Divergence and Convergence in Enzyme Evolution

Michael Y. Galperin and Eugene V. Koonin

Comparative analysis of the sequences of enzymes encoded in a variety of prokaryotic and eukaryotic genomes reveals convergence and divergence at several levels. Functional convergence can be inferred when structurally distinct and hence non-homologous enzymes show the ability to catalyze the same biochemical reaction. In contrast, as a result of functional diversification, many structurally similar enzyme molecules act on substantially distinct substrates and catalyze diverse biochemical reactions. Here, we present updates on the ATP-grasp, alkaline phosphatase, cupin, HD hydrolase, and N-terminal nucleophile (Ntn) hydrolase enzyme superfamilies and discuss the patterns of sequence and structural conservation and diversity within these superfamilies. Typically, enzymes within a superfamily possess common sequence motifs and key active site residues, as well as (predicted) reaction mechanisms. These observations suggest that the strained conformation (the entatic state) of the active site, which is responsible for the substrate binding and formation of the transition complex, tends to be conserved within enzyme superfamilies. The subsequent fate of the transition complex is not necessarily conserved and depends on the details of the structures of the enzyme and the substrate. This variability of reaction outcomes limits the ability of sequence analysis to predict the exact enzymatic activities of newly sequenced gene products. Nevertheless, sequence-based (super)family assignments and generic functional predictions, even if imprecise, provide valuable leads for experimental studies and remain the best approach to the functional annotation of uncharacterized proteins from new genomes.

The availability of complete genome sequences of numerous bacteria, archaea, and eukaryotes has fundamentally transformed modern biology. With complete genomes, it is a realistic goal to catalog all proteins that are responsible for every essential cellular function, i.e. to create a "genomic parts list." Comparative genomics revealed a surprising flexibility of the key metabolic pathways, including numerous biochemical reactions catalyzed by highly diverged or previously uncharacterized enzyme forms (1). Due to the combination of computational and experimental approaches, many such "missing" enzymes have been identified and characterized in some detail, uncovering many cases in which consecutive steps of the pathway are catalyzed by enzymes with different evolutionary histories (reviewed in Refs. 1–5). In addition, it has been shown that many key biochemical steps can be catalyzed by two or more diverse, often unrelated enzyme forms (6,7), a phenomenon known as non-orthologous gene displacement. This patchy complex distribution of enzymes reflects a long evolutionary history of the enzymes with numerous events of gene duplication, followed by diversification, gene loss, and non-orthologous gene displacement, often via horizontal gene transfer (1,3). Here, we consider the two key processes in enzyme evolution, namely sequence divergence, which leads to functional diversification within the same protein superfamily, and functional convergence, which results in members of distinct superfamilies being recruited to catalyze the same metabolic reaction. We also briefly discuss how sequence comparison can assist the experimental research in enzymology.

Functional Diversification of Protein Superfamilies

Historically, proteins were unified in families based on sequence similarity (8). Protein families were combined into superfamilies based on similar catalytic activities, sequence motifs, and other conserved features (9,10). The rapid growth of protein structural data, brought about in part by the structural genomics initiatives, has put identification of protein superfamilies on a firm(er) basis. The current classifications of protein structural (super)families, implemented in the popular SCOP, CATH, and Dali databases, are generally compatible with each other despite the differences between the underlying methodologies (11–13). Furthermore, these superfamilies often correspond to sequence-based domain families (or clans) in the Pfam database (14) and contain conserved sequence motifs that are represented in such databases as InterPro (15). Therefore, proteins within the same superfamily can be confidently inferred to have evolved from a common ancestor, even though they might have dramatically different enzymatic activities or no (known) activity at all (16–18). Most studies on enzyme evolution consider evolution only within families of closely related enzymes, which typically involves changes in the enzymatic specificity without any major changes in protein structure. Here, we focus instead on the evolution of functional diversity within large protein superfamilies that are unified by common sequence motifs and structural cores. Table 1 lists catalytic activities and three-dimensional structures, where known, for members of five representative protein superfamilies that we discuss in this minireview and that, to the best of our knowledge, have not been recently reviewed from an evolutionary standpoint. These superfamilies span a wide range of sequence and structure conservation and provide multiple examples of divergence and convergence in the evolution of enzymes. We use these examples as leads for a general discussion of evolutionary trends in enzymes (see Refs. 10, 17, and 19–25 for in-depth reviews of several other enzyme superfamilies).
TABLE 1

Common features of proteins from several structural superfamilies

| ATP-grasp superfamily | Common traits of superfamilies |
|------------------------|-------------------------------|
| Glutathione synthetase (1gsq, 2hsq), d-alanine-d-Ala ligase (1dov), d-Ala-d-Lactate ligase (1e1e), biotin carboxylase (1dv1), carbamoyl-phosphate synthase (1jbd), pyruvate-phosphate dikinase (1dpk), phosphoribosylamine-glycine ligase PurD (1gsq), phosphoribosylglycinamidase formyltransferase PurT (1eyn), N2-carboxyanhydrinominodiacid ribonucleotide synthase PurK (1b6s), 5-formiminotriose-4-carboxamide ribonucleotide synthase PurP (2f7k), tubulin-tyrinosine ligase, tubulin glycase, tubulin polyglutamyase, ribosomal protein S6-glutamate ligase Rmk, succinate-CoA ligase (1jik), ATP-citrate synthase (3mwd), malate-CoA ligase, synapsin (1aux), α-α-mannosidase (1aap), N-acetylaspartylglutamate synthase, β-citrylglycine synthase, nikkomycin biosynthesis carboxylase SanS, intrinsol, 1,3-trisphosphate 5/6-kinase (1z2n), mycosporine glycin synthetase SynA (3856) | Common structural core (<4.3 Å Ca r.m.s.d. on ≥230 aa); common ATP-binding residues, which include two conserved lys/Arg residues that bind α- and β-phosphates of ATP, Gln/Asp residue that interacts with adenine amino group and N6 atom, hydrophobic residues that bind adenine ring, and three Gx/Ax residues that coordinate Mg2+ ions; common catalytic mechanism that includes formation of phosphoryl intermediate |

| AlkP superfamily | Common structural core (<3.6 Å Ca r.m.s.d. on ≥220 aa); conserved metal (Zn2+, Mn2+, or Mg2+)-binding His and Asp residues; common catalytic mechanism that includes phosphorylation (sulfatation) of active site Ser/Thr/Gly residue |
|-------------------|---------------------------------------------------------------|
| Alkaline phosphatase (1alk), phosphorylase mutase (1o98, 2zkt), oxalate oxidase (1fi2), oxalate decarboxylase (1uw8), gentisate 1,2,5,3-acetylaspartylglutamate synthase, ATP-citrate synthase (3mwd), malate-CoA ligase, synapsin (1aux), α-α-mannosidase (1aap), N-acetylaspartylglutamate synthase, β-citrylglycine synthase, nikkomycin biosynthesis carboxylase SanS, intrinsol, 1,3-trisphosphate 5/6-kinase (1z2n), mycosporine glycin synthetase SynA (3856) | Conserved structural core (<4.6 Å Ca r.m.s.d. on >99 aa); partly conserved metal (Mn2+, Fe2+, Cu2+, Ni2+, or Zn2+)-binding His and Asp residues; common catalytic mechanism that includes binding of dioxygen to metal atom and substrate with formation of peroxydic intermediate |

| Cupin superfamily | Common structural core (<3.6 Å Ca r.m.s.d. on >105 aa); conserved metal (Mn2+, Mg2+, Co2+, or Fe2+)-binding His and Asp residues organized into H2X1–5HD, 5-exoribonuclease YhaM, uridylyl-removing enzyme GlnD, myo-inositol oxidase MioX (2huo) |
|-------------------|---------------------------------------------------------------|
| Oxalate oxidase (1fz2), oxalate decarboxylase (1uw8), gentisate 1,2-dioxygenase (2dfo), homogentisate 1,2-dioxygenase (1eyz), 3-hydroxyanthranilate 3,4-dioxygenase (1lyu), cytochrome dioxygenase (3e1n), quinolinate 2,3-dioxygenase (1lajn), acetylacetone dioxygenase (3b0l), 1,2-dihydropyridine-3-keto-5-thiophenepentone (acireductone) dioxygenase (1v1s), 1-hydroxy-2-naphthoate dioxygenase, phosphomannomannose isomerase (1iq7), glucose-6-phosphate isomerase (1qy4), D-lyxose isomerase, 5-keto-4-deoxyuronyl isomerase, Dihd1 (1xru), ATDP-4-dehydrodromannose 3,5-epimerase RmcI (1dzr), ATDP-4-keto-6-deoxyglucose 5-epimerase Evad (1o1d), ATDP-6-deoxy-3,4-ketohexulose isomerase FdtA (2pa7), ectoine synthase, ureidoglycolate hydrolase (1vqc), hydroxypropylphosphonic acid epoxidase (2bnm), dimethylsulfoniopropionate lyase DddL, phaseolin (2phl), carnosine synthase, N-acetyltransferase (2x1c), -L-glutamate ligase MptN, N-acetylaspartylglutamate synthase, β-citrylglycine synthase, nikkomycin biosynthesis carboxylase SanS, intrinsol, 1,3-trisphosphate 5/6-kinase (1z2n), mycosporine glycin synthetase SynA (3856) | Conserved structural core (<4.6 Å Ca r.m.s.d. on >99 aa); partly conserved metal (Mn2+, Fe2+, Cu2+, Ni2+, or Zn2+)-binding His and Asp residues; common catalytic mechanism that includes binding of dioxygen to metal atom and substrate with formation of peroxydic intermediate |

| HD domain phosphohydrolase superfamily |
|----------------------------------------|
| 3'-5'-cAMP/cGMP phosphodiesterase (2bd1), [pppGpp]hydrolase (1v77), cyclic di-GMP phosphodiesterase (3m8t), exopolyphosphatase (1udw), dNMP 5'-nucleotidase YbhR (2p2r), dNTP triphosphohydrolase (2d6q), dGTPase (3bg2), cytidine hydrolase, 7,8-dihydo-2-deoxy-2-norotepin 2',3'-cyclic phosphate phosphodiesterase M(19873), 2',3'-cAMP/cGMP hydrolase, 3'-5' exoribonuclease YhaM, uridylyl-removing enzyme GlnD, myo-inositol oxidase MioX (2huo) | Conserved structural core (<3.6 Å Ca r.m.s.d. on >105 aa); conserved metal (Mn2+, Mg2+, Co2+, or Fe2+)-binding His and Asp residues organized into H2X1–5HD, 5-exoribonuclease YhaM, uridylyl-removing enzyme GlnD, myo-inositol oxidase MioX (2huo) |

| Ntn hydrolase superfamily |
|---------------------------|
| Penicillin acylase (1pnl), glutamine 5-phosphoribosyl-1-phosphate amidotransferase (1acc), proteasome subunit (1plm), glycerol-3-phosphate synthase (1x2f), protease HslV (1mv9), aspartylglycaminidase (1aap), γ-glutamyltranspeptidase (2dg5), asparagine synthetase B (1ct9), β-lactam synthetase (1lgt), glutamate synthase (1e4a), t-asparaginase (2zez), threonine aspartase (2a8h), acyl-CoA:lysophosphatidic acid acyltransferase (2x1c), bile salt hydrolase (2zhe), N-acetylamidase lactone acylase PtdQ (2wyb), acid ceramidase, IMP cyclohydrolase PurO (2ntk) | Common structural core (<4.1 Å Ca r.m.s.d. on >96 aa) decorated with variety of structural elements; sequence conservation limited to N-terminal β-hairpin that contains catalytic Ser, Cys, or Thr residue |

ATP-grasp—The original description of the ATP-grasp superfamily featured five enzymes with very similar three-domain structures (each featuring an α+β-sandwich) and several other enzymes assigned to that superfAMILY based solely on conserved sequence motifs (26, 27). Since then, crystal structures of some of these enzymes have been solved, confirming sequence-based predictions and expanding the ATP-grasp superfamily to include, among others, a variety of peptide synthetases (amino acid ligases) (28); tubulin glycase and tubulin polyglutamyase, which regulate ciliary motility; and the synthetases of carnosine and N-acetylasparglumylate, dipeptides that are abundant in muscle and brain tissues, respectively (Table 1). In addition, the ATP-grasp fold and the conserved mode of ATP binding, albeit without an apparent enzymatic activity, have been identified in synapsin I, a regulator of neurotransmitter release.
The diverse members of the ATP-grasp superfamily share the conserved structural fold, typically retain a similar arrangement of the active site residues, and appear to have a common reaction mechanism that includes interaction of ATP with a carboxyl group of one substrate, followed by formation of a phosphoacyl intermediate and nucleophilic attack by an amino group of the second substrate (27). Until recently, the only deviations from this pattern were succinyl-CoA synthetase and pyruvate-phosphate dikinase, which form phosphohistidine intermediates that are attacked by a thiol or carbonyl group, respectively. However, two newly described ATP-grasp enzymes act on substrates that contain hydroxyl groups instead of carboxyl groups. Inositol-1,3,4-trisphosphate 5/6-kinase catalyzes phosphorylation of a hydroxyl group at position 5 or 6 of the inositol ring with a likely involvement of a phosphohistidine intermediate (29), whereas mycosporine glycine synthetase has a 4-deoxygadusol substrate that contains a hydroxyl group attached to a aromatic ring (30). These examples reveal substantial plasticity of the ATP-grasp fold that allows its members to evolve a variety of specific activities while preserving the key features of the superfamily.

Alkaline Phosphatases—Enzymes of the AlkP (alkaline phosphatase) superfamily share a core domain that consists of an eight-strand β-sheet surrounded by α-helices (31, 32). The recent expansion of this superfamily included both new enzymes and identification of new enzymatic activities in the known members of the superfamily. The family prototype, Escherichia coli AlkP, has been shown to possess phosphodiesterase, phosphonate monoesterase, and even phosphophosphate-dependent hydrogenase activities, in addition to its well-known phosphatase and sulfatase activities (33). Conversely, the enzyme originally characterized as a phosphonate monoesterase has been shown to have also phosphatase, sulfatase, sulfonate monoesterase, and phosphodiesterase activities (34). This catalytic promiscuity appears to be a characteristic feature of the AlkP superfamily enzymes (35). However, this is not a property of the entire superfamily, as a recently described member appears to be a highly specific inorganic pyrophosphatase (36). Other highly specific members of the AlkP superfamily are phosphotransferases that transfer phosphoglycerol, phosphothanolamine, and phosphocholine moieties of the respective phospholipids to such acceptors as bacterial lipopolysaccharide or eukaryotic glycosylphosphatidylinositol (37). A particularly important example is the lipoteichoic acid synthase (phosphoglycerol transferase) LtaS, an essential enzyme in Gram-positive bacteria and a potential drug target (38, 39).

Despite the variety of their catalytic activities, AlkP superfamily members share a conserved structural fold (decorated with a variety of additional structural elements), similarly organized active sites, and the general catalytic mechanism that includes phosphorylation (or sulfation) of the active site residue, which can be Ser, Thr, or formylglycine (formed post-translationally from Cys or Ser). In phosphopentomutase, the phosphorylated Thr residue appears to be present in the ground state, leading to the suggestion that the substrate enters this enzyme at a different point in the catalytic cycle than in AlkP (40). A phosphorylated Thr residue has also been reported in the active site of LtaS (38).

Cupins—The cupin superfamily, together with the 2-ketoglutarate- and iron-dependent dioxygenase superfamily, belongs to the double-stranded β-helix fold, and members of both superfamilies have been occasionally referred to as cupins (41, 42). However, even cupins sensu stricto are extremely diverse, ranging from metal-binding proteins with dioxygenase, hydroxylase, and other activities to sugar isomerases (epimerases), some of which are metal-dependent and some not, to catalytically inactive seed storage and sugar-binding proteins. A recent analysis of the evolution of this fold suggested an early divergence of metal-dependent and metal-independent cupins with subsequent re-emergence of metal binding in various lineages (43). For metal-dependent cupins, the proposed reaction mechanisms typically include sequential binding of the substrate and dioxygen to the catalytic divalent metal cation (44).

HD Domain Phosphohydrolases—Members of the HD domain superfamily were originally described as (putative) metal-dependent phosphatases and phosphodiesterases (45). However, this superfamily also included a Zn2+-dependent cyanamide hydratase (urea hydro-lyase), which suggested that it might possess additional catalytic activities (45). In the past several years, the HD domain has been identified in several phosphohydrolases, including the widespread HD-GYP domain phosphodiesterase that specifically hydrolyzes bacterial second messenger cyclic di-GMP (46–48). In addition, structural comparisons unexpectedly identified this domain in the iron-dependent enzyme myo-inositol oxygenase (49). Structures of more than a dozen HD domain-containing enzymes have been solved by structural genomics projects. However, few of these enzymes have been biochemically characterized, so the full range of catalytic activities evolved in this superfamily remains unknown. A plausible catalytic mechanism has been proposed for the 5′-nucleotidase (50) and might prove applicable to the whole superfamily.

N-terminal Nucleophile Hydrolases—The N-terminal nucleophile (Ntn)2 hydrolyase superfamily unifies diverse amidohydrolases that share a four-layered αβαβ structure and a common catalytic mechanism but do not have recognizable sequence similarity (51). Members of this superfamily are typically synthesized as catalytically inactive precursors that undergo autocatalytic processing to generate active enzymes. Their common reaction mechanism includes deprotonation of the hydroxyl or thiol group of the side chain of the N-terminal residue (Ser, Thr, or Cys) of the enzyme molecule by the free amino group of the same residue (52, 53). This stage is followed by nucleophilic attack on the carbonyl carbon of the amide bond of the substrate, formation of an acyl-enzyme intermediate coupled with the release of an amino group-containing part of the substrate, and subsequent hydrolysis of the acyl-enzyme, leading to the release of the carbonyl group-containing portion of the substrate (reviewed in Refs. 53 and 54).

The Ntn hydrolyase-like fold is also present in the archaeal IMP cyclohydrolase PurO, which catalyzes the final step of purine biosynthesis (55). This enzyme retains all the structural

---

2 The abbreviations used are: Ntn, N-terminal nucleophile; NISE, non-homologous isofunctional enzyme; AMP-PCP, adenosine 5′-(β,γ- methylene)diphosphonate.
features of the Ntn hydrolase superfamily but is not proteolytically processed, lacks a nucleophilic residue at the N terminus, and does not function as an amidohydrolase. Accordingly, the SCOP database assigns it to a separate superfamily (11). This enzyme is found only in a small set of methano- and haloarchaea and represents an unusual variant of extreme divergence within the common structural core.

Two other enzymes, DmpA (l-aminopeptidase p-Ala-esterase/amidase) and ornithine acetyltransferase, share with the Ntn hydrolase superfamily the αβα-structure and catalyze a similar amidohydrolase reaction; the former enzyme also undergoes proteolytic activation. However, these proteins display a substantially different directionality and connectivity of the structural elements, indicating that their similarity to Ntn hydrolases results from convergent rather than divergent evolution (56).

### Did Evolution Favor Conservation of Entatic State?

Although the abovementioned enzyme superfamilies have been defined based primarily on the structural similarity of their members, most of these members share additional properties beyond the structural fold. Such conserved features include the overall organization of the active sites, conservation of certain (although not all) active site residues, and (where known) common reaction intermediates (Table 1). To discuss the interplay of common and unique features among enzymes, it is instrumental to consider the concept of the entatic state that was originally proposed by Vallee and Williams in 1968 (57) and developed in greater detail in subsequent reviews (58, 59). The term “entatic,” meaning a stretched (or otherwise stressed) state, was used to describe “a catalytically poised state intrinsic to the active site.” This concept implied “the possibility that enzymes might be poised for catalytic action in the absence of substrate, i.e. are in an entatic state” (57). The authors acknowledged the difficulties in the interpretation of the potential indications of the entatic state, such as an exceptionally high reactivity or anomalous pK_a values of particular amino acid side chains, for most (non-metallo)enzymes (57) and concentrated on demonstrating the existence of the entatic state for catalytically active metal atoms (58, 59). As a result, entatic state is often viewed as a specific property of metalloenzymes, despite well documented instances of steric strain and perturbed pK_a values in a variety of enzyme active sites (60–62).

The concept of entatic state helps to define the characteristic features of an enzyme superfamily and explain their evolutionary conservation. Each member of the superfamily has its own range of substrates that need to be tightly bound, attacked, brought to the transition complex stage, and finally converted into the products. Although some amino acid residues, apparently those responsible for the unique specificity of the enzyme, vary from one enzyme family to another, certain residues are conserved within the superfamily as a whole. Although conservation of certain residues, e.g. glycines in the various cupin domains, appears to be related to the unique folding patterns of the respective proteins, the most conserved residues in the ATP-grasp superfamily are responsible for binding ATP; in the AlkP, cupin, and HD domain superfamilies for binding active site metal ions; and in the Ntn hydrolase superfamily for providing the N-terminal nucleophile and the oxyanion hole (Fig. 1 and Table 1). In these five superfamilies, sequence conservation apparently extends to the residues that directly participate in the initial attack on the substrate and stabilization of the transition complex. The proper positioning of these residues is provided by a variety of conserved structural elements. In ATP-grasp enzymes, for example, these include a helix-turn-helix structure connecting the first two domains (63, 64); a conserved flexible loop with a sharp turn (designated the T-loop by Thoden et al. (65)), which follows the ATP/α-phosphate-binding Lys/Arg residue (Lys-136 in Fig. 1A); a cis-peptide bond in the backbone just upstream of that Lys/Arg residue; and other rare structural features. As a result, members of these superfamilies typically share the initial stages of the catalytic process. On the other hand, the breakdown of the transition complex in different enzymes (or, as discussed by Jencks (66), in the same enzyme under different conditions) can follow a number of different paths, yielding, for example within the AlkP superfamily, substrate hydrolysis, isomerization, or phosphate group transfer (see also Refs. 10 and 19–21).

One could argue that the emergence of each distinct entatic state configuration was a major evolutionary event, opening the door to the utilization of new classes of substrates or to the catalysis of new classes of reactions. During the subsequent evolution, major changes in protein structure were restricted by the likelihood of the formation of toxic (or inactive) misfolded molecules (67). Thus, only those sequence changes would prove viable that preserved the structural fold and accordingly the mechanism of formation of the same entatic state. These constraints led to the formation of series of structurally and catalytically (albeit not necessarily functionally) related protein molecules, which later evolved into the current superfamilies. A somewhat similar conclusion was reached by Warshel and Florián (68), who singled out pre-oriented dipoles as the source of the catalytic power of enzymes and argued that evolutionary optimization of enzymes increased their “preorganization effect”, i.e. the ability of enzymes to “minimize the reorganization energy associated with the formation of the charged transition state.”

### Practical Aspects of Superfamily Assignment

The conservation of catalytic elements within enzyme superfamilies makes sequence analysis an extremely useful tool in enzymology: assignment of a poorly characterized enzyme to a specific superfamily immediately predicts the structural fold, active site residues, a range of its potential catalytic activities, and even the likely catalytic mechanism. This could be particularly valuable for enzymes with complex substrates, for which direct assays are complicated and cumbersome. Thus, measuring the activity of tubulin-modifying enzymes, which are involved in tumor progression and have a vital role in neuronal organization, is certainly not an easy task. The assignment of tubulin-tyrosine ligase to the ATP-grasp superfamily (27) led to prediction of its active site residues and suggested a plausible catalytic mechanism for this enzyme (69). Likewise, assignment of the glycosylphosphatidylinositol phosphoethanolamine transferase PIG-N (McD4) to the AlkP superfamily was instrumental for the studies of this and related enzymes (37).

---

3 M.Y. Galperin, unpublished data.
Superfamily assignments proved most valuable when used for the analysis of metabolic pathways where the nature of the "missing" enzyme could be used to look for suitable candidates among uncharacterized genes (3, 4). Reconstruction of purine biosynthesis in archaea, which included characterization of the PurP and PurO gene products (55, 70), provides an impressive example of the power of the integrative approach that combines sequence analysis with biochemical assays and structural studies (reviewed in Ref. 71).

Superfamily assignments could also be useful for the functional annotation of newly sequenced genes that do not show clear sequence similarity to any well characterized enzymes. In such cases, searching new gene products against superfamily-specific sequence profiles (available, for example, in the NCBI Conserved Domain Database (72)) provides hints that can be used for generic functional prediction and as guidance for subsequent experiments, e.g. by predicting catalytic residues that are targets of choice for site-specific mutagenesis. For example, identification of the cupin domain in the sequence of the vitamin K-dependent γ-glutamyl carboxylase (residues 524–625 of VKGC_HUMAN (15)) could open new avenues for studying this important but still enigmatic enzyme.

**Convergent Evolution: Similar Active Sites in Analogous Enzymes**

Diversification of enzyme families can result in functional overlap when members of two or more distinct families end up catalyzing the same biochemical reaction. In some cases, such
enzyme isoforms are distantly related, and the low sequence similarity conceivably stems from rapid divergence of homologous protein sequences that accompanies adaptation to different environmental conditions. Examples of such enzyme pairs include the thermostable and mesophilic forms of β-glucosidase and adenylate kinase, which have retained very similar structures but share only a limited number of conserved residues (7).

There are cases, however, in which distinct enzyme forms catalyzing the same reaction share no detectable sequence similarity or even belong to two or more distinct structural superfamilies or folds (6, 7). The best known examples include superoxide dismutase, for which four distinct structural forms have been described, and cellulase, which is found in at least five structurally distinct forms. For such analogous (as opposed to homologous) enzymes, adoption of different structural folds indicates independent evolutionary origins; we have recently proposed a more precise designation for these enzymes, non-homologous isofunctional enzymes (NISEs) (7).

As in the textbook example of trypsin and subtilisin, diverse enzymes that act on related substrates might still share similarities in the organization of their active sites. Such similarities have been noted, for example, in the similar configurations of the ATP-binding residues in the ATP-grasp enzyme D-Ala-D-Ala ligase and enzymes that adopt two other folds, cAMP-dependent protein kinase and ribonucleotide reductase (73). A subsequent comparison of the adenine-binding sites revealed a common structural framework with similar polar and hydrophobic interactions in representatives of eight different folds (74). A similar pattern of structural convergence of evolutionarily unrelated enzymes has been revealed in the organization of pyridoxal phosphate-interacting residues of pyridoxal phosphate-dependent enzymes representing five distinct folds (75). Similar examples of functional convergence can be seen in NISEs, which, by definition, act on the same substrates. A recent comparison of the enzyme-substrate complexes of the phosphorylated chemotaxis protein CheY with two structurally distinct phosphatases, CheZ and CheX, revealed a very similar organization of the catalytic residues involved in the dephosphorylation of phospho-CheY (76).

An interesting evolutionary feature is the often skewed phylogenetic distribution of distinct isoforms of the same enzyme. For example, the archaeal shikimate kinase (77) has not yet been detected outside of the archaeal domain, whereas the other form of shikimate kinase is found in bacteria and eukaryotes. Similarly, the recently described cupin form of glucose-6-phosphate isomerase is found only in certain bacteria and archaea, whereas the other form of this enzyme, a member of the sugar isomerase family, is widespread. For the cases in which a particular enzyme is confined to a certain taxonomic group, recent evolutionary emergence from an enzyme of a different specificity seems to be the easiest explanation (7). Archaeal shikimate kinase, for example, is a member of the GHMP kinase superfamily and could have evolved from homoserine kinases or similar enzymes (77).


text continues...

Acknowledgments—We thank Drs. Armen Mulkidjanian, Alexey Murzin, Andrei Osterman, and R. J. P. Williams for helpful discussions.

REFERENCES
1. Koonin, E. V., and Galperin, M. Y. (2002) Sequence-Evolution-Function: Computational Approaches in Comparative Genomics, Kluwer Academic Publishers, Boston
2. Galperin, M. Y., and Koonin, E. V. (1999) Genetica 106, 159–170
3. Osterman, A., and Overbeek, R. (2003) Curr. Opin. Chem. Biol. 7, 238–251
MINIREVIEW: Divergence and Convergence in Enzyme Evolution

76. Pazy, Y., Motaleb, M. A., Guarnieri, M. T., Charon, N. W., Zhao, R., and Silversmith, R. E. (2010) Proc. Natl. Acad. Sci. U.S.A. 107, 1924–1929
77. Daugherty, M., Vonstein, V., Overbeek, R., and Osterman, A. (2001) J. Bacteriol. 183, 292–300
78. Jeffery, C. J. (2009) Mol. BioSyst. 5, 345–350
79. Huberts, D. H., and van der Klei, I. J. (2010) Biochim. Biophys. Acta 1803, 520–525
80. Piatigorsky, J. (2003) J. Struct. Funct. Genomics 3, 131–137
81. Sriram, G., Martinez, J. A., McCabe, E. R., Liao, J. C., and Dipple, K. M. (2005) Am. J. Hum. Genet. 76, 911–924
82. Wang, Y., Addess, K. J., Chen, J., Geer, L. Y., He, J., He, S., Lu, S., Madej, T., Marchler-Bauer, A., Thiessen, P. A., Zhang, N., and Bryant, S. H. (2007) Nucleic Acids Res. 35, D298–D300
83. Gibrat, J. F., Madej, T., and Bryant, S. H. (1996) Curr. Opin. Struct. Biol. 6, 377–385
84. Murphy, J. E., Stec, B., Ma, L., and Kantrowitz, E. R. (1997) Nat. Struct. Biol. 4, 618–622
85. Simmons, C. R., Krishnamoorthy, K., Granett, S. L., Schuller, D. J., Dominy, J. E., Jr., Begley, T. P., Stipanuk, M. H., and Karplus, P. A. (2008) Biochemistry 47, 11390–11392