Understanding Enzyme Superfamilies

CHEMISTRY AS THE FUNDAMENTAL DETERMINANT IN THE EVOLUTION OF NEW CATALYTIC ACTIVITIES

Patricia C. Babbitt and John A. Gerlt

From the Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94143-0466 and Department of Biochemistry, School of Chemical Sciences, University of Illinois, Urbana, Illinois 61801

Prior to the discovery in 1990 that mandelate racemase (MR) and muconate-lactonizing enzyme (MLE) are structurally similar enzymes that catalyze different overall reactions (1), structurally related enzymes were assumed to catalyze identical chemical reactions but, perhaps, with distinct substrate specificities. For example, all of the members of the serine protease superfamily were known to catalyze the same chemistry, hydrolyses of peptide bonds, although their peptide substrates varied. As described by Craik and Perona in the previous minireview (2), evolutionary accommodation of these differences in substrate specificity can result in major reorganization of the associated structures.

In this minireview, we discuss four recently discovered enzyme superfamilies in which an alternate theme predomnates; within each of these superfamilies, the member proteins share a common structural scaffold but catalyze different overall reactions. For each of the superfamilies described, the active sites are contained within a single homologous domain. Although they represent several distinct family folds, the enzyme functions in each superfamily are related to their respective structural scaffolds in the same way; the proteins within each superfamily utilize a common mechanistic strategy for lowering the free energies of the rate-limiting transition states in the reactions they catalyze. The existence of several examples of such superfamilies lends further credence to the principle that the evolution of new catalytic activities involves the incorporation of new catalytic groups within an active site while retaining those groups necessary to catalyze the partial reaction common to all of them (3–5). As a consequence, the range of catalytic functions that can be accommodated by a single structural scaffold is considerably broader than had been previously suspected. Further, the diversity of function that each superfamily represents allows an economy in the number of unique protein folds required to support life and, as a result, undoubtedly has “simplified” the course of metabolic evolution.

The Enolase Superfamily: Abstraction of the α-Protons of Carboxylic Acids

We recently described the enolase superfamily, the members of which catalyze at least 11 different chemical reactions, including racemization, epimerization, and both syn and anti β-elimination reactions involving water, ammonia, or an intramolecular carboxylate group as leaving group (5).

Despite broad differences in substrate structures and the overall reactions they catalyze, all of the reactions of the enolase superfamily are initiated by a common partial reaction, metal-assisted, general base-catalyzed abstraction of the α-proton of a carboxylate anion to generate a stabilized enolate anion intermediate (Reaction 1). However, the fate of the intermediate (protonation in the case of racemization and epimerization reactions and vinylogous β-elimination in the others) must be determined by the different functional groups that “surround” the intermediate in the active site. The common partial reaction is thermodynamically difficult: the pKₐ values of the α-protons in the substrates range from 29 to 32 whereas the pKₐ values of the conjugate acids of the active site bases accepting the protons are 7. The enzyme-active sites must destabilize the enzyme-substrate complex and/or stabilize the enzyme-intermediate complex so that the free energy of the transition state for α-proton abstraction can be lowered sufficiently to be consistent with the observed k_cat values.

\[
\begin{align*}
\text{Enr}^-- & \rightarrow \text{Enr}^-\text{B}^+ \\
\text{REACTION 1}
\end{align*}
\]

The members of the enolase superfamily can be assigned to subgroups based upon primary sequence alignments: the MR subgroup, the MLE subgroup, and the enolase subgroup; coincidentally, high resolution x-ray structures are available for these three enzymes (MR, Protein Data Bank entry 1mnr (6); MLE I; enolase, Protein Data Bank entry 1ebh (7)). Each is composed of an N-terminal domain (βα4) of ~140 residues that precedes a βα barrel domain (TIM barrel) of ~200 residues. In all three enzymes, the analogous domains have globally similar structures, supporting the hypothesis that these enzymes are related by divergent evolution.

The three subgroups are distinguished by 1) the identities and sequence contexts of the three carboxylate ligands for the essential divalent metal ion; 2) the number and identities of the general base catalysts that initiate the reactions by abstraction of the α-proton; and 3) the identities of the electrophilic catalysts that stabilize the enolate anion intermediate by electrostatic and/or hydrogen-bonding interactions.

In the MR, MLE, and enolase structures, the observed differences in the carboxylate metal ion ligands may be correlated with differing coordination geometries of their various substrates to the divalent metal ion (5). The observed differences in the identities and numbers of the general basic catalysts, one

2 M. Hasson, I. Schlichting, J. Moulai, D. Ringe, and G. Petsko, unpublished observations.

3 In enolase the βα barrel has a βαβαββαββα structure in MR and MLE the βα barrel has a βαββαββα structure.
in the active sites of the enolase subgroup, MLEs, and galactonate dehydratase and two in the active sites of the remaining characterized members of the superfamily, allow stereospecific, stereoselective, or stereorandom formation of the stabilized intermediate followed by partitioning of the intermediate to products via differing mechanisms.

The differences in the overall reactions can be rationalized explicitly in the context of the β/α barrel structure that contains the residues directly involved in catalysis. The active sites are located, as expected, at the C-terminal ends of the barrel domains in these proteins. Each functional group involved in catalysis is presented on a separate structural unit within the barrel domain, either at the C-terminal end of a β-sheet strand or in the loop that connects a β-sheet with the following α-helix (Fig. 1). So organized, the structures of β/α barrels provide an important evolutionary advantage: the ability to deliver functional groups from each of the eight possible octants surrounding the bound substrates and intermediates. The functional groups, along with the associated secondary structural elements, can evolve independently, thereby allowing new catalytic activities to be generated while retaining the common partial reaction.

Recognition and description of this structure-function paradigm provide a rational basis for understanding the chemistry of all of the superfamily members. Comparative analysis of these proteins has led to important new observations that would have been difficult to obtain from the focused study of these proteins. Comparative analysis of all of the superfamily members. Comparative analysis of these proteins has led to important new observations that would have been difficult to obtain from the focused study of these proteins. The identities and importance of these residues have not been investigated, but their proximity to the active site suggests that the general strategy used by nature to evolve the members of the enolase superfamily may also have been used in the evolution of the NAL superfamily.

Also, in parallel with the enolase superfamily, identification of the common mechanistic strategy associated with the NAL superfamily architecture offers important insights into the reactions catalyzed by as-of-yet uncharacterized members of the superfamily. One superfamily member, the mosA gene product (Rhizobium meliloti), functions in rhizopine biosynthesis, a biosynthetic pathway that has not been fully characterized (11). Whereas studies of a mosA mutant suggest that the MosA protein catalyzes methylation of a hydroxyl group (12), this predicted function is unexpected since it cannot be readily associated with a mechanism involving a protonated Schiff base functioning as an electron sink, e.g. the use of α-ketobutyrate as a methyl group donor is unprecedented in enzymology.

The Crotonase Superfamily: Coenzyme A Hydratases/Isomerases/Dehalogenases/Hydrolases

The members of the crotonase superfamily (13, 14) catalyze several reactions of coenzyme A thioesters that require stabilization of an oxanion intermediate. An enolate ion is generated by abstraction of the α-proton of the thioester (Fig. 3A).
in reactions resulting in β-addition/elimination of water (crotonase and carnitine racemase), 1,3-proton transfer (3,2-transenyol-CoA isomerase), and carbon-carbon bond formation (naphthoate synthase, intramolecular addition of the enolate anion to a carbonyl group). An enolate anion is also generated by nuclophilic aromatic addition in the reaction catalyzed by 4-chlorobenzyol-CoA dehalogenase (Fig. 3B).

In addition, an enzyme that hydrolyzes a thioester, β-hydroxyisobutyryl-CoA hydrolase, was recently discovered to be a member of the crotonase superfamily (15). Whereas this reaction involves formation of a tetrahedral alkoxide anion (Fig. 3C) rather than the enolate anion common to the other members of the superfamily (Fig. 3A and B), kinetic competence in each reaction type is dependent on stabilization of their respective intermediates. Conversion of a thioester to either type of intermediate involves a large change in the proton affinity of the carbonyl oxygen (pK_a of the conjugate acid ≤ 0) as it is converted to either an enolate anion (pK_a of the conjugate acid, −11) or the oxygen of an alkoxide anion (pK_a of the conjugate acid, −13). Thus, a common strategy can be used to stabilize both types of intermediates; a significant increase in hydrogen bond strength between the thioester carbonyl oxygen (hydrogen bond acceptor) and active site proton donors as the enzyme–substrate complex is converted to the enzyme–intermediate complex via a transition state is likely to be similar in structure and energy to the intermediate (16–18).

High resolution structures have been described for 4-chlorobenzyol-CoA dehalogenase (19) and crotonase (20). These proteins are not β/α barrels; instead, they represent a “spiral fold” composed of two distinct domains, with the active site located entirely in the N-terminal α + β domain. In each enzyme, the thioester carbonyl group is located in an “oxyanion hole” in which it is hydrogen-bonded to two peptide N–H groups. These interactions provide the structural basis for significant stabilization of the enolate anion or tetrahedral alkoxide intermediates as charge is localized on the oxygen.

One interesting difference between crotonase and the dehalogenase is that the critical catalytic steps occur at different distances from the thioester functional group in the structures of their respective substrates; hence, the functional groups mediating the β-elimination and nucleophilic aromatic substitution reactions are located on different secondary structure elements in the enzyme-active sites. The syn β-elimination reaction catalyzed by crotonase is apparently general acid-base catalyzed by a pair of glutamate residues, Glu-144 and -164, with Glu-164 appropriately positioned to initiate the β-elimination by abstraction of the α-proton. Using the same basic architecture, the nucleophilic aromatic substitution reaction catalyzed by the dehalogenase uses a spatially different carboxylate group, that of Asp-145, as a nucleophile to form an arylated enzyme intermediate; this, in turn, is hydrolyzed using His-90 as a general basic catalyst to yield the final product (21).

The reactions catalyzed by carnitine racemase and 3,2-transenyol-CoA isomerase are also initiated by abstraction of the α-proton of the thioester substrate; the sequence alignment predicts that the active sites of these enzymes contain homologs of Glu-164 in crotonase. Interestingly, the reaction attributed to naphthoate synthase is also initiated by abstraction of the α-proton of a thioester of the aliphatic carboxylate group, but the sequence alignment fails to identify a homolog of Glu-164 in crotonase, suggesting another evolutionary refinement of the structural solution. Also, the sequence alignment suggests that the active site of the thioester hydrolase contains a homolog of Glu-144, but not Glu-144, in crotonase and of Asp-145, but not His-90, in the dehalogenase (alignment not shown). The mechanisms of the reactions catalyzed by these enzymes are unexplored, so a detailed molecular understanding of evolution of catalytic function in this superfamily cannot yet be specified.

That the substrate for each of the enzymes in the crotonase superfamily is a coenzyme A thioester could imply that substrate binding and not chemistry is the critical factor in the evolution of this superfamily. However, the diversity in the overall reactions catalyzed and the accompanying variation in the identities of the general acidic and basic catalysts that mediate those chemistries reflect a common structural solution to the problem of stabilization of reactive oxyanionic intermediates using an “oxyanion hole.” Thus, although all of the superfamly members bind a thioester substrate, a mechanistic analysis of the reactions supports our conlusion that the difficult chemical step of stabilization of a reactive intermediate has dominated the evolution of this family.

**The Vicinal Oxygen Chelate Fold: Metal-dependent Dioxygenases/Glutathione Transferases/Glyoxalases**

The members of the vicinal oxygen chelate (VOC) fold (22, 23) include 2,3-dihydroxybiphenyl dioxygenase and related meta-cleavage dioxygenases (Fe^{2+} or Mn^{2+}-dependent; oxidation of catechols to α-hydroxymuconate semialdehydes), glyoxalase I (Zn^{2+}-dependent; a 1,2-proton transfer that converts the thiohemiacetal of methylglyoxal with glutathione to lacticglythahionile), and fosfomycin resistance protein (Mn^{2+}-dependent; nucleophilic addition of glutathione to the oxirane of fosfomycin) (Fig. 4). In each of these very different reactions, catalysis involves stabilization of transition states containing vicinal anionic oxygen atoms.

The only structure currently available for this superfamily is that of 2,3-dihydroxybiphenyl 1,2-dioxygenase (24). Each polypeptide of the tetrameric protein is an eight-stranded β-sheet structure composed of two domains approximately equal in size. Although the sequence identity is less than 20% for the two domains, their three-dimensional structures are nearly superimposable. However, only the C-terminal domain contains ligands for the Fe^{2+} (two His and one Glu). Both glyoxalase I and the fosfomycin resistance protein are approximately one-half the size of the dioxygenase, and sequence alignments suggest that these proteins possess two histidines and one glutamate that function as the ligands for the divalent metal ions.

The common mechanistic strategy mediated by the similar structures of the VOC superfamily is the ability to exploit...
metal ions to facilitate their respective catalytic functions. Interestingly, each of the three currently identified members of the superfamily utilizes a different metal ion for catalysis. In the dioxygenase, the Fe^{2+} likely coordinates the vicinal hydroxyl groups of the catechol substrate, binds the dioxygen cosubstrate, and participates in redox chemistry that facilitates the ring cleavage that accompanies oxidation. In glyoxalase I, the Zn^{2+} likely provides electrostatic stabilization of the enediolate intermediate involved in the 1,2-proton transfer reaction by direct coordination. In the fosfomycin resistance protein, the Mn^{2+} likely binds to the oxirane oxygen, thereby acting as a Lewis acid catalyst that facilitates the nucleophilic attack of glutathione on an oxirane carbon. Whereas the identities and catalytic functions of the divalent metal ions are distinct, the domains that contain the metal ion ligands are homologous. Thus, in the VOC superfamily nature evidently accommodates the attack of glutathione on an oxirane carbon. Interestingly, each of the three currently identified members of the superfamily utilizes a different metal ion for catalysis. In the VOC superfamily analysis has the potential to offer insights that cannot be obtained even from the most elegant studies of a single enzyme.

Note Added in Proof—The structure of human glyoxalase I has been reported (27); the domain structure as well as the identities and positions of the metal ion ligands are homologous to those reported for 2,3-dihydroxybiphenyl 1,2-dioxygenase (24).

Acknowledgments—We thank Professors Richard N. Armstrong, Vanderbilt University School of Medicine, and Vernon E. Anderson, Case Western Reserve School of Medicine, for valuable discussions.

**REFERENCES**

1. Neidhart, D. J., Kenyon, G. L., Gerlt, J. A., and Petsko, G. A. (1990) *Nature* 347, 692–693
2. Perona, J. J., and Craik, C. S. (1997) *J. Biol. Chem.* 272, 29987–29990
3. Petsko, G. A., Kenyon, G. L., Gerlt, J. A., Ringe, D., and Rozzarich, J. W. (1993) *Trends Biochem. Sci.* 18, 372–376
4. Babitt, P. C., Mrachko, G. T., Hasson, M. S., Lies, M. A., Reed, G. H., Rayment, I., Ringe, D., Kenyon, G. L., and Gerlt, J. A. (1996) *Biochemistry* 35, 16489–16501
5. Neidhart, D. J., Howells, P. L., Petsko, G. A., Powers, V. M., Li, R. S., Kenyon, G. L., and Gerlt, J. A. (1991) *Biochemistry* 30, 9264–9273
6. Wedekind, J. E., Reed, G. H., and Rayment, I. (1995) *Biochemistry* 34, 4325–4330
7. Palmer, D. R. J., and Gerlt, J. A. (1996) *J. Am. Chem. Soc.* 118, 10323–10324
8. Izard, T., Lawrence, M. C., Malby, R. L., Lilley, G. G., and Colman, P. M. (1994) *Structure* 2, 361–369
9. Mirwaldt, C., Kornfelder, L., and Huber, R. (1995) *J. Mol. Biol.* 246, 227–239
10. Murphy, P. J., Trenz, S. P., Grzemski, W., De Bruijn, F. J., and Schell, J. (1993) *J. Bacteriol.* 175, 5193–5204
11. Anj, J. P., Grzemski, W., and Murphy, P. J. (1995) *Microbiology* 141, 1683–1690
12. Babitt, P. C., Kenyon, G. L., Martin, B. M., Charest, H., Silvestre, M., Scholtens, J. D., Chang, K.-H., Liang, P.-H., and Dunaway-Marcano, D. (1992) *Biochemistry* 31, 5594–5604
13. Dunaway-Marcano, D., and Babitt, P. C. (1994) *Bioscience* 5, 259–276
14. Haws, J. W., Jakubiec, J., Shimomura, Y., Huang, B., Bunting, J., Harper, E. T., and Harris, R. (1992) *Biol. Chem.* 271, 14248–14253
15. Gerlt, J. A., and Gassman, P. G. (1993) *Biochemistry* 32, 11943–11952
16. Cieplak, V. R., Kornfelder, L., and Huber, R. (1996) *J. Mol. Biol.* 194, 673–690
17. Shanks, S., and Herschlag, D. (1996) *Proc. Natl. Acad. Sci. U.S.A.* 93, 14474–14479
18. Benning, M. M., Taylor, K. L., Liu, R.-Q., Yang, G., Xiang, H., Wensenberg, G., Dunaway-Marcano, D., and Holden, H. M. (1996) *Biochemistry* 35, 8103–8109
19. Engel, C. K., Mathieu, M., Zeelen, P. J., Hiltunen, J. K., and Wierenga, R. K. (1996) *EMBO J.* 15, 5135–5145
20. Yang, G., Liu, R.-Q., Taylor, K. L., Xiang, H., Price, J., and Dunaway-Marcano, D. (1996) *Biochemistry* 35, 10879–10885
21. Armstrong, R. N. (1997) *Chem. Rev. Toxicol.* 10, 2–18
22. Bernat, B. A., Laughlin, L. T., and Armstrong, R. N. (1997) *Biochemistry* 36, 3050–3055
23. Han, S., Eltis, L. D., Timmis, K. N., Muchmore, S. W., and Bolin, J. T. (1995) *Science* 267, 976–980
24. Horowitz, N. H. (1945) *Proc. Natl. Acad. Sci. U.S.A.* 31, 153–157
25. Ferrin, T. E., Huang, C. C., Jarvis, L. E., and Langridge, R. (1988) *J. Mol. Graphics* 6, 13–27
26. Cameron, A. D., Olin, B., Ridderstrom, M., Mannervik, B., and Jones, T. A. (1997) *EMBO J.* 16, 3386–3395

*FIG. 4. Reactions of the VOC fold superfamily.*

---

4 This differs from the hypothesis first proposed by Horowitz (25) that binding a common substrate is the critical constraint in the evolution of new enzymes from a particular structural scaffold.

2) The catalytic activity of a newly sequenced but uncharacterized open reading frame cannot necessarily be inferred from the overall reactions catalyzed by homologous enzymes. Rather, the chemical step common to the superfamily scaffold must be identified and correlated with conserved structural features.

These conclusions should be useful in developing new strategies for solving problems of current interest in mechanistic enzymology and structural biology as well as in the emerging disciplines of bioinformatics. For example, determination of the principles that govern the structure-function correlations for any particular superfamily will likely play an important role in assigning catalytic function to unknown sequences. Finally, by providing a more contextual basis for understanding both the rapid rates and mechanisms of enzyme-catalyzed reactions, superfamily analysis has the potential to offer insights that cannot be obtained even from the most elegant studies of a single enzyme.