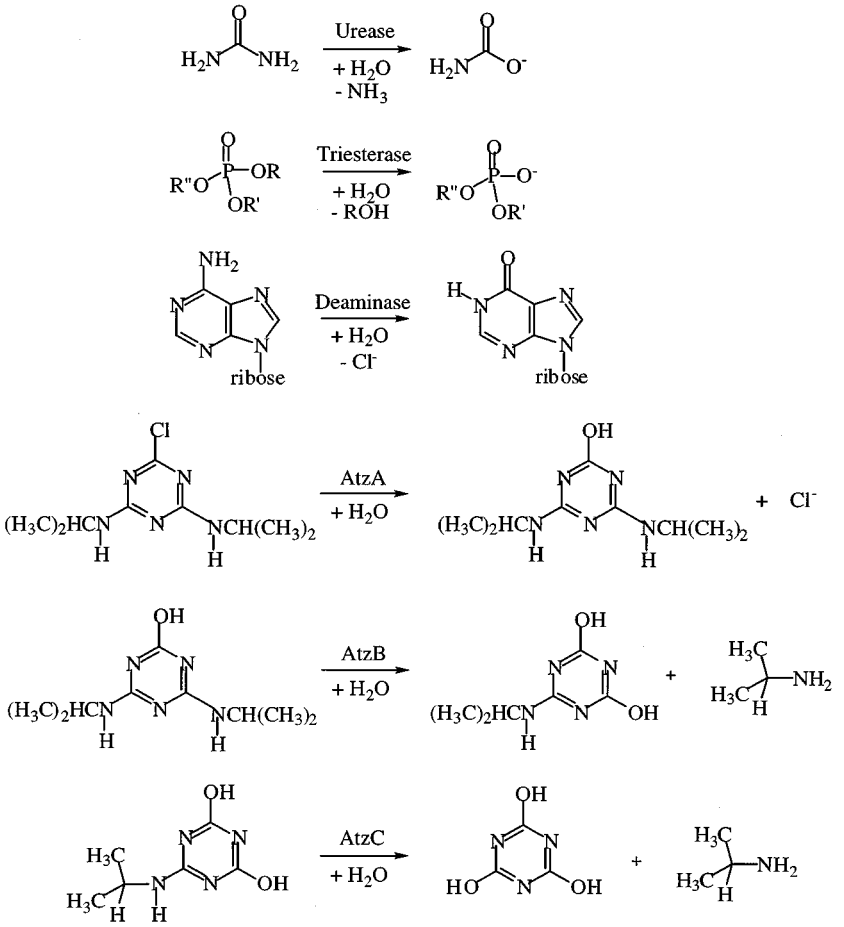

Figure 3 Reactions catalyzed by members of the amidohydrolase superfamily.

substrate-free enzyme, a hydroxide group is bridged between the two Zn(II); upon substrate binding, the nonesterified phosphoryl oxygen displaces the hydroxide ligand from one metal ion, thereby simultaneously generating the nucleophilic water and polarizing the P=O bond for nucleophilic attack (52). Thus, divergent evolution of function retained the involvement of one metal ion to enhance the acidity/nucleophilicity of a directly coordinated water molecule but elaborated the ability to activate a variety of substrates for hydrolysis.

Holm & Sander extended membership in this family by database searches and sequence alignments that revealed the presence of homologs for the ligands that bind metal ions in the structurally characterized members (49). Interestingly, for such functionally identified members, virtually all catalyze the hydrolyses

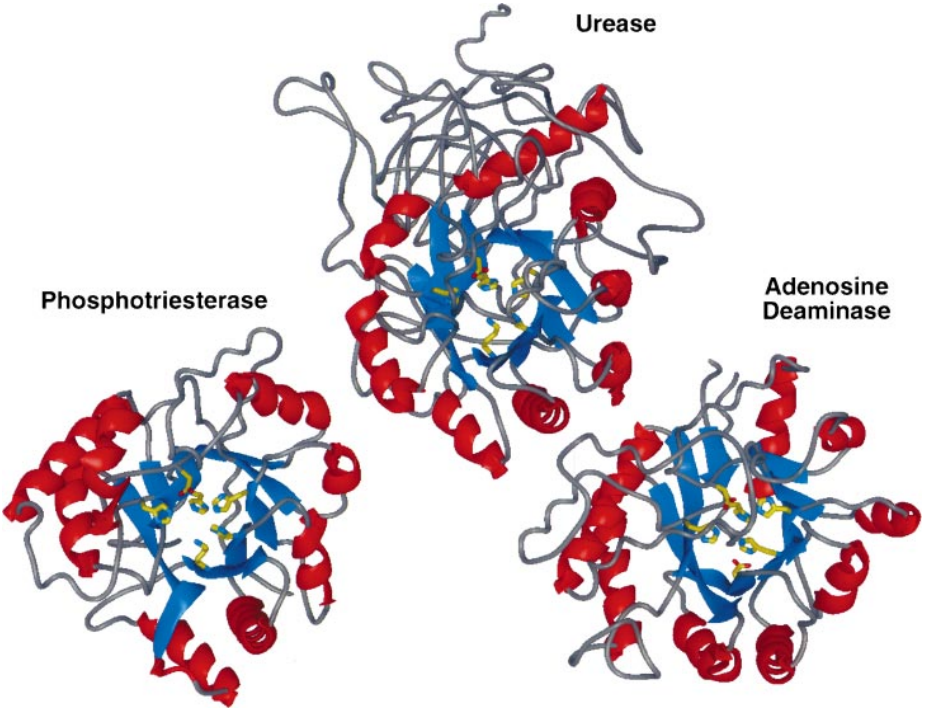

Figure 4 The $(\beta/\alpha)_8$ (TIM) barrel domains of urease, phosphotriesterase, and adenosine deaminase.

of C-N bonds of either nucleic acid bases (adenine, cytosine, and AMP deaminases) or amides (dihydroorotases, hydantoinases, imidazolonepropionase, and D-aminoacylases), suggesting that the structure-function relationships established for the deaminase, urease, and phosphotriesterase can be expected to apply to the mechanism of these reactions. Holm & Sander further pointed out that several members of this superfamily occur in nucleotide catabolism and anabolism, supporting the idea that pathways can evolve by the recruitment of progenitor enzymes that catalyze mechanistically related reactions.

An interesting extension of the scissile bond specificity is found in atrazine hydrolase (or AtzA) that catalyzes the hydrolysis of the C-Cl bond in atrazine to generate hydroxyatrazine (Figure 3). The further degradation of hydroxyatrazine involves sequential hydrolysis reactions catalyzed by two other members of this superfamily (AtzB and AtzC) that catalyze hydrolyses of C-N bonds. This suggests that the pathway for atrazine degradation was assembled by two separate gene duplication and divergent evolution processes; of course, the mechanisms of the reactions catalyzed by AtzA, AtzB, and AtzC can be expected to be similar (53–55).

Finally, a member of this superfamily of unknown function (56) and with closest sequence similarity to phosphotriesterase has been structurally characterized

(57). Structural superpositions and the presence of conserved sequence motifs clearly places it within the superfamily. Although the substrate specificity of this protein has not yet been determined (a particularly difficult task in mechanistically divergent superfamilies because of the inability to distinguish specificity from among the broad range of possibilities that could be associated with each), the structure-function paradigm described for this superfamily leads to a prediction that catalysis by this enzyme will also require a divalent metal ion(s) to enhance the acidity/nucleophilicity of coordinated water molecules. The structural investigation of this protein shows that it binds two zinc ions per monomer and has a binuclear zinc center in the presumptive active site.

Thiyl Radical Superfamily

Barrel structures are not restricted to the $(\beta/\alpha)_8$ barrel fold: Recent structural studies have produced the initially surprising result that several enzymes that use thiyl radicals share a $(\beta/\alpha)_{10}$ barrel fold (58). The structurally characterized members of this superfamily include the aerobic Fe(II)-dependent (class I) ribonucleotide reductase (RNRase), the aerobic adenosylcobalamin-dependent (class II) RNRase, the anaerobic formate-dependent (class III) RNRase, and the anaerobic pyruvate formate lyase (PFL). Reactions catalyzed by members of this superfamily are shown in Figure 5; the $(\beta/\alpha)_{10}$ barrel folds of members of the superfamily are shown in Figure 6. The members of this superfamily have been well characterized mechanistically, so the findings that they use a conserved fold with a conserved thiyl radical can be placed in a correlated structure-function context. Not only are the overall reactions and mechanisms catalyzed by these enzymes diverse, but the strategies that nature has used to generate the catalytically essential thiyl radicals are also diverse.

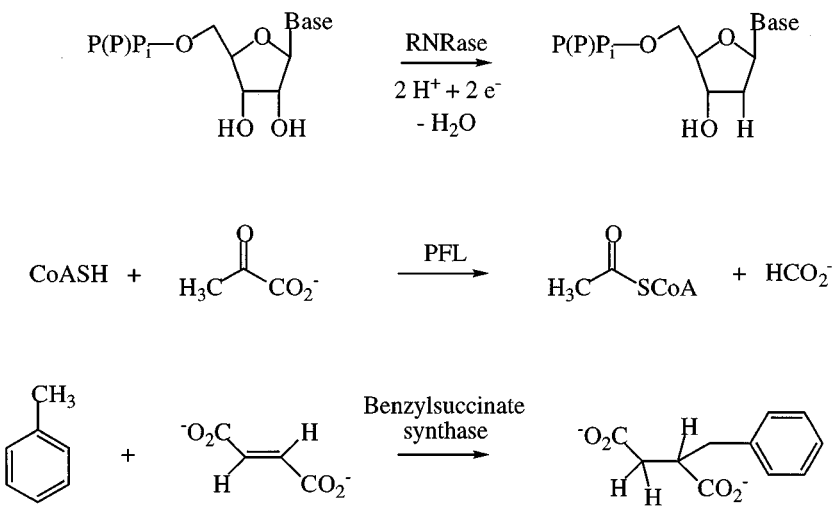

Figure 5 Reactions catalyzed by members of the thiyl radical superfamily.

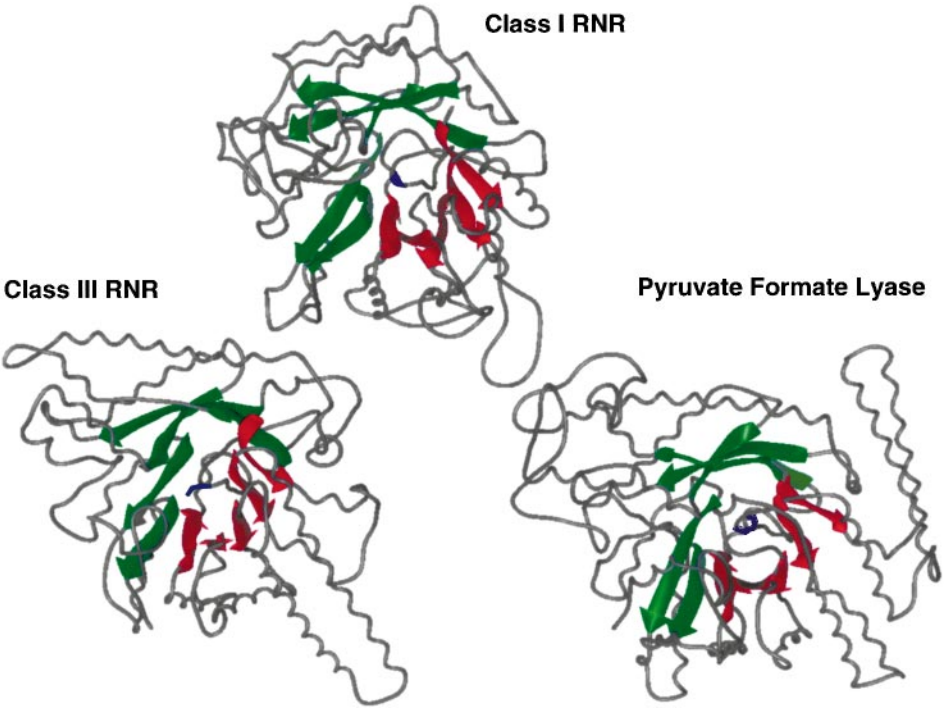

Figure 6 The $(\beta/\alpha)_{10}$ barrel domains of the class I and class III ribonucleotide reductases and PFL. The β-sheets of the two halves of the barrel are colored red and green; the active site cysteines are blue.

The first structure for a member of this superfamily was that for the thiyl radical containing polypeptide of the class I RNRase from *E. coli* (59, 60). Biochemical studies performed by Stubbe and her coworkers have established the catalytic importance of a thiyl radical derived from Cys 439 (61, 62). In this RNRase, the thiyl radical is generated by long-range electron transfer to a tyrosyl radical located and stabilized on the R2 polypeptide. The structure of the $(\beta/\alpha)_{10}$ barrel revealed that Cys 439 is located on a long loop that is sandwiched between and joins two half-barrels containing five parallel strands that are joined in an antiparallel fashion.

Pyruvate formate lyase catalyzes an unusual reaction in anaerobic metabolism in *E. coli*: the dismutation of pyruvate into acetyl CoA and formate. Mechanistic studies by Knappe (63–65) and Kozarich (66–68) and their coworkers had implicated Cys 418 and Cys 419 in catalysis, with both likely participating as thiyl radicals during the catalytic cycle. In the PFL-catalyzed reaction, the thiyl radical(s) is (are) generated by a strategy distinct from that used by the class I RNRase: An activating enzyme containing an iron-sulfur center uses *S*-adenosylmethionine (SAM) as a cosubstrate to generate a radical derived from Gly 734 (69, 70), which biochemical experiments suggest then generates the thiyl radical. The structure

of PFL revealed that Cys 418 and Cys 419 are located on a loop sandwiched in the interior of a $(\beta/\alpha)_{10}$ barrel structure that is homologous to the structure of the class I RNRase (58, 71). The sulfur of Cys 419 is located spatially proximal to the α-carbon of Gly 734, providing a structural confirmation of a previously proposed pathway for generation of the thiyl radical.

The disclosure of the structure of the G580A mutant of the anaerobic (class III) RNRase from bacteriophage T4 (NrdD) showed it to belong to the same fold class as that observed in the class I RNRases and PFL (72). The active site is located within a $(\beta/\alpha)_{10}$ barrel domain, and the active site Cys 290 is located at the tip of the loop sandwiched in the interior of the 10-stranded barrel. Biochemical experiments had identified Gly 580 as the progenitor of the thiyl radical derived from Cys 290; the structure was determined for the G580A mutant to avoid an O_2-dependent cleavage of the polypeptide. In analogy with the structure of the active site of PFL, C_α of Ala 580 is located within 5.2 Å of S_γ of Cys 290, consistent with a short-range transfer of the hydrogen of Cys 290 to the radical derived from Gly 580. Again, in analogy with PFL, the glycyl radical is generated by the action of a specific activating enzyme (NrdG) that uses SAM as a cosubstrate (73, 74).

Interestingly, the reductive cosubstrate for the class III RNRase is formate anion, the product of the PFL-catalyzed reaction (75). Thus, a possible mechanism involves abstraction of the 3′ hydrogen from the nucleotide substrate by the thiyl radical derived from Cys 290 in analogy with the class I RNRase. Perhaps a second cysteine in the active site of this RNRase quenches the substrate-derived radical, and this thiyl radical, in turn, abstracts the hydrogen from formate anion, generating formate radical anion that is the actual strong reductant.

Both Fontecave (76) and Benner (77) and their coworkers independently isolated RNRases from archaeal sources that upon functional characterization were determined to be class II RNRases that require adenosylcobalamin. The sequences of both of these enzymes, one from *Thermoplasma acidophila* and the second from *Pyrococcus furiosus*, show similarities to both class I and class III RNRases; previous attempts to detect sequence homology between the class I and class III enzymes failed as a result of their distant evolutionary relationship. As described previously, the class I and class III enzymes are structurally homologous. Thus, when the structure of a class II RNRase, the adenosylcobalamin-dependent enzyme from *Lactobacillus leichmanii*, was solved by Drennan & Stubbe (unpublished information), it came as no surprise that the class II RNRase shared the fold previously found for other members of this superfamily: a $(\beta/\alpha)_{10}$ barrel in which the active site cysteine (identified by Stubbe and coworkers in the biochemical experiments) was located on a loop anchored by the halves of the 10-stranded barrel.

Although not yet structurally characterized, recent experiments have shown that benzylsuccinate synthase from *Thauera aromatica*, which catalyzes the condensation of toluene and fumarate, shows significant sequence conservation with the active site cysteines (Cys 489) and glycines (Gly 825) in PFL and the class III RNRases (79). The mechanism of this condensation reaction likely involves initial abstraction of a hydrogen from the methyl group of toluene to form a radical

that adds to the double bond in fumarate; the radical so formed is likely quenched by Cys 489 to regenerate the starting form of the enzyme. There is little doubt that the structure of this enzyme will reveal a $(\beta/\alpha)_{10}$ barrel in which Cys 489 is located on a loop that is restrained by the barrel structure to be spatially proximal to Gly 825.

Thus, despite the diverse reactions described in this section, structural and sequence analyses as well as biochemical studies clearly confirm the existence of a mechanistically diverse superfamily in which various mechanisms are used to generate a thiyl radical positioned in an active site contained in a unique $(\beta/\alpha)_{10}$ barrel; depending upon the reaction that is catalyzed, the thiyl radical initiates a diverse range of reactions.

Crotonase Superfamily

In analogy to the functional flexibility of diverse barrel folds described in the previous sections, other folds provide active site architectures that can be modified to support different overall reactions related by a common mechanistic feature. One such structure is found in a superfamily that has been biochemically and structurally characterized, the crotonase (or enoyl CoA hydratase) superfamily: The members catalyze reactions that use an oxyanion hole to stabilize enolate anions derived from thioesters or tetrahedral intermediates formed in the hydrolyses of peptide bonds (15, 16, 80). Other than the oxyanion hole, no active site functional groups are conserved in members of the superfamily. Reactions catalyzed by members of the crotonase superfamily are shown in Figure 7.

The fold that provides the oxyanion hole is formed by two layers of β-sheets that are perpendicular to one another and surrounded by α-helices. In those members of the superfamily that use CoA esters as substrates, the oxyanion hole is at the mouth of the active site, with the acyl group of the substrate embedded in a cavity that contains the determinants for substrate specificity as well as the catalytic groups that generate the enolate anions that occur on the reaction pathways in a variety of different reactions. At present, four structures are available for members of the superfamily that use CoA esters as substrates: 4-chlorobenzoyl CoA dehalogenase (81, 82), the prototypic rat mitochondrial crotonase (enoyl CoA hydratase) (83, 84), $\Delta^{3,5}$, $\Delta^{2,4}$-dienoyl CoA isomerase (85), and methylmalonyl CoA decarboxylase (86). The structures of the polypeptides in these enzymes are shown in Figure 8.

In addition, the structural studies of the ClpP protease from *E. coli* (the serine protease of the bacterial analog of the eukaryotic proteasome) unexpectedly revealed that it shares the same fold as the CoA ester–utilizing members of the superfamily (87, 88). Although the mechanism of the reaction catalyzed by ClpP protease has not been investigated in detail, the oxyanion holes for ClpP and the CoA ester–utilizing members are structurally homologous, providing persuasive evidence that the underlying mechanistic feature that unites the members of this superfamily is stabilization of oxyanion intermediates, albeit enolate

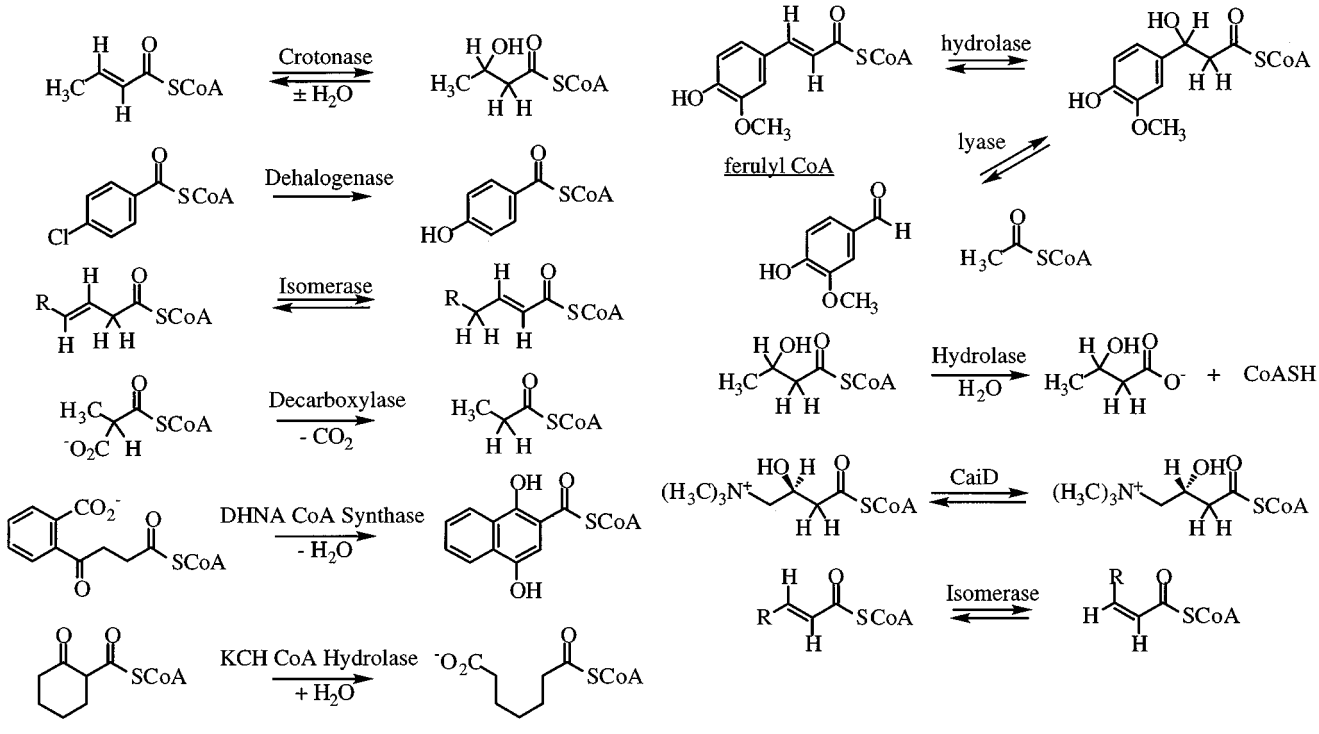

Figure 7 Reactions catalyzed by members of the crotonase superfamily. Dihydroxynaphthoate (DHNA); 2-ketocychlohexyl (KCH); carnitinyl CoA epimerase (CaiD).

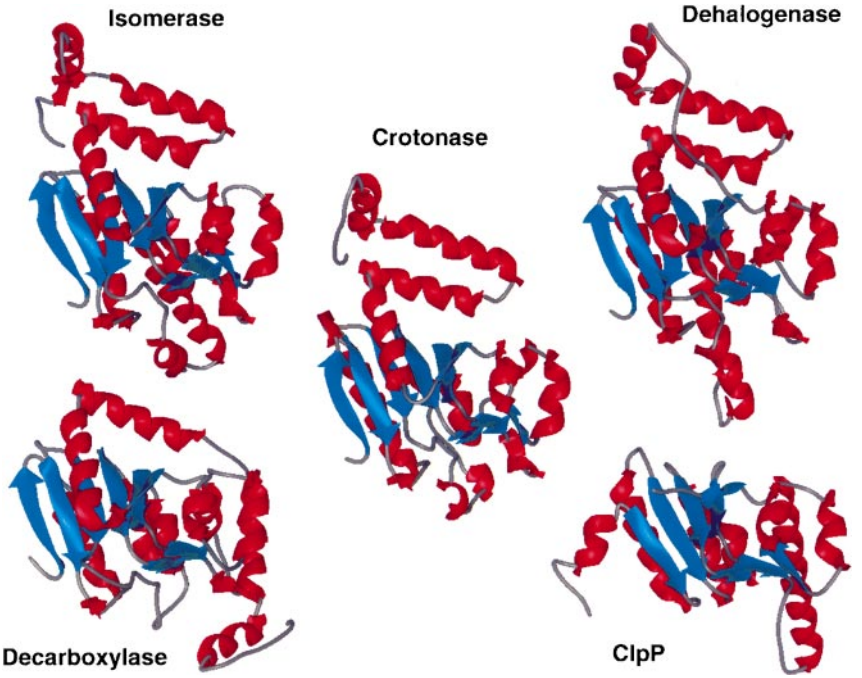

Figure 8 Structures of crotonase, dehalogenase, $\Delta^{3,5}$, $\Delta^{2,4}$-dienoyl CoA isomerase, methylmalonyl CoA decarboxylase, and ClpP protease.

anions derived from CoA thioesters or tetrahedral intermediates derived from peptides.

The mechanism of the reaction catalyzed by rat mitochondrial crotonase is the best studied in the superfamily (89–91): Glu 144 and Glu 164 catalyze the *syn* addition of water to enoyl CoA esters to yield 3S-OH acyl CoA esters, with Glu 144 the general base that assists the attack of water on the β-carbon and Glu 164 the general acid that delivers a proton to the α-carbon. Although kinetic isotope effect investigations of the mechanism suggest that the reaction is concerted (89), the fact that the members of this superfamily are united by the existence of an oxyanion hole argues that the addition and protonation reactions are either stepwise or decidedly asynchronous (21).

Although not structurally characterized, 3,2-*trans*-enoyl CoA isomerase that catalyzes 1,3-proton transfer reaction contains a homolog of Glu 164 but not Glu 144 in crotonase that is thought to mediate suprafacial proton transfer reaction (92). The structurally characterized $\Delta^{3,5}$, $\Delta^{2,4}$-dienoyl CoA isomerase that catalyzes a 1,5-proton transfer reaction also contains a homolog of Glu 164 (Glu 196) but not Glu 144 (85); in this case, the active site geometry reveals the presence of Asp 204 that may be appropriately located to mediate proton transfers to/from carbon-6 while Glu 196 mediates proton transfers to/from carbon-2 of

the substrate. Unfortunately, neither of these reactions has been characterized in sufficient (structural or mechanistic) detail to allow an unequivocal description of how these reactions are catalyzed.

The structurally characterized 4-chlorobenzoyl CoA dehalogenase catalyzes a nucleophilic aromatic substitution reaction in which an attack by Asp 145 (the homolog of Asp 204 in dienoyl CoA isomerase) on the 4-carbon of the substrate leads to formation of a Meisenheimer complex (a structural analog of an enolate anion) that is stabilized by the oxyanion hole (93–97). Subsequent decomposition of this intermediate by expulsion of the 4-chlorine yields an arylated enzyme intermediate; this intermediate is hydrolyzed with the assistance of a His. Dunaway-Mariano and colleagues have used insights into the architectural design principles represented by this superfamily to rationally engineer the 4-chlorobenzoyl CoA dehalogenase template to catalyze a low level of enoyl CoA hydratase activity (98).

The conclusion that stabilization of enolate anion intermediates derived from CoA esters is the underlying mechanistic principle provides restrictions on the reactions catalyzed by unknown members of the superfamily discovered in genome sequencing projects. We used this clue in our identification of an ORF in the *E. coli* genome as a novel biotin-independent methylmalonyl CoA decarboxylase (99). The *E. coli* genome encodes seven paralogs of the crotonase superfamily. One of these is located in an operon that includes Sbm, a homolog of methylmalonyl CoA mutase, and a CoA transferase. Although Sbm was not known to catalyze the methylmalonyl CoA mutase reaction, we were able to demonstrate this activity. Using this reaction as an additional clue, we hypothesized that the crotonase homolog would catalyze a reaction requiring stabilization of an enolate anion derived from methylmalonyl CoA, e.g. epimerization or decarboxylation. The latter activity was demonstrated and found to have a value of k_{cat}/K_m consistent with this being the biological function. Although the operon that contains methylmalonyl CoA mutase, methylmalonyl CoA decarboxylase, and a propionyl CoA:succinate CoA transferase is sufficient to encode a pathway for the decarboxylation of succinate to propionate, the physiological context of this ability is unknown. The structure of methylmalonyl CoA decarboxylase with an inert thioether analog of methylmalonyl CoA was solved (86): The active site includes Tyr 140, which likely orients the carboxylate group of methylmalonyl CoA orthogonal to the plane for the thioester carbonyl group and His 66; Glu 113 is hydrogen bonded to Arg 87, suggesting that Glu 113 is not directly involved in the decarboxylation reaction, although it is a structural and sequence homolog of Glu 144 in crotonase.

A number of other reactions are known to be catalyzed by members of the crotonase superfamily, including the Dieckman and retro-Dieckman reactions catalyzed by 1,4-dihydroxynapthoyl CoA synthase (100) and 2-ketocyclohexylcarboxyl CoA hydrolase (101), respectively, feruloyl CoA hydratase/lyase (102), and carnitinyl CoA epimerase (103). Although we and others have initiated biochemical studies of these enzymes, no correlations between structure and function are yet possible, so the identities of the functional groups and their mechanistic roles are unknown.

An interesting member of the superfamily is 3-OH isobutyryl CoA hydrolase, which hydrolyzes a CoA thioester bond (104). At least two mechanisms can be envisaged for this reaction: (*a*) hydrolysis involving the transient formation of a tetrahedral intermediate and (*b*) abstraction of the α-proton to yield an enolate anion intermediate that decomposes to a ketene and CoA, with the former captured by water to generate the carboxylate product. Both mechanisms involve the formation of an oxyanion intermediate that would be stabilized by the conserved oxyanion hole. At present, insufficient information is available to distinguish between these possible mechanisms.

If the tetrahedral intermediate mechanism is demonstrated for the hydrolase, this enzyme will provide a transition between the members of the superfamily that use CoA esters as substrates and the ClpP protease. Murzin first recognized that the fold of ClpP is homologous to that of crotonase (87); Flanagan and his coworkers had solved the structure of ClpP (88) and noted that this was the fifth family of serine proteases in which a catalytic triad (Ser 97, His 122, and Asp 171) had evolved by convergent evolution. As noted previously, the oxyanion hole that is required to stabilize the tetrahedral intermediate is structurally conserved with the oxyanion hole in the enzymes that use CoA esters as substrates. The residues of the catalytic triad are located in a cleft that is distinct from the cavity that binds the acyl moieties of CoA esters, emphasizing that divergent evolution retained only the underlying structural feature that is essential for stabilizing intermediates that otherwise would not be kinetically competent.

Murzin further noted that position-sensitive iterative basic local alignment sequence tool (PSI-BLAST) analyses suggest that the protease IV superfamily and the carboxyltransferase subunits of acetyl CoA carboxylase appear to be related to ClpP and crotonase by conservation of residues involved in formation of the oxyanion hole (87). Confirmation of this prediction by structural and functional studies would define the most amazing mechanistically diverse superfamily.

Vicinal-Oxygen-Chelate Superfamily

The $\beta\alpha\beta\beta\beta$ fold is found in the vicinal-oxygen-chelate (VOC) (15, 105) superfamily. The members of this superfamily catalyze a remarkable range of divalent metal ion–dependent reactions that appear to require stabilization of oxyanion intermediates (106): various extradiol dioxygenases that require Fe(II), glyoxalases I that require Zn(II), a glutathione S-transferase that inactivates fosfomycin by opening of the epoxide ring (FosA) and requires Mn(II), and methylmalonyl CoA epimerases that require Co(II). These reactions are shown in Figure 9. A fifth member of the superfamily, the bleomycin resistance protein, has no known enzymatic function and does not bind divalent metal ions (107, 108). Sequence alignments reveal the presence of three or four conserved ligands for the required divalent metal ion in all members of the superfamily.

High-resolution structures are available for two dioxygenases (109, 110) as well as glyoxalase I from both *E. coli* (111) and humans (112, 113). The structures of

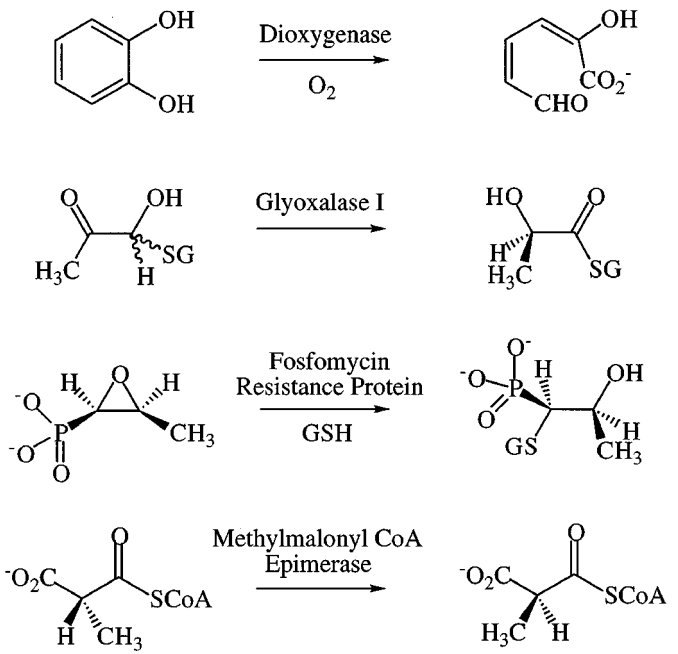

Figure 9 Reactions catalyzed by members of the VOC superfamily.

glyoxalase I and 2,3-dihydroxybiphenyl 1,2-dioxygenase are shown in Figure 10. These, together with the sequence alignments, reveal a remarkable structural strategy for binding the divalent metal ions and delivering the catalytic groups (106, 114). The various polypeptides are formed from either two or four copies of the $\beta\alpha\beta\beta\beta$ motif, with two copies of the $\beta\alpha\beta\beta\beta$ motif forming a metal-binding/active site, so the evolution of these proteins involved a series of gene duplication events. The pair of $\beta\alpha\beta\beta\beta$ motifs are arranged with pseudo–C2 symmetry to form a U-shaped cavity, with each having the potential to supply two metal ion ligands. In the case of the structurally characterized glyoxalases I, each $\beta\alpha\beta\beta\beta$ motif provides two ligands for the Fe(II) in the substrate-free enzyme; in the structurally characterized dioxygenases and by homology in FosA, one $\beta\alpha\beta\beta\beta$ motif provides two ligands and the other provides a single ligand.

The observed metal ligation in the ligand-free glyoxalase I was mechanistically enigmatic (112): Biochemical experiments suggested the presence of two general acid/base catalysts in the active site because diastereomeric thiohemiacetal adducts of methylglyoxal with glutathione could be converted to the single (S)-lactoylglutathione product (115), but the structure of the ligand-free enzyme revealed no candidates for these catalysts in the vicinity of the metal ion–binding site. A solution to this functional problem was suggested by the structure of a complex of human glyoxalase with a mimic of the enediolate intermediate analog

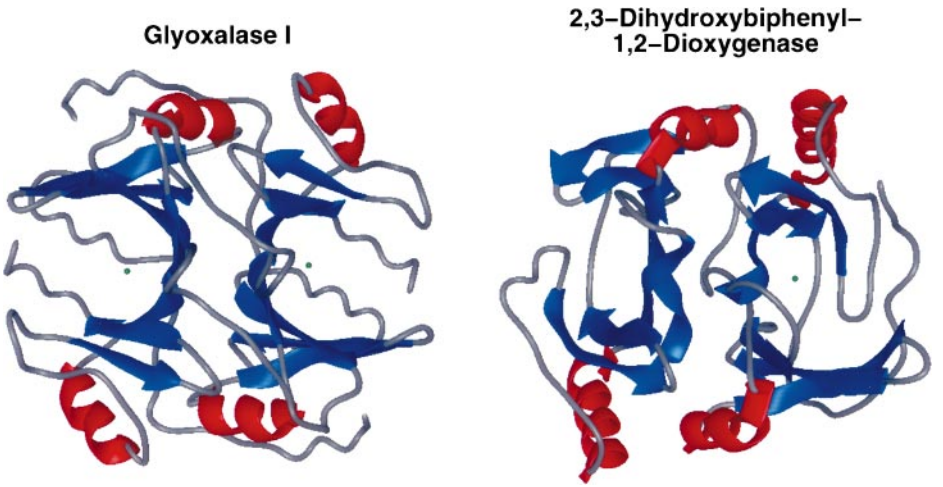

Figure 10 Structures of glyoxalase I and 2,3-dihydroxybiphenyl 1,2-dioxygenase.

(113): One of the metal ligands, Glu 172, was displaced, thereby allowing it to abstract a proton from the thiohemiacetal carbon of the (S)-diastereomer. Given the pseudosymmetry of the active site, the second acid/base catalyst required to process the (R)-diastereomer could be the symmetrically related Glu 99 provided by the second $\beta\alpha\beta\beta\beta$ motif; Glu 99 might be released from metal ion ligation when the (R)-diastereomer of the thiohemiacetal binds (106). An analogous metal ion ligation and proton transfer mechanism could be operative in the methylmalonyl CoA epimerase–catalyzed reaction, for which biochemical evidence also suggests the participation of two acid/base catalysts (116, 117).

The tridentate coordination observed for Fe(II) in the dioxygenases provides sites for two substrate-derived ligands as well as diatomic oxygen; perhaps electron transfer from the Fe(II) to O_2 provides a superoxide radical anion that can initiate the oxygen insertion reaction that forms a lactone intermediate that ultimately is hydrolyzed by the metal-coordinated hydroxide anion (106). Although the mechanism of the FosA-catalyzed reaction is distinct from those of the dioxygenases, sequence alignments suggest tridentate coordination of the Mn(II). Armstrong has proposed a mechanism for this reaction in which the Mn(II) coordinates geminal phosphonate oxyanions, and a water in the third available coordination site participates as a general acid catalyst to assist in opening of the epoxide ring by the glutathione anion (106).

The pseudosymmetric structure provided by the pair of $\beta\alpha\beta\beta\beta$ motifs has provided a versatile template for the evolution of diverse metalloenzymes. In contrast to the barrel folds and others described here, the remarkable structural flexibility offered by the varied domain architectures within this superfamily adds to the opportunities for functional variation within the superfamily. A case in point is

the recent functional identification and assignment of a 2,6-dichlorohydroquinone dioxygenase, a ring cleavage enzyme required for pentachlorophenol degradation, to the VOC superfamily (118). Although functionally a dioxygenase, the mechanism of catalysis for this enzyme must vary from that of the well-characterized 1,2-extradiol dioxygenases because its *para* arrangement of the OH groups obviates metal ion chelation characteristic of all other extradiol dioxygenases assigned to the superfamily (105, 119). The discovery of this new theme within the structure-function paradigm represented by the VOC superfamily is supported by sequence analysis that shows the proteins of the 2,6-dichlorohydroquinone dioxygenase subgroup to be different in size than other known dioxygenases and likely to possess alternative domain arrangements containing conserved active site motifs typical of dioxygenases. Understanding how superfamily-consistent dioxygenase chemistry is accommodated by these predicted structural variations will require structural characterization of this enzyme.

Fe-Dependent Oxidases and Oxygenases

Recent studies of enzymes involved in the biosynthesis of β-lactam antibiotics, penicillins, and cephalosporins have revealed a remarkable superfamily of Fe(II)-dependent oxidases, some of which require 2-ketoglutarate (2KG) as a cosubstrate and some of which do not. Three members of this superfamily have been structurally characterized: the 2KG-independent isopenicillin *N* synthase (IPNS) that catalyzes the formation of both the β-lactam and thiazolidine rings from L-δ-(α-aminoadipoyl)-L-cysteine-D-valine (ACV; a four-electron oxidation) (120–122); the 2KG-dependent deacetoxycephalosporin synthase (DAOCS) that catalyzes expansion of the thiazolidine ring of penicillins to produce cephalosporins (a two-electron oxidation) (123); and the 2KG-dependent clavaminic acid synthase (CAS) that catalyzes three separate steps (hydroxylation, cyclization, and desaturation) in the biosynthesis of the β-lactam clavulanic acid (each two-electron oxidations) (124). The reactions catalyzed by these enzymes are shown in Figure 11; their structures are shown in Figure 12.

The structure determination of IPNS revealed a jelly-roll fold in which the active site is buried within a β-barrel formed by eight β-strands, substantially different, again, from other folds described above. These strands provide four ligands for the essential Fe(II). The Fe(II) binds dioxygen, and a ferryl intermediate is produced by the oxidative closure of the β-lactam ring; the ferryl intermediate then abstracts a hydrogen that leads to oxidative closure of the thiazolidine ring. The IPNS-catalyzed biscyclization reaction is a four-electron oxidation, so the diatomic oxygen substrate can be reduced by four electrons without the need for an external reductant.

The structure of DAOCS reveals homologs of the ligands for Fe(II) in IPNS in an analogous jelly-roll fold. Additionally, a binding site for 2KG was located, with an Arg that is conserved in other 2KG-dependent oxidases

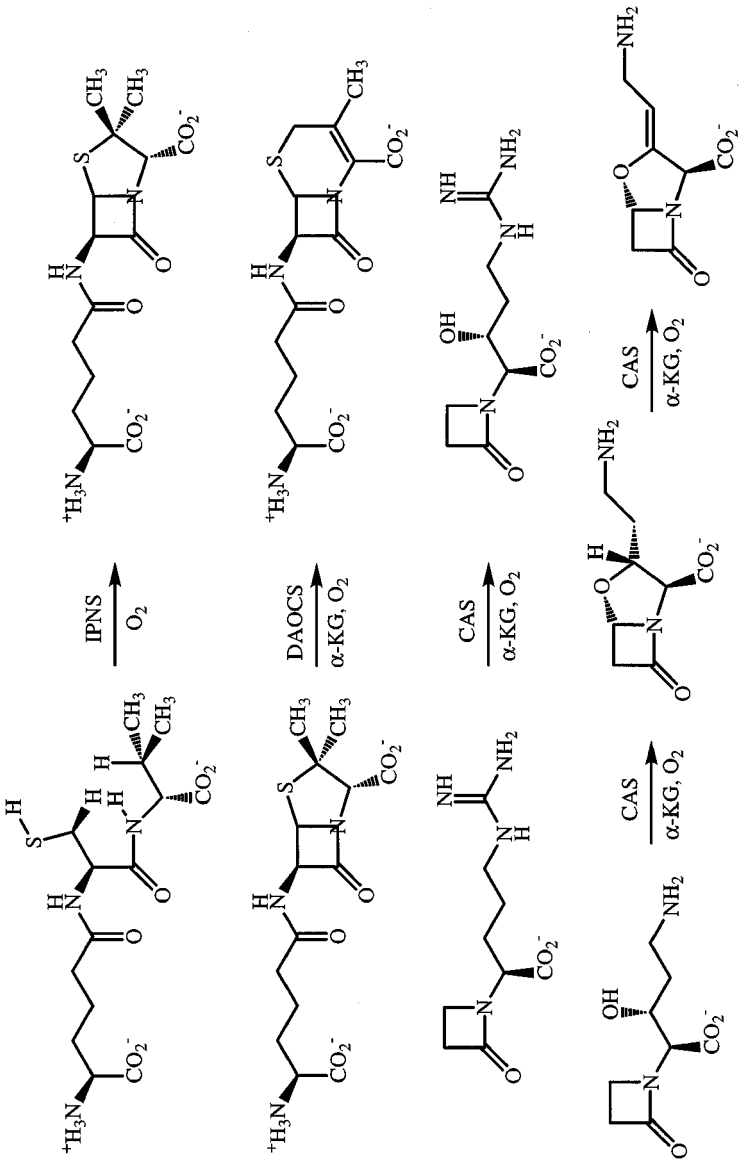

Figure 11 Reactions catalyzed by members of the Fe-dependent oxidase superfamily.

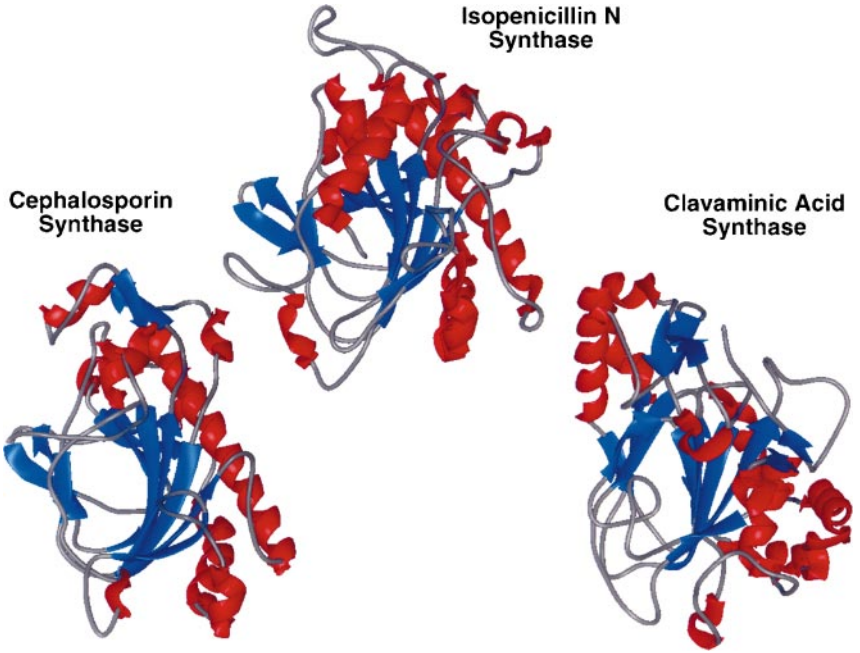

Figure 12 Structures of isopenicillin *N* synthase, deacetoxycephalosporin synthase, and clavaminic acid synthase.

forming a salt bridge with the 5-carboxylate group of the cosubstrate. In this case, the ferryl intermediate is formed by oxidative decarboxylation of 2KG; this intermediate then initiates the ring expansion reaction by hydrogen abstraction.

The structure of CAS again revealed homologs of the residues required for Fe(II) coordination and the Arg required for 2KG binding. In each of the three reactions catalyzed by CAS, the ferryl intermediate is formed by the oxidative decarboxylation of 2KG, with the ferryl intermediate mediating three different reactions in the biosynthetic pathway.

Sequence alignments suggest that the same jelly-roll fold and ligands for metal ion binding are retained in other 2KG-dependent oxidases, including prolyl-4-hydroxylase. Thus, divergent evolution has been used to generate a mechanistically diverse superfamily in which reactive ferryl intermediates are generated and used in a remarkable number of different oxidation and oxygenation reactions.

FUNCTIONALLY DISTINCT SUPRAFAMILIES

The existence of these and other mechanistically diverse enzyme *superfamilies* provides ample evidence that retention of the chemical mechanism can direct the divergent evolution of proteins. The existence of functionally distinct *suprafamilies*

would provide support for the other mechanisms that we outlined earlier, e.g. retention of substrate specificity and exploitation of active site architecture to catalyze different reactions. Both of these mechanisms for the divergent evolution of enzymatic function are illustrated in the three functionally distinct suprafamilies we describe in the following sections.

As described above, Horowitz's proposal that metabolic pathways evolved backwards requires that both an ancestral enzyme and those evolved from that scaffold share the ability to bind the same substrate/product molecule (13, 14). This proposal has been described both as retrograde evolution and substrate-driven evolution. Although few examples of divergent members of such superfamilies have been reliably identified thus far, pairs of enzymes in tryptophan and histidine biosynthesis provide two examples of substrate-driven evolution.

Successive Enzymes in Tryptophan Biosynthesis: PRAI and InGPS

Structures are available for the last three enzymes in the tryptophan biosynthetic pathway: the bifunctional phosphoribosylanthranilate isomerase:indoleglycerol phosphate synthase (PRAI:InGPS) from *E. coli* (125), the monofunctional PRAI from *Thermotoga maritima* (126), and the bifunctional tryptophan synthase (TS; the α-polypeptide releases indole from indoleglycerol phosphate, and the β-polypeptide forms tryptophan from indole and serine) from *Salmonella typhimurium* (127–130). PRAI, InGPS, and the α-polypeptide of TS share the $(\beta/\alpha)_8$ fold; although the α-polypeptide TS does not share significant sequence identity with PRAI or InGPS, InGPS and PRAI share 22% sequence identity. The observation of the shared $(\beta/\alpha)_8$ fold does not necessarily provide evidence for divergent retrograde evolution of the tryptophan biosynthetic pathway, given the widespread occurrence of this fold in many other enzymes; however, the sequence identity relating PRAI and InGPS does provide persuasive evidence that these are related by divergent evolution. The hypothesis that these diverged by retrograde evolution is also supported by the finding that no other divergent members of this superfamily are readily identified by sequence analysis. The reactions catalyzed by PRAI and InGPS are shown in Figure 13.

Although the structures of PRAI and InGPS were determined in the absence of active site ligands, the product of PRAI is the substrate for InGPS, so common structural elements in the $(\beta/\alpha)_8$ barrel likely are responsible for the conserved substrate/product binding. This implies that loops that connect the β-sheets in the barrels with the following α-helices contain the functional groups that direct the differing mechanisms for the two reactions. Fersht and his coworkers used this analysis as the basis for directed evolution experiments in which the barrel of InGPS domain from the *E. coli* bifunctional enzyme was used as the starting structure; combinatorial libraries of randomized sequences were introduced for the loops at the end of three β-sheets, and members of the libraries that could catalyze the PRAI reaction were identified by complementing a mutant of *E. coli*

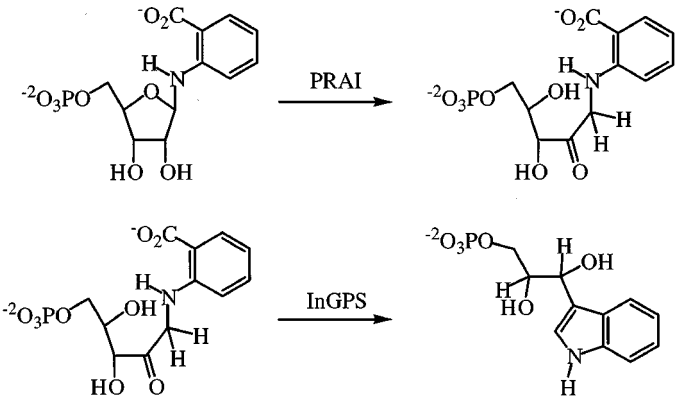

Figure 13 Reactions catalyzed by PRAI and InGPS.

that lacks the gene encoding PRAI (131). New, in vitro–evolved proteins were identified that could catalyze the PRAI reaction, with the best having a value for k_{cat}/K_m sixfold greater than that of the natural PRAI domain, although the value for k_{cat} was only 50% that of the natural PRAI; the evolved PRAI did not catalyze the InGPS reaction. The sequence of the evolved PRAI is 90% identical to the starting InGPS, in contrast to the 22% identity that describes the relationship between the enzymes evolved by nature, and 28% identical to the natural PRAI.

Although this experiment does not prove that the natural PRAI evolved from an InGPS progenitor by divergent evolution, it demonstrates that such evolution is feasible and provides strong evidence for the mechanism of divergent evolution in which substrate specificity is retained.

Successive Enzymes in Histidine Biosynthesis: PFACRI and ImGPS

Another striking example of evidence for the applicability of Horowitz's proposal is found in the histidine biosynthetic pathway with phosphoribosyl-formimino-5-aminoimidazole carboxamide ribonucleotide isomerase (PFACRI; HisA) and imidazole-3-glycerol phosphate synthase (ImGPS; HisF). Although PFACRI and ImGPS share 25% sequence identity (132, 133), only recently has structural evidence become available that convincingly argues for their relationship by divergent evolution. Like PRAI, PFACRI catalyzes an Amadori rearrangement; unlike InGPS, ImGPS catalyzes a Gln-dependent reaction that results in formation of the imidazole glycerol phosphate and AICAR, an intermediate in purine biosynthesis. The reactions catalyzed by PFACRI and ImGPS are shown in Figure 14.

The structures of the monofunctional PFACRI and ImGPS from *Thermotoga maritima* reveal that they share the $(\beta/\alpha)_8$ fold (134). Although the structures were determined in the absence of substrates or substrate analogs, the active sites can

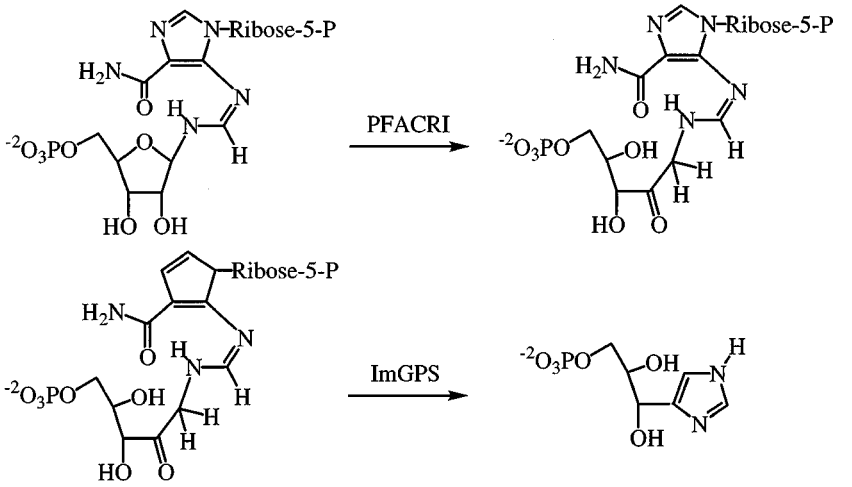

Figure 14 Reactions catalyzed by PFACRI and ImGPS.

be assigned to the usual position at the C-terminal ends of the barrels based on the locations of invariant residues as well as the positions of phosphate ions in the structure of ImGPS. The most striking features are two-fold repeated structural patterns in each barrel in which the first and last four strands contain symmetrically disposed loops of equal size and conformation. The substrates for both enzymes contain two phosphate groups, and the phosphate ions in the structure of ImGPS are bound to the loops following β-strands 3/4 and 7/8, emphasizing the utility of the two-fold symmetry in the barrel. Furthermore, several amino acids are conserved in all four half-barrels of PFACRI and ImGPS, an Asp in strands 1/5, a Val following strands 1/5, and three Gly (one connecting the α-helix 1 with β-strand 2, and two in a GG motif that follows β-strand 3 in each half-barrel), confirming the divergent sequence and structural relationships that relate not only the half-barrels but also PFACRI and ImGPS.

Surprisingly, ImGPS catalyzes, albeit weakly, the PFACRI reaction, although PFACRI does not catalyze the more chemically complex ImGPS reaction. Thus, there is little doubt that these enzymes are related by divergent evolution; the observed cross-reactivity is consistent with ImGPS being the progenitor of PFACRI by virtue of a catalytic promiscuity that has not been lost.

That both PRAI (TrpF) and PFACRI (HisA) catalyze Amadori rearrangements on structurally similar substrates (substituted 5′-phosphoribosyl amines) also raises the possibility that these are related to one another by divergent evolution, with the implication that divergent evolution is responsible for diverse chemistries in different metabolic pathways. That this scenario likely is the case is suggested by the observation that a single mutation at an active site allows PFACRI to catalyze the PRAI reaction (135).

OMPDC/R5PE/HUMPS Suprafamily: Different Enzymes in Different Metabolic Pathways

The previous sections describe a number of examples of divergent evolution of enzyme function in which retention of chemical mechanism (in mechanistically diverse superfamilies) and of substrate specificity (in retrograde evolution of metabolic pathways to produce functionally distinct suprafamilies) dominate the divergent evolution of new enzymatic activities. An important question that remains is whether a new enzyme can naturally evolve from a progenitor using the same active site functional group(s) in different mechanistic and metabolic contexts. An example of such divergent evolution would suggest that nature can make sizable functional leaps as it explores the structure-function landscape and that the number of progenitor enzymes can be more limited than previously imagined; for example, perhaps most, if not all, $(\beta/\alpha)_8$ barrels are the products of divergent evolution rather than independent evolutionary processes.

In unpublished studies, we recently discovered an example of such divergent evolution, although structural and functional studies are still at an early stage. Orotidine 5′-phosphate decarboxylase (OMPDC), the enzyme that catalyzes the final step in the de novo biosynthesis of UMP, recently has attracted much attention in the mechanistic enzymological community, given the realization that its reaction is the most proficient of any enzyme studied to date (136). As a result of this interest, structural information now has been obtained for OMPDCs from four species: *Saccharomyces cerevisiae* (137), *Methanobacterium thermoautotrophicum* (138), *B. subtilis* (139), and *E. coli* (140). All of these enzymes exhibit a dimeric structure in which the active sites are located at the C-terminal ends of $(\beta/\alpha)_8$ domains. Despite prior indications to the contrary, the active sites of the OMPDCs do not contain a divalent metal ion, and the reactions do not require divalent metal ions (141, 142). Although the sequences of these enzymes are highly diverged, each active site contains two conserved Lys and two Asp residues. The likely mechanism is an S_E2 reaction in which a destabilizing electrostatic interaction between one active site Asp and the substrate carboxylate group initiates loss of carbon dioxide; as the bond between the carboxylate group and carbon-6 is broken, a proton is transferred to carbon-6 from an active site Lys, thereby avoiding the formation of a vinyl anion intermediate that would be too unstable to be kinetically competent. One active site Lys is located at the end of the second β-sheet in the $(\beta/\alpha)_8$ barrel; the other three functional groups are found in an Asp-X-Lys-X-X-Asp motif located at the end of the third β-sheet.

PSI-BLAST and Shotgun searches of the sequence databases using the sequences of OMPDCs disclose a number of homologs, some of which have been functionally characterized. These include the biochemically characterized hex-3-ulose monophosphate synthase (HUMPS) that contains a structurally conserved homolog of the Asp-X-Lys-X-X-Asp in OMPDC, as noted by others (143), as well as the more distantly related but structurally characterized ribulose 5-phosphate epimerase (R5PE) (144). HUMPS is found in methylotropic organisms that use it

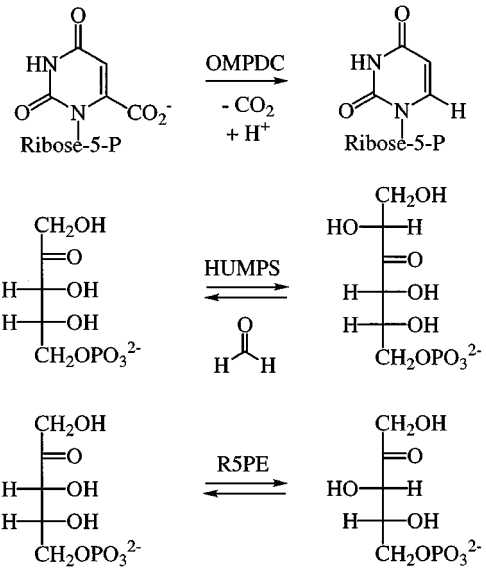

Figure 15 Reactions catalyzed by members of the OMPDC suprafamily.

to fix formaldehyde generated from methylamine or methanol by specific dehydrogenases/oxidases; R5PE catalyzes a familiar reaction in the ubiquitous pentose phosphate pathway. In contrast to OMPDC, HUMPS requires a divalent metal ion, Mg(II), for catalytic activity (145, 146); R5PE apparently does not require a divalent metal ion for activity (147, 148). Reactions catalyzed by OMPDC, HUMPS, and R5PE are shown in Figure 15.

The available biochemical and structural evidence suggests that even though the reactions catalyzed by HUMPS and R5PE share a common substrate, D-ribulose 5-phosphate, and can be envisaged to proceed via similar mechanisms, enolization of a ketose, these are members of a functionally distinct suprafamily. Equally surprising is the conclusion that although HUMPS and OMPDC share a sequentially conserved Asp-X-Lys-X-X-Asp motif, the reactions they catalyze use distinct substrates and necessarily proceed via distinct mechanisms. Thus, OMPDC, HUMPS, and R5PE are members of a functionally distinct suprafamily in which neither mechanism nor substrate specificity is retained.

This suprafamily serves as a cogent reminder that nature is not restricted to chemically and metabolically sensible strategies for divergent evolution; instead, divergent evolution is opportunistic, selecting progenitors that possess an active site architecture that can be used in mechanistically distinct catalytic functions. Given that the paradigms described by either retrograde evolution or chemistry-based evolution do not easily account for the functional diversity observed among the known members of the protein universe, jumps in enzyme evolution such as those suggested by the OMPDC suprafamily likely occurred multiple times. Members of the OMPDC suprafamily are under study in our laboratories.

The discovery of the OMPDC suprafamily further underscores the versatility of the $(\beta/\alpha)_8$ barrel fold in the divergent evolution of enzyme function: Active site scaffolds in one metabolic context may be useful in another context. This provides evidence in support of the hypothesis that enzymes that possess $(\beta/\alpha)_8$ barrel folds and catalyze different reactions may, in fact, be derived from a common or limited number of progenitor enzymes.

The existence of mechanistically diverse superfamilies and functionally distinct suprafamilies that share the $(\beta/\alpha)_8$ barrel fold further suggests that modification of this structure by directed evolution may allow the creation of new enzymes that may have applications in the synthesis of compounds that are not found in nature, including the construction of new metabolic pathways that will allow microorganisms to catabolize otherwise inert compounds (e.g. environmental contaminants) or synthesize compounds that have biomedical importance.

CONCLUSIONS

Although specificity-diverse superfamilies are familiar to biochemists, ample evidence exists for mechanistically diverse superfamilies and functionally distinct suprafamilies whose members do not catalyze the same chemical reactions. The discovery of these groups of enzymes that are related by divergent evolution has several important consequences:

1. In the case of mechanistically diverse superfamilies, the structural elements that are common to all members of the superfamily allow a succinct identification of the underlying structural strategy used to catalyze reactions that share a common mechanistic attribute (partial reaction, intermediate, or transition state). The synergistic study of several members of a mechanistically diverse superfamily can be expected to provide a more efficient solution to this goal than the more traditional approach of studying a single enzyme at a time.
2. Functional diversity is more complex than sequence diversity. Functional annotations of proteins identified in genome sequencing projects based on sequence similarity without regard to the existence of mechanistically diverse superfamilies or functionally distinct suprafamilies will lead to many assignments that are either misleading or incorrect.
3. Functional diversity is more complex than structural diversity. As such, knowledge of the three-dimensional structures of the complete repertoire of protein folds will be insufficient to assign specific functions to homologs that possess a given fold. This conclusion suggests that the ongoing ambitious efforts in structural genomics may have a more limited impact on determination and prediction of specific function than expected.
4. Genomic enzymology requires the interplay of both chemical (functional) and structural (and/or sequence) analysis, as neither structure-only-based

nor function-only-based investigations can provide an accurate picture of how enzyme structures evolve and deliver new functions. Thus, only through an explicit mapping between conserved elements of structure and conserved characteristics of function can the underlying evolutionary principles in enzyme superfamilies be discovered and verified.

Visit the Annual Reviews home page at www.AnnualReviews.org

LITERATURE CITED

1. Neidhart DJ, Kenyon GL, Gerlt JA, Petsko GA. 1990. *Nature* 347:692–94
2. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. 1997. *Nucleic Acids Res.* 25:3389–402
3. Karplus K, Barrett C, Hughey R. 1998. *Bioinformatics* 14:846–56
4. Pegg SC, Babbitt PC. 1999. *Bioinformatics* 15:729–40
5. Tatusov RL, Koonin EV, Lipman DJ. 1997. *Science* 278:631–37
6. Galperin MY, Walker DR, Koonin EV. 1998. *Genome Res.* 8:779–90
7. Murzin AG, Brenner SE, Hubbard T, Chothia C. 1995. *J. Mol. Biol.* 247:536–40
8. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM. 1997. *Structure* 5:1093–108
9. Pearl FM, Lee D, Bray JE, Sillitoe I, Todd AE, et al. 2000. *Nucleic Acids Res.* 28:277–82
10. Zhang C, DeLisi C. 1998. *J. Mol. Biol.* 284:1301–5
11. Govindarajan S, Recabarren R, Goldstein RA. 1999. *Proteins* 35:408–14
12. Wolf YI, Grishin NV, Koonin EV. 2000. *J. Mol. Biol.* 299:897–905
13. Horowitz NH. 1945. *Proc. Natl. Acad. Sci. USA* 31:153–57
14. Horowitz NH. 1965. In *Evolving Genes and Proteins*, ed. V Bryson, HJ Vogel, pp. 15–23. New York: Academic
15. Babbitt PC, Gerlt JA. 1997. *J. Biol. Chem.* 272:30591–94
16. Gerlt JA, Babbitt PC. 1998. *Curr. Opin. Chem. Biol.* 2:607–12
17. O'Brien PJ, Herschlag D. 1999. *Chem. Biol.* 6:R91–105
18. Babbitt PC, Mrachko GT, Hasson MS, Huisman GW, Kolter R, et al. 1995. *Science* 267:1159–61
19. Babbitt PC, Hasson MS, Wedekind JE, Palmer DR, Barrett WC, et al. 1996. *Biochemistry* 35:16489–501
20. Gerlt JA, Gassman PG. 1993. *Biochemistry* 32:11943–52
21. Gerlt JA, Gassman PG. 1993. *J. Am. Chem. Soc.* 115:11552–69
22. Landro JA, Gerlt JA, Kozarich JW, Koo CW, Shah VJ, et al. 1994. *Biochemistry* 33:635–43
23. Hasson MS, Schlichting I, Moulai J, Taylor K, Barrett W, et al. 1998. *Proc. Natl. Acad. Sci. USA* 95:10396–401
24. Helin S, Kahn PC, Guha BL, Mallows DG, Goldman A. 1995. *J. Mol. Biol.* 254:918–41
25. Larsen TM, Wedekind JE, Rayment I, Reed GH. 1996. *Biochemistry* 35:4349–58
26. Kenyon GL, Gerlt JA, Petsko GA, Kozarich JW. 1995. *Acc. Chem. Res.* 28:178–86
27. Palmer DR, Gerlt JA. 1996. *J. Am. Chem. Soc.* 118:10323–24
28. Palmer DR, Hubbard BK, Gerlt JA. 1998. *Biochemistry* 37:14350–57
29. Gulick AM, Palmer DR, Babbitt PC, Gerlt JA, Rayment I. 1998. *Biochemistry* 37:14358–68
30. Gulick AM, Hubbard BK, Gerlt JA, Rayment I. 2000. *Biochemistry* 39:4590–602
31. Wieczorek SJ, Kalivoda KA, Clifton JG, Ringe D, Petsko GA, Gerlt JA. 1999. *J. Am. Chem. Soc.* 121:4540–41

32. Tokuyama S, Hatano K. 1995. *Appl. Microbiol. Biotechnol.* 42:884–89
33. Tokuyama S, Hatano K. 1995. *Appl. Microbiol. Biotechnol.* 42:853–59
34. Palmer DR, Garrett JB, Sharma V, Meganathan R, Babbitt PC, Gerlt JA. 1999. *Biochemistry* 38:4252–58
35. Deleted in proof
36. Thompson TB, Garrett JB, Taylor EA, Meganathan R, Gerlt JA, Rayment I. 2000. *Biochemistry* 39:10662–76
37. Deleted in proof
38. Benning MM, Kuo JM, Raushel FM, Holden HM. 1994. *Biochemistry* 33:15001–7
39. Benning MM, Kuo JM, Raushel FM, Holden HM. 1995. *Biochemistry* 34:7973–78
40. Vanhooke JL, Benning MM, Raushel FM, Holden HM. 1996. *Biochemistry* 35:6020–25
41. Jabri E, Carr MB, Hausinger RP, Karplus PA. 1995. *Science* 268:998–1004
42. Jabri E, Karplus PA. 1996. *Biochemistry* 35:10616–26
43. Park IS, Michel LO, Pearson MA, Jabri E, Karplus PA, et al. 1996. *J. Biol. Chem.* 271:18632–37
44. Pearson MA, Michel LO, Hausinger RP, Karplus PA. 1997. *Biochemistry* 36:8164–72
45. Sharff AJ, Wilson DK, Chang Z, Quiocho FA. 1992. *J. Mol. Biol.* 226:917–21
46. Wang ZM, Quiocho FA. 1998. *Biochemistry* 37:8314–24
47. Wilson DK, Rudolph FB, Quiocho FA. 1991. *Science* 252:1278–84
48. Wilson DK, Quiocho FA. 1993. *Biochemistry* 32:1689–94
49. Holm L, Sander C. 1997. *Proteins* 28:72–82
50. Taylor TC, Andersson I. 1997. *J. Mol. Biol.* 265:432–44
51. Karplus PA, Pearson MA, Hausinger RP. 1997. *Acc. Chem. Res.* 30:330–37
52. Raushel FM, Holden HM. 2000. *Adv. Enzymol. Relat. Areas Mol. Biol.* 74:51–93
53. Boundy-Mills KL, de Souza ML, Mandelbaum RT, Wackett LP, Sadowsky MJ. 1997. *Appl. Environ. Microbiol.* 63:916–23
54. de Souza ML, Seffernick J, Martinez B, Sadowsky MJ, Wackett LP. 1998. *J. Bacteriol.* 180:1951–54
55. Sadowsky MJ, Tong Z, de Souza M, Wackett LP. 1998. *J. Bacteriol.* 180:152–58
56. Scanlan TS, Reid RC. 1995. *Chem. Biol.* 2:71–75
57. Buchbinder JL, Stephenson RC, Dresser MJ, Pitera JW, Scanlan TS, Fletterick RJ. 1998. *Biochemistry* 37:5096–106
58. Eklund H, Fontecave M. 1999. *Struct. Fold Des.* 7:R257–62
59. Eklund H, Eriksson M, Uhlin U, Nordlund P, Logan D. 1997. *Biol. Chem.* 378:821–25
60. Uhlin U, Eklund H. 1996. *J. Mol. Biol.* 262:358–69
61. Stubbe J, Riggs-Gelasco P. 1998. *Trends Biochem. Sci.* 23:438–43
62. Licht S, Gerfen GJ, Stubbe J. 1996. *Science* 271:477–81
63. Knappe J, Elbert S, Frey M, Wagner AF. 1993. *Biochem. Soc. Trans.* 21:731–34
64. Plaga W, Vielhaber G, Wallach J, Knappe J. 2000. *FEBS Lett.* 466:45–48
65. Knappe J, Wagner AF. 1995. *Methods Enzymol.* 258:343–62
66. Reddy SG, Wong KK, Parast CV, Peisach J, Magliozzo RS, Kozarich JW. 1998. *Biochemistry* 37:558–63
67. Parast CV, Wong KK, Kozarich JW, Peisach J, Magliozzo RS. 1995. *Biochemistry* 34:5712–17
68. Parast CV, Wong KK, Lewisch SA, Kozarich JW, Peisach J, Magliozzo RS. 1995. *Biochemistry* 34:2393–99
69. Broderick JB, Henshaw TF, Cheek J, Wojtuszewski K, Smith SR, et al. 2000. *Biochem. Biophys. Res. Commun.* 269:451–56
70. Kulzer R, Pils T, Kappl R, Huttermann J, Knappe J. 1998. *J. Biol. Chem.* 273:4897–903
71. Leppanen VM, Merckel MC, Ollis DL,

Wong KK, Kozarich JW, Goldman A. 1999. *Struct. Fold Des.* 7:733–44
72. Logan DT, Andersson J, Sjoberg BM, Nordlund P. 1999. *Science* 283:1499–504
73. Sun X, Eliasson R, Pontis E, Andersson J, Buist G, et al. 1995. *J. Biol. Chem.* 270:2443–46
74. Young P, Andersson J, Sahlin M, Sjoberg BM. 1996. *J. Biol. Chem.* 271:20770–75
75. Sun XY, Ollagnier S, Schmidt PP, Atta M, Mulliez E, et al. 1996. *J. Biol. Chem.* 271:6827–31
76. Riera J, Robb FT, Weiss R, Fontecave M. 1997. *Proc. Natl. Acad. Sci. USA* 94:475–78
77. Tauer A, Benner SA. 1997. *Proc. Natl. Acad. Sci. USA* 94:53–58
78. Deleted in proof
79. Leuthner B, Leutwein C, Schulz H, Horth P, Haehnel W, et al. 1998. *Mol. Microbiol.* 28:615–28
80. Xiang H, Luo LS, Taylor KL, Dunaway-Mariano D. 1999. *Biochemistry* 38:7638–52
81. Benning MM, Wesenberg G, Liu R, Taylor KL, Dunaway-Mariano D, Holden HM. 1998. *J. Biol. Chem.* 273:33572–79
82. Benning MM, Taylor KL, Liu RQ, Yang G, Xiang H, et al. 1996. *Biochemistry* 35:8103–9
83. Engel CK, Kiema TR, Hiltunen JK, Wierenga RK. 1998. *J. Mol. Biol.* 275:847–59
84. Engel CK, Mathieu M, Zeelen JP, Hiltunen JK, Wierenga RK. 1996. *EMBO J.* 15:5135–45
85. Modis Y, Filppula SA, Novikov DK, Norledge B, Hiltunen JK, Wierenga RK. 1998. *Structure* 6:957–70
86. Benning MM, Haller T, Gerlt JA, Holden HM. 2000. *Biochemistry* 39:4630–39
87. Murzin AG. 1998. *Curr. Opin. Struct. Biol.* 8:380–87
88. Wang J, Hartling JA, Flanagan JM. 1997. *Cell* 91:447–56
89. Bahnson BJ, Anderson VE. 1989. *Biochemistry* 28:4173–81
90. D'Ordine RL, Bahnson BJ, Tonge PJ, Anderson VE. 1994. *Biochemistry* 33:14733–42
91. Hofstein HA, Feng Y, Anderson VE, Tonge PJ. 1999. *Biochemistry* 38:9508–16
92. Muller-Newen G, Janssen U, Stoffel W. 1995. *Eur. J. Biochem.* 228:68–73
93. Yang G, Liang PH, Dunaway-Mariano D. 1994. *Biochemistry* 33:8527–31
94. Taylor KL, Liu RQ, Liang PH, Price J, Dunaway-Mariano D, et al. 1995. *Biochemistry* 34:13881–88
95. Yang G, Liu RQ, Taylor KL, Xiang H, Price J, Dunaway-Mariano D. 1996. *Biochemistry* 35:10879–85
96. Taylor KL, Xiang H, Liu RQ, Yang G, Dunaway-Mariano D. 1997. *Biochemistry* 36:1349–61
97. Xiang H, Dong J, Carey PR, Dunaway-Mariano D. 1999. *Biochemistry* 38:4207–13
98. Xiang H, Luo L, Taylor KL, Dunaway-Mariano D. 1999. *Biochemistry* 38:7638–52
99. Haller T, Buckel T, Retey J, Gerlt JA. 2000. *Biochemistry* 39:4622–29
100. Sharma V, Suvarna K, Meganathan R, Hudspeth ME. 1992. *J. Bacteriol.* 174:5057–62
101. Pelletier DA, Harwood CS. 1998. *J. Bacteriol.* 180:2330–36
102. Gasson MJ, Kitamura Y, McLauchlan WR, Narbad A, Parr AJ, et al. 1998. *J. Biol. Chem.* 273:4163–70
103. Eichler K, Bourgis F, Buchet A, Kleber HP, Mandrand-Berthelot MA. 1994. *Mol. Microbiol.* 13:775–86
104. Hawes JW, Jaskiewicz J, Shimomura Y, Huang B, Bunting J, et al. 1996. *J. Biol. Chem.* 271:26430–34
105. Laughlin LT, Bernat BA, Armstrong RN. 1998. *Chem. Biol. Interact.* 111:41–50
106. Armstrong RN. 2000. *Biochemistry* 39:13625–32
107. Kawano Y, Kumagai T, Muta K, Matoba

Y, Davies J, Sugiyama M. 2000. *J. Mol. Biol.* 295:915–25
108. Dumas P, Bergdoll M, Cagnon C, Masson JM. 1994. *EMBO J.* 13:2483–92
109. Han S, Eltis LD, Timmis KN, Muchmore SW, Bolin JT. 1995. *Science* 270:976–80
110. Kita A, Kita S, Fujisawa I, Inaka K, Ishida T, et al. 1999. *Struct. Fold Des.* 7:25–34
111. He MM, Clugston SL, Honek JF, Matthews BW. 2000. *Biochemistry* 39:8719–27
112. Cameron AD, Olin B, Ridderstrom M, Mannervik B, Jones TA. 1997. *EMBO J.* 16:3386–95
113. Cameron AD, Ridderstrom M, Olin B, Kavarana MJ, Creighton DJ, Mannervik B. 1999. *Biochemistry* 38:13480–90
114. Bergdoll M, Eltis LD, Cameron AD, Dumas P, Bolin JT. 1998. *Protein Sci.* 7:1661–70
115. Landro JA, Brush EJ, Kozarich JW. 1992. *Biochemistry* 31:6069–77
116. Fuller JQ, Leadlay PF. 1983. *Biochem. J.* 213:643–50
117. Leadlay PF, Fuller JQ. 1983. *Biochem. J.* 213:635–42
118. Xu L, Resing K, Lawson SL, Babbitt PC, Copley SD. 1999. *Biochemistry* 38:7659–69
119. Eltis LD, Bolin JT. 1996. *J. Bacteriol.* 178:5930–37
120. Roach PL, Clifton IJ, Fulop V, Harlos K, Barton GJ, et al. 1995. *Nature* 375:700–4
121. Roach PL, Clifton IJ, Hensgens CM, Shibata N, Long AJ, et al. 1996. *Eur. J. Biochem.* 242:736–40
122. Roach PL, Clifton IJ, Hensgens CM, Shibata N, Schofield CJ, et al. 1997. *Nature* 387:827–30
123. Valegard K, van Scheltinga AC, Lloyd MD, Hara T, Ramaswamy S, et al. 1998. *Nature* 394:805–9
124. Zhang Z, Ren J, Stammers DK, Baldwin JE, Harlos K, Schofield CJ. 2000. *Nat. Struct. Biol.* 7:127–33
125. Wilmanns M, Priestle JP, Niermann T, Jansonius JN. 1992. *J. Mol. Biol.* 223:477–507
126. Hennig M, Sterner R, Kirschner K, Jansonius JN. 1997. *Biochemistry* 36:6009–16
127. Rhee S, Miles EW, Davies DR. 1998. *J. Biol. Chem.* 273:8553–55
128. Hyde CC, Miles EW. 1990. *Bio-Technology* 8:27–32
129. Hyde CC, Ahmed SA, Padlan EA, Miles EW, Davies DR. 1988. *J. Biol. Chem.* 263:17857–71
130. Brzovic PS, Hyde CC, Miles EW, Dunn MF. 1993. *Biochemistry* 32:10404–13
131. Altamirano MM, Blackburn JM, Aguayo C, Fersht AR. 2000. *Nature* 403:617–22
132. Fani R, Lio P, Lazcano A. 1995. *J. Mol. Evol.* 41:760–74
133. Fani R, Tamburini E, Mori E, Lazcano A, Lio P, et al. 1997. *Gene* 197:9–17
134. Lang D, Thoma R, Henn-Sax M, Sterner R, Wilmanns M. 2000. *Science* 289:1546–50
135. Jurgens C, Strom A, Wegener D, Hettwer S, Wilmanns M, Sterner R. 2000. *Proc. Natl. Acad. Sci. USA* 97:9925–30
136. Radzicka A, Wolfenden R. 1995. *Science* 267:90–93
137. Miller BG, Hassell AM, Wolfenden R, Milburn MV, Short SA. 2000. *Proc. Natl. Acad. Sci. USA* 97:2011–16
138. Wu N, Mo YR, Gao JL, Pai EF. 2000. *Proc. Natl. Acad. Sci. USA* 97:2017–22
139. Appleby TC, Kinsland C, Begley TP, Ealick SE. 2000. *Proc. Natl. Acad. Sci. USA* 97:2005–10
140. Harris P, Poulsen JCN, Jensen KF, Larsen S. 2000. *Biochemistry* 39:4217–24
141. Cui W, DeWitt JG, Miller SM, Wu W. 1999. *Biochem. Biophys. Res. Commun.* 259:133–35
142. Miller BG, Smiley JA, Short SA, Wolfenden R. 1999. *J. Biol. Chem.* 274:23841–43
143. Traut TW, Temple BR. 2000. *J. Biol. Chem.* 275:28675–81
144. Kopp J, Kopriva S, Suss KH, Schulz GE. 1999. *J. Mol. Biol.* 287:761–71

145. Yasueda H, Kawahara Y, Sugimoto S. 1999. *J. Bacteriol.* 181:7154–60
146. Kato N, Ohashi H, Tani Y, Ogata K. 1978. *Biochim. Biophys. Acta* 523:236–44
147. Chen YR, Larimer FW, Serpersu EH, Hartman FC. 1999. *J. Biol. Chem.* 274:2132–36
148. Chen YR, Hartman FC, Lu TY, Larimer FW. 1998. *Plant Physiol.* 118:199–207