Protein Science (1995), *4*:268-274. Cambridge University Press. Printed in the USA.

Divergent evolution of a β/α-barrel subclass: Detection of numerous phosphate-binding sites by motif search

PEER BORK,[1,2] JOHANNES GELLERICH,[3] HOLGER GROTH,[3] ROB HOOFT,[1] AND FALK MARTIN[3]

[1] European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany
[2] Max-Delbrück-Center, 13189 Berlin-Buch, Germany
[3] Department of Biology, Humboldt University, Berlin, Germany

(RECEIVED October 7, 1994; ACCEPTED November 21, 1994)

Abstract

Study of the most conserved region in many β/α-barrels, the phosphate-binding site, revealed a sequence motif in a few β/α-barrels with known tertiary structure, namely glycolate oxidase (GOX), cytochrome b_2 (Cyb2), tryptophan synthase α subunit (TrpA), and the indoleglycerolphosphate synthase (TrpC). Database searches identified this motif in numerous other enzyme families: (1) IMP dehydrogenase (IMPDH) and GMP reductase (GuaC); (2) phosphoribosylformimino-5-aminoimidazol carboxamide ribotide isomerase (HisA) and the cyclase-producing D-erythro-imidazole-glycerolphosphate (HisF) of the histidine biosynthetic pathway; (3) dihydroorotate dehydrogenase (PyrD); (4) glutamate synthase (GltB); (5) ThiE and ThiG involved in the biosynthesis of thiamine as well as related proteins; (6) an uncharacterized open reading frame from *Erwinia herbicola*; and (7) a glycerol uptake operon antiterminator regulatory protein (GlpP). Secondary structure predictions of the different families mentioned above revealed an alternating order of β-strands and α-helices in agreement with a β/α-barrel-like topology. The putative phosphate-binding site is always found near the C-terminus of the enzymes, which are all at least about 200 amino acids long. This is compatible with its assumed location between strand 7 and helix 8. The identification of a significant motif in functionally diverse enzymes suggests a divergent evolution of at least a considerable fraction of β/α-barrels. In addition to the known accumulation of β/α-barrels in the tryptophan biosynthetic pathway, we observe clusters of these enzymes in histidine biosynthesis, purine metabolism, and apparently also in thiamine biosynthesis. The substrates are mostly heterocyclic compounds. Although the marginal sequence similarities do not allow a reconstruction of the barrel spreading, they support the idea of pathway evolution by gene duplication.

Keywords: evolution of pathways; homology; protein evolution; sequence analyses; structure prediction

At present, about 10% of all enzymes with known tertiary structures contain at least one domain that has the β/α-barrel fold: a typically eight-stranded β-barrel surrounded by eight α-helices (Chothia, 1988; Farber & Petsko, 1990; Brändén, 1991; Farber, 1993). The well-studied family of β/α-barrels is mainly found in metabolic enzymes and extracellular glycohydrolases; only the recently identified extracellular plant protein narbonin is still in search of a function (Henning et al., 1992). In spite of their structural similarity, the β/α-barrels are apparently able to support diverse functional activities (for a recent collection, see Farber, 1993). This functional diversity and the poor sequence relationships among the different barrels motivated an intensive study of the evolution of their topology. Until now, the descent of the barrel is, however, still under debate and arguments for both convergent (Lesk et al., 1989; Brändén, 1991; Lindqvist et al., 1991) and divergent evolution of the β/α-barrels (Farber & Petsko, 1990; Wilmanns et al., 1991) have been presented. A few subfamilies that include different enzymatic activities have been assumed to be the result of divergent evolution (Farber & Petsko, 1990; Wilmanns et al., 1991; Scrutton, 1994). Nevertheless, it has been found to be very difficult to identify common signals at the sequence level (Pickett et al., 1992), although certain constraints exist (Wilmanns et al., 1991; Scheerlinck et al., 1992; Wilmanns & Eisenberg, 1993).

Many enzymes with a β/α-fold have a C-terminal phosphate-binding site in common that involves the region between strand 7 and helix 8 (Farber & Petsko, 1990; Brändén, 1991; Wilmanns

Reprint requests to: Peer Bork, EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany; e-mail: bork@embl-heidelberg.de.

Abbreviations: ORF, open reading frame; for gene product designations see Table 2.

et al., 1991). These functional constraints are reflected at the sequence level in a few enzymes with known three-dimensional structure and their homologues; the phosphate-binding site always belongs to the most conserved regions when comparing proteins across species.

A sequence family that includes different enzymatic activities, such as glycolate oxidase, L-lactate dehydrogenase, lactate-2-monooxidase, mandelate racemase, and cytochrome b_2, has recently been summarized (Scrutton, 1994). The members of this family (hereafter the GOX family) share an FMN-binding site, can be readily aligned, and have thus been proposed to be the result of divergent evolution. This is supported by two distantly related enzymes of the family that have very similar three-dimensional structures (Lindqvist et al., 1991): glycolate oxidase (Brookhaven Protein Data Bank code 1GOX) and cytochrome b_2 (1FCB).

The sequence variety of the GOX family prompted us to explore the conserved phosphate-binding part of the FMN site in order to find a connection to other β/α-barrels at the sequence level. Using a variety of database search methods (Koonin et al., 1994), we indeed found numerous protein sequences that resemble the phosphate-binding site of the GOX family, among them tryptophan synthase α subunit (TrpA) and indoleglycerolphosphate synthase (TrpC) with known three-dimensional structure (Fig. 1). The accumulation of β/α-barrels in several pathways and their detectable sequence similarity support a divergent evolution of at least a considerable fraction of the β/α-barrels.

Results and discussion

We show here the results of two standard database search methods, Blastp (Altschul et al., 1990) and Profilesearch (Gribskov et al., 1987). Other tools such as Fasta (Pearson & Lipman, 1988), Propat (Rohde & Bork, 1993), and Most (Tatusov et al., 1994) give similar results (data not shown). Sequence database searches with the members of the GOX family using Blastp indicate (1) a significant similarity to an ORF in *Erwinia herbicola* (probability of matching by chance is $P = 5.1\ 10^{-6}$, which is significant provided that segments biased in amino acid composition are filtered out [Altschul et al., 1994]) and (2) considerable but not yet significant sequence similarities to glutamate synthases ($P = 1.9\ 10^{-2}$) and IMPDHs ($P = 1.9\ 10^{-3}$). Whereas the *Erwinia* ORF has an overall sequence similarity to the GOX family, the glutamate synthases and IMPDHs only match the conserved phosphate-binding site. The NADP-dependent glutamate synthases contain iron-sulfur clusters and probably flavins and FMNs as prosthetic groups; a similarity to the FMN-binding site in cytochrome b_2 has already been reported (Sakakibara et al., 1991). Subsequent Blastp searches were carried out with the complete sequences of all IMPDHs, the related GMP reductases, and with the glutamate synthases. In particular, the IMPDHs have probabilities of matching by chance below $P = 10^{-2}$ to a couple of additional families (Fig. 2). Again, the matching regions only comprise the phosphate-binding site (Figs. 1, 3). In summary, Blastp always detects the same segments in a diverse set of database proteins. Although the similarities are not significant by pairwise matching, they are at least indicative of a homology. In addition, proteins from the other families displayed in Table 1 also matched with scores better than the random background of unrelated proteins (data not shown).

In order to use an independent method that takes advantage of the conservation within sequence families, iterative profile searches starting with the region around the phosphate-binding site in the GOX family were carried out (Table 1). In a first round, IMPDHs, related GMP reductases, but also glutamate synthases were identified well above the random background of other database proteins with significant Z scores (Gribskov et al., 1987) of 6.11 and higher (Table 1). The second search was based on a multiple alignment of the GOX family and the newly identified sequence families. This second iteration identified the respective region in ThiG, a third round revealed the motif in HisA, and subsequent searches successively collected the families that were already indicated by the Blastp iterations (Table 1; Fig. 2). Similar results, obtained by different sequence analysis methods, clearly point to a homology of all families shown in Figures 2 and 3.

Homologous proteins usually have at least a whole domain in common. In order to verify the sequence similarities and to support the assumption that all these families form β/α-barrels, we predicted the secondary structures independently for families for which a global alignment can be constructed (Table 2; Fig. 4). The method used (Rost & Sander, 1993) has an average accuracy of more than 70% and specifies regions with higher

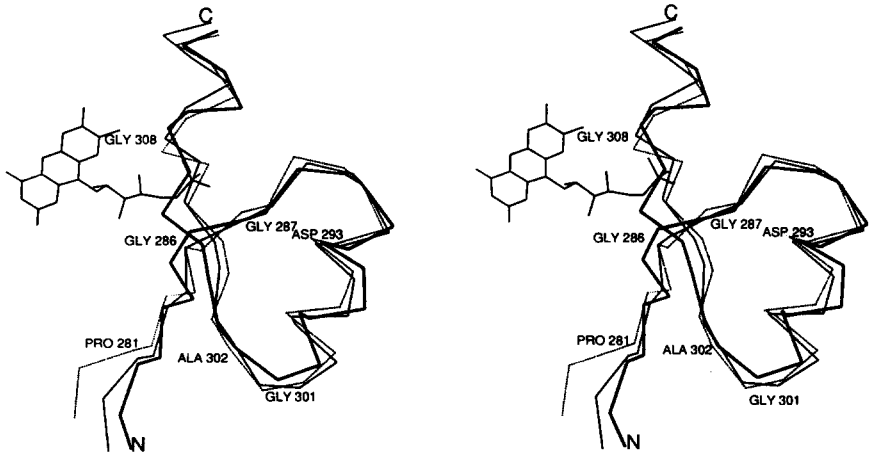

Fig. 1. Stereo view of superimposed phosphate-binding sites in different β/α-barrels of known structure that contain the sequence motif. The Cα-chains of the phosphate-binding motifs of GOX (Brookhaven Protein Data Bank code 1GOX; Lindqvist, 1989), TrpC (1PII; Wilmanns et al., 1992), and TrpA (1WSY; Hyde et al., 1988) are compared. The FMN with the terminal phosphate group was taken from the 1GOX data set (Lindqvist, 1989). The labeled residues and the numbering correspond to the 1GOX structure (GOX_SPIOL in the SWISS-PROT protein sequence database and in Fig. 3). These residues are conserved in numerous enzymes of this β/α-barrel sequence family (bold in Fig. 3).

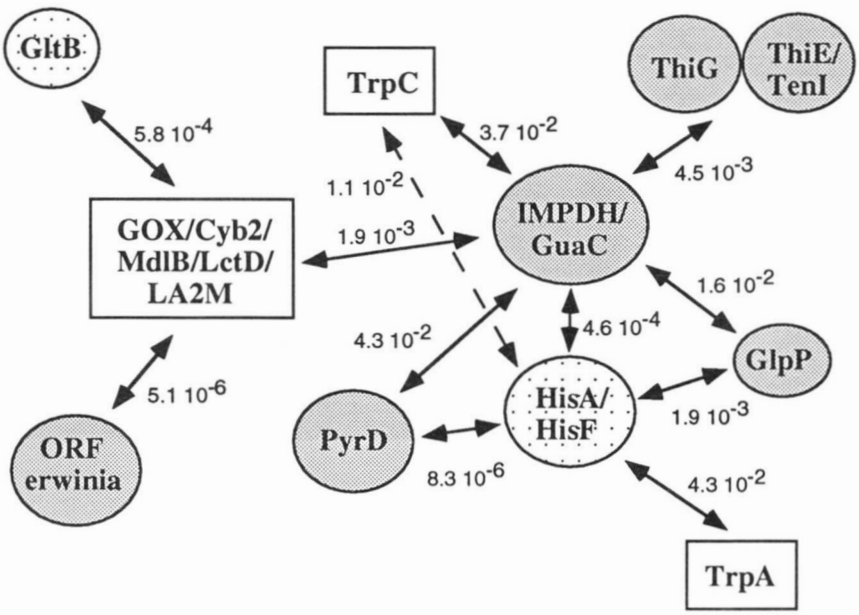

Fig. 2. Representation of sequence similarities between proteins containing the motif. The best similarities between sequence families as indicated by Blastp *P*-values are displayed. Families (circled) are called groups of sequences that have an overall similarity to each other and that have already been described. The names of the gene products are used to describe the families; if not available, capitals as used in the SWISS-PROT database (Bairoch & Boeckmann, 1993) abbreviate the respective proteins. Squares mark sequence families that contain at least one member with known three-dimensional structure. The shading indicates the degree of new information: dark, phosphate-binding motif and topology proposed in this work; light, either phosphate-binding site or topology have already been reported by others.

accuracy. Thus, the alternation of predicted β-strands and α-helices as shown in Figure 4 is highly supportive of a barrel-like topology. In all cases, the lengths of the putative domains that contain the sequence motif are sufficiently long to form a β/α-barrel. The motifs are always located near the C-terminus of the respective domains (Fig. 4; Table 2). The putative phosphate-binding site is among the most conserved regions in each family (data not shown). When mapping the motifs onto the predicted secondary structure (Fig. 4), the position of the motif within the sequences is in considerable agreement with a location between β-strand 7 and α-helix 8 (see Fig. 1).

It is not only coincidence that exactly the phosphate-binding sites of TrpA and TrpC were found by sequence searches alone (Figs. 2, 3). The motif is the structurally most similar region among GOX, TrpC, and TrpA, and also the phosphate can be superimposed well (Fig. 1). However, the phosphate group belongs to different compounds: FMN in GOX and probably (1-indol-3-YL)-glycerol-3-phosphate in TrpA and TrpC. A phosphate-containing heterocyclic ligand or substrate is apparently the only functional feature that all metabolic enzymes found by the sequence searches have in common; the overall functions of the detected enzymes vary considerably (Table 2). They are scattered in various metabolic pathways even though a few clusters in particular biochemical pathways that act on heterocyclic intermediates can be observed.

HisA and HisF are located next to each other in the operon for histidine biosynthesis. They catalyze the fourth and sixth step in histidine synthesis. HisA and HisF are 25% identical to each other and their homology has recently been described (Fani et al., 1994). Independently, Wilmanns and Eisenberg (1993) predicted a β/α-barrel for HisA based on contact profiles derived from known three-dimensional structures. As for this latter prediction, an independently developed technique was used; it provides further support of our findings.

IMPDH converts inosine-5′-phosphate into xanthosine-5′-phosphate as a major step in purine biosynthesis; the closely related GMP reductase catalyzes the formation of inosine-5′-phosphate from guanosine-5′-phosphate in a salvage pathway. Although in different pathways, IMPDH and GMP reductase catalyze consecutive steps in purine metabolism (Weber et al., 1992). Whereas the latter protein uses NADP as a cofactor, the former requires NADH for catalysis (Table 2). Dihydroorotate dehydrogenase (PyrD) also catalyzes an NADH-dependent step in nucleotide metabolism: the formation of orotate, a precursor in pyrimidine biosynthesis (see Nagy et al., 1992).

The genes of two *Escherichia coli* proteins, ThiG and ThiE, are located in the operon for the de novo biosynthesis of vitamin B1, thiamine pyrophosphate (Van der Horn et al., 1993). Both gene products are required for the synthesis of a thiazole, one of the precursors of thiamine pyrophosphate (Van der Horn et al., 1993). Two *Bacillus subtilis* proteins, one of unknown function (Fig. 1) and TenI, a negative regulator for the production of extracellular enzymes (Pang et al., 1991), have sequence similarity over the entire length to ThiE and can thus be grouped into this family. Although the biosynthesis of thiazole is not yet completely understood, ThiG and ThiE appear to catalyze two steps in this pathway that finally leads to thiamine pyrophosphate.

The three clusters of reactions catalyzed by β/α-barrels show that the accumulation of β/α-barrels found in tryptophan biosynthesis (Wilmanns et al., 1991; Wilmanns & Eisenberg, 1993) is apparently not an exception, but that formation of biochemical pathways by gene duplication has frequently occurred during molecular evolution. Some of the less-characterized proteins found by the described procedure might turn out to be enzymes in the future; for example, TenI has a clear overall similarity to ThiE of thiamine biosynthesis. We also predict an enzymatic function for GlpP, described as a transcription enhancer in the presence of glycerol-3-phosphate (Beijer et al., 1993).

Although we assume that all enzymes with the common sequence motif have evolved by divergent evolution, it remains unclear whether the present variety of β/α-barrels is derived from a common ancestor. Even within the superfamily presented here, there are some uncertainties about the spread of the domain. Re-

```
 gox_spiol  279 RIPVFLDGGVRRGTDVFKAL.ALGAAGVFIGRPVVFSLAAEGEAGVKK  P05414
cyb2_yeast  483 KLEVFVDGGVRRGTDVLKAL.CLGAKGVGLGRPFLYANSCYGRNGVEK  P00175
mdlb_psepu  297 GKPVLIDSGFRRGSDIVKAL.ALGAEAVLLGRATLYGLAARGETGVDE  P20932
lctd_ecoli  300 DIAILADSGIRNGLDVVRMI.ALGADTVLLGRAFLYALATAGQAGVAN  P33232
la2m_mycsm  314 DTPVLFDSGIRTGADVVKAL.AMGASAVGIGRPYAWGAALGGSKGIEH  P21795
orf5/erwhe  272 QMPLIASGGIKNGVDAAKAL.RLGACMVGQAAAVLGSAGVSTEKVIDH  M87280
glsn_medsa 1208 RTTLQTDGQLKTGRDVAIAA.LLGAEEYGFSTAPLITL...GCIMMRK  Q03460
gltb_ecoli 1092 KIRLQVDGGLKTGVDIIKAA.ILGAESFGFGTGPMVAL...GCKYLRI  P09831
glsf_maize 1190 RVVLRVDGGFRSGQDVLIAA.AMGADEYGFGSVAMIAT...GCVMARI  P23225
impl_human  358 GVPIIADGGIQTVGHVVKAL.ALGASTVMMGSLLAATTEAPGEYFFSD  P20837
 imp_leido  354 GVPCTADGGLRQVGDICKAL.AIGANCAMLGGMLSGTTETPGEYFFKG  P21620
 imp_acica  330 QIPLIADGGIPFSGDMAKAI.GAGASTIMVGSLLAGTEEAPGEVEFFQ  P31002
 imp_bacsu  335 GKTIIADGGIKFSGDITKAL.AAGGHAVMLGSLLAGTSESPGETEIYQ  P21879
guac_asclu  216 NGHVMSDGGCTNPGDVAKAF.GGGADFVMIGGLLAGHDQCGGEVVEKD  P27442
guac_ecoli  213 GGMIVSDGGCTTPGDVAKAF..ARADFVMLGGMLAGHEESGGRIVEEN  P15344
thig_ecoli  243 TVPVVVDAGIGVPSHAAQAL.EMGADAVLVNTAIAVA....DDPVNMA  P30139
teni_bacsu  148 SIPVIAIGGMT.PDRLRDVK.QAGADGIAVMSGIFSS....AEPLEAA  P25053
 orf/bacsu  187 SIPIVGIGGIT.IDNAAPVI.QAGADGVSMISAISQA....EDPESAA  X73124
thie_ecoli  158 DYPTVAIGGIS.LARAPAVI.ATGVGSIAVVSAITQA....ADWRLAT  P30137
pyrd_human  327 RVPIIGVGGVSSGQNALEKI.RAGASLVQLYTALPFW....GPPVVGK  Q02127
pyrd_arath  392 KIPLIGCGGVSSGEDAYKKI.RAGATLVQLYTGFAYG....GPAPSHK  P32746
pyrd_yeast  245 EIKVIGTGGIKSGKDAFEHL.LCGASMLQIGTELQKE....GVKIFER  P28272
pyrd_bacsu  235 NIPIIGMGGVQTAEDALEFL.LAGASAVAVGTANFVN....PFACPEI  P25996
trpa_metvo  211 KNKLYVGFGISNGEHAEKII.ENGADGVIVGSAFVDIIKEYGDSNETI  P14637
trpa_bacsu  198 TVPVAVGFGISNREQVIKMN.EIGSDGVVVGSALVRKIEELKDRLISA  P07601
 trp_yeast  207 DTPLAVGFGVSTREHFQSV..GSVADGVVIGSKIVTLC...GDAPEGK  P00931
trpa_ecoli  205 AAPPLQGFGISAPDQVKAAI.DAGAAGAISGSAIVKIIEQHINEPEKM  P00928
trpc_halvo  205 DVLLVAESGVQTVDDARRMR.EAGADALLVGTAIMD.....GDVRQNT  P18304
trpg_yeast  437 DVLLIALSGITTRDDAEKYK.KEGVHGFLVGEALMK.....STDVKKF  P00937
trpc_bacsu  202 ESLLVSESGIGSLEHLTFVK.EHGARAVLIGESLMR.....QTSQQKA  P03964
trpc_ecoli  208 NVTVISESGINTYAQVREL..SHFANGFLIGSALMA.....HDDLHAA  P00909
glpp_bacsu  146 GIPIFAGGFIRTEEDVEQAL.KAGAVAVTTSNTKLW.....KKYENFL  P30300
his6_ecoli  201 HVPLIASGGAGTMEHFLEAFRDADVDGALAASVFHK.....QIINIGE  P10373
his6_lacla  197 NIPVVASGGCGKISDIVEVFQNTRSDAALFASLFHY.....GEEQLMK  Q02133
his5_yeast  493 KIPVIASSGAGVPEHFEEAFLKTRADACLGAGMFHR.....GEFTVND  P33734
his4_metvo  191 NIPIIASGGVTTIEDLIEFK.KIGVAGVVVGSALYK.....NNFKLQD  P05325
his4_ecoli  196 QVAFQSSGGIGDIDDVAALR.GTGVRGVIVGRALLE.....GKFTVKE  P10371
his4_lacla  192 SLNVIISGGVKDNSDIQRAT.RSDFYGIIVGKAYYE.....GKINLEK  Q02131
consensus:      phPhhh GGhp  pDhhph    GA hhhhGphhh       G
```

Fig. 3. Multiple alignment of selected sequence motifs that comprise the phosphate-binding site. SWISS-PROT codes (first column) are used if available (underscore within the name). The position of the motif in the respective sequences is shown in the second column. Accession numbers are given in the last column. The bottom line shows a consensus: capitals, amino acids conserved in at least 60% of the sequences; h, mainly hydrophobic; p, mainly polar. Residues in boldface as described in Figure 1.

cently, Fani et al. (1994) reported an internal duplication in the HisA and HisF proteins. They proposed that an ancestor protein with a length of about 120 amino acids has been duplicated to form the ancestor of HisA and HisF, which are the result of a second duplication. Because the N- and C-terminus of some HisA have up to 33% sequence identity to each other and regions conserved in HisA and HisF match very well, this duplication might indeed have happened. The formation of a β/α-barrel by simple duplication of a smaller, probably planar, β/α-protein with a Rossmann fold strengthens arguments of convergent evolution. On the other hand, there is a likely scenario for divergent evolution that incorporates the formation of HisA or HisF by duplication: the spreading of the β/α-barrel family described here could have originated from a close rela-

Table 1. *Iterated profile searches starting with the GOX family*[a]

Profile	GOX		IMP		GltB		ThiG		HisF		TrpA		HisA		TrpC		PyrD		Proteins ranking next		
	Z	*T*	*Z*	*T*	*Z*	*T*	*Z*	*T*	*Z*	*T*	*Z*	*T*	*Z*	*T*	*Z*	*T*	*Z*	*T*		*Z*	*T*
GOX	23.55	36.03	7.58	17.52	6.20	16.35	3.91	33.82	5.45	13.27	5.20	13.26	3.76	11.20	3.04	10.58	3.16	11.59	YHBG_PSEPU	5.36	12.36
+IMP+GltB	14.84	16.30	18.34	19.56	7.69	11.42	6.78	9.80	5.47	8.41	5.55	8.64	5.45	8.26	4.55	7.77	5.30	8.64	YHBG_PSEPU	5.97	8.31
+ThiG	13.52	12.89	16.98	15.66	6.98	9.16	9.17	9.70	6.69	7.62	6.03	7.47	5.90	7.11	5.46	6.99	5.81	7.52	YHBG_PSEPU	5.92	6.76
+HisF	11.13	9.66	11.26	10.08	5.59	7.01	7.51	7.36	7.71	7.12	4.80	5.65	5.09	5.67	4.03	5.28	3.98	5.49	YHDG_ECOLI	4.58	5.76
+TrpA	10.50	8.52	10.67	8.90	4.90	6.01	7.38	6.69	7.34	6.36	5.67	5.68	5.27	5.31	4.10	4.82	3.87	4.99	YHDG_ECOLI	4.71	5.36
+HisA	10.15	7.26	9.67	7.30	5.73	5.61	7.67	6.04	6.70	5.09	5.08	4.72	5.32	4.74	4.89	4.62	4.05	4.44	Y19K_BACSU	4.34	3.93
+TrpC	9.06	6.89	9.44	7.23	4.86	5.25	6.93	5.78	6.22	5.25	5.94	5.23	6.97	5.46	6.39	5.31	5.26	5.02	YKK_CAEEL	4.98	4.25
+PyrD	8.52	6.51	8.84	6.86	3.66	4.59	7.13	5.69	5.53	4.68	5.39	4.80	6.61	5.18	5.79	4.78	5.56	4.97	YKK_CAEEL	5.38	4.19

[a] Only the region corresponding to the phosphate-binding site was used as input for the set of aligned sequences. *Z* (*Z*-scores) and *T* (total scores) are given for each protein. Because *Z*-scores (Gribskov et al., 1987) high-weight short sequences, and because β/α-barrels structurally characterized so far have at least a length of about 200 amino acids, hits with a length shorter than 150 amino acids were neglected. Only the best scoring new sequence family (indicated by underlines) has been added to the multiple alignment per iteration. Exceptions were the glutamate synthases, which scored nearly as high as the IMPDHs in the first iteration and which are already known to resemble an FMN-binding site similar to cytochrome b_2 (Sakakibara et al., 1991). Proteins with the best ranking that have not been detected by Blastp (last columns) are all not well characterized and might also form β/α-barrels.

tive of HisA/HisF. Because the molecular clocks and functional constraints are different for each protein, the duplication might have been blurred in the other protein sequences of the described family during evolution.

Materials and methods

The protein sequences were retrieved from public sequence databases such as SWISS-PROT, PIR, EMBL, and Genbank; several sequence analysis methods were used as summarized previously (Koonin et al., 1994). Initial database searches were performed using the Blast series of programs (Altschul et al., 1990, 1994). The BLOSUM 62 matrix (Henikoff & Henikoff, 1993) was applied. Filters for amino acid composition bias allowed a relatively sensitive scan (Altschul et al., 1994). Putative relatives in databases with a *P*-value below 0.1 were studied in detail. The candidate sequences were extracted and subjected to multiple alignment procedures (Higgins et al., 1992). The most conserved regions were identified and characterized using the program

Table 2. *Summary of detected proteins containing the putative phosphate-binding site*

Protein	Gene product	EC number	Length	Pathway	Ligands containing PO_4^{3-}
Glycolate oxidase	GOX	1.1.3.15	369 (*Spinacia oleracea*)	Several	FMN
Cytochrome b_2	Cyb2	1.1.2.3	591 (*Saccharomyces cerevisiae*)	Carbohydrate metabolism	FMN
L-lactate dehydrogenase	LctD	1.1.2.3	396 (*Escherichia coli*)	Carbohydrate metabolism	FMN
L-lactate-2-monooxidase	LA2M	1.13.12.4	393 (*Mycobacterium smegmatis*)	Carbohydrate metabolism	FMN
Mandelate dehydrogenase	MdlB	1.-.-.-	393 (*Pseudomonas putida*)	Mandelate synthesis	FMN
IMP dehydrogenase	IMPDH	1.1.1.205	511 (*E. coli*)	Purine synthesis	Substrate, NAD^+
GMP reductase	GuaC	1.6.6.8	346 (*E. coli*)	Purine metabolism	Substrate, $NADP^+$
Dihydroorotate dehydrogenase	PyrD	1.3.3.1	336 (*E. coli*)	Pyrimidine synthesis	FAD^+
Glutamate synthase	GltB	1.4.1.13	1514 (*E. coli*)	Several	FMN, $NADP^+$
Glutamate synthase	GLSF	1.4.7.1	1616 (*Zea mays*)	Several	FMN
Glutamate synthase	GLSN	1.4.1.14	2194 (*Medicago sativa*)	Several	FMN, NAD^+
P-ribosylformimino-5-aminoimidazol carboxamide ribotide isomerase	HisA	5.3.1.16	245 (*E. coli*)	Histidine synthesis	Substrate
Cyclase	HisF		258 (*E. coli*)	Histidine synthesis	Substrate
ThiG protein (thiazol synthesis)	ThiG		324 (*E. coli*)	Thiamine synthesis	Substrate
ThiE protein (thiazol synthesis)	ThiE		211 (*E. coli*)	Thiamine synthesis	Substrate
Putative regulatory protein TenI	TenI		205 (*Bacillus subtilis*)	?	?
Putative regulatory protein GlpP	GlpP		192 (*B. subtilis*)	?	?
Tryptophan synthase α-chain	TrpA	4.2.1.20	268 (*E. coli*)	Tryptophan synthesis	Substrate
Indoleglycerolphosphate synthase	TrpC	4.1.1.48	256 (1st domain, *E. coli*)	Tryptophan synthesis	Substrate

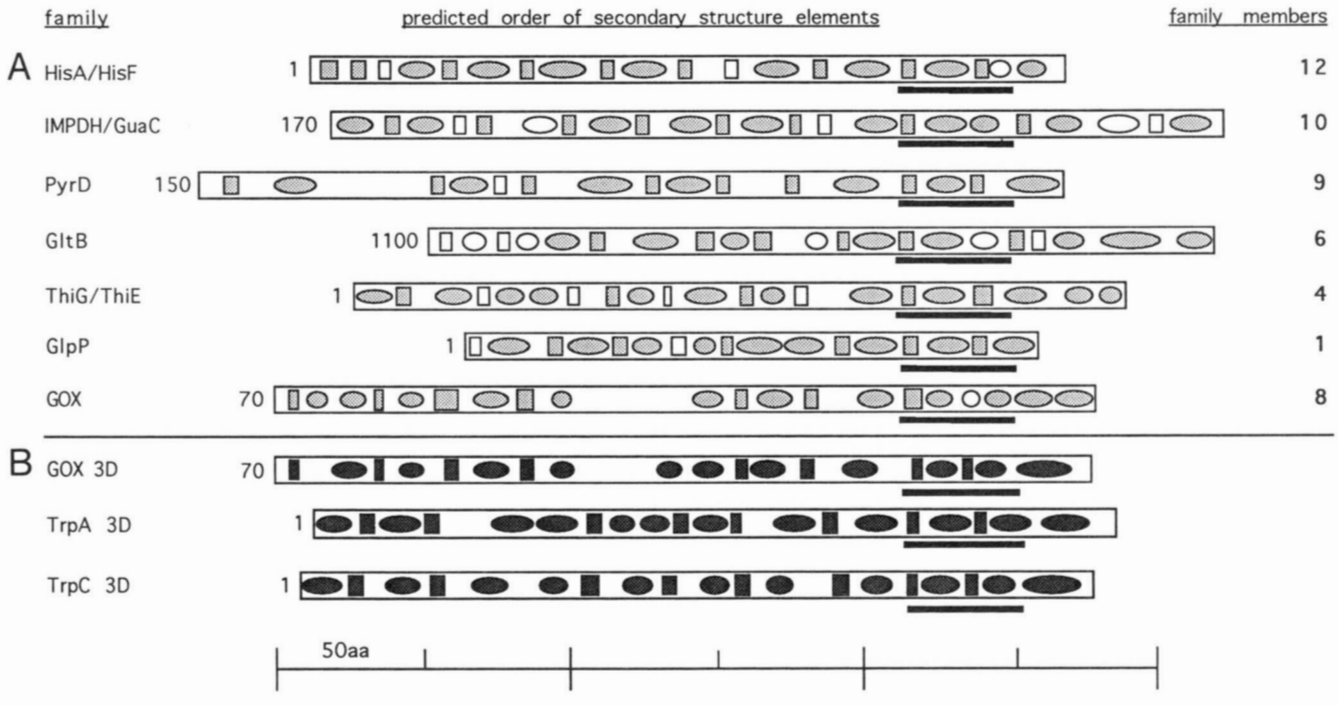

Fig. 4. Summary of secondary structure predictions for different enzyme families that contain the putative phosphate-binding site (**A**) compared with three known β/α structures (**B**). The number preceding the secondary structure elements denotes the beginning of the putative domains within the proteins. The number of aligned family members used for secondary structure prediction is shown in the last column. The location of the motif within the domain is indicated by thick black bars. Squares symbolize predicted β-strands, ovals are predicted α-helices. Shaded secondary structure elements are predicted with an expected accuracy better than 82% (Rost & Sander, 1993); dark-shaded ones are taken from the structures.

MACAW (Schuler et al., 1990). Furthermore, region underlined in Figure 4 was subjected to iterative profile searches (Gribskov et al., 1987); the results of a database search were used for improving the multiple alignment, which was then used for the next round of database searches. The multiple alignments of subfamilies with clear overall similarity were used as input for the secondary structure predictions (Rost & Sander, 1993). The WHATIF program (Vriend, 1990) was used for automatic superposition of the known structures of this family and for preparation of Figure 1.

Acknowledgments

We are indebted to Liisa Holm for information on the three-dimensional structures of β/α-barrels and to Eugene Koonin, Christos Ouzounis, and Rebecca Wade for helpful discussions and critical reading of the manuscript.

References

Altschul SF, Boguski MS, Gish W, Wootton JC. 1994. Issues in searching molecular sequence databases. *Nature Genet 6*:119–163.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol 215*:403–410.

Bairoch A, Boeckmann B. 1993. The SWISS-PROT protein sequence data bank, recent developments. *Nucleic Acids Res 21*:3093–3096.

Beijer L, Nilsson RP, Holmberg C, Rutberg L. 1993. The glpP and glpF genes of the glycerol regulon in *Bacillus subtilis*. *J Gen Microbiol 139*:349–359.

Brändén CI. 1991. The TIM barrel—The most frequently occurring folding motif in proteins. *Curr Opin Struct Biol 1*:978–983.

Chothia C. 1988. Protein structure: 14th barrel rolls out. *Nature 333*:598–599.

Fani R, Liò P, Chiarelli I, Bazzicalupo M. 1994. The evolution of histidine biosynthetic genes in prokaryotes: A common ancestor for *hisA* and *hisF* genes. *J Mol Evol 38*:489–495.

Farber GK. 1993. An α/β-barrel full of evolutionary trouble. *Curr Opin Struct Biol 3*:409–412.

Farber GK, Petsko GA. 1990. The evolution of α/β barrel enzymes. *Trends Biochem Sci 15*:228–234.

Gribskov M, McLachlan AD, Eisenberg D. 1987. Profile analysis: Detection of distantly related proteins. *Proc Natl Acad Sci USA 84*:4355–4358.

Henikoff S, Henikoff J. 1993. Performance evaluation of amino acid substitution matrices. *Proteins Struct Funct Genet 17*:49–61.

Henning M, Schlesier B, Dauter Z, Pfeffer S, Betzel C, Höhne WE, Wilson KS. 1992. A TIM barrel protein without enzymatic activity? *FEBS Lett 306*:80–84.

Higgins DG, Bleasby AJ, Fuchs R. 1992. CLUSTAL V: Improved software for multiple sequence alignment. *Comput Appl Biosci 8*:189–191.

Hyde CC, Ahmed SA, Padlan EA, Miles EW, Davies DR. 1988. Three-dimensional structure of the tryptophan synthase alpha2 beta2 multienzyme complex from *Salmonella typhimurium*. *J Biol Chem 263*: 17857–17871.

Koonin EV, Bork P, Sander C. 1994. Yeast chromosome III: New gene functions. *EMBO J 13*:493–503.

Lesk AM, Brändén CI, Chothia C. 1989. Structural principles of α/β barrel proteins. *Proteins Struct Funct Genet 5*:139–148.

Lindqvist Y. 1989. Refined structure of spinach glycolate oxidase at 2 Ångstroms resolution. *J Mol Biol 209*:151–166.

Lindqvist Y, Brändén CI, Mathews FS, Lederer F. 1991. Spinach glycolate oxidase and yeast flavocytochrome b_2 are structurally homologous and evolutionarily related enzymes with distinctly different function and flavin mononucleotide binding. *J Biol Chem 266*:3198–3207.

Nagy M, Lacroute F, Thomas D. 1992. Divergent evolution of pyrimidine biosynthesis between anaerobic and aerobic yeasts. *Proc Natl Acad Sci USA 89*:8966–8970.

Pang AS, Nathoo S, Wong SL. 1991. Cloning and characterization of a pair of novel genes that regulate production of extracellular enzymes in *Bacillus subtilis*. *J Bacteriol 173*:46–54.

Pearson WR, Lipman DJ. 1988. Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA 85*:2444–2448.

Pickett SD, Saqi MAS, Sternberg MJE. 1992. Evaluation of sequence tem-

plate method for protein structure prediction: Discrimination of the (β/α)8-barrel fold. *J Mol Biol 228*:170–187.

Rohde K, Bork P. 1993. A fast, sensitive pattern-matching approach for protein sequences. *Comput Appl Biosci 9*:183–189.

Rost B, Sander C. 1993. Prediction of protein secondary structure at better than 70%. *J Mol Biol 232*:584–599.

Sakakibara H, Watanabe M, Hase T, Sugiyama T. 1991. Molecular cloning and characterization of complementary DNA encoding for ferredoxin-dependent glutamate synthase in maize leaf. *J Biol Chem 266*:2028–2035.

Scheerlinck JPY, Lasters I, Claessens M, De Maeyer M, Pio F, Delhaise P, Wodak SJ. 1992. Recurrent α/β loop structures in TIM barrel motifs show a distinct pattern of conserved structural features. *Proteins Struct Funct Genet 12*:299–313.

Schuler GD, Altschul SF, Lipman D. 1990. A workbench for multiple sequence alignment construction and analysis. *Proteins Struct Funct Genet 9*:180–190.

Scrutton NS. 1994. α/β barrel evolution and the modular assembly of enzymes: Emerging trends in flavin oxidase/dehydrogenase family. *Bioessays 16*:115–122.

Tatusov RL, Altschul SF, Koonin EV. 1994. Detection of distantly related proteins by iterative motif search. *Proc Natl Acad Sci USA 91*:12091–12095.

Van der Horn PB, Backstrom AD, Stewart V, Begley TP. 1993. Structural genes for the thiamine biosynthetic enzymes (thiCEFGH) in *E. coli* K-12. *J Bacteriol 175*:982–992.

Vriend G. 1990. WHATIF, a molecular modelling and drug design program. *J Mol Graphics 8*:52–56.

Weber G, Nakamura H, Natsumeda Y, Szekeres T, Nagai M. 1992. Regulation of GTP biosynthesis. *Adv Enzyme Regul 32*:57–69.

Wilmanns M, Eisenberg D. 1993. Three-dimensional profiles from residue-pair preferences: Identification of sequences with β/α-barrel fold. *Proc Natl Acad Sci USA 90*:1379–1383.

Wilmanns M, Hyde CC, Davies DR, Kirschner K, Jansonius JN. 1991. Structural conservation in parallel β/α-barrel enzymes that catalyze three sequential reactions in the pathway of tryptophan biosynthesis. *Biochemistry 30*:9161–9169.

Wilmanns M, Priestle JP, Niermann T, Jansonius JN. 1992. Three-dimensional structure of the bifunctional enzyme phosphoribosylanthranilate isomerase:indoleglycerolphosphate synthase from *Escherichia coli* refined at 2.0 Å resolution. *J Mol Biol 223*:477–507.