The Enolase Superfamily: A General Strategy for Enzyme-Catalyzed Abstraction of the α-Protons of Carboxylic Acids†

Patricia C. Babbitt,*,‡ Miriam S. Hasson,§,|| Joseph E. Wedekind,⊥,# David R. J. Palmer,▽ William C. Barrett,▽ George H. Reed,⊥ Ivan Rayment,⊥ Dagmar Ringe,§ George L. Kenyon,‡ and John A. Gerlt*,▽

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446, Departments of Biochemistry and Chemistry and Rosenstiel Center for Basic Biomedical Research, Brandeis University, Waltham, Massachusetts 02154-9110, The Institute for Enzyme Research and Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706, and Department of Biochemistry, University of Illinois, Urbana, Illinois 61801

Received July 5, 1996⊗

ABSTRACT: We have discovered a superfamily of enzymes related by their ability to catalyze the abstraction of the α-proton of a carboxylic acid to form an enolic intermediate. Although each reaction catalyzed by these enzymes is initiated by this common step, their overall reactions (including racemization, *β*-elimination of water, *β*-elimination of ammonia, and cycloisomerization) as well as the stereochemical consequences (*syn vs anti*) of the *β*-elimination reactions are diverse. Analysis of sequence and structural similarities among these proteins suggests that all of their chemical reactions are mediated by a common active site architecture modified through evolution to allow the enolic intermediates to partition to different products in their respective active sites *via* different overall mechanisms. All of these enzymes retain the ability to catalyze the thermodynamically difficult step of proton abstraction. These homologous proteins, designated the "enolase superfamily", include enolase as well as more metabolically specialized enzymes: mandelate racemase, galactonate dehydratase, glucarate dehydratase, muconate-lactonizing enzymes, *N*-acylamino acid racemase, *β*-methylaspartate ammonia-lyase, and *o*-succinylbenzoate synthase. Comparative analysis of structure−function relationships within the superfamily suggests that carboxyphosphonoenolpyruvate synthase, another member of the superfamily, does not catalyze the reaction proposed in the literature but catalyzes an enolase-like reaction instead. The established and deduced structure−function relationships in the superfamily allow the prediction that other apparent members of the family for which no catalytic functions have yet been assigned will also perform chemistry involving abstraction of the α-protons of carboxylic acids.

Understanding the role of active site structure in determining both the mechanisms and rates of enzyme-catalyzed reactions is an important problem in enzymology. Until recently, the interdependence of structure, mechanism, and rates has been deduced primarily by examining the structural and catalytic properties of a single enzyme of interest, typically with the combined approaches of X-ray crystallography, site-directed mutagenesis, and both kinetic and mechanistic analyses. An alternative approach is now possible due to the rapidly expanding structural and sequence data bases for proteins.

The presence of conserved amino acid residues in enzymes that catalyze the *same* reaction permits recognition of an enzyme family and may allow identification of active site residues. The availability of increasing numbers of three-dimensional structures now provides an opportunity to expand families defined on the basis of conserved catalytic properties to include enzymes that catalyze *different* overall reactions yet share the same structural framework. Comparison of structures in the context of sequence information allows the requirements for catalysis to be identified in a way that cannot be obtained by the focused study of a single enzyme alone. In particular, those features that are important for shared chemical steps can be distinguished from those that relate to the specific chemistry of an individual enzyme. This information can then be used to predict the functions of homologs of unknown function from their sequences alone. Such a general approach is necessary to interpret the explosion of sequence data being generated from the ongoing genome-sequencing projects, which will, in principle, define all the enzymatic activities required for the survival of these organisms. These projects have already identified the sequences for a large number of proteins of unknown function. For example, 1743 predicted coding regions were identified in the genome of *Haemophilus influenzae* Rd (Fleischmann *et al.*, 1995), whereas functions could be assigned to only 1007 (58%) by sequence comparisons.

The power inherent in such a comparative approach is also exemplified by its broad utility for additional applications; thus, it can also be used to elucidate enzyme mechanisms,

† This research was supported by Grants GM-40570 (to J.A.G., G.L.K., and Gregory A. Petsko), GM-35752 (to G.H.R.), AR-35186 (to I.R.), and GM-52594 (to J.A.G.) from the National Institutes of Health and Fellowship DRG-1194 of the Cancer Research Fund of the Damon Runyon-Walter Winchell Foundation (to M.S.H.). The Computer Graphics Laboratory at the University of California, San Francisco, was supported by NIH Grant RR-01081.

‡ University of California.

§ Brandeis University.

|| Present address: Department of Biological Sciences, Purdue University, West Lafayette, IN 47907-1392.

⊥ University of Wisconsin.

Present address: Department of Structural Biology, Fairchild Research Center, Stanford University School of Medicine, Stanford, CA 94305-5400.

▽ University of Illinois.

⊗ Abstract published in *Advance ACS Abstracts,* November 1, 1996.

S0006-2960(96)01641-8 CCC: $12.00

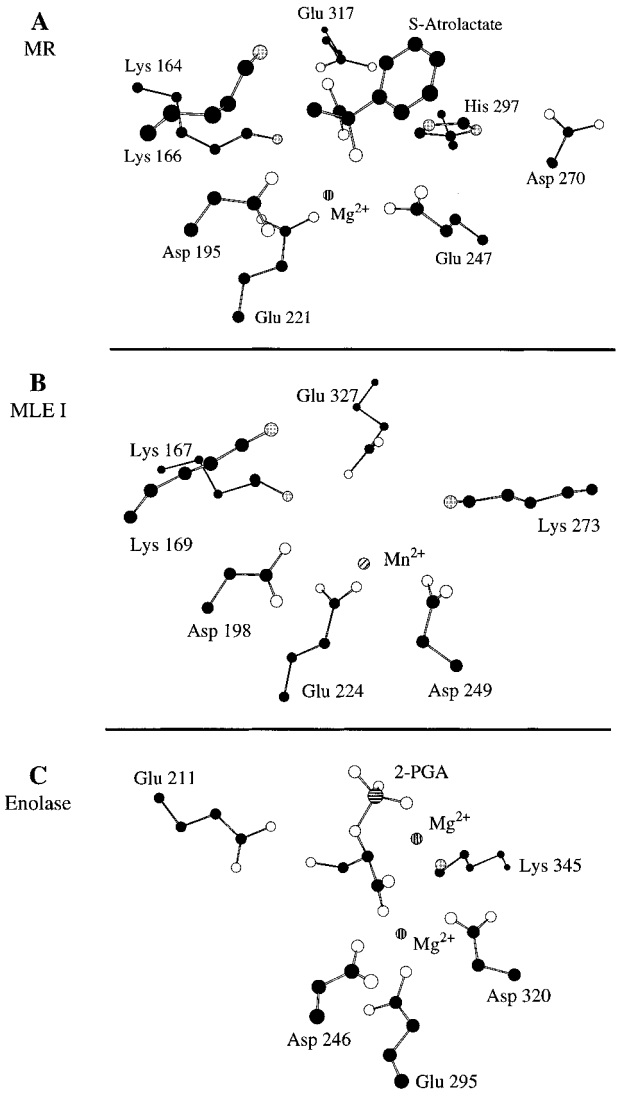
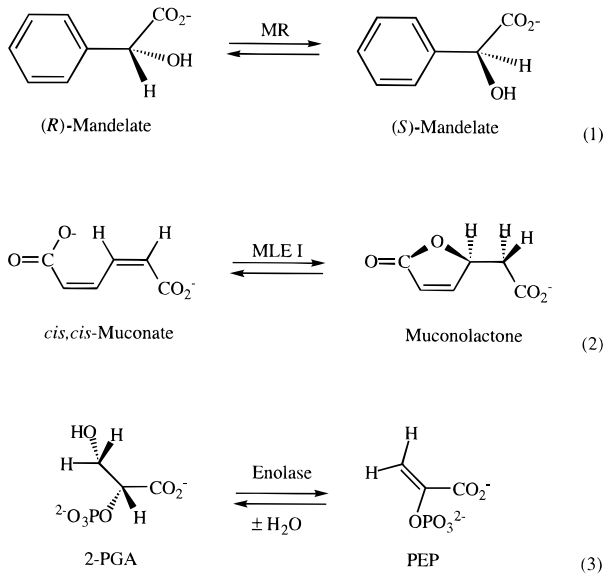


FIGURE 1: Active site of MR with bound (*S*)-atrolactate (A; PDB entry 1MNS; Landro *et al.*, 1994), MLE I (B; M. Hasson, I. Schlichting, J. Moulai, D. Ringe, and G. Petsko, unpublished observations), and enolase with bound 2-PGA (C; PDB entry 1ONE; Larsen *et al.*, 1996).

to correct misassigned functions of enzymes, and to predict additional catalytic functions for enzymes of previously assigned function. Conversely, the increasing availability of complete sequences of multiple genomes helps define the range of chemical functions compatible with a given structural template. In this larger framework, this approach promises to provide insights into how nature has both optimized catalysis and evolved new enzymatic activities.

We report here the application of this comparative approach to the "enolase superfamily" of enzymes, structurally related proteins sharing the common ability to catalyze abstraction of the α-protons of carboxylic acids. This superfamily is designated by the name of the family member whose function is most central to living organisms.

Although the reactions catalyzed by the members of the enolase superfamily are ubiquitous in biochemistry, an understanding of how the weakly acidic protons of the substrates can be abstracted by weakly basic active site functional groups has been incomplete (Gerlt & Gassman, 1992, 1993a,b). The three proteins of this superfamily that have been examined structurally, mandelate racemase (MR[1]), muconate-lactonizing enzyme I (MLE I), and enolase, are members of the large $(\beta/\alpha)_8$ β-barrel structural class (Farber & Petsko, 1990).[2] While the overall reactions catalyzed by these enzymes differ markedly, each is initiated by abstraction of the α-proton of a carboxylic acid. Mandelate racemase (MR) catalyzes the interconversion of the (*R*)- and (*S*)-enantiomers of mandelate (eq 1); muconate-lactonizing enzyme I (MLE I) catalyzes the reversible cycloisomerization of *cis,cis*-muconate to muconolactone (eq 2), and enolase catalyzes the β-elimination reaction of water from 2-phosphoglycerate (2-PGA) to form PEP in the glycolytic pathway (eq 3):



These enzymes possess a similar active site architecture (Figure 1) (Lebioda & Stec, 1988; Neidhart *et al.*, 1990; Wedekind *et al.*, 1995) that mediates the enolization of a carboxylic acid (Figure 2) despite broad differences in their substrates and the overall reactions they catalyze. The activity of each enzyme is dependent upon a divalent metal ion. This feature provides a critical link for understanding the interdependence of their structures and functions.

In this paper, we describe the discovery of *at least seven and perhaps as many as eleven* additional members of the enolase superfamily (Table 1) that are homologous to MR, MLE I, and enolase. Each enzyme catalyzes a distinct overall reaction initiated by a common step, abstraction of the α-proton of a carboxylic acid. The new enzymes were

[1] Abbreviations: CPEP, carboxyphosphonoenolpyruvate; CPEPS, carboxyphosphonoenolpyruvate synthase; GalD, galactonate dehydratase; GlucD, glucarate dehydratase; β-MAL, β-methylaspartate ammonia lyase; MBP1, c-Myc promoter binding protein; MLE, muconate-lactonizing enzyme; MR, mandelate racemase; NAAAR, *N*-acylamino acid racemase; NCBI, National Center for Biotechnology Information; orf, open reading frame; OSBS, *o*-succinylbenzoate synthase; PDB, Protein Data Bank; PEP, phosphoenolpyruvate; 2-PGA, 2-phosphoglycerate; SynORF, unidentified orf from *Synechocystis* sp.; TIM, triose phosphate isomerase.

[2] The β-barrel in enolase has an unusual $\beta_2\alpha_2(\beta/\alpha)_6$ structure; however, the overall superposition of the structures of MR, MLE I, and enolase from yeast provides persuasive evidence that these are related by divergent evolution (see Figure 2). Enolase from yeast is a dimer (Lebioda & Stec, 1988); both MR (Neidhart *et al.*, 1990) and MLE I (Helin, 1995) are tetramers of dimers. However, the enolases from *Bacillus megaterium* (Singh & Setlow, 1978), *Thermatoga maritima* (Schurig *et al.*, 1995), and *Zymomonas mobilis* (Pawluk *et al.*, 1986) are octamers (presumably tetramers of dimers).

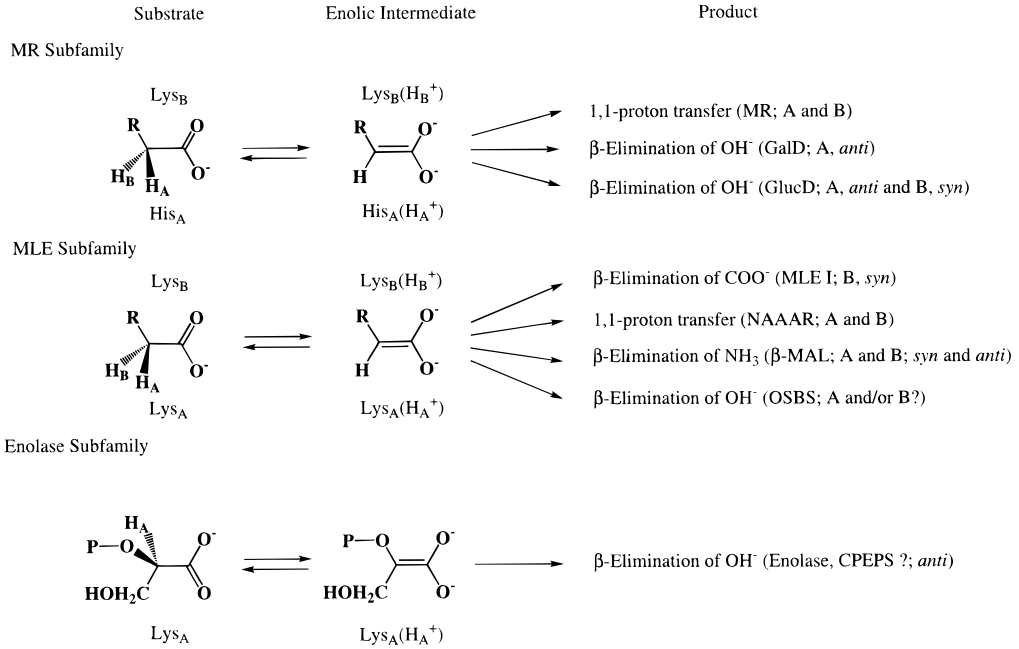


FIGURE 2: Involvement of active site general bases in the reactions catalyzed by the MR, MLE I, and enolase subgroups of the enolase superfamily. The general bases designated A are on the same relative face of the active site, and those designated B are on the opposite face. In MR, His_A is H297 and Lys_B is K166; in MLE I, Lys_A is K273 and Lys_B is K169, and in enolase, Lys_A is K345.

Table 1: Members of the Enolase Superfamily

protein	species	accession no.[a]	ref
mandelate racemase (MR)	*Pseudomonas putida*	SP, P11444[b]	Ransom *et al.*, 1988
galactonate dehydratase (GalD)	*Escherichia coli*	SP, P31458[b]	Burland *et al.*, 1993
glucarate dehydratase (GlucD)	*P. putida*	SP, P42206[b]	unpublished
	Bacillus subtilis	SP, P42238	Ogawa *et al.*, 1995
	E. coli[c]	GB, U29581	unpublished
RspA	*E. coli*	SP, P38104[b]	Huisman & Kolter, 1994
Spa2	*Streptomyces ambofaciens*	SP, P32426[b]	Schneider *et al.*, 1993
rTS-α	*Homo sapiens*	GB, X67098[b]	Dolnick, 1993
rTS-*β*	*H. sapiens*	GB, X89602	unpublished
muconate-lactonizing enzyme I (MLE I)	*P. putida*	GB, U12557[b,d]	Houghton *et al.*, 1995
	Acinetobacter calcoaceticus	GB, M76991	Shanley *et al.*, 1994
muconate-lactonizing enzyme II (MLE II)	*P. putida*	SP, P11452[b]	Ghosal & You, 1989
	Alcaligenes eutrophus	PIR, JQ0175	Ghosal & You, 1989
	Pseudomonas sp. P51	SP P27099	van der Meer *et al.*, 1991
N-acylamino acid racemase (NAAAR)	*Amycolatopsis* sp.	GB, D30738[b]	Tokuyama & Hatano, 1995b
β-methylaspartate ammonia lyase (*β*-MAL)	*Clostridium tetanomorphum*	SP, Q05514[b]	Goda *et al.*, 1992
o-succinylbenzoate synthase (OSBS)	*E. coli*	SP, P29208	Sharma *et al.*, 1993
	Haemophilus influenzae	SP, P44961[b]	Fleischmann *et al.*, 1995
SynORF	*Synechocystis* sp.	GB, D64001[b]	Kaneko *et al.*, 1995
enolase	*Haloarcula marismortui*	SP, P29201[b]	Kroemer & Arndt, 1991
	Saccharomyces cerevisiae	SP, P00924[b]	Holland *et al.*, 1981
carboxyphosphonoenolpyruvate synthase (CPEPS)	*Streptomyces hygroscopicus*	GB, D37878[b]	Lee *et al.*, 1995
cMyc promoter binding protein (MBPI)	*H. sapiens*	SP, P22712[b]	Ray & Miller, 1991

[a] GB, GenBank; SP, Swiss-Protein; PIR, Protein Identification Resource. [b] Sequences used to construct the sequence alignment in Figure 4. [c] Two orfs: one full length and one fragment. [d] The sequence used in the alignment in Figure 4 differs slightly from GB U12557 (W. A. Barrett, M. Tomaska, and J. A. Gerlt, unpublished observations).

initially recognized as members of this superfamily from sequence similarities, *e.g.*, by the presence of carboxylate groups homologous to the ligands for the essential divalent metal ions required by all three structurally characterized enzymes (Table 2A). Once the metal ion ligands were aligned, the candidate general bases responsible for α-proton abstraction could be identified. With this information, detailed mechanisms for the reactions catalyzed by the newly discovered members of the family can be proposed, thereby demonstrating the power of our comparative approach for understanding enzyme mechanisms in the context of active site architecture.

MATERIALS AND METHODS

Superposition of the N-Terminal and β-Barrel Domains. The coordinates used to superimpose the α-carbon backbones of MR (PDB entry 2MNR), MLE I (M. Hasson, I. Schlichting, J. Moulai, D. Ringe, and G. Petsko, unpublished observations), and enolase (PDB entry 1EBH) are from the authors' laboratories. These refined structures each contain a single bound divalent cation in the active site with no other bound ligands.

These enzymes contain two major domains: an N-terminal domain and a *β*-barrel domain. The N-terminal domains include residues 5−128 of MR, residues 5−129 of MLE I,

Table 2: Active Site Residues of the Members of the Enolase Superfamily[a]

A enzyme	ligand 1	ligand 2	ligand 3
	MR Subfamily		
MR	D195*	E221*	E247*
GalD	D183	E209	E235
GlucD	D241	E266	?
RspA	D212	E238	E264
Spa2	D210	E236	E262
rTS	D168	E194	E223
	MLE Subfamily		
MLE	D198*	E224*	D249*
NAAAR	D189	E214	D240
β-MAL	D250	E273	D307
OSBS	D186	E215	D239
SynORF	D172	E204	D229
	Enolase Subfamily		
enolase	D246*	E295*	D320*
CPEPS	D166	E208	D235
MBP1	D147	E195	?

B enzyme	electrophile	general base 1	general base 2
	MR Subfamily		
MR	K164*	K166*	H297*
GalD	K144	(N146)	H285
GlucD	K211	K213	H345
RspA	?	?	H314
Spa2	?	?	H312
rTS	K138	K140	H273
	MLE Subfamily		
MLE I	K167*	K169*	(K273)*
NAAAR	K161	K163	K263
β-MAL	? [R219/K224]	? [R221/R226]	K331
OSBS	K156	K158	K260
SynORF	K141	K143	K253
	Enolase Subfamily		
enolase	−	−	K345*
CPEPS	−	−	K256
MBP1	−	−	K244

[a] (A) Metal ion ligands and (B) electrophilic and general basic residues. Residues whose roles have been confirmed or characterized experimentally are designated by asterisks (*). All of the other active site residues are predicted from this analysis and the alignment shown in Figure 4. Residues in parentheses designate residues that are predicted to be structurally homologous but not functionally involved in proton abstraction. Residues that could not be clearly predicted are shown in brackets.

and residues 1−134 of enolase. The β-barrel domains include residues 135−318 of MR, residues 131−320 of MLE I, and residues 143−398 of enolase.

The α-carbon coordinates of the N-terminal and β-barrel domains of MR and enolase were positioned onto MLE I with the interactive graphics program O (Jones *et al.*, 1991) and the algorithm (LSQKAB) of Kabsch (1978; CCP4, 1994); 71 atoms of the β-barrel domains and 76 atoms of the N-terminal domains were chosen for the comparison. The N- and C-terminal domains were superimposed independently (Figure 3). The average rms displacements for superpositions with MLE I are as follows: 1.53 Å for MR and 1.26 Å for enolase (N-terminal domains) and 1.04 Å for MR and 1.35 Å for enolase (β-barrel domains).

Sequence Alignments. The BLAST server (Altschul *et al.*, 1990) at the National Center for Biotechnology Information (accessed *via* the Internet) and utilities provided by the Sequence Analysis and Consulting Service of the Computer Graphics Laboratory (University of California, San Francisco) were used to identify superfamily members. FASTA was also used to search for homologs of selected superfamily members (Pearson & Lipman, 1988). Up to 499 scores were recorded for each search and manually screened for structural and chemical characteristics that might allow them to be included in the superfamily. Homology was determined from conservation of functionally important residues (Table 2) as given in the multiple alignment shown in Figure 4. The homologs found by this procedure are listed in Table 1.

The sequence alignment of all members of the superfamily was prepared as a composite of alignments of various subgroups generated by the PILEUP function of the GCG software package (Genetic Computer Group, Inc., Madison, WI). The β-barrel domains for MR, MLE I, and enolase were determined from their structures; the boundaries of the putative β-barrel domains of the remaining superfamily members were assigned from comparisons to those structures. Where gaps had to be introduced to resolve conflicts among different PILEUP alignments, they were introduced systematically at the same positions within the sequences of all members of a subgroup. Positions of gaps introduced without precedent from any of the initial PILEUP alignments were chosen to occur at secondary-structure boundaries as determined from the known structures of MR, MLE I, or enolase. Finally, the main elements of the resulting alignment were corrected with the structural superposition. This was performed through an interactive interface designed for facile comparison of multiple sequence alignments and superimposed tertiary structures (M. M. Young, P. C. Babbitt, and T. E. Ferrin, unpublished observations). Except for one small element resulting from visual inspection alone, the introduction of a small gap in the β-MAL sequence to align the homolog for the metal binding ligand D195 in MR, the alignment was generated entirely from the PILEUP solutions.

RESULTS AND DISCUSSION

We have assigned the members of the enolase superfamily to three subgroups on the basis of the identities of their active site general base(s): (1) the MR subgroup (one Lys and/or one His residue), (2) the MLE I subgroup (two Lys residues), and (3) the enolase subgroup (one Lys residue). The following analysis and discussion supports this classification and describes the relationships of the structural and functional properties of the proteins both within and among the three subgroups.

Structural Superposition and Active Site Geometries

The structural superposition of MR, MLE I, and enolase (Figure 3) confirms the results of earlier pairwise superpositions (Lebioda & Stec, 1988; Neidhart *et al.*, 1990; Wedekind *et al.*, 1995) and reveals the striking global homology among these enzymes. Combination of this information with the identity of the catalytic residues derived from the high-resolution structures of MR, MLE I, and enolase with bound substrates and/or substrate analogs described below formed the basis for alignment of the superfamily. The structure of MR complexed with the competitive inhibitor (*S*)-atrolactate (Landro *et al.*, 1994) was the foundation for proposing a general acid−general base-catalyzed mechanism for this enzyme that involves the formation of a stabilized enolic intermediate (Figure 2) (Gerlt *et al.*, 1992; Kenyon *et al.*,

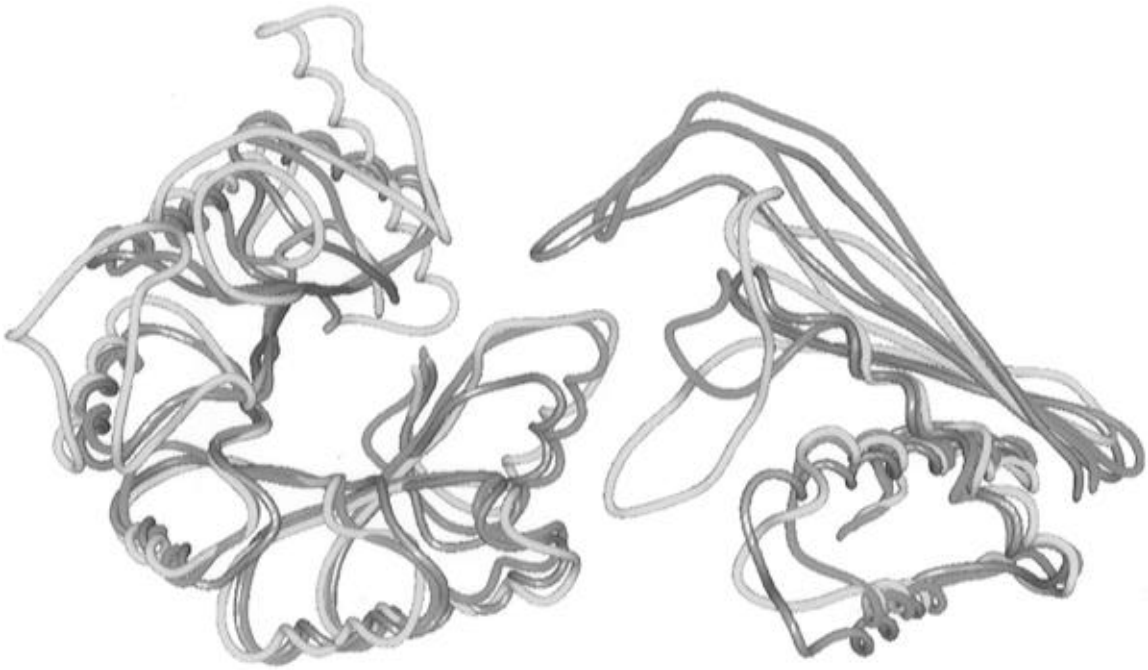


FIGURE 3: Superposition of the α-carbon backbones of MR (gray), MLE I (cyan), and enolase (yellow). The β-barrel domains are on the left and the N-terminal domains on the right.

1995). In this mechanism, K164 was identified as the electrophilic catalyst that neutralizes the anionic charge of the carboxylate group of the substrate, and K166 was identified as the (*S*)-specific general base that abstracts the α-proton from (*S*)-mandelate (Figure 1A and Table 2). K164 and K166 form a KXK motif that is easily recognized in sequence alignments. This motif appears to be diagnostic of the presence of functional groups that are involved as electrophilic and general basic catalysts, respectively (*vide infra*). H297 is the (*R*)-specific general base that abstracts the α-proton from (*R*)-mandelate (associated with D270 in a His-Asp dyad). E317 is the general acid catalyst that evidently donates a proton to the carboxyl group of the substrate as the α-proton is abstracted (not given in Table 2). These functions have been verified by site-directed mutagenesis, X-ray crystallography, and mechanistic analyses of mutant proteins (Landro *et al.*, 1991; Mitra *et al.*, 1995; Kallarakal *et al.*, 1995; Schafer *et al.*, 1996). The required Mg^{2+} is coordinated to the carboxylate groups of D195, E221, and E247 as well as to one carboxylate oxygen and the α-hydroxyl group of the bound inhibitor. The ϵ-ammonium group of K164 (electrophile in Table 2B) is hydrogen-bonded to the same carboxylate oxygen of the inhibitor that is coordinated to the Mg^{2+}.

Both structural (Neidhart *et al.*, 1990; Petsko *et al.*, 1993; Wedekind *et al.*, 1995) and primary sequence comparisons (Tsou *et al.*, 1990; Babbitt *et al.*, 1995) of MR with MLE I from *Pseudomonas putida* reveal the presence of homologs for the metal ion ligands found in the active site of MR (Figure 1B and Table 2A) as well as the presence of functional homologs of K164, K166 (Table 2B), and E317. By analogy to the mechanism established for the MR-catalyzed reaction, a general acid−general base-catalyzed mechanism can be written for the equilibration of muconolactone with *cis,cis*-muconate catalyzed by MLE I (Figure 2). K167 is an electrophilic catalyst, K169 the general basic catalyst, and E327 the general acidic catalyst. While both enzymes catalyze the abstraction of the α-proton of a substrate to form enolic intermediates, the fates of these intermediates differ. Protonation occurs in the active site of MR, while vinylogous elimination occurs in the active site of MLE I (Avigad & Englard, 1969; Chari *et al.*, 1987). This difference can be partially associated with the absence of a His in MLE I that is homologous to H297 in the active site of MR.

Comparison of the structure and chemical mechanism of enolase with those of MR and MLE I further supports a claim for the common ancestry of all three proteins. The high-resolution structure of enolase from *Saccharomyces cerevisiae* complexed with the substrate 2-PGA (Larsen *et al.*, 1996) reveals the presence of homologs for the metal ion ligands found in the active sites of MR and MLE I (Figure 1C and Table 2A). Mutagenesis studies of enolase suggest that K345 is the general base that abstracts the α-proton from (*R*)-2-PGA (Table 2B) and that E211 is likely to be the general acid that facilitates the departure of the β-hydroxide leaving group (Poyner *et al.*, 1996). The functional group of K345 is spatially homologous to the functional groups of K273 in the active site of MLE I and H297 in the active site of MR (Figure 1). The α-carbon of K345 can be structurally superimposed on the α-carbons of K273 in MLE I and of D270 in MR. As might be expected from the differences among their overall chemical mechanisms, E211 has no homolog in the active site of either MR or MLE I. The structure of enolase also reveals the presence of a second Mg^{2+} that is required for catalysis (Faller *et al.*, 1977; Poyner & Reed, 1992; Wedekind *et al.*, 1994). The first Mg^{2+} (that conserved in all of the superfamily members) coordinates to both carboxylate oxygens of 2-PGA; one of these oxygens is also coordinated to a second Mg^{2+}. However, the interactions of the substrate carboxylate group with the electrophilic groups in the active site of enolase (two Mg^{2+} ions) differ in significant detail from those of MR and MLE I (one Mg^{2+} ion and one electrophilic Glu residue). The proposed mechanism for enolase involves initial abstraction of the α-proton by K345 to generate an enolic intermediate

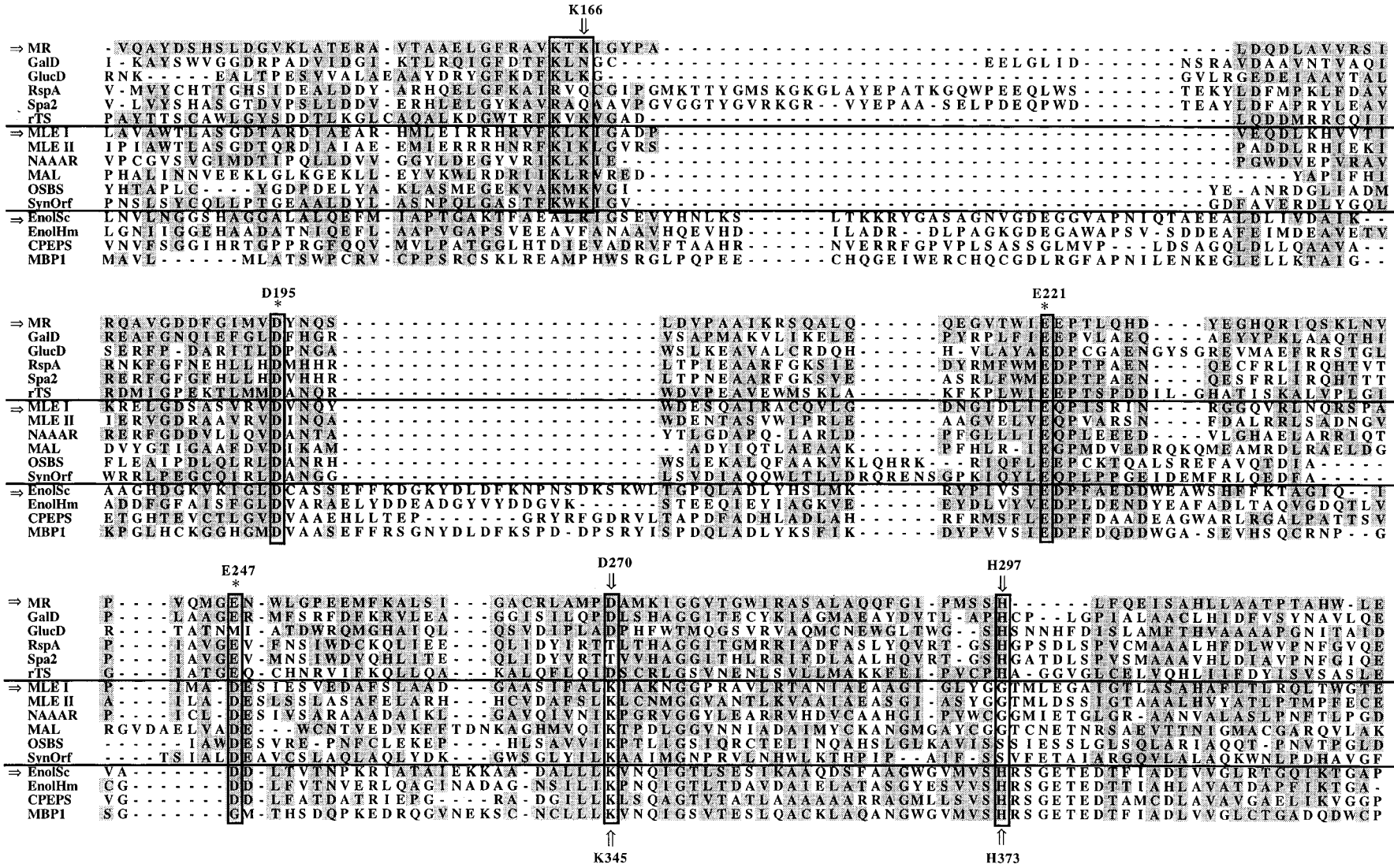


FIGURE 4: Sequence alignment of the β-barrel domains of the members of the enolase superfamily. Division into subgroups is indicated by horizontal lines. The proteins for which tertiary structural information is available are marked by horizontal arrows: MR, MLE I, and enolase. Sequence similarity is indicated by shading as provided by the SeqVu program (Garvan *et al.*, 1995). Important active site residues (Table 2) are boxed and labeled with respect to residue numbering for MR (above) and enolase (below) at each block in the alignment. The positions of the metal binding ligands are labeled by residue number, boxed, and marked with asterisks. Accession numbers for the sequences used in the alignment are given in Table 1.

that is stabilized by the two Mg^{2+} ions followed by vinylogous β-elimination of the hydroxide in a reaction that may be general acid-catalyzed by E211 [Figure 2 (Poyner *et al.*, 1996; Larsen *et al.*, 1996)].

Primary Sequence Alignment

Alignment of the sequences of enolases with those of MR and MLE I has been difficult because of limited sequence similarity. This problem now has been overcome using the enolase sequence from the archaebacterium *Haloarcula marismortui* (Kroemer & Arndt, 1991). The archaebacterial sequence is slightly more similar to those of other members of the superfamily, allowing us to "bootstrap" the sequence alignment to the yeast (and other eukaryotic) enolases.

The sequences of the putative β-barrel domains of both the established enzymes and predicted open reading frames (orfs) of the enolase superfamily (summarized in Table 1) are aligned in Figure 4. All of the superfamily members were initially identified with at least two different superfamily members in data base searches, usually by the presence of the conserved metal ion ligands (starred in Figure 4). Inspection of the alignment reveals that these proteins are globally homologous despite their great structural and functional divergence.

In addition to the sequence for MR, one representative sequence each for MLE I and MLE II, and two enolase sequences, the alignment presented in Figure 4 includes representative sequences for each of the five other homologs whose catalytic functions have been assigned definitively:[3] galactonate dehydratase (GalD), glucarate dehydratase (GlucD), β-methylaspartate ammonia-lyase (β-MAL), *N*-acylamino acid racemase (NAAAR), and *o*-succinylbenzoate synthase (OSBS). Four orfs having unassigned catalytic functions were also discovered to be members of the superfamily (the first three were reported previously; Babbitt *et al.*, 1995): RspA, Spa2, rTS, and SynORF (Figure 4; see Table 1 for a key for abbreviations). An additional homolog, carboxyphosphonoenolpyruvate synthase (CPEPS), having an uncertain catalytic function was also identified. Finally, a DNA-binding protein from *Homo sapiens* designated MBP1 [cMyc promoter-binding protein (Ray & Miller, 1991)] was identified as a homolog of enolase, although no catalytic function has been associated with this protein.

Since three-dimensional structures are available only for MR, MLE I, and enolase, the alignments of the structurally uncharacterized sequences with these proteins are provisional but are sufficient to deduce (1) the common strategies that are used to catalyze the multiple reactions that are found in the superfamily and (2) rudimentary phylogenetic relationships among these proteins. These relationships can be inferred from the clustering of the sequences into the subgroups shown in Figure 4. The ordering of the various sequences within each subgroup is based upon the known or predicted importance of the active site residues in catalysis[4] and is consistent with the degree of similarity among each pair of sequences in the alignment as evaluated by their percent identities (data not shown). More sequence and structural information is necessary to assign phylogenetic relationships with greater confidence. However, assignment of the ligands for the divalent metal ion necessary for catalysis appears to be secure (with the exception of one ligand in GlucD; Table 2A). The conservation of these functional groups allows the prediction that a divalent metal ion is essential for the function of all of the proteins in this proposed superfamily. As discussed below, the alignment is also sufficient to identify or predict the general bases in the various active sites (Table 2B). However, the alignment is not sufficient to identify the homologs, if any, of the general acid catalyst (E317 in MR) in many of the other members of the superfamily (not included in Table 2B). Accordingly, this article focuses on the identities and properties of the putative general bases in the active sites of the newly identified members of the superfamily.

MR Subgroup

Those superfamily members with a conserved His presumably acting as an active site base at the position of H297 of MR were assigned to the MR subgroup. These include the enzymes galactonate dehydratase (GalD) and glucarate dehydratase (GlucD) and three reading frames of unknown function: RspA, Spa2, and rTS (Table 1 and Figure 5). Although the members of the enolase subgroup also exhibit a functionally important conserved His that aligns with H297, these residues apparently do not have direct roles in a proton abstraction step itself (*vide infra*).

Galactonate Dehydratase. The function of GalD, a previously unassigned orf in the *Escherichia coli* chromosome, was recently proposed on the basis of its homology with MR and MLE I (Babbitt *et al.*, 1995). The sequence alignment of GalD with MR reveals the presence of homologs for the metal ion ligands found in the active site of MR as well as a homolog for H297 (Table 2). Although no structure is yet available for GalD, the sequence alignment unequivocally demonstrates that a homolog of K166 in the active site of MR is not present in the active site of GalD since a diagnostic KXK sequence motif cannot be located. (K164 and K166 in the active site of MR and K167 and K169 in the active site of MLE I are located on the same strands of β-sheet and, therefore, are separated by one amino acid residue.) In GalD, this motif appears to be replaced with a KXN motif in which the N146 may be inferred to be a nonfunctional replacement for K166 in the active site of MR. By extension, K144 may be the homolog of the K166 in the active site of MR that interacts with the carboxylate oxygen of the substrate that is also coordinated to the essential Mg^{2+}. The presence of H285, the assigned homolog of H297, the (*R*)-specific base in the active site of MR, is in accord with the stereochemistry at carbon-2 of the substrate [also (*R*)] and supports the proposal that this residue initiates the β-elimination of water by abstraction of the α-proton. The subsequent vinylogous β-elimination reaction is expected to be facilitated by a general acidic catalyst, but the identity of this group cannot now be specified in the absence of

[3] In addition to the sequences shown in the alignment, the data bases were found to contain one additional sequence for an MLE I (MLE I utilizes unsubstituted *cis*,*cis*-muconate as a substrate; *Acinetobacter calcoaceticus*) and two additional sequences for MLE IIs (MLE II utilizes halogenated *cis*,*cis*-muconates as a substrate; *Alcaligenes eutrophus* and pseudomonad strain P51). Enolase sequences from a large range of organisms are available in the NCBI data bases. Two other proteins of identified function (GlucD and OSBS) are represented by sequences from several species (Table 1).

[4] The ordering of sequences given in Figure 4 differs slightly from that obtained with the PileUp algorithm in the GCG software package.

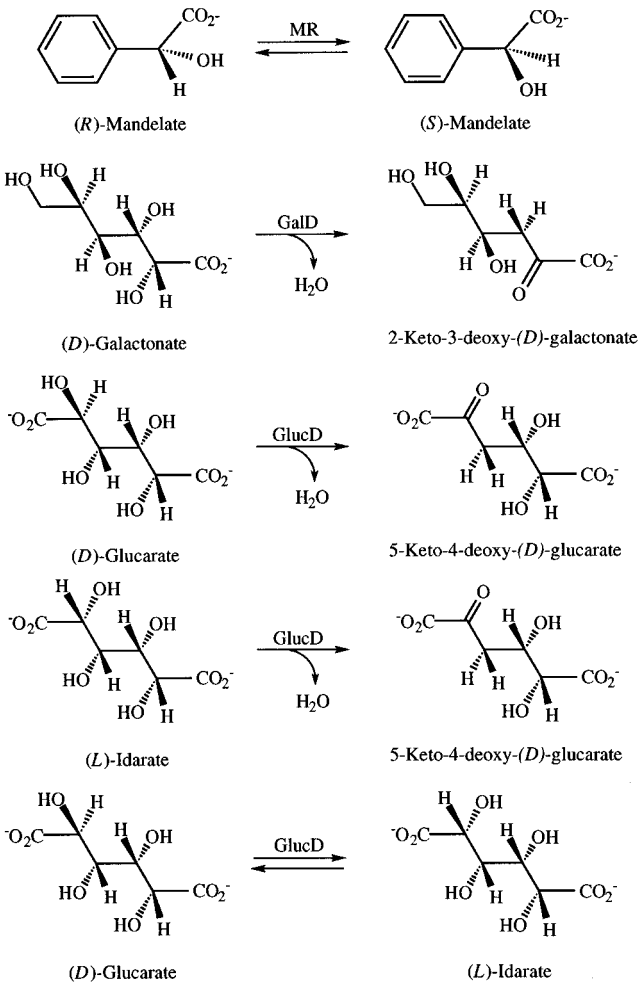


FIGURE 5: Reactions of the MR subgroup. The dehydration reactions catalyzed by GalD and GluD are shown as irreversible because the keto tautomer of the product is greatly favored relative to the enol tautomer.

structural information. Thus, by analogy to MR, we propose that the mechanism of the GalD-catalyzed reaction is initiated by abstraction of the α-proton to form a stabilized enolic intermediate in a general acid−general base-catalyzed reaction (Figure 2; Babbitt *et al.*, 1995). If the substrate binds in the active site in the extended conformation shown in Figure 5, the stereochemical course of the β-elimination reaction is expected to be *anti*.

Glucarate Dehydratase. We have confirmed the tentative functional identification of the GlucD from *P. putida* (K. C. Backman, personal communication) by expressing the gene in *E. coli*, purifying the protein to homogeneity, and verifying that it does, in fact, have GlucD activity (Palmer & Gerlt, 1996). GlucD contains homologs for two of the three metal ion ligands found in the active site of MR (Table 2A, Figure 4). The identity of the third putative metal ion ligand cannot be unequivocally assigned on the basis of the sequence alignment.

The mechanism of GlucD-catalyzed reaction was more difficult to rationalize on the basis of the deduced active site structure than that of the presumably similar GalD-catalyzed reaction. The dehydration reaction catalyzed by GlucD is initiated by abstraction of the proton from carbon-5 (Jeffcoat *et al.*, 1969). We expected that the active site of GlucD would contain a single general basic catalyst, the homolog of K166 in the active site of MR, since the absolute configuration of carbon-5 of D-glucarate is (*S*). However, the structural model that can be inferred from the sequence alignment suggests a more complex situation. The alignment reveals that GlucD contains homologs for *both* the (*S*)-specific (K213, the homolog of K166) and the (*R*)-specific (H345, the homolog of H297) bases in MR. A homolog for the K164 in the active site of MR (K211) is also present as shown by a KLK sequence in the appropriate portion of the primary sequence (Table 2B). Given the substantial divergence among the members of this superfamily, conservation of *both* active site bases in GlucD suggested that both would be likely to play a functional role. Contrary to our initial chemical model, the presence of a H297 homolog suggested that this enzyme might be competent to abstract a proton from an (*R*)-substrate as well (Figure 5).

We have tested this prediction and determined that GlucD can catalyze dehydration of both D-glucarate and L-idarate as well as their interconversion (Palmer & Gerlt, 1996; Figure 5);[5] D-glucarate and L-idarate are epimers at carbon-5, 5-(*S*) and 5-(*R*), respectively. The dehydration product obtained from D-glucarate results from β-elimination of water across the 5−4 bond. The dehydration product obtained from L-idarate also results from the β-elimination of water across the 5−4 bond [L-idarate has an axis of symmetry so the "top" (carbon-2 and -3) and "bottom" (carbon-4 and -5) halves of the molecule are stereochemically equivalent]. Thus, the GlucD-catalyzed dehydration products derived from both D-glucarate and L-idarate are identical. The k_{cat}s for dehydration of D-glucarate [reaction initiated by abstraction of the 5-(*S*) proton] and for L-idarate [reaction initiated by abstraction of the 5-(*R*) proton] are also approximately equal (∼4 s^{-1}), demonstrating that this enzyme is stereorandom with respect to abstraction of the α-proton from diastereomeric substrates. Consistent with the ability of GlucD to catalyze dehydration of both D-glucarate and L-idarate, the enzyme is also able to catalyze their epimerization in competition with dehydration.

From these observations, we can propose a mechanism for the GlucD-catalyzed reaction that is analogous to that proposed for GalD except for the additional ability to utilize both diastereomeric substrates (Figure 5). A likely consequence of the stereorandom nature of the reactions catalyzed by GlucD is that it catalyzes both *syn* (with D-glucarate as the substrate) and *anti* (with L-idarate as the substrate) β-elimination reactions, assuming that each substrate binds in an extended conformation. This lack of stereospecificity is an unusual feature since the stereochemical courses of enzyme-catalyzed β-elimination reactions of carboxylate substrates are almost always *anti* (Chari *et al.*, 1987; Creighton & Murthy, 1990).

Reading Frames with Unassigned Function: RspA, Spa2, and rTS. Three orfs of unassigned function (Table 1) were identified as members of the MR subgroup of the enolase superfamily on the basis of the presence of metal ion ligands homologous to those found in MR as well as a His homologous to H297 in MR (Figure 4). RspA is encoded by a gene in *E. coli* (*rspA*) that is involved in the onset of the stationary phase when metabolite depletion occurs (Huisman & Kolter, 1994). Spa2 is encoded by an orf in a

[5] A partially purified GlucD from *E. coli* has been reported to utilize L-idarate as a substrate at a rate $^1/_2$ of that observed for D-glucarate (Blumenthal, 1966).

region of the *Streptomyces ambofaciens* chromosome that is associated with "chromosomal instability" (Schneider *et al.*, 1993). The primary sequence of Spa2 is 65% identical to that of RspA, suggesting that these two proteins may catalyze the same reaction. rTS is a putative protein encoded by an mRNA from *H. sapiens* that accumulates in tumor cells and has been associated with methotrexate resistance in those cells (Dolnick, 1993). While the (*R*)-specific His base appears to be conserved in all three of these orfs, the sequence alignment is equivocal regarding the presence of homologs for K164 and K166 (Table 2B) for both RspA and Spa2. rTS, however, does contain a KXK motif representing homologs for K164 and K166 in the active site of MR. From the conservation of functionally important residues, we propose that all three of these proteins are enzymes and that they catalyze the abstraction of the α-proton of an α-hydroxy acid to form an enolic intermediate that is stabilized by hydrogen-bonding interactions with an electrophilic catalyst. Neither their specific chemical reactions nor their substrates are known at this time.

MLE I Subgroup

Members of the MLE I subgroup include MLE II, *N*-acylamino acid racemase (NAAAR), *β*-methylaspartate ammonia-lyase (*β*-MAL), *o*-succinylbenzoate synthase (OSBS), and a sequence of unknown function, SynORF (Table 1 and Figure 6). All MLE I subgroup members contain a KX(K/R) motif. In contrast to the members of the MR subgroup, however, no member of the MLE I subgroup contains a His-Asp dyad pair homologous to H297 and D270 in the active site of MR that function together to abstract the proton from an (*R*)-substrate (Schafer *et al.*, 1996). Instead, as shown in Figure 4, these proteins contain Gly/Ser in place of the H297 and exclusively Lys in place of D270.

N-Acylamino Acid Racemase. On the basis of the identities of the active site basic functional groups (*vide infra*), NAAAR belongs to the MLE I subgroup rather than to the MR subgroup.[6] Thus, from these considerations, one might expect NAAAR to catalyze α-proton abstraction of only the (*S*)-substrate using a homolog of K169, in analogy to the mechanism of MLE I-catalyzed reaction. But NAAAR is a racemase and, like MR, catalyzes a 1,1-proton transfer reaction that is necessarily stereorandom with respect to abstraction of the α-proton of the substrate; the reported k_{cat}s for the (*R*)- to (*S*)- and (*S*)- to (*R*)-directions are comparable (100 *vs* 65 s^{-1}, respectively; Tokuyama & Hatano, 1995a).

We propose that NAAAR contains two bases in its active site positioned for proton abstraction from either the (*R*)- or (*S*)-enantiomers of the substrate. The (*S*)-specific base is likely K163 that aligns with K166 of MR and K169 of MLE I. The (*R*)-specific base is likely K263 that aligns with K273 in MLE I, even though K273 is not thought to be involved directly in α-proton abstraction/delivery in MLE I (W. C. Barrett and J. A. Gerlt, unpublished results). Thus, in NAAAR, the (*R*)-specific base (Table 2B) could be considered to be a *functional* homolog of the H297 in the active site of MR but a *structural* homolog of the K273 in the active site of MLE I. This proposal is consistent with the observation that the functional group of K273 in MLE I can be superimposed on the functional group of H297 in MR, even though the α-carbons of these residues reside on different *β*-sheets and thus do not align in the primary structures.

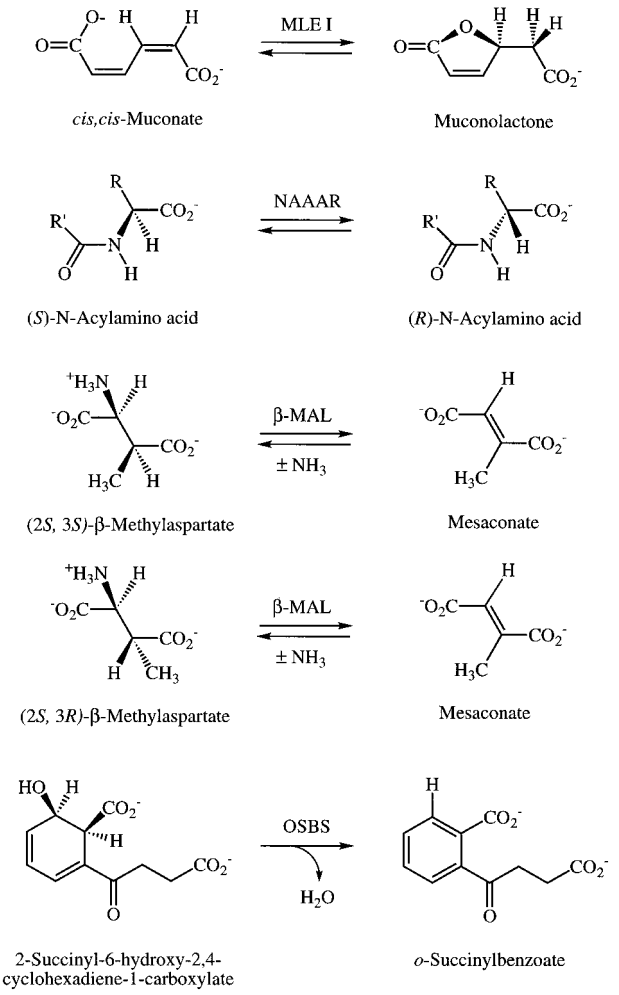


FIGURE 6: Reactions of the MLE subgroup. The dehydration reaction catalyzed by OSBS is shown as irreversible because the reaction results in aromatization of the carbocyclic ring.

β-Methylaspartate Ammonia-Lyase. In contrast to the other *β*-elimination reactions catalyzed by members of this superfamily that involve oxygen leaving groups, *β*-MAL (Goda *et al.*, 1992) catalyzes the *β*-elimination of ammonia. This enzyme utilizes either of the diastereomers of 3-methyl-(2*S*)-aspartate, although the k_{cat} for the (3*S*)-diastereomer exceeds that of the (3*R*)-diastereomer by a factor of 38 (490 *vs* 13 s^{-1}, respectively; Archer *et al.*, 1993). In each reaction, the geometry of the unsaturated product (mesaconate) is the same, indicating that the (3*S*)-substrate is converted to product *via* an *anti* *β*-elimination mechanism while the (3*R*)-substrate is converted to product *via* a *syn* *β*-elimination mechanism. Thus, *β*-MAL is a stereoselective, not a stereospecific, catalyst and, like GlucD, is capable of catalyzing either *syn* or *anti* *β*-elimination reactions depending upon the structure of the substrate.

Although *β*-MAL contains a homolog of K273 in MLE I (K331; Figure 4), the sequence alignment is equivocal with respect to the identity of the homolog of the active site base, K169, in MLE I; the sequence does not contain the diagnostic KXK motif. We assume that an (*S*)-specific base is necessarily present to accommodate the stereoselective nature

[6] Overall, NAAAR is also slightly more similar to MLE I than to MR: 30% identical *vs* 26% identical, respectively.

of the reaction. We note that both an RDR and a KLR sequence are present nearby in the primary sequence and might provide candidates for this functional role (Table 2B). The possibility that one of these Arg residues could function in this regard is supported by mutagenesis experiments with MR in which the k_{cat} of the K166R mutant is reduced approximately 10^3-fold relative to wild type MR (Kallarakal *et al.*, 1995). Thus, the guanidino functional group of an Arg residue could participate as a general basic catalyst in β-MAL as well. We expect that the β-elimination of ammonia is initiated by abstraction of the α-proton by an active site base to form a stabilized enolic intermediate in a general acid−general base-catalyzed reaction analogous to β-elimination reactions catalyzed by other superfamily members (Figure 2).

o-Succinylbenzoate Synthase. Like GalD and GlucD, OSBS (Sharma *et al.*, 1993) catalyzes a β-elimination of water that is initiated by abstraction of an α-proton. OSBS contains homologs for both K169 (K158; within a KVK motif) and K273 (K260) found in the active site of MLE I (Figure 4). Virtually nothing is known about either the substrate specificity or the mechanism for the OSBS-catalyzed reaction (Weische *et al.*, 1987), but our analysis would infer a catalytic mechanism that is likely to be similar to those described for MLE I, GalD, and GlucD. Given the structure of the substrate for OSBS, the orientation of the substrate (and, therefore, the position of the α-proton) relative to the ϵ-amino groups of both active site lysines is unknown, so the general basic catalyst that initiates the reaction cannot yet be specified. The presence of potential general bases on both sides of the active site may suggest that either (1) OSBS could be either a stereoselective or a stereorandom catalyst that utilizes both epimers at the α-carbon of the substrate or (2) in analogy with MLE I either a K169 or a K273 homolog stereospecifically abstracts the α-proton of the substrate while the other Lys residue is a spectator to catalysis.

SynORF. The sequence of SynORF was determined in the chromosomal-sequencing project for *Synechocystis* strain PCC6803 (Kaneko *et al.*, 1995). No function was assigned to SynORF. SynORF is most homologous to OSBS and MLE II from *P. putida*; however, at this time, there is insufficient information to predict the chemistry catalyzed by SynORF.

Enolase Subgroup

The sequence alignment in Figure 4 contains two sequences for enolases, one from *H. marismortui* (Kroemer & Arndt, 1991) and the other of the structurally characterized enzyme from *S. cerevisiae* (Holland *et al.*, 1981). On the basis of both high sequence similarity with many other enolases and conservation of residues known to be associated with enolase function, we have also assigned CPEPS and cMyc to the enolase subgroup (Table 1 and Figure 7).

The active sites of enolases contain a single general base, K345 in the enzyme from *S. cerevisiae*, that initiates the dehydration reaction by abstraction of the α-proton from 2-PGA (Figure 2). This Lys residue is homologous to K273 in MLE I. No (general basic) residue homologous to K169 in MLE I (or K166 in MR) is present in enolase as deduced from either the sequence alignment or three-dimensional structure (Table 2B); nor would such a homolog be expected, given the differences in chemical mechanism between MLE I and enolase.

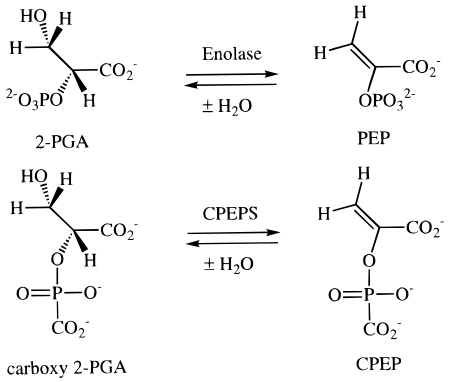
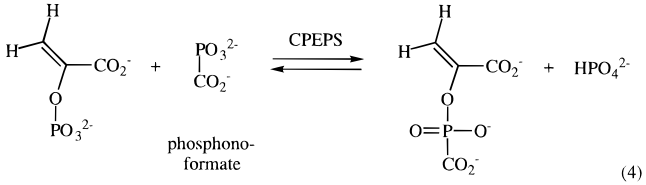


FIGURE 7: Reactions of the enolase subgroup.

That the functional groups of H297 in MR, K273 in MLE I, and K345 in enolase can be superimposed in the three-dimensional structures suggests that the active sites of the members of the superfamily can be modified to retain overall catalytic function but without retaining precise residue identity, position, and function. (Recall that both K273 in MLE I and K345 in enolase are structurally homologous to D270 in MR but their functional groups are spatially homologous to the functional group of H297 in MR.) We also note that H297 in MR and H373 in enolase can be aligned in their primary sequences (Figure 4), although these residues have different catalytic roles in the two enzymes. In the active site of MR, H297 functions as the (*S*)-specific base, whereas in the active site of enolase, H373 apparently positions the γ-carboxylic acid group of E211, the putative general acid catalyst. H373 is contained within an enolase "fingerprint" motif that is highly conserved in all known enolases (P. C. Babbitt, unpublished observations).

Carboxyphosphonoenolpyruvate Synthase. The function of CPEPS (Lee *et al.*, 1995) proposed in the literature is the transesterification reaction in which carboxyphosphonoenolpyruvate (CPEP) is generated from PEP and phosphonoformate (Hidaka *et al.*, 1990) as shown in eq 4. This reaction is mechanistically very distinct from those catalyzed by enolase, MR, MLE I, or any of the members of the superfamily previously discussed.



(4)

If, as suggested by our comparative studies, the interdependence of structure and function throughout the superfamily arises from a common chemical step, then the reaction proposed for CPEPS is likely incorrect. While we hypothesize that the CPEP product is correct, we propose that neither the proposed substrates nor the proposed mechanism is correct. Instead, we hypothesize that CPEP is generated by the β-elimination of water from carboxy 2-PGA as shown in Figure 7. This hypothesis is based on our expectation that all members of this superfamily will catalyze abstraction of the α-proton of a carboxylic acid to form an enolic intermediate (*vide infra*).

Our proposal for an alternate chemistry for CPEPS more consistent with its assignment to the enolase superfamily is supported by the available sequence data. For example, CPEPS contains a homolog for the general basic catalyst K345 in yeast enolase as well as a homolog for H373 within the signature sequence found in all enolases. In enolase, H373 is spatially proximal to the carboxylate group of E211, the putative general acidic catalyst in enolase; CPEPS does not appear to contain a homolog for E211.[7]

The alternate reaction we propose for CPEPS is also consistent with reported biochemical characterization of this enzyme. Using a cell-free extract (*vide infra*), carboxy 2-PGA could be generated from PEP and phosphonoformate after an initial enolase-catalyzed hydration of PEP to generate 2-PGA. Further, 2-PGA is a more logical substrate for a transesterification reaction than the previously suggested PEP since this strategy would avoid the requirement that CPEPS stabilize the enol of pyruvate in the active site. This enzyme has not been purified to homogeneity and characterized; instead, unfractionated cell-free preparations were observed to synthesize CPEP from PEP and phosphonoformate in the original description of this enzyme. Our alternate proposal is currently being investigated (M. A. Lies and J. A. Gerlt, work in progress).

MBP1. An orf encoding a DNA-binding protein, the cMyc promoter-binding protein (Ray & Miller, 1991), was identified on the basis of the presence of two metal ion ligands homologous to two of those found throughout the enolase superfamily. No homolog for the D247 in MR (D320 in enolase) could be located (Table 2A). While MBP1 contains a homolog for K345 in enolase, no homolog for E211 could be located (Table 2B).

MBP1 is globally similar to a large number of enolases, although (1) it is truncated at the N terminus (data not shown) and (2) both a metal ion ligand and a general acid that would be expected to assist the departure of a hydroxide ion leaving group apparently are missing from the sequence (Figure 4). Perhaps these presumed differences in active site structure result from sequencing errors, and MBP1 does, in fact, have a catalytic activity in addition to its reported DNA-binding activity. Alternatively, these alterations would be expected to make MBP1 devoid of catalytic activity as a result of the evolution of a DNA binding function. However, the conservation of "signature sequences" for enolase structure and function argues that MBP1 does, in fact, possess an enolase-like catalytic function that is yet to be determined. Interestingly, yeast enolase has some ability to bind single-stranded DNA (Al-Giery & Brewer, 1992).

Number of Metal Ions in the MR and MLE I Subfamilies: Predictions Based upon the Relationships to Enolase

With the discovery of the salient structure−function relationships characteristic of the enolase superfamily, we are now in a position to predict additional properties of the superfamily members, thereby providing a crucial test of the usefulness of our comparative approach. One important hypothesis arising from our characterization of the relationships between enolases and the MR and MLE I subgroups concerns the number of metal ions involved in their catalytic mechanisms. Although only one metal ion binding site is observed in MR and its mutants in the presence of ligands (Landro *et al.*, 1994) and in MLE I in the absence of a substrate or inhibitor (Helin *et al.*, 1995; M. Hasson, I. Schlichting, J. Moulai, D. Ringe, and G. A. Petsko, unpublished observations), the sequence and mechanistic homologies relating the MR and MLE I subgroups to the enolase subgroup suggest that members of the MR and MLE I subgroups may have a binding site for a second divalent metal ion. While no mechanistic role for a second metal ion is obvious for the 1,1-proton transfer reactions catalyzed by MR and NAAAR, the β-elimination of hydroxyl leaving groups catalyzed by GalD, GlucD, and OSBS and of the carboxylate group catalyzed by MLE I could reasonably be assisted by a divalent metal ion (*i.e.*, a Lewis acid). Although the second Mg^{2+} in the active site of enolase functions as an electrostatic/electrophilic catalyst, divergent evolution of the MR and MLE I subgroups from enolase could have allowed the second Mg^{2+} to acquire a new catalytic role as a Lewis acid catalyst. For example, a Lewis acid catalyst is consistent with the ability of GlucD to catalyze both *syn* and *anti* β-elimination reactions that are initiated by proton abstraction from carbon atoms having opposite configurations; *i.e.*, no catalytically unproductive proton transfers are possible between a divalent metal and the general basic catalyst that abstracts the α-proton. This would allow the general basic catalyst and the metal ion to be positioned either on the same or on the opposite faces of the active site (Gerlt & Gassman, 1992).

We note that the available crystal forms of MR were obtained from concentrated solutions of $(NH_4)_2SO_4$, conditions that interfered with the binding of the second Mg^{2+} to the active site of enolase (G. H. Reed, unpublished observations). For MLE I, no structure containing bound substrates or substrate analogs is available, obviating the opportunity to observe binding of a second metal ion in MLE I structures. We are currently investigating whether members of the MR and MLE I subgroups bind a second divalent metal ion that may be directly involved in catalysis.

Other Enzyme Superfamilies That Catalyze Abstraction of the α-Protons from Carboxylic Acids

A number of other enzyme superfamilies catalyze abstraction of the α-protons from carboxylic acids. The members of the fumarase superfamily (Williams *et al.*, 1992) catalyze the β-elimination of various leaving groups from carboxylic acids: fumarase (OH^-), aspartase (NH_3), arginosuccinate lyase (the guanidino group of arginine), adenylosuccinate lyase (the 6-amino group of adenine), and 3-carboxyl-*cis,cis*-muconate-lactonizing enzyme (intramolecular β-elimination of a carboxylate group). In each case, the product of the reaction is an α,β-unsaturated carboxylic acid (fumarate in the biproduct reactions and a substituted fumarate in the lactonizing enzyme) and the leaving group. The members of the 6-phosphogluconate dehydratase superfamily catalyze the β-elimination of OH^- from carboxylic acids: 6-phosphogluconate dehydratase and dihydroxy-acid dehydratase. The members of the aconitase superfamily also catalyze the

[7] Analysis of the reported sequence of CPEPS suggests the existence of multiple errors in the sequence, resulting in an apparent absence of a homolog for E211. First, the GC content is ~75%. Second, the N-terminal region is about 70 amino acids shorter than that of enolase which eliminates the ligands for the second Mg^{2+}. Third, the DNA sequence upstream of the putative initiation codon for CPEP synthase encodes an out-of-frame peptide sequence that is highly similar to the N-terminal sequence of enolase (P. C. Babbitt and J. A. Gerlt, unpublished observations).

β-elimination of OH^- from carboxylic acids: aconitase and 3-isopropylmalate dehydratase. Thus, while a number of superfamilies have evolved to accomplish α-proton abstraction from carboxylic acids, the enolase superfamily appears to be the most versatile in terms of the range of overall reactions that can be catalyzed.

Structural Basis for Divergent Evolution in the Enolase Superfamily

The active sites of β-barrel enzymes are found in cavities at their C-terminal ends (Brändén, 1980); the cavity is formed by the C-terminal ends of the β-strands and the loops that connect the β-strands to the following α-helices. In the context of understanding the diverse reactions catalyzed by members of the superfamily, this location for the active site undoubtedly is important. For example, in the active site of MR, K166 is located at the C-terminal end of the second β-strand, D195 is located at the end of the third β-strand, E221 is located at the end of the fourth β-strand, E247 is located on the loop that connects the fifth β-strand with the fifth α-helix, D270 is located at the end of the sixth β-strand, and H297 is located at the end of the seventh β-strand. (K164 is located within the second β-strand, and E317, the putative electrophilic catalyst, is located within the eighth β-strand; these residues lie on the "floor" of the active site cavity.) In the active sites of MLE I and enolase, the functional groups involved in catalysis are similarly located. Thus, the available "hot spots" for evolution of new functions in the enolase superfamily (M. S. Hasson, I. Schlichting, J. Moulai, K. Taylor, W. Barrett, G. L. Kenyon, P. C. Babbitt, J. A. Gerlt, G. A. Petsko, & D. Ringe, submitted for publication) are located in separate structural units that surround the substrates, thereby allowing chemistry to be redirected from a variety of geometric orientations. This architectural "master plan" provides a structurally flexible active site capable of evolving new functions *via* localized alteration of functionally important residues accompanied by structural reorganization of the cavity to accommodate varying substrates. At the same time, the design preserves a relative constancy in the overall fold, as reflected in the remarkably similar tertiary structures superimposed in Figure 3.

Implications for the Evolution of Catalytic Function

Our study of the enolase superfamily lends considerable weight to the proposal that new enzyme activities evolve by gene duplication followed by alterations in substrate specificity and/or mechanisms for processing a common intermediate (Petsko *et al.*, 1993). The origin of the ancestral enzyme molecules is a matter for continued speculation. However, the observations reported in this manuscript strongly corroborate the suggestion (Petsko *et al.*, 1993; Babbitt *et al.*, 1995) that chemistry, and not substrate specificity, is the more challenging problem to solve during evolution. The principle we espouse is in direct conflict with that initially suggested by Horowitz (1945) that enzymes are conscripted for new catalytic functions based on a common ability to bind a particular substrate.

CONCLUSIONS

We have identified a new superfamily of enzymes whose members catalyze the formation of an enolic intermediate *via* abstraction of the α-proton of a carboxylic acid. The enolic intermediate then partitions to different products *via* different chemistries in the various active sites, establishing that a common active site architecture can be modified to catalyze new reactions while retaining the ability to accomplish the facile abstraction of the α-proton of a carboxylic acid. Our conclusions provide persuasive support for the hypothesis that new enzyme activities evolve by recruitment of an existing enzyme catalyzing the necessary chemistry but differing in substrate specificity (Petsko *et al.*, 1993). Recognition of this principle coupled with our comparative approach led to our earlier prediction of the function of an unknown reading frame as GalD (Babbitt *et al.*, 1995). From a comparative analysis of the sequences and only a few available structures of the superfamily members, important additional predictions about the chemical mechanisms of the reactions catalyzed by many of the members can be inferred. Our analysis also suggests new possibilities in the chemistries of some of these enzymes (*e.g.*, lack of stereospecificity for GlucD, alternative function for CPEPS, and additional metal ligands in MR and MLE) that had not been imagined from previous, sometimes extensive, investigations focused on the single enzymes alone.

ACKNOWLEDGMENT

We thank Dr. Keith C. Backman for his generous gift of the plasmid containing the gene encoding GlucD from *P. putida*, Professor Debra Dunaway-Mariano for first pointing out the likely error in the CPEPS reaction as proposed in the literature, and Professor Gregory A. Petsko for helpful discussions. Figure 3 was generated using the MidasPlus program (Ferrin *et al.*, 1988) at the Computer Graphics Laboratory, University of California, San Francisco.

REFERENCES

al-Giery, A. G., & Brewer, J. M. (1992) *Biochim. Biophys. Acta 1159*, 134.
Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990) *J. Mol. Biol. 215*, 403.
Archer, C. H., Thomas, N. R., & Gani, D. (1993) *Tetrahedron: Asymmetry 4*, 1141.
Avigad, G., & Englard, S. (1969) *Fed. Proc. 28*, 345 (Abstract 486).
Babbitt, P. C., Mrachko, G. T., Hasson, M. S., Huisman, G. W., Kolter, R., Ringe, D., Petsko, G. A., Kenyon, G. L., & Gerlt, J. A. (1995) *Science 267*, 1159.
Blumethal, H. J. (1966) *Methods Enzymol. 9*, 660.
Brändén, C.-I. (1980) *Q. Rev. Biophys. 13*, 317.
Burland, V. D., Plunkett, G., Daniels, D. L., & Blattner, F. R. (1993) *Genomics 16*, 551.
CCP4 (1994) *Acta Crystallogr., Sect. D: Biol. Crystallogr. 50*, 760.
Chari, R. V. J., Whitman, C. P., Kozarich, J. W., Ngai, K.-L., & Ornston, L. N. (1987) *J. Am. Chem. Soc. 109*, 5514.
Creighton, D. J., & Murthy, N. S. R. K. (1990) *Enzymes 19*, 323.
Dolnick, B. J. (1993) *Nucleic Acids Res. 21*, 1747.
Faller, L. D., Baroudy, B. M., Johnson, A. M., & Ewall, R. X. (1977) *Biochemistry 16*, 3864.
Farber, G. K., & Petsko, G. A. (1990) *Trends Biochem. Sci. 15*, 228.
Ferrin, T. E., Huang, C. C., Jarvis, L. E., & Langridge, R. (1988) *J. Mol. Graphics 6*, 13.
Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J.-F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., Fitzhugh, W., Fields, C., Gocayne, J. D., Scott, J. D., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips,

C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O., & Venter, J. C. (1995) *Science 269*, 496.
Gardner, J. (1995) *SeqVu*, The Garvan Institute of Medical Research, Sydney, Australia.
Gerlt, J. A., & Gassman, P. G. (1992) *J. Am. Chem. Soc. 114*, 5928.
Gerlt, J. A., & Gassman, P. G. (1993a) *J. Am. Chem. Soc. 115*, 11552.
Gerlt, J. A., & Gassman, P. G. (1993b) *Biochemistry 32*, 11943.
Gerlt, J. A., Kenyon, G. L., Neidhart, D. J., Petsko, G. A., & Powers, V. M. (1992) *Curr. Opin. Struct. Biol. 2*, 736−742.
Ghosal, D., & You, I. S. (1989) *Gene 83*, 225.
Goda, S. K., Minton, N. P., Botting, N. P., & Gani, D. (1992) *Biochemistry 31*, 10747.
Helin, S., Kahn, P. C., Guha, B. L., Mallows, D. G., & Goldman, A. (1995) *J. Mol. Biol. 254*, 918.
Hidaka, T., Hara, O., Imai, S., Anzai, H., Murakami, T., Nagaoka, K., & Seto, H. (1990) *Agric. Biol. Chem. 54*, 2121.
Holland, M. J., Holland, J. P., Thill, G. P., & Jackson, K. A. (1981) *J. Biol. Chem. 256*, 1385.
Horowitz, N. H. (1945) *Proc. Natl. Acad. Sci. U.S.A. 31*, 153.
Houghton, J. E., Brown, T. M., Appel, A. J., Hughes, E. J., & Ornston, L. N. (1995) *J. Bacteriol. 177*, 401.
Huisman, G. W., & Kolter, R. (1994) *Science 265*, 537.
Jeffcoat, R., Hassall, H., & Dagley, S. (1969) *Biochem. J. 115*, 969.
Jones, T. A., Zou, J. Y., Cowan, S. W., & Kjeldgaard, M. (1991) *Acta Crystallogr., Sect. A: Fundam. Crystallogr. 47*, 110.
Kabsch, W. (1978) *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. 32*, 922.
Kallarakal, A. T., Mitra, B., Kozarich, J. W., Gerlt, J. A., Clifton, J. R., Petsko, G. A., & Kenyon, G. L. (1995) *Biochemistry 34*, 2788.
Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N., Sugiura, M., & Tabata, S. (1995) *DNA Res. 2*, 153.
Kenyon, G. L., Gerlt, J. A., Petsko, G. A., & Kozarich, J. W. (1995) *Acc. Chem. Res. 28*, 178.
Kroemer, W. J., & Arndt, E. (1991) *J. Biol. Chem. 266*, 24573.
Landro, J. A., Kallarakal, A. T., Ransom, S. C., Gerlt, J. A., Kozarich, J. W., Neidhart, D. J., & Kenyon, G. L. (1991) *Biochemistry 30*, 9274.
Landro, J. A., Gerlt, J. A., Kozarich, J. W., Koo, C. W., Shah, V. J., Kenyon, G. L., Neidhart, D. J., Fujita, S., & Petsko, G. A. (1994) *Biochemistry 33*, 635.
Larsen, T. M., Wedekind, J. E., Rayment, I., & Reed, G. H. (1996) *Biochemistry 35*, 4349.
Lebioda, L., & Stec, B. (1988) *Nature 333*, 683.
Lee, S.-H., Hidaka, T., Hankashita, H., & Seta, H. (1995) *Gene 153*, 143.
Mitra, B., Kallarakal, A. T., Kozarich, J. W., Gerlt, J. A., Clifton, J. R., Petsko, G. A., & Kenyon, G. L. (1995) *Biochemistry 34*, 2777.
Neidhart, D. J., Kenyon, G. L., Gerlt, J. A., & Petsko, G. A. (1990) *Nature 347*, 692.
Ogawa, K.-I., Akagawa, E., Nakamura, K., & Yamane, K. (1995) *Microbiology 141*, 269.
Palmer, D. R. J., & Gerlt, J. A. (1996) *J. Am. Chem. Soc. 118*, 10323.
Pawluk, A., Scopes, R. K., & Griffiths-Smith, K. (1986) *Biochem. J. 238*, 275.
Pearson, W. R., & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 2444.
Petsko, G. A., Kenyon, G. L., Gerlt, J. A., Ringe, D., & Kozarich, J. W. (1993) *Trends Biochem. Sci. 18*, 372.
Poyner, R. R., & Reed, G. H. (1992) *Biochemistry 31*, 7166.
Poyner, R. R., Laughlin, L. T., Sowa, G. A., & Reed, G. H. (1996) *Biochemistry 35*, 1692.
Ransom, S. C., Gerlt, J. A., Powers, V. M., & Kenyon, G. L. (1988) *Biochemistry 27*, 540.
Ray, R., & Miller, D. M. (1991) *Mol. Cell. Biol. 11*, 2154.
Schafer, S. L., Barrett, W. C., Kallarakal, A. T., Mitra, B., Kozarich, J. W., Gerlt, J. A., Clifton, J. G., Petsko, G. A., & Kenyon, G. L. (1996) *Biochemistry 35*, 5662.
Schneider, D., Aigle, B., Leblond, P., Simonet, J.-M., & Decaris, B. (1993) *J. Gen. Microbiol. 139*, 2559.
Schurig, H., Rutkat, K., Rachel, R., & Jaenicke, R. (1995) *Protein Sci. 4*, 228.
Shanley, M. S., Harrison, A., Parales, R. E., Kowalchuk, G., Mitchell, D. J., & Ornston, L. N. (1994) *Gene 138*, 59.
Sharma, V., Meganathan, R., & Hudspeth, M. E. S. (1993) *J. Bacteriol. 175*, 4917.
Singh, R. P., & Setlow, P. (1978) *J. Bacteriol. 134*, 353.
Tokuyama, S., & Hatano, K. (1995a) *Appl. Microbiol. Biotechnol. 42*, 853.
Tokuyama, S., & Hatano, K. (1995b) *Appl. Microbiol. Biotechnol. 42*, 884.
Tsou, A. Y., Ransom, S. C., Gerlt, J. A., Buechter, D. D., Babbitt, P. C., & Kenyon, G. L. (1990) *Biochemistry 29*, 9856.
van der Meer, J. R., Eggen, R. I., Zehnder, A. J. B., & DeVos, W. M. (1991) *J. Bacteriol. 173*, 2425.
Wedekind, J. E., Poyner, R. R., Reed, G. H., & Rayment, I. (1994) *Biochemistry 33*, 9333.
Wedekind, J. E., Reed, G. H., & Rayment, I. (1995) *Biochemistry 34*, 4325.
Weische, A. W., Garvert, W., & Leistner, E. (1987) *Arch. Biochem. Biophys. 256*, 223.
Williams, S. E., Woolridge, E. M., Ransom, S. C., Landro, J. A., Babbitt, P. C., & Kozarich, J. W. (1992) *Biochemistry 31*, 9768.

BI9616413