

Homology among $(\beta\alpha)_8$ Barrels: Implications for the Evolution of Metabolic Pathways

Richard R. Copley* and Peer Bork*

*Biocomputing, European
Molecular Biology Laboratory
Meyerhofstrasse 1
69117 Heidelberg, Germany*

We provide statistically reliable sequence evidence indicating that at least 12 of 23 SCOP $(\beta\alpha)_8$ (TIM) barrel superfamilies share a common origin. This includes all but one of the known and predicted TIM barrels found in central metabolism. The statistical evidence is complemented by an examination of the details of protein structure, with certain structural locations favouring catalytic residues even though the nature of their molecular function may change. The combined analysis of sequence, structure and function also enables us to propose a phylogeny of TIM barrels. Based on these data, we are able to examine differing theories of pathway and enzyme evolution, by mapping known TIM barrel folds to the pathways of central metabolism. The results favour widespread recruitment of enzymes between pathways, rather than a “backwards evolution” model, and support the idea that modern proteins may have arisen from common ancestors that bound key metabolites.

© 2000 Academic Press

Keywords: homology; protein evolution; pathway evolution; metabolism; TIM barrels

*Corresponding authors

Introduction

Horowitz (1945) proposed that biochemical pathways may have evolved backwards (i.e. with the enzyme catalyzing the N th step of a pathway arising before the enzyme catalyzing the $N - 1$ th step), later modifying this theory (Horowitz, 1965) to suggest that enzymes within a similar pathway could be homologous. The idea has intuitive appeal. As the product of one enzyme is the substrate of the enzyme following it in a pathway, it follows that binding sites need not be reinvented if homologous enzymes are used within a pathway. There are a handful of examples where adjacent enzymes in a pathway are indeed homologous (e.g. see Belfaiza *et al.*, 1986; Bork & Rohde, 1990; Wilmanns *et al.*, 1991; Fani *et al.*, 1994). An alternative view of pathway evolution is that of enzyme recruitment, whereby enzymes catalyzing similar reactions in different pathways would be found to

be homologous (Jensen, 1976). Related to this view is that in which enzyme evolution has been driven by retention of catalytic mechanisms. There is good evidence to suggest that this has occurred within many protein families (Lawrence *et al.*, 1997; Babbitt & Gerlt, 1997; Gerlt & Babbitt, 1998; Eklund & Fontecave, 1999). In general, however, the links between a protein fold and the type of reaction catalyzed are not strong (Martin *et al.*, 1998; Hegyi & Gerstein, 1999).

To enquire more deeply into the nature of pathway evolution, we have analyzed the distribution of one particular fold in the enzymes of central metabolism. Enzymes containing triose phosphate isomerase (TIM) $(\beta\alpha)_8$ barrel-like folds adopt a particularly wide range of functions (Pujadas & Palau, 1999). Approximately 10% of enzymes may contain such a fold (Gerlt, 2000), but their history remains obscure. Although common structure is often taken as an indicator of common ancestry, the simplicity of the core fold, taken with an absence of sequence similarity, and a vast range of functions, raises the possibility of convergent evolution. Farber and Petsko (1990) have argued for common ancestry of a limited number of TIM barrels (compared to present day numbers) on the basis of similarity in the position of active sites, and other structural features. Janacek (Janecek, 1995, 1996; Janecek & Balaz, 1995) reported the

Abbreviations used: TIM, triose phosphate isomerase; FMN, flavin mononucleotide; NAD, nicotinamide-adenine dinucleotide; PLP, pyridoxal phosphate; IMP, inosine monophosphate; IMPDH, inosine monophosphate dehydrogenase; PEP, phosphoenolpyruvate.

E-mail addresses of the corresponding authors: copley@embl-heidelberg.de, bork@embl-heidelberg.de

presence of a few similar residues at approximately similar positions within various members of TIM barrel-like enzymes, and used this to infer homology. The significance of these findings, and how likely they were to have arisen by chance, is difficult to assess. The SCOP protein structure database (Murzin *et al.*, 1995) currently distinguishes 23 superfamilies of TIM barrel, where the members of a superfamily have a probable common evolutionary origin. Most recently, Thornton and co-workers have analyzed TIM barrels from a structural viewpoint, classifying nine families on the basis of hydrogen bond patterns and residue packing in the interior of the barrel (Nagano *et al.*, 1999).

In this study, we present statistically robust evidence to support the hypothesis that many families of TIM barrel-containing enzymes have diverged from a common ancestor. The increasing coverage of sequence databases, coupled with increasing sensitivity of sequence-based search methods, has enabled the discovery of sequences that provide evolutionary links between major families of TIM barrel-like structures. Taken together, the similarities provide convincing evidence for a monophyletic origin of many TIM barrel families, and a plausible route by which modern functions have arisen from ancestral ones. The ubiquity of this family in the biosynthetic pathways of central metabolism and the implications for theories of pathway evolution are analyzed.

Results

In order to establish the distribution of homologues over known pathways, it is first necessary to determine which proteins are homologous. Rather than simply take similarity in fold (i.e. purely structural data) as a criterion, our strategy has been to look for statistically justified sequence evidence. We then analyzed the structure and function of these proteins to consider the implications of the sequence evidence.

Homology among TIM barrels of known structure

Figure 1 provides an overview of selected sequence relationships detected by PSI-Blast (Altschul *et al.*, 1997) searching with representative TIM barrel sequences. Sequences and SCOP superfamilies (Lo Conte *et al.*, 2000) are described in Table 1. Central (although not critical) to the scheme is the protein encoded by the *hisA* gene of *Methanococcus voltae*. This protein shows an internal duplication (observable in other members of this family (Fani *et al.*, 1994)), which is readily detectable by PSI-Blast. This internal duplication may assist the rapid detection of distant homologues, by producing a "deeper" sequence profile during initial stages of searching (Aravind & Koonin, 1999) and so represents an ideal starting point for exploring sequence space. Hits found were used to initiate further PSI-Blast searches,

until no new family was found. Below, we outline our searches and justify our inferences of homology in more detail.

None of the proteins found during the searching stage was an obvious false positive (i.e. a significant match known to adopt a non-TIM barrel-like fold). Full PSI-Blast results files are available for inspection at <http://www.embl-heidelberg.de/~copley/TIM>.

A superfamily of phosphate-binding barrels

Many sequence families that are readily detectable from *hisA* share a characteristic phosphate-binding motif at the end of β -strand 7 (see Figure 2). This includes the superfamilies triosephosphate isomerase, ribulose phosphate-binding barrel (including tryptophan biosynthesis enzymes, and ribulose phosphate epimerase), thiamin phosphate synthase and FMN-linked oxidoreductases, which in SCOP are all noted to share a common binding site, although with no explicit statement of homology. Additionally, we detect inosine monophosphate dehydrogenase (IMPDH) from the NAD(P)-linked oxidoreductase superfamily, although this appears to be a mis-classification in SCOP, as the phosphate-binding site is occupied by IMP rather than by the NADP characteristic of other superfamily members. The recently solved structure of orotidine 5'-monophosphate decarboxylase is detected by our PSI-blast searches, and shares a similar phosphate-binding site (Appleby *et al.*, 2000). The C-terminal domain of quinolinic acid phosphoribosyl transferase consists of an incomplete TIM barrel domain, with only seven β -strands (Eads *et al.*, 1997). It is found as a significant hit in our searches, and contains a conserved phosphate-binding motif at a position equivalent to that in the other phosphate-binding barrels.

There are eight equivalent β -strands in the core $(\beta\alpha)_8$ structure. There does not appear to be an adequate explanation of the exact topological equivalence (i.e. at the end of the seventh β -strand) of phosphate-binding sites within TIM barrels, other than homology.

Related phosphate-binding sites are found in the TIM barrels of the SCOP PLP-binding barrel superfamily, and the C-terminal domain of the large subunit of Rubisco (Figure 2), although no significant global sequence link could be found to the wider family.

Aldolases

Aldolases are an important class of enzymes, catalyzing the fusion or cleavage of two carbonyl compounds to form an aldol, a common strategy for growing carbon chains during biosynthesis. Many (but not all) aldolases are known to adopt a TIM barrel-like fold. Within the $(\beta\alpha)_8$ class of aldolase folds, phosphate binding at the end of β -strand 7 is found in both type I and type II fructose bisphosphate aldolases (Figure 2).

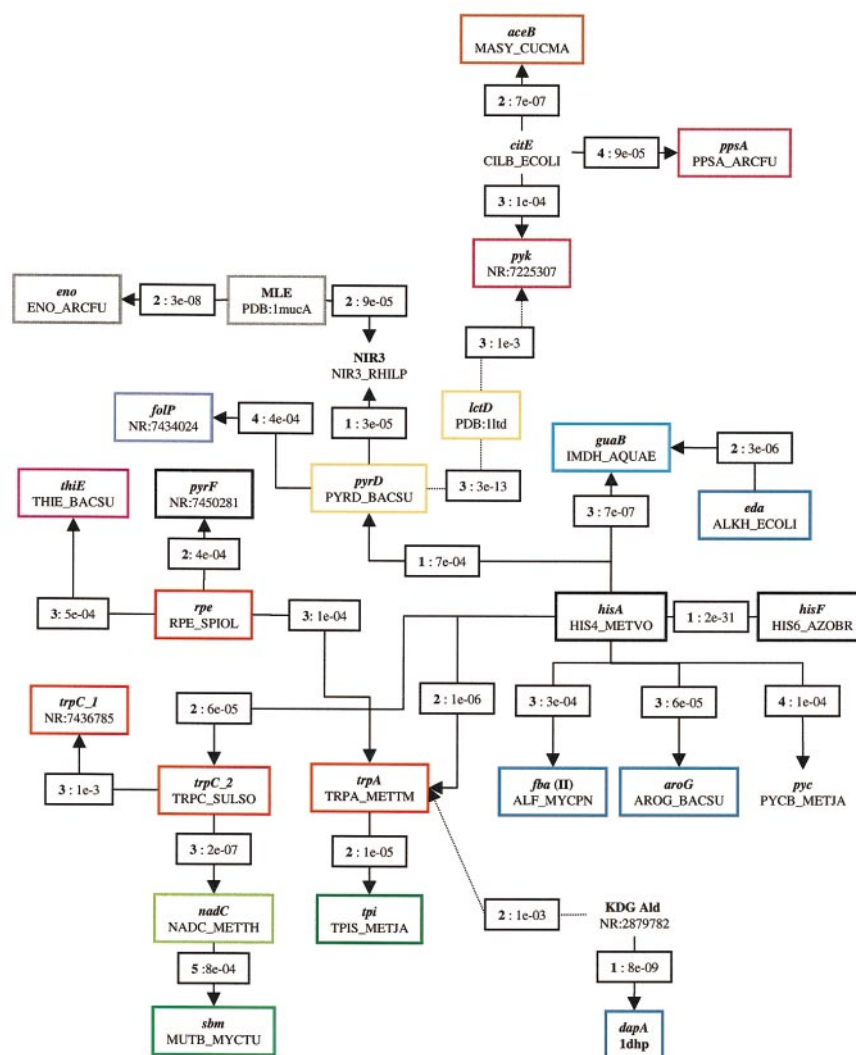


Figure 1. Overview of sequence-searching strategy used to identify homologous families using PSI-Blast. Parameters used are discussed in Methods and in the text. Gene names correspond to those in Table 1. Identifiers of sequences used to initiate searches are also given (SWISS-PROT identifiers, or NCBI NR GI accession numbers). Coloured boxes around the gene name and sequence identifier correspond to different SCOP superfamilies, with black being used for solved structures not yet in SCOP. Unboxed genes do not have a close homologue of known structure. Arrows point away from the sequence used as the seed in any particular search. The iteration number (bold face) and *E*-value at which the sequence was first found are indicated in boxes. For clarity, many links between sequences are not shown. Full search results may be obtained from <http://www.embl-heidelberg.de/~copley/TIM/>.

Type II fructose biphosphate aldolases are identified as putative homologues of the phosphate-binding superfamily by our PSI-Blast analysis, producing significant *E*-values within three iterations from searches initiated with *hisA*. Type II fructose biphosphate aldolases operate *via* a metal-dependent mechanism, as does the aldolase 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase (*aroG*), the enzyme catalyzing the first step of chorismate biosynthesis. This enzyme also adopts a TIM barrel fold (Shumilin *et al.*, 1999), and is detected in three iterations of PSI-Blast searching (Figure 1). Although this enzyme catalyzes the association of two phosphorylated substrates, neither phosphate group is found in the characteristic site at the end of β -strand 7.

However, the metal ion required for catalysis is found in the same location as that from type II fructose biphosphate aldolase.

Type I fructose biphosphate aldolases catalyze reactions *via* a lysine residue, which forms a Schiff base with the substrate. This residue is found on β -strand 6. These enzymes have been shown to be homologous to type II aldolases and deoxyribose phosphate aldolases (*deoC*) (Galperin *et al.*, 2000). We also detect significant similarity between members of the *trpA* family and a non-phosphorylated Entner-Doudoroff pathway aldolase from *Sulfolobus solfataricus* (with an *E*-value inclusion threshold of 0.01), which is in turn homologous to dihydrodipicolinate synthase and *N*-acetylneuraminate lyase, both Schiff base-dependent enzymes

Table 1. Guide to TIM barrel containing proteins discussed in the text

Gene	Protein	SCOP superfamily	Pathway	PDB	Reference
<i>fba(I)</i>	Fructose-1,6-bisphosphate aldolase (Schiff base)	Aldolase	Glycolysis	1fba	Hester <i>et al.</i> (1991)
<i>fba(II)</i>	Fructose-1,6-bisphosphate aldolase (Metal dept)	Aldolase	Glycolysis	1b57	Hall <i>et al.</i> (1999)
<i>dhnA</i>	Fructose-1,6-bisphosphate aldolase	-	Glycolysis	N/A	Dandekar <i>et al.</i> (1999)
<i>tpi</i>	Triose phosphate isomerase	TIM	Glycolysis	1tph	Zhang <i>et al.</i> (1994)
<i>eno</i>	Enolase	Enolase C-terminal domain	Glycolysis	1one	Larsen <i>et al.</i> (1996)
<i>pyk</i>	Pyruvate kinase	Phosphoenolpyruvate/pyruvate	Glycolysis	1a3w	Jurica <i>et al.</i> (1998)
<i>eda</i>	KDGP aldolase	-	Entner-Doudoroff	1kga*	Mavridis <i>et al.</i> (1982)
<i>rpe</i>	D-Ribulose-5-P 3-phosphate epimerase	Ribulose phosphate binding	Pentose phosphate	1rpx	Kopp <i>et al.</i> (1999)
<i>tal</i>	Transaldolase	Aldolase	Pentose phosphate	1onr	Jia <i>et al.</i> (1996)
<i>aroG</i>	2-Dehydro-3-deoxyphosphoheptanoate aldolase	Aldolase	Chorismate	1qr7	Shumilin <i>et al.</i> (1999)
<i>aroD</i>	3-Dehydroquinase	-	Chorismate	1qfe	Gourley <i>et al.</i> (1999)
<i>trpC_1</i>	Phosphoribosyl anthranilate isomerase	Ribulose phosphate-binding	Tryptophan	1nsj	Hennig <i>et al.</i> (1997)
<i>trpC_2</i>	Indole 3-glycerol-P synthase	Ribulose phosphate-binding	Tryptophan	1a53	Hennig <i>et al.</i> (1995)
<i>trpA</i>	Trp synthase α subunit	Ribulose phosphate-binding	Tryptophan	1ttq	Rhee <i>et al.</i> (1997)
<i>hisA</i>	P-Ribosylformimino-AICAR-P-isomerase	-	Histidine	1qo2	Lang <i>et al.</i> (2000)
<i>hisF</i>	Imidazoleglycerolphosphate synthase	-	Histidine	1thf	Lang <i>et al.</i> (2000)
<i>pyrC</i>	Dihydroorotase	Metallo dept hydrolase	Pyrimidine	N/A	Holm & Sander (1997)
<i>pyrD</i>	Dihydroorotate dehydrogenase	FMN-linked oxidoreductase	Pyrimidine	2dor	Rowland <i>et al.</i> (1998)
<i>pyrF</i>	Orotidine-5'-P decarboxylase	-	Pyrimidine	1dbt	Appleby <i>et al.</i> (2000)
<i>guaC</i>	GMP reductase	-	Purine interconversions	N/A	Bork <i>et al.</i> (1995)
<i>guaB</i>	Inosine 5'-P dehydrogenase	NAD(P)-linked oxidoreductase	Purine interconversions	1ak5	Whitby <i>et al.</i> (1997)
<i>hosc*</i>	Homocitrate synthase	-	Lysine (α -amino adipic acid)	N/A	-
<i>dapA</i>	Dihydrodipicolinate synthase	Aldolase	Lysine (diaminopimelic acid)	1dhp	Mirwaldt <i>et al.</i> (1995)
<i>ppsA</i>	Phosphoenolpyruvate synthase	Phosphoenolpyruvate/pyruvate	Pyruvate conversions	1dik	Herzberg <i>et al.</i> (1996)
<i>ppc</i>	Phosphoenolpyruvate carboxylase	Phosphoenolpyruvate/pyruvate	Pyruvate conversions	1fiy	Kai <i>et al.</i> (1999)
<i>pyc</i>	Pyruvate carboxylase	-	Pyruvate conversions	N/A	-
<i>leuA</i>	Isopropylmalate synthase	-	Leucine	N/A	-
<i>deoC</i>	Deoxyribose phosphate aldolase	-	Nucleotide salvage	N/A	Galperin <i>et al.</i> (2000)
<i>aceB</i>	Malate synthase	Malate synthase G	Glyoxylate cycle	1d8c	Howard <i>et al.</i> (2000)
<i>thiE</i>	Thiamin phosphate synthase	Thiamin phosphate synthase	Thiamin synthesis	2tps	Chiu <i>et al.</i> (1999)
<i>nadC</i>	Quinolate phosphoribosyl transferase	Quinolinic acid PR-transferase	NAD synthesis	1qpo	Sharma <i>et al.</i> (1998)
<i>lctD</i>	Lactate dehydrogenase	FMN-linked oxidoreductase	Respiration	1ltd	Tegoni & Cambillau (1994)
<i>folP</i>	Dihydropteroate synthase	Dihydropteroate synthase	Folic acid synthesis	1ajz	Achari <i>et al.</i> (1997)
<i>sbm</i>	Methylmalonyl-CoA mutase	Cobalmin (vit. B12) dept.	Methylmalonyl pathway	7req	Mancia <i>et al.</i> (1999)
<i>glTB</i>	Glutamate synthase	-	Glutamate synthesis	N/A	Bork <i>et al.</i> (1995)
<i>citE</i>	Citrate lyase subunit	-	-	N/A	-
<i>deoC</i>	Deoxyribose phosphate aldolase	-	Nucleotide salvage	N/A	Galperin <i>et al.</i> (2000)
<i>MLE*</i>	Muconate lactonizing enzyme	Enolase C-terminal domain	Phenylalanine metabolism	1muc	Helin <i>et al.</i> (1995)

Where possible, *Escherichia coli* gene name abbreviations have been used. Other abbreviations are indicated by asterisks. The reference given is either to a representative structure determination, or a paper predicting TIM barrel structure. The structure file for *eda* (KDGP aldolase, 1kga) is of extremely low quality. However, structural reports clearly show it adopts a TIM barrel-like fold, and it is known to be a Schiff base-dependent aldolase.

Gene	PDB	Start	β6		β7		End
<i>trpA</i>	1ttqA	173	YTYL	<30>	FALQFGIS		215
<i>guaB</i>	1ak5	308	FIKI	<42>	EVCSGGIV		362
<i>hisA</i>	1qo2A	160	EIVH	<26>	KVLAAGGIS		198
<i>rubisco</i>	1burA	325	HIHS	<46>	LPVASGGTH		383
<i>nadC</i>	1qpoA	218	LILL	<22>	MLESSGGLS		252
<i>trpC_2</i>	1a53	176	FIGI	<26>	VKVAESGIS		214
<i>thiE</i>	2tpsA	147	YVGL	<31>	FIVGIGGIT		190
<i>trpC_1</i>	1nsj	122	PILL	<27>	YLVLGGGLN		161
<i>pyrD</i>	2dorA	188	YVNS	<52>	QIIGTGGIE		252
<i>hisF</i>	hisfD	167	EILL	<26>	FIIASGGAG		205
<i>tpi</i>	1tph1	161	VLAY	<40>	RIIYGGSVT		213
<i>rpe</i>	1rpxA	143	LVLI	<34>	WIEVDGGVG		189
<i>fba (II)</i>	1b57A	217	TIAA	<38>	LNFVFHGGG		267
<i>fba (I)</i>	1fbaA	227	LLKP	<36>	VTFLSGGQS		275
<i>eda</i>	-	131	EFKF	<22>	RFCPTGGIS		165
<i>dhna</i>	-	235	IVKQ	<46>	GLINSGGAA		293
<i>dapA</i>	1dhpA	159	GIKE	<18>	FVLLSGGDA		189
<i>aroD</i>	1qfeA	168	IPKI	<27>	FVITMsmak		207

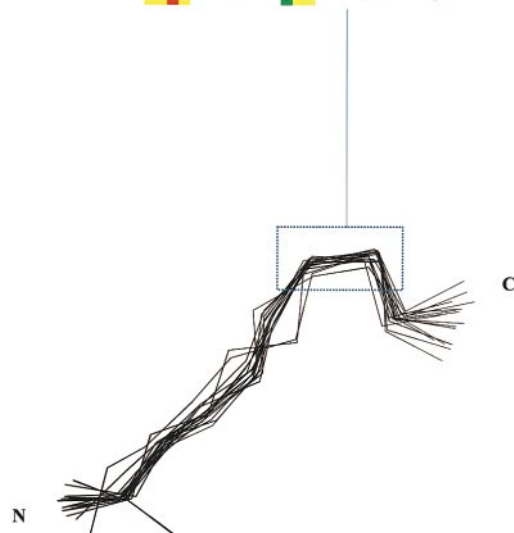


Figure 2. (a) Alignment of β -strands 6 and 7 for a variety of TIM barrels, produced using the STAMP package (Russell & Barton, 1992), where structures were available, and PSI-Blast analysis. Sequences with a conserved lysine residue in β strand 6, indicated in red, are Schiff base-dependent aldolases: β strand 7 is phosphate binding, the blue region in the alignment corresponds to the marked region in the structures. It can be seen that some sequences share both motifs, indicating that both families are likely to be related. Residues in small letters cannot be aligned accurately on the basis of structure. Yellow shading indicates conserved hydrophobic positions. The start and end residues of sequence fragments are indicated. Numbers in angle brackets represent the number of residues between the represented fragments. (b) The corresponding structural alignment, showing the distinctive bulge of the conserved residues at the end of the β -strand.

(Buchanan *et al.*, 1999). The equivalent enzyme from the phosphorylated Entner-Doudoroff pathway *eda* is also a Schiff base-dependent TIM barrel aldolase (Mavridis *et al.*, 1982), and shows significant sequence similarity to other phosphate-binding TIM barrels (Figure 1). Comparison of known structures shows that the Schiff base-forming catalytic lysine residue is found in an exactly equivalent position in type I fructose-1,6-bisphosphate aldolase, *N*-acetylneuraminase lyase, dihydrodipicolinate synthase (Lawrence *et al.*, 1997) and 3-dehydroquinase (see Figure 2). A further Schiff base-dependent enzyme, transaldolase, contains a similar residue believed to be related to the wider family by a circular permutation of the barrel structure (Jia *et al.*, 1996). Again, such exact equivalence of functional sites is most easily explained by homology. In the cases of *N*-acetylneuraminase lyase, dihydrodipicolinate synthase and dehydroquinase the glycine residues of the phosphate-binding loop at the end of β strand 7 are not conserved, indicating this particular function has been lost.

The enolase superfamily

The enolase superfamily, including enolase, muconate lactonizing enzyme and mandelate racemase share a common mechanism of catalysis and have the same two-domain architecture (Babbitt *et al.*, 1996; Lo Conte *et al.*, 2000). The C-terminal domain adopts a TIM barrel-like fold, although in the case of enolase itself, the arrangement of secondary structure is $\beta\beta\alpha(\beta\alpha)_6$ rather than the classical $(\beta\alpha)_8$. Although enolase catalyzes the formation of phosphoenolpyruvate, 2-phospho-D-glycerate, it does not bind the phosphate group at the end of β -strand 7; rather, the phosphate group is coordinated by a loop from the N-terminal domain.

The link between the enolase superfamily and the wider superfamily of phosphate-binding TIM barrels is made *via* a sequence family of unknown structure and function (Pfam name UPF0034, accession PF01207 (Bateman *et al.*, 2000)). Representative sequences of this family show significant similarity to both muconate lactonizing enzyme

and dihydroorotate dehydrogenase. The PSI-blast alignments of both dihydroorotate dehydrogenase and muconate lactonizing enzyme to the UPF0034 protein family suggests a circular permutation-like rearrangement in the lineage leading to the enolase superfamily, with the correct alignment mapping β -strand 3 of enolase to β -strand 5 of the superfamily of phosphate-binding barrels (see Figure 3). A study based solely on structural criteria has suggested a similar permutation (Sergeev & Lee, 1994). The atypical topology of the enolase barrel, beginning with a $\beta\beta\alpha$ subunit is not explained by this theory; however, our result suggests that this region of sequence is not homologous to the $\alpha\beta\alpha\beta$ region that begins standard $(\beta\alpha)_8$ barrels.

The proposed alignment equivalences metal-binding residues of the enolase superfamily with the glycine-rich, phosphate-binding loop of the phosphate-binding superfamily, and a catalytic lysine residue of dihydroorotate dehydrogenase is aligned with a further metal-binding residue of the enolase family. It seems likely that the backbone conformations of these sites make them favourable locations for functional residues, even if the molecular functions of the residues themselves are not conserved.

The UPF0034 proteins are widespread in both bacteria and eukaryotes, and although not experimentally characterized, have been predicted to adopt TIM barrel-like folds (Doerks *et al.*, 1998).

They align in phase to the dihydroorotate dehydrogenase family. The active-site cysteine residue of dihydroorotate dehydrogenase is present, although only one glycine residue in the phosphate-binding region of dihydroorotate dehydrogenase is conserved.

The pyruvate kinase family

Pyruvate kinase adopts a $(\beta\alpha)_8$ structure with an additional inserted domain at the end of β -strand 3, and a further domain at the C terminus of the barrel. The enzyme binds ATP and, like enolase, phosphoenolpyruvate and Mg^{2+} .

Within the context of our exploration of sequence space, we initiated a PSI-Blast search with the sequence of flavocytochrome *b2*. (The C-terminal domain of this enzyme is a member of the canonical phosphate-binding set of TIM barrel structures). With an increased inclusion threshold of 0.01, we detected pyruvate kinase from *Streptomyces coelicolor* with reliable significance (iteration 3, *E*-value 0.001, Figure 1). The phosphate-binding residues of flavocytochrome *b2* are not conserved (although equivalent β -strand 7 regions of the proteins are aligned). Inspection of available structures shows that this region does not bind phosphate in pyruvate kinase. However, the region is functionally important, containing a conserved threonine residue that may be involved in catalysis (Larsen

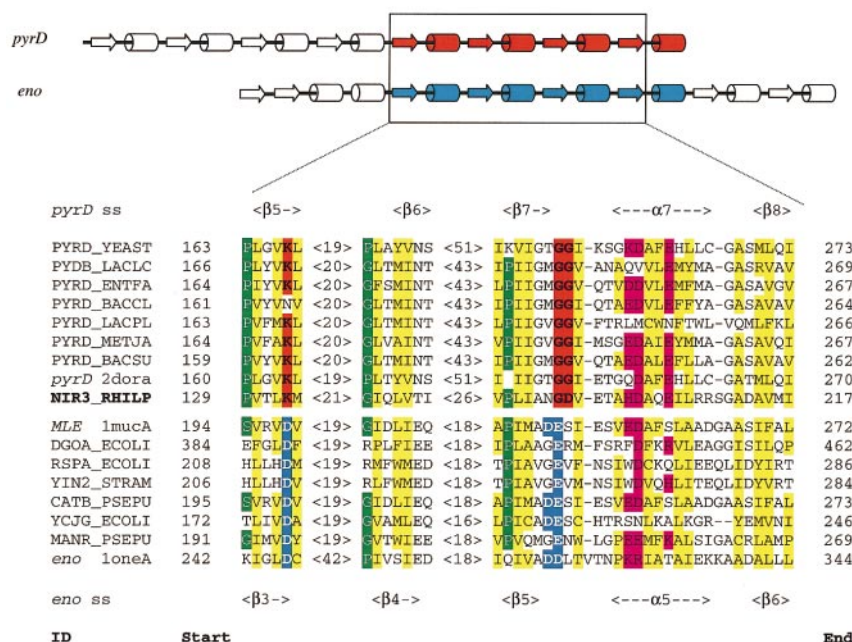


Figure 3. Alignment of permuted sequences of members of the *pyrD* family and enolase superfamily, as implied *via* the link to members of the UPF0034 family (see the text). Alignment of 1mucA to loneA, within the enolase superfamily, produced using STAMP. Within the *pyrD* family, the catalytic lysine residue and the glycine-rich, phosphate-binding loop are coloured in red. Within the enolase superfamily, metal-binding residues are coloured in blue. Conserved hydrophobic positions are shown in yellow, conserved charged in magenta, and conserved small in green. The ID column shows SWISS-PROT identifiers, or gene name and PDB identifier. Start and end residues are indicated. Numbers in angle brackets represent the number of residues between the represented fragments. Secondary structures of *pyrD* and *eno* are shown above and below the alignment, respectively.

et al., 1998). Moreover, the conformation of the protein backbone region of β -strand 7 and the following loop is remarkably similar to that found in the characteristic phosphate-binding loops (data not shown).

As we had relaxed the *E*-value threshold for this PSI-blast search, we further evaluated the relationship of pyruvate kinase to other ($\beta\alpha$)₈ barrels by performing structural comparisons, and calculating the significance of the sequence identity of the structural alignment (Murzin, 1993). The sequence identity of the structural alignment of pyruvate kinase and IMPDH is 19.1%. This corresponds to a significance of 8e-04, calculated using the method of Murzin (Murzin, 1993; Russell & Barton, 1992), suggesting that the two are likely homologues.

Further PSI-Blast searches showed significant relationships between pyruvate kinase and malate synthase of the glyoxylate cycle, *via* the sequence of citrate lyase. This latter sequence is also homologous to phosphoenolpyruvate synthase, and pyruvate carboxylase (see Figure 1).

The HMGL-like family

The proteins isopropylmalate synthase and homocitrate synthase, the first steps in leucine and lysine (α -amino adipate pathway) biosynthesis, respectively, contain a homologous region that is shared with pyruvate carboxylase, oxaloacetate decarboxylase and 4-hydroxy-2-oxovalerate aldolase (Pfam PF00682, HMGL-like (Bateman *et al.*, 2000)). Members of this family are detected in the fourth iteration of PSI-Blast searching starting from the *hisA* seed (e.g. the β subunit of pyruvate carboxylase from *Methanococcus janischii*, round 4 *E*-value 1e-04), indicating that a TIM barrel-like fold is adopted by this family. Secondary structure prediction (Jones, 1999) of this sequence clearly predicted eight β -strands alternating with α helices (Figure 4). Inspection of the alignment to *hisA* shows that a characteristic HxH motif of the HMGL-like family occurs at the end of the predicted β -strand 7 (Figure 4). It thus seems likely that this motif is involved in metal binding. As

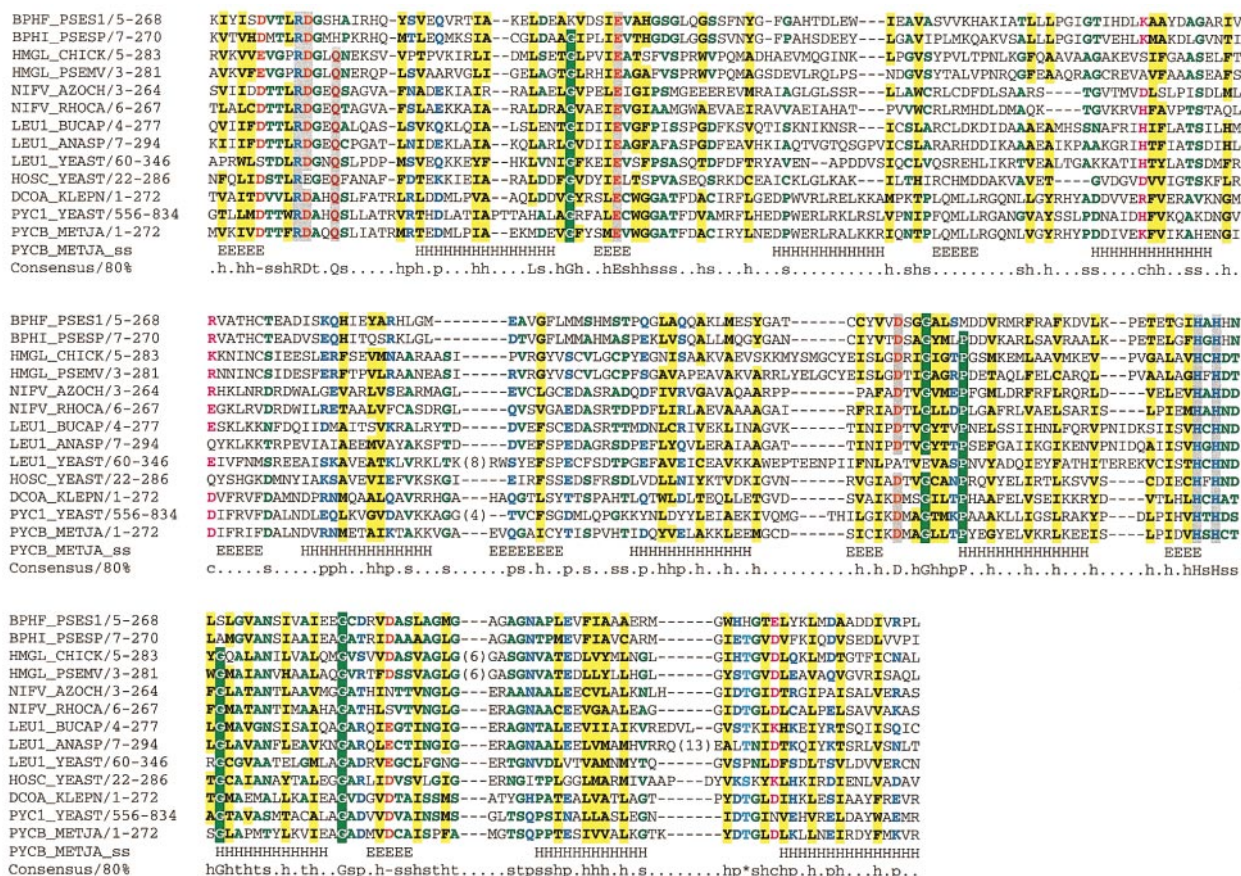


Figure 4. Multiple sequence alignment and secondary structure prediction for the HMGL-like family of proteins. Representative proteins were chosen from PFAM, and realigned, following slight modification of domain boundaries. SWISS-PROT sequence identifiers are used. The psipred (Jones, 1999) secondary structure prediction initiated with PYCB_METJA is superimposed on the alignment (two predictions of helices with fewer than four residues are not shown). The colour scheme for conserved residues is: yellow background, hydrophobic; green background, conserved glycine or proline; grey background, conserved single amino acid; white backgrounds: green, small; magenta, charged; blue, polar. Conservation is defined as 80% consensus. The Figure was produced using Chroma (L. Goodstadt & C.P. Ponting, unpublished).

some members of this family are aldolases, it is tempting to speculate that they may be related to the metal-dependent members of the aldolase superfamily, although a precise characterization must await the determination of the structure of a representative member.

Other links

In addition to the homologies reported above, we detected significant similarities from our set of homologues to two further SCOP superfamilies, characterized by dihydropteroate synthase and the N-terminal domain of methylmalonyl-CoA mutase (Figure 1 and Table 1), in addition to further sequence families (see <http://www.embl-heidelberg.de/~copley/TIM>) of proteins not involved in central metabolism.

Discussion

A phylogeny of TIM barrel-containing enzymes

The sequences of the enzymes found to be homologous in the previous section are too divergent to allow the reliable reconstruction of phylogeny from a sequence alignment, and techniques for phylogenetic tree inference from structures (as opposed to simply clustering similar structures) are not well developed (May, 1999). However, by analyzing their sequences, structures and functions, it is possible to discern some likely features of the branching order of modern TIM barrel-containing enzymes. Here, we concentrate only on those enzymes found in central metabolism, for which a structure is known.

As has been observed, the sequence of the protein encoded by *hisA* shows a clear internal duplication (Fani *et al.*, 1994). As this duplication is not easily observable in other closely related homologous TIM barrel structures, it follows that *hisA* has retained ancestral features that have been obscured in other structures, suggesting that the common ancestor of modern barrels was most reminiscent of the protein encoded by *hisA*. Moreover, the internal sequence duplication conserves the functional glycine-rich, phosphate-binding site (Thoma *et al.*, 1998), indicating that it was present in the putative ancestral half-barrel structure and that phosphate-binding at the C terminus of β -strand 7 is likely to be an ancestral, rather than derived feature in modern TIM barrels (Lang *et al.*, 2000). Of the extant TIM barrels, it appears that *hisA* is closest in appearance to the putative ancestor (although not necessarily sharing the same function).

The protein encoded by *hisA* catalyzes the ring opening of a ribose sugar moiety. An analogous reaction is catalyzed by PRA isomerase (*trpC_1*). However, analysis of the remaining phosphate-binding proteins shows very different chemical activities. Rather, the function common to all

appears to be the use of a phosphate group as a molecular anchor, by which to bind ligands to the protein, and bring them into proximity with diverse catalytic residues (Figure 5). However, of particular interest is the observation that transaldolase, the α -subunit of tryptophan synthase, triose phosphate isomerase and fructose-1,6-bisphosphate aldolase all catalyze reactions where D-glyceraldehyde 3-phosphate is one of the products (Figure 5). Such a common product may be indicative of a common D-glyceraldehyde 3-phosphate-binding ancestor for these enzymes. An analogous situation is found with phosphoenolpyruvate, which is a product or substrate of enolase, pyruvate kinase, phosphoenolpyruvate carboxylase and phosphoenolpyruvate synthase.

The presence of the phosphate-binding site in the putative ancestor of TIM barrels suggests that TIM barrels homologous to this family that do not contain an equivalent phosphate-binding site are derived from phosphate-binding ancestors. Hence, we conclude that 3-dehydroquinase, N-acetylneuraminic lyase and dihydrodipicolinate synthase have all arisen from an ancestor having a phosphate-binding site and a catalytic Schiff base-forming lysine residue on β -strand 6, such as fructose-1,6-bisphosphate aldolase.

The relationships of the enolase and pyruvate kinase superfamilies to the main group of enzymes are more problematic. Global structure alignment using the STAMP program (Russell & Barton, 1992) indicates that the closest relative of the pyruvate kinase barrel domain is IMPDH (*guaB*), which is in turn closest structurally to the FMN-binding family of barrels characterized by dihydroorotate dehydrogenase (*pyrD*). Indeed, the structural alignment equivalences the catalytic lysine residue of dihydroorotate dehydrogenase with an essential catalytic lysine of pyruvate kinase (Bollenbach *et al.*, 1999). We also place enolase close to the FMN-binding family, as the closest sequence link is to *pyrD*. Solution of the structure of a representative member of the PFAM UPF0034 family may help to resolve this position further.

Both pyruvate kinase and enolase bind phosphoenolpyruvate and magnesium, and neither shows the standard phosphate-binding motif characteristic of other TIM barrel-containing enzymes. Although the mode of binding of PEP is different in both cases (after allowing for the proposed permutation of enolase), these differences may have been caused by the later acquisition of additional domains. We suggest that both may have arisen from a common PEP and Mg^{2+} -binding ancestor.

Figure 6 shows a synthesis of the arguments above, presented as a phylogeny of TIM barrel-containing enzymes. We have tried to reconcile the nature of functional residues and details of bound ligands with a clustering of similar structures. The pathway to which an enzyme belongs is shown on the phylogeny, but has not influenced our reconstruction.

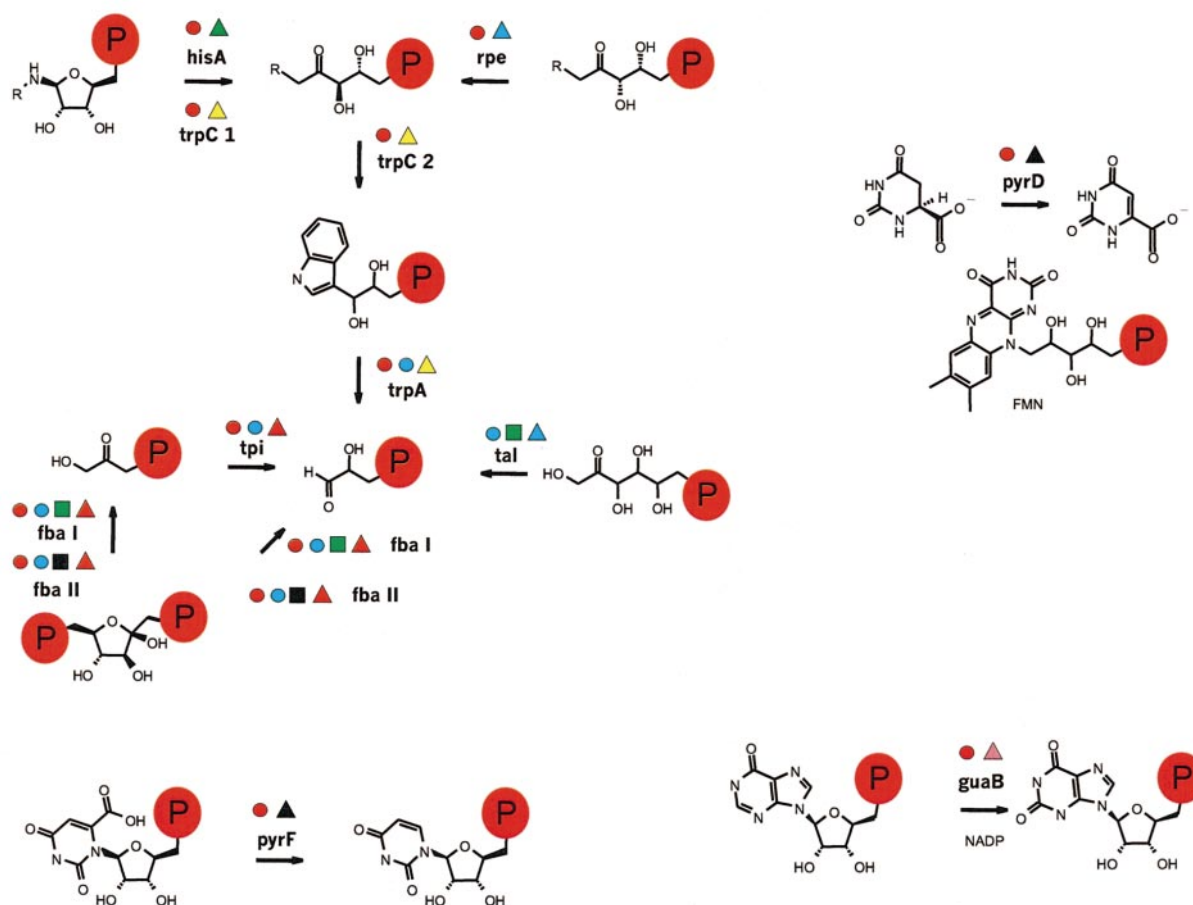


Figure 5. Diverse reactions catalyzed by canonical phosphate-binding TIM barrel enzymes. Coloured symbols as for Figure 6

TIM barrel-containing proteins and the evolution of pathways

Known TIM barrel folds within the enzymes of central metabolism are shown in Figure 7. Of these we have shown all but dihydroorotase (*pyrC*) are likely to be homologous. It is clear that homologues are distributed widely both within and between pathways. When multiple homologues occur within the same pathway, they are not necessarily found to occur in adjacent enzymes. Such a picture points to large-scale recruitment of enzymes between different pathways, rather than the duplication within pathways model emphasized by Horowitz.

TIM barrel-containing enzymes occur at adjacent pathway steps in tryptophan biosynthesis, histidine biosynthesis and glycolysis. A corollary of the Horowitz theory of pathway evolution is that, where a set of adjacent enzymes in a pathway is homologous, the enzyme of the latest step should be the deepest branching (i.e. ancestral) when a phylogeny is reconstructed. Our results do not provide strong evidence for this. On the evidence of the previous section, we suggest that it is more likely that a *hisA*-like protein was ancestral to the

modern *hisF* protein. Similarly, we propose that *trpA* had a *trpC_1* or *hisA*-like ancestor, but not *vice versa*. Such ancestral proteins could have had the ability to catalyze more than one reaction, albeit less efficiently, or they could have had essentially their modern functions, with intervening reactions being catalyzed by alternative methods that have since been displaced.

Recently, protein engineering followed by artificial selection has been used to generate *trpC_1*-like function from a *trpC_2*-like scaffold (i.e. in the direction suggested by Horowitz) (Altamirano *et al.*, 2000). However, while alternative evolutionary routes are untested, the result sheds little light on evolutionary history, besides showing the possibility of such functional conversion.

Intriguingly, our results suggest that both class I and class II fructose-1,6-bisphosphate aldolases may have arisen relatively late (compared to histidine biosynthesis). This idea is given some support by the fact that there is considerable plasticity in the upper half of the glycolytic pathway; alternative pathways are found in some species, and non-orthologous displacement of enzymes is more common, suggesting that it arose later in evolution (Koonin *et al.*, 1996; Romano & Conway, 1996;

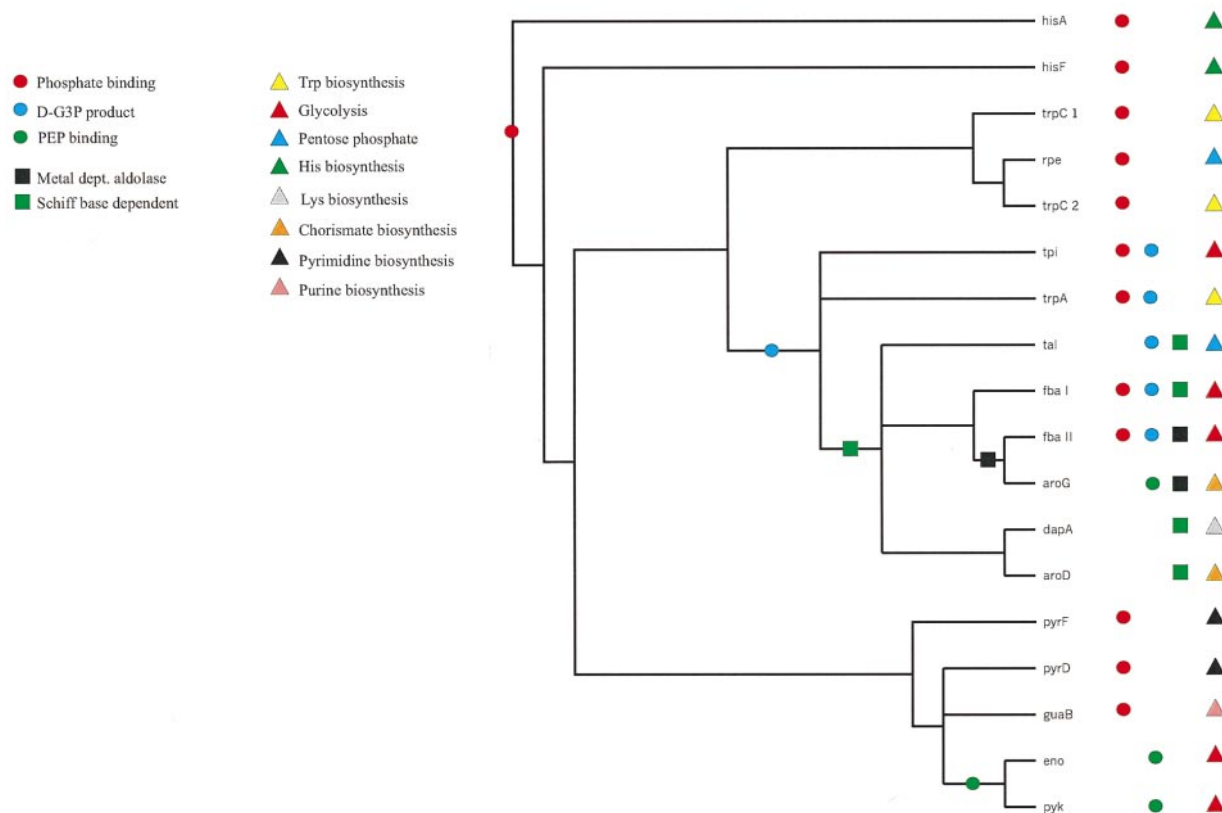


Figure 6. A proposed phylogeny for the TIM barrel-containing enzymes of central metabolism. See the text for detailed discussion.

Dandekar *et al.*, 1999). Moreover, of the four Schiff base-dependent TIM barrel enzymes found in central metabolism, two (3-dehydroquinase and fructose-1,6-bisphosphate aldolase) show examples of non-orthologous enzymes performing the same function in different species (Galperin *et al.*, 1998; Gourley *et al.*, 1999), and a third (dihydrodipicolinate synthase) forms part of the diaminopimelic acid pathway of lysine biosynthesis, which is replaced by the alternative α -amino adipic acid pathway in some species (Michal, 1999).

Concluding Remarks

The weight of available evidence strongly favours divergent evolution for the proteins discussed here. Our results suggest that many very different functions can evolve on a common molecular scaffold, and that it may be possible to find extant links between very different functional families. Certain backbone positions seem particularly favourable for the evolution of catalytic residues (e.g. metal-binding residues in enolase occurring in a position equivalent to that of the phosphate-binding loop of other enzymes).

Looking beyond the functions of individual families, we have analyzed the homologous family presented here to determine the implications for pathway evolution. Modern pathways may not

reflect the state of ancestral pathways. The scale of replacement by analogous enzymes during the course of evolution is not known, but so far, examples in central metabolism are limited (Galperin *et al.*, 1998). Even if the proteins of a particular biochemical pathway have changed over time, the question of how contemporary pathways evolved can still be addressed.

We have found no strong evidence to support the idea that pathways may have evolved backwards and enzymes within a pathway are likely to be homologous, even in cases where this has previously been suggested (e.g. tryptophan biosynthesis). Given that the $(\beta\alpha)_8$ barrel is probably the most common enzyme fold (Hegyi & Gerstein, 1999), it seems unlikely that any such evidence will emerge in the future, at least for the pathways of central metabolism. It is possible there is increased likelihood of homologues at certain key intermediates (for instance D-glyceraldehyde-3-phosphate and PEP). Equally, notwithstanding previous work (Gerlt & Babbitt, 1998), over long enough time-scales, the catalytic mechanism of enzymes does not appear to be conserved. Although members of the enolase superfamily all share similarities in mechanism, we have shown that they are likely to be homologous to a broader superfamily that do not, and similarly with Schiff base-dependent TIM barrel enzymes.

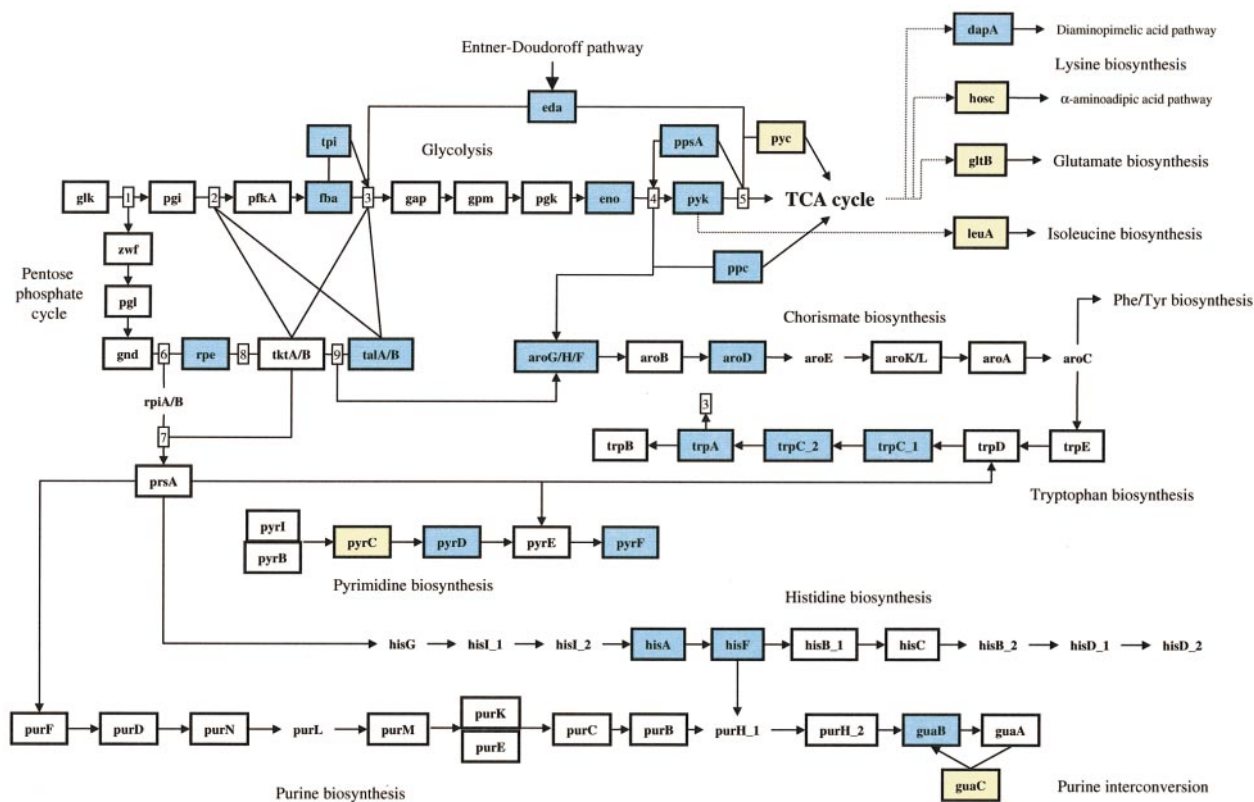


Figure 7. A schematic view showing the relative positions of TIM barrel-containing enzymes in the pathways of central metabolism. *E. coli* gene names (Karp *et al.*, 2000) have been used with the exception of *hosc* for homocitrate synthase (see Table 1 for full descriptions of gene names). Where two enzymatic steps are encoded by the same gene, we have labelled them *gene_1* and *gene_2*, in the order that they occur in the pathway as drawn. Proteins of known or predicted structure are boxed. Known TIM barrel-containing proteins are coloured blue. Predicted TIM barrel-containing proteins are yellow. Non-TIM barrel-containing proteins are white. Dotted connecting lines indicate more than one step between proteins. Arrows indicate the conventional view of the direction of the pathway. Some key compounds are shown for clarity: (1) D-glucose; (2) β -D-fructose 6-phosphate; (3) D-glyceraldehyde 3-phosphate; (4) phosphoenolpyruvate; (5) pyruvate; (6) D-ribulose 5-phosphate; (7) D-ribose 5-phosphate; (8) xyulose 5-phosphate; (9) erythrose 4-phosphate. The reaction of *aroD* is also catalyzed by a non-TIM barrel-containing enzyme (see the text for details). Abbreviations of non-TIM barrel-containing enzymes (see Table 1 for those containing TIM barrels): *glk*, glucokinase; *pgi*, phosphoglucose isomerase; *pfkA*, 6-phosphofructokinase-1; *gap*, glyceraldehyde 3-phosphate dehydrogenase; *gpm*, phosphoglycerate mutase; *pgk*, phosphoglycerate kinase; *zwf*, glucose 6-phosphate-1-dehydrogenase; *pgl*, 6-phosphogluconolactonase; *gnd*, 6-phosphogluconate dehydrogenase; *tktA*, transketolase; *prsA*, phosphoribosylpyrophosphate synthase; *aroB*, 3-dehydroquinate synthase; *aroE*, shikimate dehydrogenase; *aroK/L*, shikimate kinase I/II; *aroA*, 3-phosphoshikimate-1-carboxyvinyltransferase; *aroC*, chorismate synthase; *trpE*, anthranilate synthase; *trpD*, anthranilate phosphoribosyl transferase; *trpB*, tryptophan synthase β -subunit; *pyrI/B*, aspartate carbonyltransferase; *pyrE*, orotate phosphoribosyl transferase; *hisG*, ATP phosphoribosyltransferase; *hisI_1*, phosphoribosyl-ATP pyrophosphatase; *hisI_2*, phosphoribosyl AMP cyclohydrolase; *hisB_1*, imidazoleglycerolphosphate dehydratase; *hisC*, histidine-phosphate aminotransferase; *hisB_2*, histidinol-phosphatase; *hisD_1*, histidinol dehydrogenase; *hisD_2*, histidinal dehydrogenase; *purF*, amidophosphoribosyl transferase; *purD*, P-ribosylamine-glycine ligase; *purN*, phosphoribosylglycinamide formyltransferase; *purL*, phosphoribosylformylglycinamide synthase; *purM*, phosphoribosylformylglycinamide cyclo-ligase; *purK*, *purE*, phosphoribosylaminoimidazole carboxylase; *purC*, phosphoribosylaminoimidazole-succinocarboxamide synthase; *purB*, adenylosuccinate lyase; *purH_1*, AICAR transformylase; *purH_2*, IMP cyclohydrolase; *guaA*, GMP synthase.

Instead, our results favour general recruitment of enzymes with similar functions (either a particular catalytic activity, or specifically binding particular groups). Such recruitment need not necessarily be from the same pathway, as discussed by Jensen (1976). Such a view points to ancient enzymes having very broad specificities, but also hints that it may be possible to resolve the order in which pathways have evolved, if it is possible to construct robust phylogenies for their constituent proteins.

Methods

Sequence database searching

The SCOP database (Lo Conte *et al.*, 2000), in conjunction with ASTRAL (Brenner *et al.*, 2000) was used to select representative TIM barrel sequences at the 50% sequence identity level. Protein families for which no structures were available were selected if a clear relationship existed to a protein known to adopt a TIM barrel fold. The sequences were used to initiate PSI-Blast

searches of the NR protein sequence database from the NCBI (Altschul *et al.*, 1997). We use the term seed to refer to these initiation sequences. PSI-Blast uses an iterative procedure, constructing position-dependent weight-matrices on the basis of alignments generated by a Blast database search. Sequences are used in the construction of the weight matrix if they generate an expectation (*E*) value more significant than a given inclusion threshold. These weight matrices are then used to score the next iteration of database searching. The NR database was prefiltered for low complexity, coiled coil and transmembrane regions using the *pfilt* program from the method of Jones (1999). PSI-Blast searches were run to convergence, with an inclusion threshold of *E* < 0.001, unless otherwise stated. In all cases the output was inspected for the presence of obvious false positives, including low-complexity regions or proteins of known non-TIM barrel structure.

The use of a database of 50% identity, in conjunction with a sensitive algorithm such as PSI-Blast leads to considerable redundancy in the search process in many cases. (i.e. relationships may be discovered many times over). However, PSI-Blast does show sensitivity to the protein sequence chosen to seed it, and as such starting with a range of initial sequences can be advantageous (Aravind & Koonin, 1999). In our results, we preferentially report relationships detected with few iterations.

In some cases, sequences that were common to the results of two PSI-Blast runs were used to link families, in a similar manner to the work reported by Koretke *et al.* (1999).

Structure comparison

Global TIM barrel structure comparisons were performed using the method by Russell & Barton (1992). Structures were obtained from the RCSB protein data base (Berman *et al.*, 2000), with the exception of *hisA* and *hisF* (M. Wilmanns, personal communication). Secondary structure was predicted using the method of Jones (1999).

Metabolic pathways

We define central metabolism to include glycolysis, the tricarboxylic acid (TCA) cycle, the pentose phosphate pathway, amino acid biosynthesis and nucleotide biosynthesis. Information on metabolic pathways, enzymes and substrates was taken from Michal (1999).

Acknowledgements

We are very grateful to Matthias Wilmanns (EMBL Hamburg) for providing us with the coordinates of *hisA* and *hisF*, and to Reinhard Sterner for supplying us with preprints of his work. We thank Martijn Huynen, Berend Snel, Warren Lathe, Steffen Schmidt and Thomas Dandekar for many helpful discussions, and Jörg Schultz, for assistance with an early analysis which led to this work. This work was funded by the BMBF ELSA project.

References

Achari, A., Somers, D. O., Champness, J. N., Bryant, P. K., Rosemond, J. & Stammers, D. K. (1997). Crys-

tal structure of the anti-bacterial sulfonamide drug target dihydropteroate synthase. *Nature Struct. Biol.* **4**, 490-497.

- Altamirano, M. M., Blackburn, J. M., Aguayo, C. & Fersht, A. R. (2000). Directed evolution of new catalytic activity using the α/β -barrel scaffold. *Nature*, **403**, 617-622.
- Altschul, S. F. L. M. T., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389-3402.
- Appleby, T. C., Kinsland, C., Begley, T. P. & Ealick, S. E. (2000). The crystal structure and mechanism of orotidine 5'-monophosphate decarboxylase. *Proc. Natl Acad. Sci. USA*, **97**, 2005-2010.
- Aravind, L. & Koonin, E. V. (1999). Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. *J. Mol. Biol.* **287**, 1023-1040.
- Babbitt, P. C. & Gerlt, J. A. (1997). Understanding enzyme superfamilies. *J. Biol. Chem.* **272**, 30591-30594.
- Babbitt, P. C., Hasson, M. S., Wedekind, J. E., Palmer, D. R. J., Barrett, W. C., Reed, G. H., Rayment, I., Ringe, D., Kenyon, G. L. & Gerlt, J. A. (1996). The enolase superfamily: a general strategy for enzyme-catalyzed abstraction of the α -protons of carboxylic acids. *Biochemistry*, **35**, 16489-16501.
- Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L. & Sonnhammer, E. L. (2000). The PFAM protein families database. *Nucl. Acids Res.* **28**, 263-266.
- Belfaiza, J., Parsot, C., Martel, A., Bouthier de la Tour, C., Margarita, D., Cohen, G. N. & Saint-Girons, I. (1986). Evolution of biosynthetic pathways: two enzymes catalyzing consecutive steps in methionine biosynthesis originate from a common ancestor and possess a similar regulatory region. *Proc. Natl Acad. Sci. USA*, **83**, 867-871.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). The protein data bank. *Nucl. Acids Res.* **28**, 235-242.
- Bollenbach, T. J., Mesecar, A. D. & Nowak, T. (1999). Role of lysine 240 in the mechanism of yeast pyruvate kinase catalysis. *Biochemistry*, **38**, 9137-9145.
- Bork, P. & Rohde, K. (1990). Sequence similarities between tryptophan synthase β subunit and other pyridoxal-phosphate-dependent enzymes. *Biochem. Biophys. Res. Commun.* **171**, 1319-1325.
- Bork, P., Gellerich, J., Groth, H., Hooft, R. & Martin, F. (1995). Divergent evolution of a $\beta\alpha$ -barrel subclass: detection of numerous phosphate-binding sites by motif search. *Protein Sci.* **4**, 268-274.
- Brenner, S. E., Koehl, P. & Levitt, M. (2000). The ASTRAL compendium for protein structure and sequence analysis. *Nucl. Acids Res.* **28**, 254-256.
- Buchanan, C. L., Connaris, H., Danson, M. J., Reeve, C. D. & Hough, D. W. (1999). An extremely thermostable aldolase from *Sulfolobus solfataricus* with specificity for non-phosphorylated substrates. *Biochem. J.* **343**, 563-570.
- Chiu, H. J., Reddick, J. J., Begley, T. P. & Ealick, S. E. (1999). Crystal structure of thiamin phosphate synthase from *Bacillus subtilis* at 1.25 Å resolution. *Biochemistry*, **38**, 6460-6470.
- Dandekar, T., Schuster, S., Snel, B., Huynen, M. & Bork, P. (1999). Pathway alignment: application to the

- comparative analysis of glycolytic enzymes. *Biochem. J.* **343**, 115-124.
- Doerks, T., Bairoch, A. & Bork, P. (1998). Protein annotation: detective work for function prediction. *Trends Genet.* **14**, 248-250.
- Eads, J. C., Ozturk, D., Wexler, T. B., Grubmeyer, C. & Sacchettini, J. C. (1997). A new function for a common fold: the crystal structure of quinolinic acid phosphoribosyltransferase. *Structure*, **15**, 47-58.
- Eklund, H. & Fontecave, M. (1999). Glycyl radical enzymes: a conservative structural basis for radicals. *Structure*, **7**, R257-R262.
- Fani, R., Lio, P., Chiarelli, I. & Bazzicalupo, M. (1994). The evolution of the histidine biosynthetic genes in prokaryotes: a common ancestor for the *hisA* and *hisF* genes. *J. Mol. Evol.* **38**, 489-495.
- Farber, G. K. & Petsko, G. A. (1990). The evolution of α/β barrel enzymes. *Trends Biochem. Sci.* **15**, 228-234.
- Galperin, M. Y., Walker, D. R. & Koonin, E. V. (1998). Analogous enzymes: independent inventions in enzyme evolution. *Genome Res.* **8**, 779-790.
- Galperin, M. Y., Aravind, L. & Koonin, E. V. (2000). Aldolases of the DhnA family: a possible solution to the problem of pentose and hexose biosynthesis in archaea. *FEMS Microbiol. Letters*, **183**, 259-264.
- Gerlt, J. A. (2000). New wine from old barrels. *Nature Struct. Biol.* **7**, 171-173.
- Gerlt, J. A. & Babbitt, P. C. (1998). Mechanistically diverse enzyme superfamilies: the importance of chemistry in the evolution of catalysis. *Curr. Opin. Chem. Biol.* **2**, 607-612.
- Gourley, D. G., Shrive, A. K., Polikarpov, I. P., Krell, T., Coggins, J. R., Hawkins, A. R., Isaacs, N. W. & Sawyer, L. (1999). The two types of 3-dehydroquinase have distinct structures but catalyze the same overall reaction. *Nature Struct. Biol.* **6**, 521-525.
- Hall, D. R., Leonard, G. A., Reed, C. D., Watt, C. I., Berry, A. & Hunter, W. N. (1999). The crystal structure of *Escherichia coli* class II fructose-1,6-bisphosphate aldolase in complex with phosphoglycolohydroxamate reveals details of mechanism and specificity. *J. Mol. Biol.* **287**, 383-394.
- Hegyi, H. & Gerstein, M. (1999). The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. *J. Mol. Biol.* **288**, 147-164.
- Helin, S., Kahn, P. C., Guha, B. L., Mallows, D. G. & Goldman, A. (1995). The refined X-ray structure of muconate lactonizing enzyme from *Pseudomonas putida* PRS2000 at 1.85 Å resolution. *J. Mol. Biol.* **254**, 918-941.
- Hennig, M., Darimont, B., Sterner, R., Kirschner, K. & Jansonius, J. N. (1995). 2.0 Å structure of indole-3-glycerol phosphate synthase from the hyperthermophile *Sulfolobus solfataricus*: possible determinants of protein stability. *Structure*, **3**, 1295-1306.
- Hennig, M., Sterner, R., Kirschner, K. & Jansonius, J. N. (1997). Crystal structure at 2.0 Å resolution of phosphoribosyl anthranilate isomerase from the hyperthermophile *Thermotoga maritima*: possible determinants of protein stability. *Biochemistry*, **36**, 6009-6016.
- Herzberg, O., Chen, C. C. H., Kapadia, G., McGuire, M., Carroll, L., Noh, S. J. & Dunaway-Mariano, D. (1996). Swivelling-domain mechanism for enzymatic phosphotransfer between remote reaction sites. *Proc. Natl Acad. Sci. USA*, **93**, 2652-2657.
- Hester, G., Brenner-Holzach, O., Rossi, F. A., Stuck-Donatz, M., Winterhalter, K. H., Smit, J. D. & Piontek, K. (1991). The crystal structure of fructose-1,6-bisphosphate aldolase from *Drosophila melanogaster* at 2.5 Å resolution. *FEBS Letters*, **292**, 237-242.
- Holm, L. & Sander, C. (1997). An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. *Proteins: Struct. Funct. Genet.* **28**, 72-82.
- Horowitz, N. H. (1945). On the evolution of biochemical syntheses. *Proc. Natl Acad. Sci. USA*, **31**, 153-157.
- Horowitz, N. H. (1965). The evolution of biochemical syntheses - retrospect and prospect. In *Evolving Genes and Proteins* (Bryson, V. & Vogel, H. J., eds), pp. 15-23, Academic Press, New York.
- Howard, B. R., Endrizzi, J. A. & Remington, S. J. (2000). Crystal structure of *Escherichia coli* malate synthase G complexed with magnesium and glyoxylate at 2.0 Å resolution. *Biochemistry*, **39**, 3156-3168.
- Janecek, S. (1995). Similarity of different β -strands flanked in loops by glycines and prolines from distinct (α/β)₈-barrel enzymes: chance or a homology? *Protein Sci.* **4**, 1239-1242.
- Janecek, S. (1996). Invariant glycines and prolines flanking in loops the strand β_2 of various (α/β)₈-barrel enzymes: a hidden homology? *Protein Sci.* **5**, 1136-1143.
- Janecek, S. & Balaz, S. (1995). Functionally essential, invariant glutamate near the C terminus of strand β_3 in various (α/β)₈-barrel enzymes as a possible indicator of their evolutionary relatedness. *Protein Eng.* **8**, 809-813.
- Jensen, R. A. (1976). Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **30**, 409-425.
- Jia, J., Huang, W., Schorken, U., Sahn, H., Sprenger, G. A., Lindqvist, Y. & Schneider, G. (1996). Crystal structure of transaldolase B from *Escherichia coli* suggests a circular permutation of the α/β barrel within the class I aldolase family. *Structure*, **4**, 715-724.
- Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195-202.
- Jurica, M. S., Mesecar, A., Heath, P. J., Shi, W., Nowak, T. & Stoddard, B. L. (1998). The allosteric regulation of pyruvate kinase by fructose-1,6-bisphosphate. *Structure*, **6**, 195-210.
- Kai, Y., Matsumura, H., Inoue, T., Terada, K., Nagara, Y., Yoshinga, T., Kihara, A., Tsumura, K. & Izui, K. (1999). Three-dimensional structure of phosphoenolpyruvate carboxylase: a proposed mechanism for allosteric inhibition. *Proc. Natl Acad. Sci. USA*, **96**, 823-828.
- Karp, P. D., Riley, M., Saier, M., Paulsen, I. T., Paley, S. M. & Pellegrini-Toole, A. (2000). The ecocyc and metacyc databases. *Nucleic Acids Res.* **28**, 56-59.
- Koonin, E. V., Mushegian, A. R. & Bork, P. (1996). Non-orthologous gene displacement. *Trends Genet.* **12**, 334-336.
- Kopp, J., Kopriva, S., Suss, K. H. & Schulz, G. E. (1999). Structure and mechanism of the amphibolic enzyme D-ribulose-5 3-epimerase. *J. Mol. Biol.* **287**, 761-771.
- Koretke, K. K., Russell, R. B., Copley, R. R. & Lupas, A. N. (1999). Fold recognition using sequence and secondary structure information. *Proteins: Struct. Funct. Genet.* **37** (S3), 141-148.
- Lang, D., Thoma, R., Henn-Sax, M., Sterner, R. & Wilmanns, M. (2000). Structural evidence for evol-

- ution of the β/α barrel scaffold by gene duplication and fusion. *Science*, **289**, 1546-1550.
- Larsen, T. M., Wedekind, J. E., Rayment, I. & Reed, G. H. (1996). A carboxylate oxygen of the substrate bridges the magnesium ions at the active site of enolase: structure of the yeast enzyme complexed with the equilibrium mixture of 2-phosphoglycerate and phosphoenolpyruvate at 1.8 Å resolution. *Biochemistry*, **35**, 4349-4358.
- Larsen, T. M., Benning, M. M., Rayment, I. & Reed, G. H. (1998). Structure of the bis(Mg^{2+})-ATP oxalate complex of the rabbit muscle pyruvate kinase at 2.1 Å resolution: ATP binding over a barrel. *Biochemistry*, **37**, 6247-6255.
- Lawrence, M. C., Barbosa, J. A. R. G., Smith, B. J., Hall, N. E., Pilling, P. A., Ooi, H. C. & Marcuccio, S. M. (1997). Structure and mechanism of a sub-family of enzymes related to *N*-acetylneuraminase lyase. *J. Mol. Biol.* **266**, 381-399.
- Lo Conte, L., Ailey, B., Hubbard, T. J., Brenner, S. E., Murzin, A. G. & Chothia, C. (2000). SCOP: a structural classification of proteins database. *Nucl. Acids Res.* **28**, 257-259.
- Mancia, F., Smith, G. A. & Evans, P. R. (1999). Crystal structure of substrate complexes of methylmalonyl-CoA mutase. *Biochemistry*, **38**, 7999-8005.
- Martin, A. C. R., Orengo, C. A., Hutchinson, E. G., Jones, S., Karmirantzou, M., Laskowski, R. A., Mitchell, J. B. O., Taroni, C. & Thornton, J. M. (1998). Protein folds and functions. *Structure*, **6**, 875-884.
- Mavridis, I. M., Hatada, M. H., Tulinsky, A. & Lebiada, L. (1982). Structure of 2-keto-3-deoxy-6-phosphogluconate aldolase at 2.8 Å resolution. *J. Mol. Biol.* **162**, 419-444.
- May, A. C. (1999). Towards a more meaningful hierarchical classification of protein three-dimensional structures. *Proteins: Struct. Funct. Genet.* **37**, 20-29.
- Michal, G. (1999). *Biochemical Pathways*, John Wiley & Sons, Inc.
- Mirwaldt, C., Korndorfer, I. & Huber, R. (1995). The crystal structure of dihydrodipicolinate synthase from *Escherichia coli* at 2.5 Å resolution. *J. Mol. Biol.* **246**, 227-239.
- Murzin, A. G. (1993). Sweet-tasting protein monellin is related to the cystatin family of thiol proteinase inhibitors. *J. Mol. Biol.* **230**, 689-694.
- Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536-540.
- Nagano, N., Hutchinson, E. G. & Thornton, J. M. (1999). Barrel structures in proteins: automatic identification and classification including a sequence analysis of TIM barrels. *Protein Sci.* **8**, 2072-2084.
- Pujadas, G. & Palau, J. (1999). TIM barrel fold: structural, functional and evolutionary characteristics in natural and designed molecules. *Biologia, Bratislava*, **54**, 231-253.
- Rhee, S., Parris, K. D., Hyde, C. C., Ahmed, S. A., Miles, E. W. & Davies, D. R. (1997). Crystal structures of mutant (β K87T) tryptophan synthase $\alpha_2\beta_2$ complex with ligands bound to the active sites of the α and β -subunits reveal ligand-induced conformational changes. *Biochemistry*, **36**, 7664-7680.
- Romano, A. H. & Conway, T. (1996). Evolution of carbohydrate metabolic pathways. *Res. Microbiol.* **147**, 448-455.
- Rowland, P., Bjornberg, O., Nielsen, F. S., Jensen, K. F. & Larsen, S. (1998). The crystal structure of *Lactococcus lactis* dihydroorotate dehydrogenase A complexed with the enzyme reaction product throws light on its enzymatic function. *Protein Sci.* **7**, 1269-1279.
- Russell, R. B. & Barton, G. J. (1992). Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins: Struct. Funct. Genet.* **14**, 309-323.
- Sergeev, Y. & Lee, B. (1994). Alignment of beta-barrels in (β/α)₈ proteins using hydrogen-bonding pattern. *J. Mol. Biol.* **244**, 168-182.
- Sharma, V., Grubmeyer, C. & Sacchettini, J. C. (1998). Crystal structure of quinolinic acid phosphoribosyltransferase from *Mycobacterium tuberculosis*: a potential TB drug target. *Structure*, **6**, 1587-1599.
- Shumilin, I. A., Kretsinger, R. H. & Bauerle, R. H. (1999). Crystal structure of phenylalanine-regulated 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from *Escherichia coli*. *Structure*, **7**, 865-875.
- Tegoni, M. & Cambillau, C. (1994). The 2.6 Å refined structure of the *Escherichia coli* recombinant *Saccharomyces cerevisiae* flavocytochrome b2-sulfite complex. *Protein Sci.* **3**, 303-313.
- Thoma, R., Schwander, M., Liebl, W., Kirschner, K. & Sterner, R. (1998). A histidine gene cluster of the hyperthermophile *Thermotoga maritima*: sequence analysis and evolutionary significance. *Extremophiles*, **2**, 379-389.
- Whitby, F. G., Luecke, H., Kuhn, P., Somoza, J. R., Huete-Perez, J. A., Phillips, J. D., Hill, C. P., Fletterick, R. J. & Wang, C. C. (1997). Crystal structure of *Tritochomonas foetus* inosine-5'-monophosphate dehydrogenase and the enzyme-product complex. *Biochemistry*, **36**, 10666-10674.
- Wilmanns, M., Hyde, C. C., Davies, D. R., Kirschner, K. & Jansonius, J. N. (1991). Structural conservation in parallel β/α -barrel enzymes that catalyze three sequential reactions in the pathway of tryptophan biosynthesis. *Biochemistry*, **30**, 9161-9169.
- Zhang, Z., Sugio, S., Komives, E. A., Liu, K. D., Knowles, J. R., Petsko, G. A. & Ringe, D. (1994). Crystal structure of recombinant chicken triosephosphate isomerase-phosphoglycohydroxamate complex at 1.8 Å resolution. *Biochemistry*, **33**, 2830-2837.

Edited by G. Von Heijne

(Received 17 May 2000; received in revised form 6 September 2000; accepted 7 September 2000)