Evolution of function in (β/α)$_8$-barrel enzymes

John A Gerlt[*] and Frank M Raushel[†]

The (β/α)$_8$-barrel is the most common fold in structurally characterized enzymes. Whether the functionally diverse enzymes that share this fold are the products of either divergent or convergent evolution (or both) is an unresolved question that will probably be answered as the sequence databases continue to expand. Recent work has examined natural, designed, and directed evolution of function in several superfamilies of (β/α)$_8$-barrel containing enzymes.

Addresses
[*]Departments of Biochemistry and Chemistry, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801, USA
e-mail: j-gerlt@uiuc.edu
[†]Department of Chemistry, Texas A&M University, PO Box 30012, College Station, Texas 77842-3012, USA
e-mail: raushel@tamu.edu

Abbreviations

BLAST	basic local alignment sequence tool
DHO	dihydroorotase
HisA	phosphoribosylformimino-5-aminoimidazole carboxamide ribonucleotide isomerase
HisF	imidazole glycerolphosphate synthase
GalD	D-galactonate dehydratase
GlucD	D-glucarate dehydratase
HPS	D-arabino-hex-3-ulose 6-phosphate synthase
HYD	hydantoinase
KDGP	2-keto-3-deoxy-6-phosphogluconate
KGPDC	3-keto-L-gulonate 6-phosphate decarboxylase
MAL	3-methylaspartate ammonia lyase
MLE	muconate lactonizing enzyme
MR	mandelate racemase
OMP	orotidine 5′-monophosphate
OSBS	o-succinylbenzoate synthase
PRAI	N-(5′-phosphoribosyl)anthranilate isomerase
PSI-BLAST	position sensitive interactive BLAST
PTE	phosphotriesterase
RhamD	L-rhamnonate dehydratase
RPE	D-ribulose 5-phophate 3-epimerase
SCOP	Structural Classification of Proteins
URE	urease

Introduction

The (β/α)$_8$-barrel fold is the most common in enzymes, with ∼10% of structurally characterized proteins containing at least one domain with this fold (Figure 1) [1[•]]. The closed, parallel β-sheet structure of the (β/α)$_8$-barrel is formed from eight parallel (β/α)-units linked by hydrogen bonds that form a cylindrical core. Despite its eightfold pseudosymmetry, the packing within the interior of the barrel is better described as four (β/α)$_2$-subdomains in which distinct hydrophobic cores are located between the β-sheets and flanking α-helices [2].

The active sites are located at the C-terminal ends of the β-strands. So placed, the functional groups surround the active site and are structurally independent: the 'old' and 'new' enzymes retain functional groups at the ends of some β-strands, but others are altered to allow the 'new' activity [3]. With this blueprint, the (β/α)$_8$-barrel is optimized for evolution of new functions. With a rapidly increasing number of sequences and structures, the challenge is to understand how nature uses the (β/α)$_8$-barrel scaffold in natural evolution so that *in vitro* methods can be used to devise catalysts for new reactions.

Using both structural inspection and substrate specificity, the SCOP database (http://scop.mrc-lmb.cam.ac.uk/scop/) clusters proteins into 'superfamilies' that evolved from a common ancestor; the clustering assumes that divergent evolution retains discernible structural similarity even if sequence similarity has disappeared [4]. In the most recent release (1.61; November 2002), (β/α)$_8$-barrel proteins were clustered into 25 superfamilies. Given that the topologies of all (β/α)$_8$-barrels are similar (closed, cylindrical structures) and that retention of substrate specificity need not be important in evolutionary processes, the SCOP database risks either placing homologous enzymes into separate superfamilies or placing non-homologous enzymes into the same superfamily.

Using sequence relationships detected with PSI-BLAST, Copley and Bork [5] concluded that 12 of the superfamilies are derived from a common ancestor. Are the remaining superfamilies the result of convergent evolution? Or have all (β/α)$_8$-barrel enzymes diverged from a common ancestor? These questions can be answered as sequences and structures of 'new' proteins provide 'missing links' to connect the superfamilies.

Evolution from libraries of fractional barrels?

The (β/α)$_2$-subdomain structure allows the suggestion that (β/α)$_8$-barrels are assembled from (β/α)$_{2N}$-precursors. Mixing precursors that separately deliver binding or catalysis might produce the progenitors of superfamilies. After that process, evolution of function could occur by gene duplication followed by mutation.

Figure 1
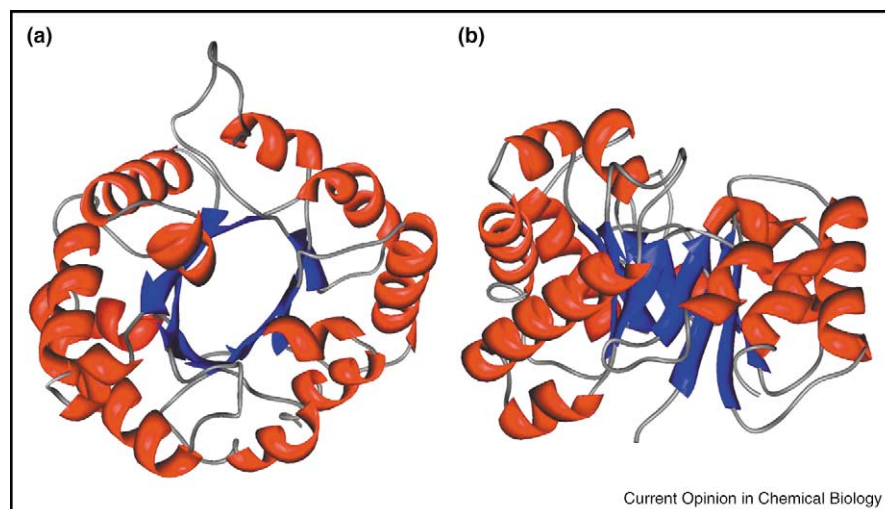
The (β/α)$_8$-barrel in yeast triosephosphate isomerase (7TIM.pdb) viewed from **(a)** the C-terminal ends of the β-strands and **(b)** from the side.

Evidence to support the formation of (β/α)$_8$-barrels from (β/α)$_{2N}$-subdomains includes:

1. HisA and HisF, which catalyse successive reactions in histidine biosynthesis, are homologous in both sequence and structure [6•]. Each were formed by fusion of two equivalent (β/α)$_4$-units. The N- and C-terminal (β/α)$_4$-units in HisF are stable polypeptides that form both homodimeric and heterodimeric structures in solution [7•].

2. With few exceptions, barrels have the (β/α)$_8$-fold. The pertinent exceptions are enolase with a βααβ(β/α)$_6$-barel fold [8] and quinolinate phosphoribosyltransferase with a (β/α)$_6$-fold [9]. These irregular barrels can be explained by modular construction that deviates in detail from the usual (β/α)$_8$-barrel.

3. A search of the Protein Data Bank with the (β/α)$_4$-half barrels of HisA and HisF revealed sequence and structure similarity with a (β/α)$_5$-flavodoxin-like domain found in several proteins [10]. The discovery of this homologous substructure supports the hypothesis that a progenitor (β/α)$_4$-structure existed.

Irrespective of whether the formation of (β/α)$_8$-barrels occurs *in toto* or in pieces, the strategies that nature uses in divergent evolution of intact (β/α)$_8$-barrels have important implications for genome annotation and *de novo* design of catalysts.

Natural evolution

A reasonable pathway for evolution of function involves duplication of the gene encoding the progenitor so that the original function can be retained while the duplicate undergoes divergence of sequence and function.

How is the progenitor chosen? Two criteria can be envisaged:

1. The progenitor 'accidentally' catalyzes the desired new reaction, probably at a low but sufficient rate to provide selective advantage, immediately after gene duplication. Such catalytic promiscuity apparently is wide-spread (e.g. a protein that catalyzes the *o*-succinylbenzoate synthase (OSBS) reaction in menaquinone biosynthesis was first identified by its *N*-acylamino acid racemase activity [11]).

2. The progenitor does not catalyse the desired new reaction, but a limited number of mutations (one?) yields the 'new' function, presumably at the expense of the original reaction. When this mutation occurs, selection is possible and optimization can proceed. Until recently, no example of such evolutionary potential had been discovered in enzymes that possess the (β/α)$_8$-barrel fold [12••].

Three strategies can be envisaged for the relationship between the reactions catalyzed by the progenitor and 'new' enzyme [13]:

1. Chemical mechanism is dominant. The mechanism of progenitor can be used to catalyse the 'new' reaction. In the simplest example, the progenitor and 'new' enzyme catalyse the same chemical reaction but differ in substrate specificity. In more complex examples, the reactions share a structural strategy to stabilize a common intermediate or transition state. The 'enolase C-terminal domain-like' and 'metallo-dependent hydrolases' superfamilies provide examples of this strategy.

2. Ligand specificity is dominant. Metabolic pathways can evolve 'linearly' (from start to finish or *vice versa*) so the progenitor and 'new' enzymes share a substrate/product. Enzymes in histidine and tryptophan biosynthesis are examples of this strategy. The SCOP database assumes such a strategy in the 'ribulose-phosphate binding' and 'phospho*enol* pyruvate/pyruvate domain' superfamilies.

3. Active site architecture is dominant. Nature selects a progenitor with functional groups that can be used to catalyse a mechanistically distinct 'new' reaction. The first example of this opportunistic strategy is provided by orotidine 5′-monophosphate (OMP) decarboxylase and 3-keto-L-gulonate 6-phosphate decarboxylase (KGPDC), mechanistically distinct decarboxylases that have homologous sequences and structures.

Enolase superfamily

The 'enolase C-terminal domain-like' superfamily in SCOP contains those enzymes that had been recognized as homologues using strict structural and mechanistic criteria. This superfamily is the paradigm for superfamilies whose members catalyse different reactions sharing a common partial reaction: general-base catalyzed formation of an enolic intermediate that is stabilized by coordination to an active site Mg^{2+}; depending on the identity of the reaction, the intermediate is partitioned to different products using either conserved or reaction-specific catalysts [14]. The members share two conserved domains: an N-terminal domain that determines substrate specificity; and a $(\beta/\alpha)_8$-barrel domain that contains the catalytic groups. Conservation of the bidomain structure provides compelling evidence that the enolase superfamily represents divergent evolution from a common progenitor.

The members are partitioned into three groups on the basis of the identities of the three metal ion ligands (located at the ends of the third, fourth, and fifth β-strands) and the acid/base catalysts. The enolases contain a lysine (at the end of the sixth β-strand) that catalyzes formation of the enolic intermediate; the general acid is located on a long loop inserted after the second β-strand. Other members are similar to muconate lactonizing enzyme (MLE) and contain lysine acid/base catalysts on opposite sides of the active site (at the ends of the second and sixth β-strands). Others are most similar to mandelate racemase (MR) and contain a His–Asp dyad (at the ends of the seventh and sixth β-strands, respectively) and often, but not always, a lysine on the opposite side of the active site (at the end of the second β-strand).

An understanding of the range of functions catalyzed by members of the enolase superfamily is essential for dissecting the structural changes involved in divergent evolution.

Recently, homologues of MLE encoded by the *Escherichia coli* and *Bacillus subtilis* genomes were assigned a novel function. Although MLE catalyzes cycloisomerization and the homologous OSBS catalyzes dehydration, the OSBS from *Amycolaptosis* also catalyzes racemization of *N*-acyl-amino acids [11]. Genome context as well as the 1,1-proton transfer reaction catalyzed by the promiscuous OSBS were used to deduce the L-Ala-D/L-Glu epimerase function [15,16]. L-Ala-D-Glu is a constituent of the murein peptide, and the epimerase, endopeptidases, and binding proteins encoded by proximal genes apparently are involved in utilization of the murein peptide as carbon source.

The 'canonical' acid/base catalysts found in enolase, MLE, and MR are not sufficient to catalyse the range of reactions associated with the enolase superfamily. D-Galactonate dehydratase (GalD) from *E. coli* catalyzes *anti*-dehydration involving base-catalyzed proton abstraction from C-2 and acid-catalyzed departure of the hydroxide leaving group from C-3. GalD contains the conserved His–Asp general base but is lacking the acidic lysine at the end of the second β-strand. Instead, His185, a 'new' general acid, and Asp183, a Mg^{2+} ligand, are located at the end of the third β-strand [17].

A *syn*-dehydration is catalyzed by D-glucarate dehydratase (GlucD) from *E. coli*; a single acid/base, the His339–Asp313 dyad, has been implicated as the base for proton abstraction as well as the acid for hydroxide departure [18,19]. L-Rhamnonate dehydratase (RhamD) from *E. coli* also catalyzes *syn*-dehydration. In this case, the available data point to base-catalyzed proton abstraction from C-2 by the conserved His329–Asp302 dyad. However, the 'new' acid is probably His281, located at the end of the fifth β-strand (following the Asp280 ligand for the essential Mg^{2+}). Why two enzymes that catalyse *syn*-dehydration use different acid catalysts is unknown; however, this mystery underscores the difficulty in deducing function from sequence or structure alone.

3-Methylaspartate ammonia lyase (MAL), which catalyzes *anti*-deamination, was proposed to be a member of the enolase superfamily on the basis of limited sequence homology: only four residues, the ligands for the essential Mg^{2+} and the lysine general base, are conserved in MAL and other members [14]. Structural studies of two MALs revealed that these share the bidomain structure characteristic of the enolase superfamily, confirming their membership in the superfamily [20,21]. MAL, like enolase, which catalyzes *anti*-dehydration, utilizes a lysine at the end of the sixth β-strand as the base, but no acid group is located on the opposite face of the active site; the ammonia leaving group does not require protonation for departure.

Eleven reactions now have been associated with the enolase superfamily: enolase; MR and five acid sugar

dehydratases (GalD, GlucD, RhamD, D-altronate dehydratase, D-gluconate dehydratase; C Millikin and JA Gerlt, unpublished results) that constitute the MR subgroup; MLE, OSBS, and L-Ala-D/L-Glu epimerase that constitute the MLE subgroup; and MAL (Figure 2). The databases contain > 380 members of the superfamily, excluding the ubiquitous enolases. In some cases, 'old' (orthologous) functions can be assigned from sequence identity and genome context. However, the identities of most are unknown, and these often contain active site motifs that differ from those established for the 'old' functions. The challenge is assignment of 'new' functions that expand our understanding of the limits of natural evolution.

Amidohydrolase superfamily

The 'metallo-dependent hydrolase' superfamily in SCOP was designated as the 'amidohydrolase superfamily' by Holm and Sander [22]. Subsequent analyses have revealed that members of this superfamily catalyse the hydrolysis of a wide variety of substrates at tetrahedral phosphorus and trigonal carbon centres. A partial list of substrates recognized by members of this superfamily of enzymes is graphically depicted in Figure 3.

The hallmark for this superfamily of enzymes is a mononuclear or binuclear metal centre that functions to activate a hydrolytic water molecule for nucleophilic attack. In some instances the metal centre also serves to enhance the electrophilic character of the substrate. In all members of this superfamily the metal centre is found at the C-terminal end of the central (β/α)₈-barrel domain. The most common structural motif identified thus far for enzymes with two divalent cations is the one represented by dihydroorotase (DHO), as illustrated in Figure 4a [23•]. Virtually identical binuclear metal centres are found in phosphotriesterase (PTE), urease (URE) and hydantoinase (HYD). In this particular example, the two metal ions in DHO are separated by about 3.5 Å. The more solvent-shielded metal ion (M_α) is ligated to two histidine residues from the first β-strand and an aspartic acid from the eighth β-strand. The more solvent-exposed metal ion (M_β) is ligated to two histidine residues from the fifth and sixth β-strands. In addition to these five ligands, the metal ions are bridged by a carbamate functional group formed from the post-translational carboxylation of the ε-amino group of a lysine from the fourth β-strand. Finally the two metal ions are bridged by a hydroxide from solvent.

The functional significance of the carboxylated lysine residue that bridges the two metal ions is not clear. This particular post-translational modification might serve as a control mechanism for the regulation of catalytic activity or it might simply be an evolutionary relic. *In vitro* experiments with PTE have demonstrated that the binuclear metal centre will self-assemble in the presence of bicarbonate and the appropriate divalent cation [24]. However, a protein catalyst for the carboxylation reaction

has not been identified thus far. In at least one enzyme, designated as the PTE homology protein (PHP), the bridging lysine residue has been replaced by a glutamate residue as illustrated in Figure 4b [25]. Although the function of this protein is unknown, this structural perturbation suggests that a carbamate functional group is not required to bridge the two metal ions.

With adenosine and cytosine deaminase [26], the bridging ligand is missing altogether and the protein apparently functions with a single divalent cation in the active site. This perturbation illustrates the structural and functional plasticity of the active site within the amidohydrolase superfamily of enzymes. The other active site ligands (four histidines and the aspartic acid) are present in the subfamily but the catalytic functions have been altered. The structure of the mononuclear metal centre is illustrated in Figure 4c for cytosine deaminase [26]. Here the single divalent cation is ligated to two histidine residues from the first β-strand and an aspartic acid from the eighth β-strand. However, the histidine from the fifth β-strand now serves as the final ligand to the lone metal ion. The histidine from the fourth β-strand does not ligate to the metal ion but functions as a general base catalyst for the activation of the nucleophilic water molecule.

The most structurally diverse example identified to date is found within the active site of human renal dipeptidase [27]. This enzyme binds two divalent cations but the bridging lysine residue has been replaced by a glutamic acid. Moreover, one of the histidine residues from the first β-strand and the aspartic acid from the eighth β-strand have been replaced by a single aspartic acid from the first β-strand. The structure of this binuclear metal centre is presented in Figure 4d.

The identity of the specific metal ion activators within the amidohydrolase superfamily has not been conserved. With URE there appears to be a requirement for Ni^{2+} [28], whereas with PTE, DHO, and HYD the 'natural' metal ion activator appears to be Zn^{2+}, although PTE functions perfectly well with Co^{2+}, Ni^{2+}, Cd^{2+} or Mn^{2+} [29]. Cytosine deaminase requires Fe^{2+} for activity, although the structurally and functional equivalent adenosine deaminase requires Zn^{2+} [26].

What is not so clear is the functional requirement for either a binuclear or mononuclear metal centre. The conservation of the metal ligands for these two subfamilies clearly indicates a divergence from a common ancestral parent. One plausible functional requirement may emanate from the reactions catalyzed by these two systems. With adenosine and cytosine deaminases, cleavage of the C–N bond requires the protonation of both products and thus the resting state of the protein may require a bound water for the supply of both protons. By contrast, with DHO and URE, only the amino group product requires protonation

Figure 2

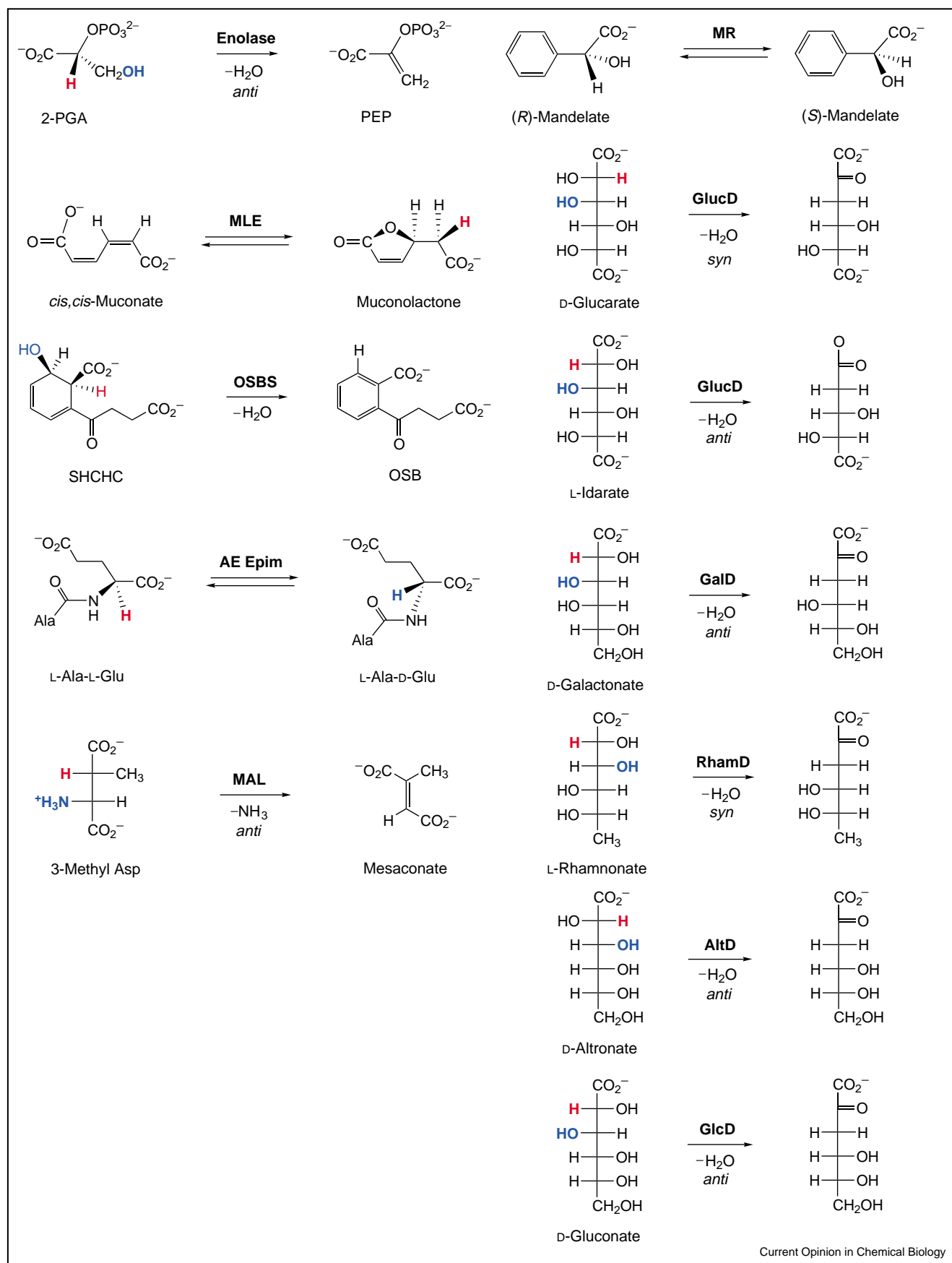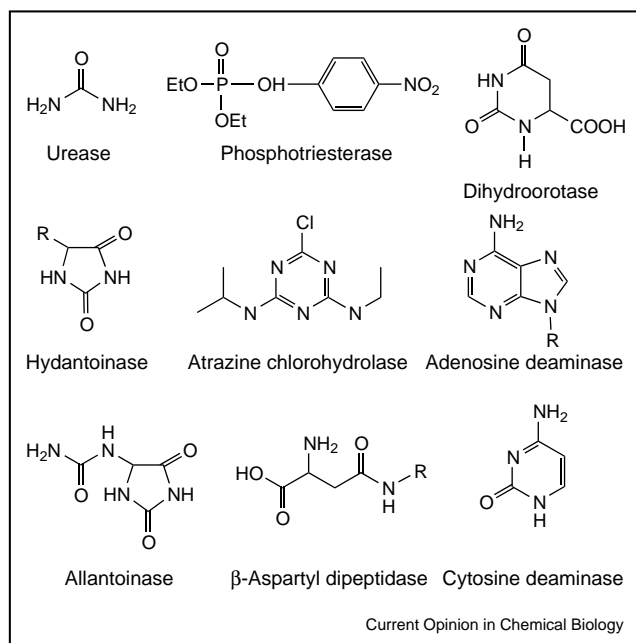

Current Opinion in Chemical Biology

Figure 3

Substrates recognized by members of the amidohydrolase superfamily.

and thus the resting state of the enzyme may only require an hydroxide bound between the two divalent cations.

OMP decarboxylase suprafamily

The 'ribulose-phosphate binding barrel' superfamily in SCOP includes enzymes that catalyse mechanistically distinct reactions:

1. Phosphoribosylformimino-5-aminoimidazole carboxamide ribonucleotide isomerase (HisA) and imidazole glycerolphosphate synthase (HisF) in histidine biosynthesis.
2. *N*-(5′-phosphoribosyl)anthranilate isomerase (PRAI), indole-3-glycerol phosphate synthase (IPGS) and the α-subunit of tryptophan synthase (cleavage of indole glycerol phosphate) in tryptophan biosynthesis.
3. D-Ribulose 5-phosphate 3-epimerase (RPE) in the pentose phosphate cycle.
4. OMP decarboxylase in pyrimidine nucleotide biosynthesis.

These are clustered on the basis of a conserved phosphate binding motif at the ends of the seventh and eighth β-strands. The recent conclusion that (β/α)₈-barrels can be formed from (β/α)₂ₙ-subdomains has the potential of invalidating the previous assumption that (β/α)₈-barrels evolve *en toto*. So, the question remains whether the members of this SCOP superfamily are true homologues. BLASTP searches of the databases provide both 'new'

members of the 'ribulose-phosphate binding barrel' superfamily as well as 'missing links' that provide persuasive evidence that OMP decarboxylase and RPE are homologous enzymes. OMP decarboxylases are homologous to both D-arabino-hex-3-ulose 6-phosphate synthase (HPS) and KGPDC, the latter incorrectly annotated as 'probable hexulose phosphate synthases' in several bacterial genomes. These constitute the members of the OMP decarboxylase suprafamily, the members of which catalyse different overall reactions with unrelated mechanisms (Figure 5) [12••,13]. KGPDC participates in a newly discovered pathway for L-ascorbate fermentation [30]. The OMP decarboxylase-catalyzed reaction is metal-ion-independent [31] and must avoid a vinyl anion intermediate that is too unstable to exist; HPS [32] and KGPDC [30] use Mg^{2+} to stabilize enediolate intermediates in aldol condensation and decarboxylation, respectively. The KGPDC from *E. coli*, like the OMP decarboxylases [33–36], is a dimer of identical polypeptides [12••]. Structural superposition reveals conservation of the polypeptide interface, confirming divergent evolution. (Conservation of the (β/α)₈-barrel fold in a single domain protein is necessary but insufficient to prove evolution from a common progenitor.)

OMP decarboxylases and KGPDC share conserved active site functional groups: in OMP decarboxylase, these are involved in a hydrogen-bonding network that probably destabilizes the substrate; in KGPDC, these provide ligands for the essential Mg^{2+} (Figure 5). Thus, divergent evolution can produce homologous enzymes that catalyse different reactions with unrelated mechanisms. When the sequences of HPSs are used to query the databases, RPEs are identified that share 28% sequence identity (Figure 5) (P Babbitt, W Novak and JA Gerlt, unpublished data). The active site groups are not strictly conserved, but their positions are conserved. RPE and HPS stabilize enediolate intermediates derived from ketose phosphates; however, the HPS-catalyzed reaction requires Mg^{2+} whereas the RPE-catalyzed reaction is independent of divalent metal ions [37,38]. Thus, divergent evolution produced homologous enzymes that stabilize enediolate intermediates by distinct structural strategies. Resolution of whether OMPDC, KGPDC, HPS and RPE are true homologues of the 'ribulose-phosphate binding' enzymes in histidine and tryptophan biosynthesis must await discovery of 'new' sequences, perhaps of 'new' enzymes, that provide the 'missing links' which allow sequence homologies to be established.

Aldolases

SCOP clusters several mechanistically distinct enzymes that catalyse aldol condensation reactions in an 'aldolase' superfamily. Not all aldolases with the (β/α)₈-barrel fold are clustered in this superfamily: HPS will be grouped in

(Figure 2 Legend) The reactions catalyzed by members of the enolase superfamily.

Figure 4
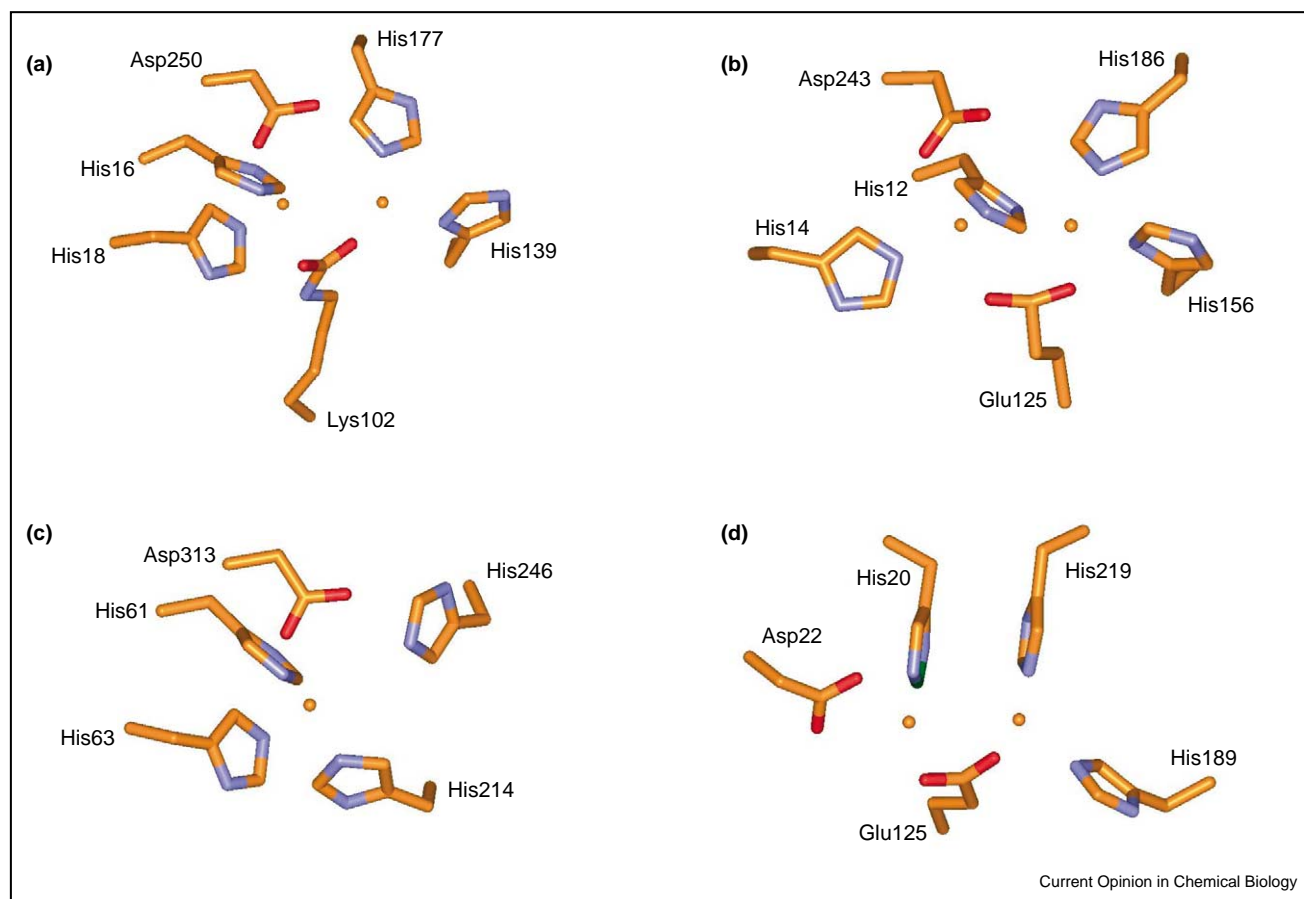
The active site motifs for members of the amidohydrolase superfamily. **(a)** DHO; **(b)** PTE homology protein; **(c)** cytosine deaminase; **(d)** human renal dipeptidase.

the 'ribulose-phosphate binding barrel' superfamily when its structure is solved, and a family of class II aldolases is included in the 'phospho*enol* pyruvate/pyruvate domain' superfamily. The diverse members of the 'aldolase' superfamily include:

1. Class I aldolases (a Schiff base with a lysine at the end of the sixth β-strand stabilizes an enediolate intermediate, although in transaldolase the lysine is at the end of the fourth β-strand)
2. Some, but not all, class II aldolases (a divalent metal ion stabilizes an enediolate intermediate).
3. Porphobilinogen synthase (Schiff base intermediates with, perhaps, two lysine residues [39]).
4. Homologous heptulosonate [40], octulosonate [41] and neuraminate phosphate synthases that use PEP as an *enol*/pyruvate source for an aldol reaction (some are metal-dependent and others metal-independent) — other enzymes that utilize *enol*/pyruvate as intermediates are placed in the 'phospho*enol*/pyruvate/pyruvate domain' superfamily.

5. Type I dehydroquinate dehydratases (a Schiff base stabilizes an enolate intermediate in dehydration [42].

Whether these diverse enzymes are related is important in delineating the bounds and limits of divergent evolution. However, using the Dali algorithm, these show no more structural homology to one another than to proteins in the 'ribulose-phosphate binding barrel' and the 'phospho*enol*-pyruvate/pyruvate domain' superfamilies. Although PSI-BLAST searches suggest some distant homology with many (β/α)$_8$-barrel proteins, most members of the superfamily do not show homology with one another or with other (β/α)$_8$-barrel proteins. Until members are discovered that show sequence homology, the assertion that these diverse enzymes constitute an 'aldolase' superfamily is intriguing but uncertain.

Designed and directed evolution

Attempts at directed evolution of enzymes with the (β/α)$_8$-barrel fold have been reported. Most altered the substrate specificity of the progenitor, with the hope of

Figure 5
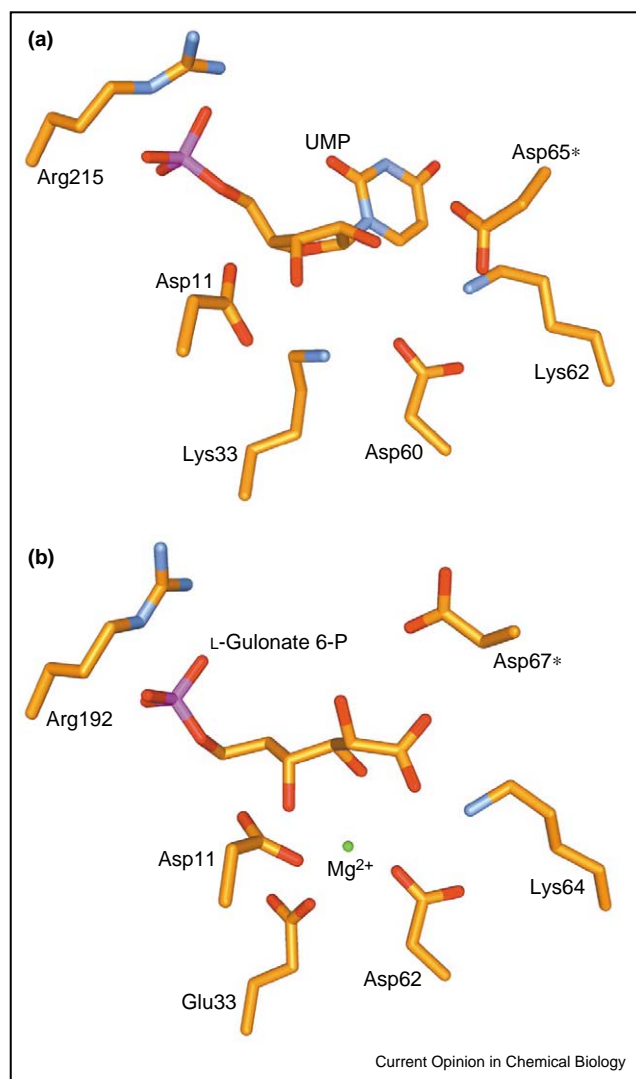
The active site motifs for **(a)** the OMP decarboxylase from *B. subtilis* and **(b)** KGPDC from *E. coli*.

using an evolved enzyme with 'improved' properties as catalysts in chemical applications. A few studies attempted alteration of the chemical reaction catalyzed by the progenitor.

Evolution of enzymes in histidine and tryptophan biosynthesis

The most noteworthy report of directed evolution of the (β/α)₈-barrel scaffold was reported by Fersht and co-workers in 2000 [43]. The claim was made that alterations in the loops connecting several of the β-stands with the following α-helices allowed indole glycerol phosphate synthase (IGPS) to efficiently catalyse the mechanistically distinct PRAI reaction, at the expense of the IGPS reaction; both enzymes participate in the tryptophan

biosynthetic pathway. Unfortunately, that claim was retracted, as subsequent characterization of the *in vitro*-evolved PRAI was not possible due to its insolubility and inability to complement an auxotrophic strain, as originally reported [44].

By contrast, exchange of substrate specificities between functionally similar enzymes in the histidine and tryptophan biosynthetic pathways has been successful [45•]. Both pathways include enzymes that catalyse Amadori rearrangements of 1′-aminonucleotides to yield aminoketoses (HisA and PRAI). Despite the lack of discernible sequence identity, their structures reveal a shared (β/α)₈-barrel fold. Using a Trp auxotroph that lacks the gene encoding PRAI as the basis for a metabolic selection, Sterner and co-workers identified two active variants of HisA containing three and four amino acid substitutions in a library of random mutants. Assays using purified proteins demonstrated that both were able to catalyse, albeit slowly, both the HisA and PRAI reactions. The seven substitutions were independently introduced into the HisA progenitor: only the Asp127Val mutant was sufficient to catalyse the PRAI reaction. Asp127 is located at the end of the fifth β-strand and is proximal to the active site; in the absence of structures for the liganded progenitors and the evolved TrpF, a detailed structural explanation of the observed change in substrate specificity is unknown.

Evolution of function in the enolase superfamily

One of our laboratories, together with Maxygen, Inc., is exploring the evolutionary potential of the (β/α)₈-barrel in the enolase superfamily. The active sites of MLE II from *Pseudomonas* sp. P51, the OSBS from *E. coli*, and the L-Ala-D/L-Glu epimerase from *E. coli* contain lysine catalysts at the ends of the second and sixth β-strands. In MLE, the lysine at the end of the second β-strand delivers a proton to the enolate intermediate formed by addition of the remote carboxylate to the double bond; in OSBS, the lysine at the end of the second β-strand both initiates the reaction by proton abstraction and protonates the hydroxide leaving group (D Klenchin, E Taylor-Ringia, JA Gerlt and I Rayment, unpublished data); and in the epimerase the lysine residues catalyse a two-base mechanism for 1,1-proton transfer. Neither the MLE nor the epimerase catalyse the OSBS reaction. However, because these are homologues, all three enzymes are derived from a common ancestor and a limited number of substitutions might allow interchange of activities. Design using the epimerase and combinatorial evolution using the MLE both yielded proteins that catalyse the OSBS reaction and differ from their progenitor by a single amino acid substitution (DMZ. Schmidt, E Mundorff, PC Babbitt, J Minshull and JA Gerlt, unpublished data). Structures are available for a wild-type MLE, the epimerase and two OSBSs, so an explanation for the changes in activity is possible. In the

design approach in which the structures of the epimerase and the OSBS, both from *E. coli*, were compared, the homologous aspartic acid at the end of the eighth β-strand of the epimerase was replaced with a glycine to generate sufficient volume so that the larger substrate for OSBS might bind in the active site. In the combinatorial approach with MLE, OSBS activity resulted from homologous substitution of a glutamic acid at the end of the eighth β-strand with a glycine. Both evolved proteins had significant levels of OSBS activity and reduced levels of the progenitor activity. That two homologous substitutions in different progenitors results in a 'new' activity emphasizes the functional plasticity of the (β/α)$_8$-barrel fold and suggests that nature may traverse different mechanisms by targeting substitutions to the ends of the various β-strands, the positions that contribute the catalytic functional groups.

Evolution of function in the amidohydrolase superfamily

Specific members of the amidohydrolase superfamily have been subjected to rational and combinatorial muta-
genesis protocols in an effort to evolve and enhance the substrate specificity of a given protein. In the first instance, the inherent stereoselectivity of PTE for pairs of chiral substrates has been enhanced, relaxed, and inverted. For example, the wild type PTE catalyzes the hydrolysis of phenyl ethyl *p*-nitrophenyl phosphate (**1**; Figures 6 and 7) with a 20-fold preference for the hydrolysis of (S_P)-**1**. It was surmised from the X-ray crystal structure that the stereoselectivity was dictated by the orientation of the side chains that form the substrate binding cavity within this enzyme. This notion was confirmed by the mutation Gly60Ala. This single change within the active site altered the stereoselectivity from 20:1 to approximately 10 000:1 by specifically reducing the turnover rate for the initial slower (R_P)-**1** [46]. Conversely, if Ile106 is converted to glycine the stereoselectivity is relaxed to the point where both enantiomers are turned over at essentially the same rate [47•]. Finally, the stereoselectivity was inverted to a ratio of 1:80 when only four mutations were made within the active site. These changes specifically enhanced the rate of the initially slower R_P-enantiomer and diminished the rate of the

Figure 6
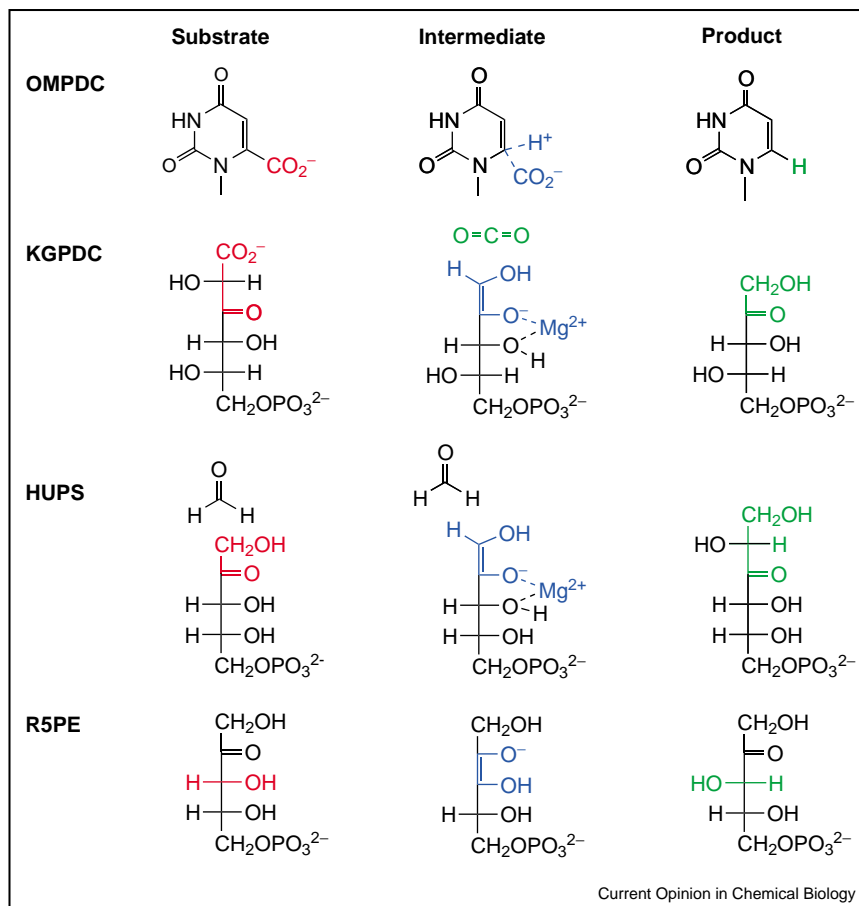
Substrates, intermediates and products of the reactions catalyzed by the members of the OMP decarboxylase suprafamily.

Figure 7
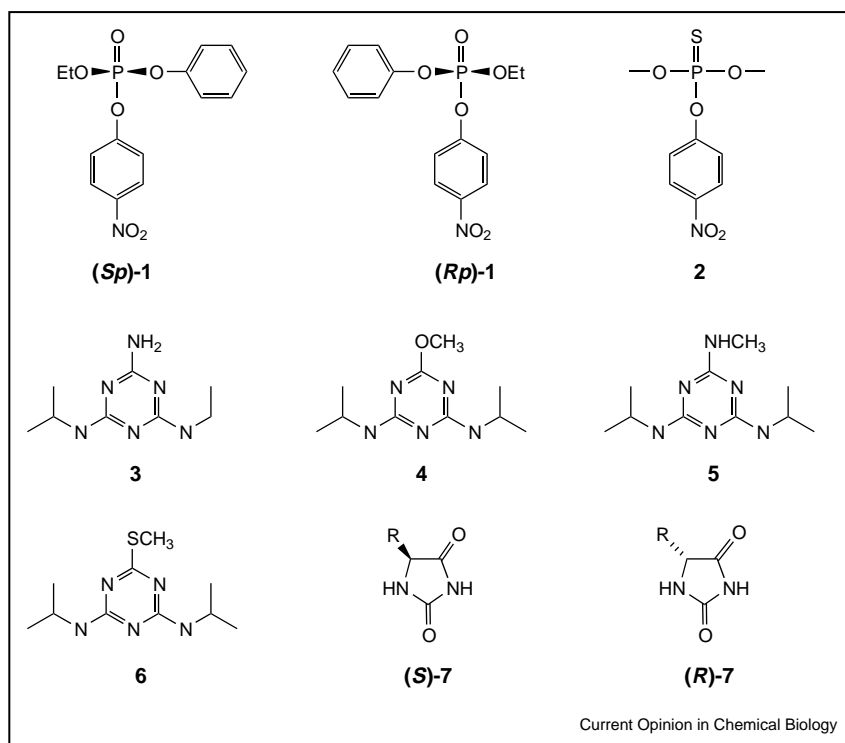
Substrates used in the modification of substrate specificity within the amidohydrolase superfamily.

initially faster S_P-enantiomer [47•]. The specific mutation was Ile106Gly/Phe132Gly/His257Tyr/Ser308Gly.

PTE has also been subjected to DNA shuffling in an attempt to improve the catalytic activity of the protein for the hydrolysis of methyl parathion [48]. The most improved variant had an increase in catalytic activity of 25-fold over the wild-type enzyme for the methyl parathion (**2**). This improved variant was shown to contain seven amino acid differences relative to the wild-type enzyme. Of these changes, only the alteration of His257 to tyrosine appears to be located directly within the active site. Therefore, alterations away from the active site can influence the catalytic activity of PTE by structural or dynamic perturbations that are not fully understood.

DNA shuffling has been used to direct the evolution of the atrazine chloro-hydrolase system (*Atz*A). In this example, the gene encoding for atrazine chloro-hydrolase was shuffled with the gene that encodes *Tri*A, an enzyme that catalyzes the deamination of substituted melamines (**3**). The objective of this endeavour was to explore substrate plasticity using a small library of substrates bearing different leaving groups and a modified triazine core [49••]. A single round of DNA shuffling was used and modified enzymes were isolated for three sub-

strates for which neither of the parents catalyzed the hydrolysis at a significant rate. These new substrates included prometon (**4**), *N*-methylamino propazine (**5**) and proetryn (**6**). There are only nine amino acid differences between *Atz*A and *Tri*A. The most versatile new mutant (active against the three new substrates) had six changes relative to *Atz*A and four changes relative to *Tri*A. Unfortunately, the structure of this enzyme is unknown and, thus, it is not yet possible to accurately map these alterations in three-dimensional space.

The HYD from *Arthrobacter* has been subjected to random and saturation mutagenesis [50]. The native enzyme preferentially hydrolyzes the D-enantiomer of a racemic mixture of D,L-methionine hydantoin (**7**) by a factor of about 1.2. After two rounds of random and saturation mutagenesis, a single clone was produced that had a preference for the L-enantiomer by a factor of approximately 5. There was a single amino acid change relative to the wild type enzyme. Ile95 was mutated to phenylalanine. Unfortunately, the structure of this HYD is not known and thus the location of mutation is undetermined.

Evolution of function in the aldolases
Despite the large number of aldolases in the sequence and structural databases, efforts to evolve their functions have been restricted to the class I (Schiff base-forming)

2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase, an enzyme in the Entner–Doudoroff pathway for glucose catabolism that yields pyruvate and D-glyceraldehyde 3-phosphate as products. KDPG aldolase has been used for synthetic applications because of its broad substrate specificity.

Wong and co-workers [51] successfully evolved the aldolase from *E. coli* so that it could utilize unphosphorylated substrates. Following an initial round of error-prone PCR, mutants with relaxed specificity were identified by screening. After two additional rounds of DNA shuffling, error-prone mutagenesis and shuffling, a variant was identified that was 600-fold more efficient that the wild-type progenitor in using the unphosphorylated substrate, although the natural substrate was still preferred by a factor of 85. This mutant with four substitutions also had a relaxed enantiomeric preference for the glyceraldehyde cosubstrate. Using the structure of the orthologueous KDPG aldolase from *Pseudomonas putida*, the substitutions were judged to be remote from the active site, thereby preventing structure–function explanation for the altered specificity.

Toone and co-workers [52•] used the novel approach of changing the position of the active site lysine to obtain a mutant that could utilize benzaldehyde as well as the isosteric pyridine carboxaldehyde as a cosubstrate. Lys133, at the end of the sixth β-strand of the $(\beta/\alpha)_8$-barrel, forms the Schiff base with the substrate. Screening a library of random mutants identified the Thr161Lys substitution as sufficient to confer the desired broadened specificity; this residue is at the end of the seventh β-strand. The Lys133Gln/Thr161Lys double mutant was constructed to 'require' Schiff base formation at an alternate position in the active site; this mutant retained modest activity on the natural substrates as well as the desired enhanced activity with benzaldehyde. This phenotype was attributed to an alteration of the size and polarity of the binding site for the aldehyde cosubstrate.

Conclusions

The recent studies of $(\beta/\alpha)_8$-barrel-containing enzymes summarized in this review highlight the range of natural functional divergence. The problem of determining the number of natural progenitors remains a challenge that will probably be solved as ongoing genome and structural genomics projects add to the sequence and structure databases. If this number is small (one?), we predict that both the design and discovery of enzymes that catalyse novel reactions on unnatural substrates will become feasible and, perhaps, routine.

Acknowledgements

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Nagano N, Orengo CA, Thornton JM: **One fold with many**
 • **functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions**. *J Mol Biol* 2002, **321**:741-765.
 A compendium of the structural features of $(\beta/\alpha)_8$-barrels.

2. Wierenga RK: **The TIM-barrel fold: a versatile framework for efficient enzymes**. *FEBS Lett* 2001, **492**:193-198.

3. Babbitt PC, Gerlt JA: **Understanding enzyme superfamilies. Chemistry as the fundamental determinant in the evolution of new catalytic activities**. *J Biol Chem* 1997, **272**:30591-30594.

4. Lo Conte L, Brenner SE, Hubbard TJ, Chothia C, Murzin AG: **SCOP database in 2002: refinements accommodate structural genomics**. *Nucleic Acids Res* 2002, **30**:264-267.

5. Copley RR, Bork P: **Homology among $(\beta\alpha)_8$ barrels: implications for the evolution of metabolic pathways**. *J Mol Biol* 2000, **303**:627-641.

6. Lang D, Thoma R, Henn-Sax M, Sterner R, Wilmanns M: **Structural**
 • **evidence for evolution of the β/α barrel scaffold by gene duplication and fusion**. *Science* 2000, **289**:1546-1550.
 Structural evidence for the construction of HisA and HisF from a single $(\beta/\alpha)_4$-half barrel.

7. Hocker B, Beismann-Driemeyer S, Hettwer S, Lustig A, Sterner R:
 • **Dissection of a $(\beta/\alpha)_8$-barrel enzyme into two folded halves**. *Nat Struct Biol* 2001, **8**:32-36.
 Biochemical evidence for the structural integrity and stability of the $(\beta/\alpha)_4$–half barrels in HisF.

8. Wedekind JE, Poyner RR, Reed GH, Rayment I: **Chelation of serine 39 to Mg^{2+} latches a gate at the active site of enolase: structure of the bis(Mg^{2+}) complex of yeast enolase and the intermediate analog phosphonoacetohydroxamate at 2.1-A resolution**. *Biochemistry* 1994, **33**:9333-9342.

9. Eads JC, Ozturk D, Wexler TB, Grubmeyer C, Sacchettini JC: **A new function for a common fold: the crystal structure of quinolinic acid phosphoribosyltransferase**. *Structure* 1997, **5**:47-58.

10. Hocker B, Schmidt S, Sterner R: **A common evolutionary origin of two elementary enzyme folds**. *FEBS Lett* 2002, **510**:133-135.

11. Palmer DR, Garrett JB, Sharma V, Meganathan R, Babbitt PC, Gerlt JA: **Unexpected divergence of enzyme function and sequence: "N-acylamino acid racemase" is o-succinylbenzoate synthase**. *Biochemistry* 1999, **38**:4252-4258.

12. Wise E, Yew WS, Babbitt PC, Gerlt JA, Rayment I: **Homologous**
 •• **(beta/alpha)8-barrel enzymes that catalyze unrelated reactions: orotidine 5′-monophosphate decarboxylase and 3-keto-L-gulonate 6-phosphate decarboxylase**. *Biochemistry* 2002, **41**:3861-3869.
 Structural evidence which confirms that divergent evolution can produce enzymes that catalyse different reactions with different mechanisms using conserved active site functional groups. OMP decarboxylase catalyzes a metal-independent reaction that must avoid a vinyl anion intermediate; KGPDC catalyzes a Mg^{2+}-dependent reaction that stabilizes an enediolate anion intermediate.

13. Gerlt JA, Babbitt PC: **Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies**. *Annu Rev Biochem* 2001, **70**:209-246.

14. Babbitt PC, Mrachko GT, Hasson MS, Huisman GW, Kolter R, Ringe D, Petsko GA, Kenyon GL, Gerlt JA: **A functionally diverse enzyme superfamily that abstracts the alpha protons of carboxylic acids**. *Science* 1995, **267**:1159-1161.

15. Schmidt DM, Hubbard BK, Gerlt JA: **Evolution of enzymatic activities in the enolase superfamily: functional assignment of unknown proteins in Bacillus subtilis and *Escherichia coli* as L-Ala-D/L-Glu epimerases**. *Biochemistry* 2001, **40**:15707-15715.

16. Gulick AM, Schmidt DM, Gerlt JA, Rayment I: **Evolution of enzymatic activities in the enolase superfamily: crystal structures of the L-Ala-D/L-Glu epimerases from** *Escherichia coli* **and** *Bacillus subtilis*. *Biochemistry* 2001, **40**:15716-15724.

17. Wieczorek SW, Kalivoda KA, Clifton JG, Ringe D, Petsko GA, Gerlt JA: **Evolution of enzymatic activities in the enolase superfamily: identification of a "new" general acid catalyst in the active site of D-galactonate dehydratase from** *Escherichia coli*. *J Am Chem Soc* 1999, **121**:4540-4541.

18. Gulick AM, Hubbard BK, Gerlt JA, Rayment I: **Evolution of enzymatic activities in the enolase superfamily: crystallographic and mutagenesis studies of the reaction catalyzed by D-glucarate dehydratase from** *Escherichia coli*. *Biochemistry* 2000, **39**:4590-4602.

19. Gulick AM, Hubbard BK, Gerlt JA, Rayment I: **Evolution of enzymatic activities in the enolase superfamily: identification of the general acid catalyst in the active site of D-glucarate dehydratase from** *Escherichia coli*. *Biochemistry* 2001, **40**:10054-10062.

20. Asuncion M, Blankenfeldt W, Barlow JN, Gani D, Naismith JH: **The structure of 3-methylaspartase from** *Clostridium tetanomorphum* **functions via the common enolase chemical step**. *J Biol Chem* 2002, **277**:8306-8311.

21. Levy CW, Buckley PA, Sedelnikova S, Kato Y, Asano Y, Rice DW, Baker PJ: **Insights into enzyme evolution revealed by the structure of methylaspartate ammonia lyase**. *Structure* 2002, **10**:105-113.

22. Holm L, Sander C: **An evolutionary treasure: unification of a broad set of amidohydrolases related to urease**. *Proteins* 1997, **28**:72-82.

23. Thoden JB, Phillips GN Jr, Neal TM, Raushel FM, Holden HM:
•   **Molecular structure of dihydroorotase: a paradigm for catalysis through the use of a binuclear metal center**. *Biochemistry* 2001, **40**:6989-6997.
The structure of DHO was determined in the presence of substrate and product to a resolution of 1.7 Å where dihydroorotate was bound to one subunit and carbamoyl aspartate was bound to the other subunit within a single dimeric species. This has provided an unprecedented view of the molecular interactions between substrate and product with the enzyme before and after the chemical transformation.

24. Shim H, Raushel FM: **Self-assembly of the binuclear metal center of phosphotriesterase**. *Biochemistry* 2000, **39**:7357-7364.

25. Buchbinder JL, Stephenson RC, Dresser MJ, Pitera JW, Scanlan TS, Fletterick RJ: **Biochemical characterization and crystallographic structure of an** *Escherichia coli* **protein from the phosphotriesterase gene family**. *Biochemistry* 1998, **37**:5096-5106.

26. Ireton GC, McDermott G, Black ME, Stoddard BL: **The structure of** *Escherichia coli* **cytosine deaminase**. *J Mol Biol* 2002, **315**:687-697.

27. Nitanai Y, Satow Y, Adachi H, Tsujimoto M: **Crystal structure of human renal dipeptidase involved in beta-lactam hydrolysis**. *J Mol Biol* 2002, **321**:177-184.

28. Karplus PA, Pearson MA, Hausinger RP: **70 years of crystalline urease: what have we learned?** *Acc Chem Res* 1997, **30**:330-337.

29. Benning MM, Shim H, Raushel FM, Holden HM: **High resolution X-ray structures of different metal-substituted forms of phosphotriesterase from** *Pseudomonas diminuta*. *Biochemistry* 2001, **40**:2712-2722.

30. Yew WS, Gerlt JA: **Utilization of L-ascorbate by** *Escherichia coli* **K-12: assignments of functions to products of the yjf-sga and yia-sgb operons**. *J Bacteriol* 2002, **184**:302-306.

31. Miller BG, Smiley JA, Short SA, Wolfenden R: **Activity of yeast orotidine-5′-phosphate decarboxylase in the absence of metals**. *J Biol Chem* 1999, **274**:23841-23843.

32. Sahm H, Schutte H, Kula MR: **Purification and properties of 3-hexulosephosphate synthase from** *Methylomonas* **M 15**. *Eur J Biochem* 1976, **66**:591-596.

33. Harris P, Navarro Poulsen JC, Jensen KF, Larsen S: **Structural basis for the catalytic mechanism of a proficient enzyme: orotidine 5′-monophosphate decarboxylase**. *Biochemistry* 2000, **39**:4217-4224.

34. Appleby TC, Kinsland C, Begley TP, Ealick SE: **The crystal structure and mechanism of orotidine 5′-monophosphate decarboxylase**. *Proc Natl Acad Sci USA* 2000, **97**:2005-2010.

35. Wu N, Mo Y, Gao J, Pai EF: **Electrostatic stress in catalysis: structure and mechanism of the enzyme orotidine monophosphate decarboxylase**. *Proc Natl Acad Sci USA* 2000, **97**:2017-2022.

36. Miller BG, Hassell AM, Wolfenden R, Milburn MV, Short SA: **Anatomy of a proficient enzyme: the structure of orotidine 5′-monophosphate decarboxylase in the presence and absence of a potential transition state analog**. *Proc Natl Acad Sci USA* 2000, **97**:2011-2016.

37. Chen YR, Hartman FC, Lu TY, Larimer FW: **D-Ribulose-5-phosphate 3-epimerase: cloning and heterologous expression of the spinach gene, and purification and characterization of the recombinant enzyme**. *Plant Physiol* 1998, **118**:199-207.

38. Kopp J, Kopriva S, Suss KH, Schulz GE: **Structure and mechanism of the amphibolic enzyme D-ribulose-5-phosphate 3-epimerase from potato chloroplasts**. *J Mol Biol* 1999, **287**:761-771.

39. Erskine PT, Coates L, Newbold R, Brindley AA, Stauffer F, Wood SP, Warren MJ, Cooper JB, Shoolingin-Jordan PM, Neier R: **The X-ray structure of yeast 5-aminolaevulinic acid dehydratase complexed with two diacid inhibitors**. *FEBS Lett* 2001, **503**:196-200.

40. Wagner T, Shumilin IA, Bauerle R, Kretsinger RH: **Structure of 3-deoxy-ᴅ-arabino-heptulosonate-7-phosphate synthase from** *Escherichia coli*: **comparison of the Mn(2+)*2-phosphoglycolate and the Pb(2+)*2-phosphoenolpyruvate complexes and implications for catalysis**. *J Mol Biol* 2000, **301**:389-399.

41. Radaev S, Dastidar P, Patel M, Woodard RW, Gatti DL: **Structure and mechanism of 3-deoxy-D-manno-octulosonate 8-phosphate synthase**. *J Biol Chem* 2000, **275**:9476-9484.

42. Gourley DG, Shrive AK, Polikarpov I, Krell T, Coggins JR, Hawkins AR, Isaacs NW, Sawyer L: **The two types of 3-dehydroquinase have distinct structures but catalyze the same overall reaction**. *Nat Struct Biol* 1999, **6**:521-525.

43. Altamirano MM, Blackburn JM, Aguayo C, Fersht AR: **Directed evolution of new catalytic activity using the α/β-barrel scaffold**. *Nature* 2000, **403**:617-622.

44. Altamirano MM, Blackburn JM, Aguayo C, Fersht AR: **Retraction. Directed evolution of new catalytic activity using the alpha/beta-barrel scaffold**. *Nature* 2002, **417**:468.

45. Jurgens C, Strom A, Wegener D, Hettwer S, Wilmanns M, Sterner R:
•   **Directed evolution of a (beta alpha)8-barrel enzyme to catalyze related reactions in two different metabolic pathways**. *Proc Natl Acad Sci USA* 2000, **97**:9925-9930.
A single mutation at the end of the fifth β-strand in HisA alters the substrate specificity so that this enzyme that catalyzes the Amadori reaction in histidine biosynthesis can catalyse the analogous Amadori reaction catalyzed by PRAI in tryptophan biosynthesis.

46. Chen-Goodspeed M, Sogorb MA, Wu F, Hong SB, Raushel FM: **Structural determinants of the substrate and stereochemical specificity of phosphotriesterase**. *Biochemistry* 2001, **40**:1325-1331.

47. Chen-Goodspeed M, Sogorb MA, Wu F, Raushel FM:
•   **Enhancement, relaxation, and reversal of the stereoselectivity for phosphotriesterase by rational evolution of active site residues**. *Biochemistry* 2001, **40**:1332-1339.
The substrate binding pockets for PTE were manipulated in a rational manner to alter the stereoselectivity. With only 4–5 amino acid changes within the active site, the stereoselectivity of the wild-type enzyme can be manipulated a factor of one-million-fold.

48. Cho CM, Mulchandani A, Chen W: **Bacterial cell surface display of organophosphorus hydrolase for selective screening of**

improved hydrolysis of organophosphate nerve agents.
Appl Environ Microbiol 2002, **68**:2026-2030.

49. Raillard S, Krebber A, Chen Y, Ness JE, Bermudez E, Trinidad R,
•• Fullem R, Davis C, Welch M, Seffernick J *et al.*: **Novel enzyme
   activities and functional plasticity revealed by recombining
   highly homologous enzymes**. *Chem Biol* 2001, **8**:891-898.
The paper is an outstanding example for the creation of novel enzyme
activities through gene shuffling methodology and characterization of
substrate profiles with a small compound library.

50. May O, Nguyen PT, Arnold FH: **Inverting enantioselectivity by
   directed evolution of hydantoinase for improved production of
   L-methionine**. *Nat Biotechnol* 2000, **18**:317-320.

51. Fong S, Machajewski TD, Mak CC, Wong C: **Directed evolution of
   D-2-keto-3-deoxy-6-phosphogluconate aldolase to new
   variants for the efficient synthesis of D- and L-sugars**.
   Chem Biol 2000, **7**:873-883.

52. Wymer N, Buchanan LV, Henderson D, Mehta N, Botting CH,
• Pocivavsek L, Fierke CA, Toone EJ, Naismith JH: **Directed
   evolution of a new catalytic site in 2-keto-3-deoxy-6-
   phosphogluconate aldolase from *Escherichia coli***.
   Structure 2001, **9**:1-9.
The position of the active site lysine residue in the class I KGPD aldolase
was moved from the sixth to the seventh β-strand to give a mutant with
enhanced substrate specificity.