

Structural Conservation in Parallel β/α -Barrel Enzymes That Catalyze Three Sequential Reactions in the Pathway of Tryptophan Biosynthesis^{†,‡}

Matthias Wilmanns,^{§,||} C. Craig Hyde,[⊥] David R. Davies,[⊥] Kasper Kirschner,[▽] and Johan N. Jansonius^{*,§}

Departments of Structural Biology and Biophysical Chemistry, Biocentre, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland, and Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892

Received December 3, 1990; Revised Manuscript Received June 6, 1991

ABSTRACT: Three successive steps in tryptophan biosynthesis are catalyzed by single-domain proteins, each folded as a parallel β/α -barrel, as observed in the crystal structures of the bienzyme (phosphoribosyl)-anthranilate isomerase:indoleglycerolphosphate synthase from *Escherichia coli* [Priestle, J. P., Grütter, M. G., White, J. L., Vincent, M. G., Kania, M., Wilson, E., Jardetzky, T. S., Kirschner, K., & Jansonius, J. N. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 5690-5694] and the α -subunit of the tetrameric bienzyme tryptophan synthase from *Salmonella typhimurium* [Hyde, C. C., Ahmed, S. A., Padlan, E. A., Miles, E. W., & Davies, D. R. (1988) *J. Biol. Chem.* 263, 17857-17871]. Recent refinement of the crystal structures of these enzymes at atomic resolution revealed that they contain a common phosphate group binding site in the β/α -barrel, created by residues of the loop between β -strand 7 and α -helix 7 and the N-terminus of an additional helix 8'. The close similarities of their β/α -barrel structures permitted the alignment of 50-75% of their respective amino acid sequences. Considerable sequence similarity was detected in the regions spanning the phosphate binding sites, whereas the percentage of identical residues was barely significant for the remaining parts of the enzymes. These observations suggest divergent evolution of these three β/α -barrel enzymes involved in tryptophan biosynthesis. The same phosphate binding site was also observed in six other β/α -barrel enzymes that are functionally unrelated to those involved in tryptophan biosynthesis: triosephosphate isomerase, ribulose-1,5-bisphosphate carboxylase/oxygenase, glycolate oxidase, flavocytochrome *b*₂, trimethylamine dehydrogenase, and tentatively also fructosebisphosphate aldolase. This observation indicates that a much larger number of β/α -barrel enzymes, not restricted to the three tryptophan biosynthesis enzymes, might be evolutionarily related. Whether these β/α -barrel enzymes or domains of enzymes evolved from the same ancestor or, in part, from each other remains open.

Examination of the rapidly increasing number of three-dimensional protein structures now available at atomic resolution reveals the existence of families of proteins with similar folding patterns (Levitt & Chothia, 1976; Finkelstein & Ptitsyn, 1988). One of these is the parallel β/α -barrel fold, first observed in triosephosphate isomerase (Banner et al., 1975) and currently exemplified by crystal structures of 18 different enzymes (Chothia, 1988; Farber & Petsko, 1990; Rouvinen et al., 1990). Most of these β/α -barrel structures consist of a central eight-stranded parallel β -barrel, surrounded by eight α -helices that connect consecutive β -strands. Two exceptions to this general folding pattern have been observed: in enolase from yeast one β -strand is antiparallel to the remaining central β -barrel strands (Lebioda et al., 1989), and in cellobiohydrolase II from *Trichoderma reesei* the central β -barrel consists of seven parallel β -strands. The first and the last β -strand, which close the central β -barrel, are connected by only one hydrogen bond and are rather separated, whereas the hydrogen-bonding network in the remainder of the β -barrel of cellobiohydrolase is as extensive as those observed in other

β/α -barrel enzymes (Rouvinen et al., 1990).

Despite an increasing understanding of the structural organization of parallel β/α -barrel enzymes (Lasters et al., 1988; Lesk et al., 1989) a challenging question remains unanswered: did this family arise by divergent evolution from a common ancestor? Arguments for evolutionary relationships have been discussed in a recent review by Farber and Petsko (1990). They divided the known β/α -barrel enzymes into four structural families. Members of these families either catalyze similar reactions or belong to common metabolic pathways, or both. One of the proposed families comprises (phosphoribosyl)anthranilate isomerase (PRAI),¹ indoleglycerolphosphate synthase (IGPS), the α -subunit of tryptophan synthase [TRPS(α)] and triosephosphate isomerase (TIM). Nevertheless, Farber and Petsko (1990) stated that their simple structural arguments cannot be used as "conclusive proof of divergent as opposed to convergent evolution".

Here we compare the high-resolution crystal structures of three complex enzymes that catalyze three of the last four steps in tryptophan biosynthesis: PRAI:IGPS from *Escherichia coli* (Priestle et al., 1987) and the α -subunit of TRPS from *Salmonella typhimurium* (Hyde et al., 1988). PRAI:IGPS is a

[†] This work was supported in part by a grant from the Swiss National Science Foundation (31-25712.88 to J.N.J.).

[‡] The atomic coordinates of PRA isomerase:IGP synthase have been deposited in the Brookhaven Protein Data Bank.

[§] Department of Structural Biology, University of Basel.

^{||} Present address: Molecular Biology Institute, Department of Chemistry, University of California—Los Angeles, 405 Hilgard Ave., Los Angeles, CA 90024-1570.

[⊥] Laboratory of Molecular Biology, NIH.

[▽] Department of Biophysical Chemistry, University of Basel.

¹ Abbreviations: PRAI, *N*-(5'-phosphoribosyl)anthranilate isomerase; IGPS, indole-3-glycerolphosphate synthase; TRPS, tryptophan synthase; TRPS(α), α -subunit of tryptophan synthase; TIM, triosephosphate isomerase; RUBISCO, ribulose-1,5-bisphosphate carboxylase/oxygenase; GAO, glycolate oxidase; FCB2, flavocytochrome *b*₂; TMDH, trimethylamine dehydrogenase; FALD, fructosebisphosphate aldolase; IPP, indolepropanol phosphate; rms, root mean square.

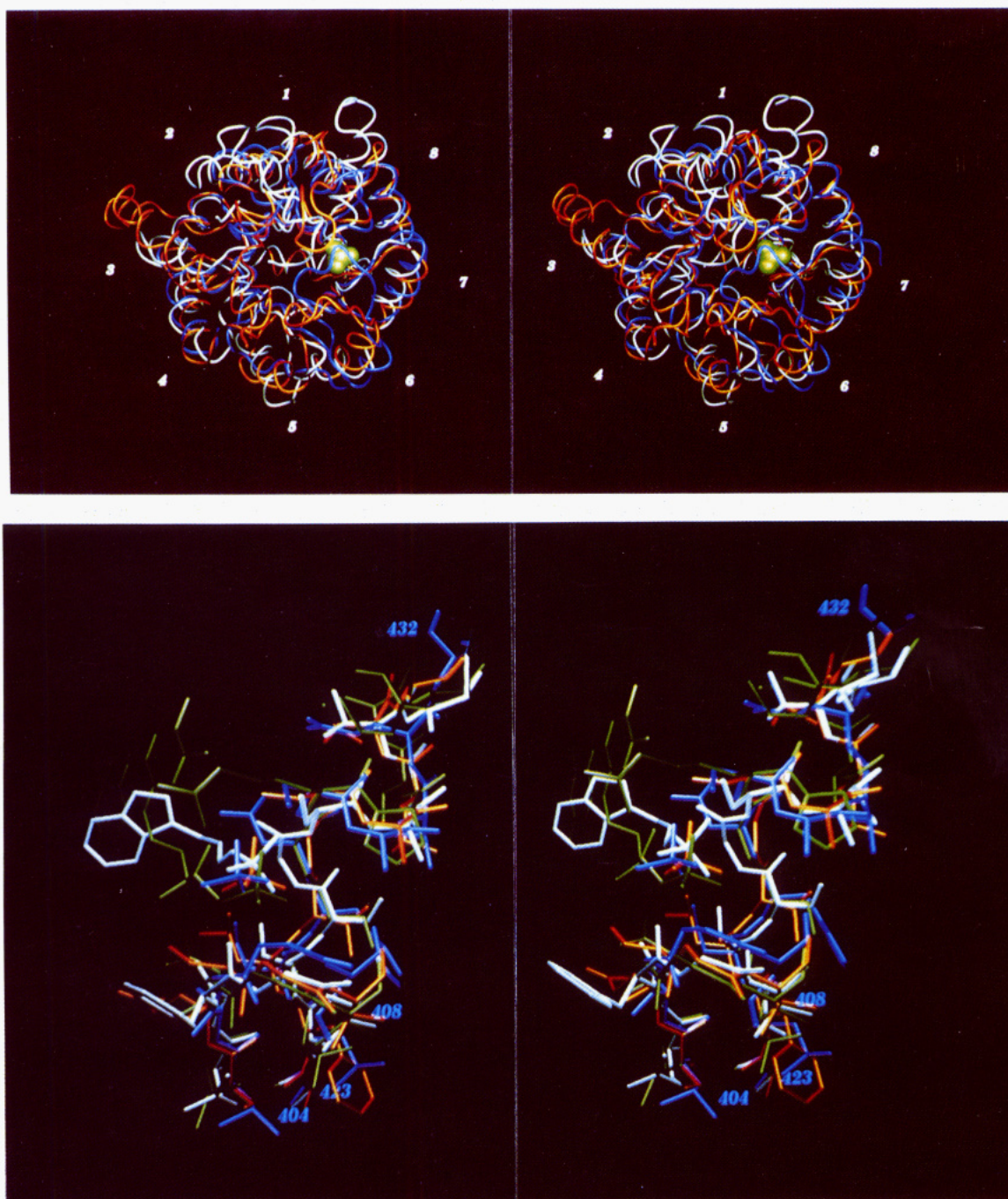


FIGURE 1: (Top) Superimposed smoothed C_{α} backbones of PRAI (blue), IGPS (red), and TRPS(α) (white). The stereo view is along the β/α -barrel axis with the C-terminal side facing the viewer. The positions of helices 1–8 are indicated. The van der Waals surface of one representative phosphate (IGPS) shows the position of the phosphate binding site. (Bottom) Superposition of the phosphate binding sites of PRAI, IGPS, TRPS(α), and GAO. The superposition is based on the main-chain atoms of the 15 residues aligned in Figure 5 and the phosphorus atoms of the phosphate ions (PRAI, IGPS) or groups [of IPP in TRPS(α) and of FMN in GAO]. The phosphate group binding site of GAO is shown in green; the color codes of the three other enzymes are as in the top panel. The bound phosphate ions of PRAI and IGPS (in the middle of this panel), the inhibitor IPP [TRPS(α)], and the cofactor FMN (GAO) in the upper left of this panel have the color of the corresponding protein moieties. The numbers relate to the PRAI sequence from *E. coli* (see Figure 5).

monomeric bienzyme that consists of two parallel β/α -barrel domains. *N*-(5'-phosphoribosyl)anthranilate is isomerized to 1-[(*o*-carboxyphenyl)amino]-1-deoxyribulose 5-phosphate by PRAI and is then converted to indole-3-glycerol phosphate in a ring-closure reaction, catalyzed by IGPS. TRPS is a tetrameric ($\alpha\beta\beta\alpha$) bienzyme complex. The α -subunit [TRPS(α)] is also a parallel β/α -barrel and catalyzes the penultimate step of tryptophan biosynthesis, in which indole-3-glycerol phosphate is cleaved to indole and glyceraldehyde 3-phosphate. The pyridoxal phosphate dependent β -subunit of tryptophan synthase, which catalyzes the replacement of the β -hydroxyl group

of L-serine by indole to produce L-tryptophan, has an unrelated fold. The crystal structure of PRAI:IGPS has now been refined to an *R*-factor of 17.2% at 2.0-Å resolution (M. Wilmanns, unpublished results). The TRPS native complex has been refined to an *R*-factor of about 19% at 2.3-Å resolution and a complex with the inhibitor indolepropanol phosphate has been partially refined to an *R*-factor of 20.1% at 2.42-Å resolution (C. C. Hyde, unpublished results). Details of the refined crystal structures will be published elsewhere.

Despite the qualitative similarity in folds, no significant amino acid sequence similarity between these three enzymes

was observed with conventional sequence alignment algorithms (programs GAP and BESTFIT, GCG program collection). The amino acid sequence identities, obtained from pairwise comparisons, range between 15% and 20%, depending on the choice of the values for the gap weight and gap length weight parameters. Taking into account widely varying propensities for different residues within a particular secondary structure (Chou & Fasman, 1974; Garnier et al., 1978) and even within three-dimensional structural microenvironments (Taylor, 1986), these amino acid sequence comparisons do not provide conclusive evidence for evolutionary relationships between these three enzymes.

In this paper we describe how the analysis of the superimposed three-dimensional structures of PRAI, IGPS, and TRPS(α) led to the discovery of an equivalent phosphate binding site. The structural similarity was also used to align both the corresponding amino acid and nucleotide sequences. The conservation of the phosphate binding site suggests that these enzymes are evolutionarily related.

EXPERIMENTAL PROCEDURES

Global Superposition of Crystal Structures. The three-dimensional structure of the bienzyme PRAI:IGPS was divided into two domains corresponding to IGPS (residues 1–255) and PRAI (residues 256–452). The three β/α -barrel structures were superimposed in pairs by using the program ALIGN (Satow et al., 1986) that iteratively minimizes the root mean square (rms) distance between corresponding atoms in two atomic coordinate sets while simultaneously and automatically assigning and optimizing the number of common atom pairs. Alignments were made on the basis of either all main-chain atom positions or only C_α atoms, thus ignoring amino acid sequence information. In order to obtain optimal superpositions, it was necessary to remove residues 1–47 in IGPS and residues 1–16 in TRPS(α).

Sequence Alignments Based on Structural Superposition. Amino acid sequence alignments between the three enzymes from *E. coli* were carried out only for peptide segments that had evidently similar secondary structure or conformation. Maximum distances between equivalent atoms of superimposed structures were restricted to 3 Å. After inspection of the superimposed β/α -barrel structures with the help of computer graphics, residues 340–351 of PRAI were manually shifted by one residue relative to the corresponding IGPS and TRPS(α) sequences in order to improve the alignment in this particular region. In the case of TRPS(α), only the structure of the enzyme from *S. typhimurium* is known (Hyde et al., 1988). The amino acid sequence from the highly homologous enzyme from *E. coli* (85% identical residues) was used for comparison with PRAI:IGPS from *E. coli*.

The GenBank (release 65.0) was searched for all available *trpA*, *trpC*, and *trpF* genes, which code for TRPS(α), IGPS, and PRAI, respectively. DNA sequences were omitted from further analysis if their translated amino acid sequences shared less than 30% identity with the *E. coli* sequence of the same enzyme or if the coding regions were interrupted by introns. Seven *trpF* sequences (*E. coli*, *S. typhimurium*, *Acinetobacter calcoaceticus*, *Caulobacter crescentus*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, and *Aspergillus niger*), 9 *trpC* sequences (*E. coli*, *S. typhimurium*, *Pseudomonas putida*, *A. calcoaceticus*, *B. subtilis*, *S. cerevisiae*, *Aspergillus nidulans*, *A. niger*, and *Halobacterium volcanii*) and 11 *trpA* sequences (*E. coli*, *S. typhimurium*, *P. putida*, *Pseudomonas aeruginosa*, *C. crescentus*, *B. subtilis*, *Methanococcus voltae*, *Klebsiella aerogenes*, *Thermus thermophilus*, *S. cerevisiae*, and *H. volcanii*) were chosen for pairwise comparison in all possible

combinations. The alignments were again based on the superimposed structures of PRAI, IGPS, and TRPS(α). The comparisons were carried out with an automated procedure (written by M. Wilmanns).

Structural Superposition of the Central β -Barrels and Their Cyclically Permuted Forms. In order to test the degree of structural conservation within the central β -barrel of each enzyme, cyclically permuted forms were constructed and aligned with the native forms of the respective other enzymes. The observation that these β -barrels contain residue layers perpendicular to the β -barrel axis (Lesk et al., 1989) helped to define four central layers (indicated in Figure 4). Each layer consists of eight residues, one from each parallel β -strand. Measurements of distances between equivalent main-chain atoms in opposite β -strands within each layer (β -strands 1–5, 2–6, 3–7, and 4–8), reveal that the minimum cross section perpendicular to the β -barrel is located between two adjacent residue layers in all three β/α -barrel structures under investigation rather than being associated with one distinct central residue layer (Lesk et al., 1989).

For each 8-fold β -barrel, a coordinate set of 32 residues was derived from the four residues of all eight β -strands that contribute to the four central layers of each central β -barrel. The rms deviations for all main-chain atoms between β -barrel coordinates and the superimposed cyclically permuted forms, which result from rotations in steps of about 45° around the β -barrel axis, were calculated by using the program LSQKAB (CCP program collection).

Interactive Graphical Analysis. The superimposed coordinate sets were inspected on an Evans & Sutherland PS 330 color graphic display system using the software FRODO (Jones, 1978). Color photographs were obtained from an IRIS 4D 120GTX workstation (Silicon Graphics). For this purpose, graphics software was developed at the Biocentre, Basel, Switzerland (C. Henn, unpublished results).

RESULTS

Comparison of the Overall β/α -Barrel Structures. The rms values for all superimposed main-chain atoms are 2.04 Å for the pair PRAI/IGPS (678 atom pairs), 2.42 Å for the pair IGPS/TRPS(α) (717 atom pairs), and 2.60 Å for the pair PRAI/TRPS(α) (676 atom pairs). The alternative use of only C_α atoms gave comparable results. The overall level of structural conservation between the three enzymes is immediately evident when the smoothed backbone structures are superimposed and inspected with computer graphics (Figure 1, top). The eight strands of the central barrels align particularly well. Some helices are found out of register by small relative translations and rotations, in particular α_4 , α_5 , and α_6 (Figure 1, top). Even greater variability is found in the loops that interconnect the strands of the central β -barrel and outer helices.

The most remarkable differences between the structures involve additional peptide segments beyond the central β -barrel. TRPS(α) and IGPS both possess extra N-terminal extensions. In TRPS(α), the 14 N-terminal residues form an α -helix (helix "0") that straddles the N-terminal side of the central β -barrel. In IGPS, 47 N-terminal residues first form a long helix (helix "0", residues 4–22), which lies across the C-terminal side of the central β -barrel, and then a long loop that wraps around the β/α -barrel. In the bienzyme from *E. coli*, the PRAI domain does not contain an N-terminal section preceding the β/α -barrel. In TRPS(α) a long insert is found in the loop following β -strand 2. This 26-residue segment contains an additional α -helix (helix 2') that interacts with the TRPS(β) subunit and provides an active-site residue (Asp

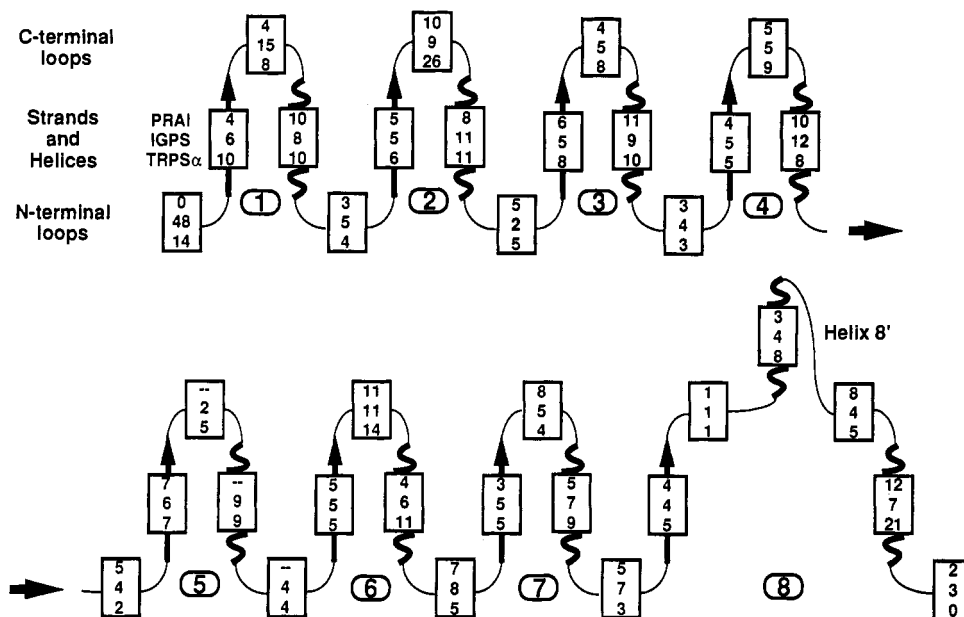


FIGURE 2: Comparison of the lengths of secondary structural elements and loops in PRAI, IGPS, and TRPS(α). The three enzymes comprise eight β/α units (labeled 1–8). The alternating strands (indicated by arrows) and helices are connected by loops that precede each strand (“N-terminal loops”) and loops that follow each strand (“C-terminal loops”). The length of each secondary structural element and loop, as determined with the program DSSP (Kabsch & Sander, 1983), is indicated in a box for PRAI (top), IGPS (middle), and TRPS(α) (bottom). In the boxes of helix 5 and of the loops preceding and following this helix no numbers are given for PRAI, because this helix is replaced in that structure by a long loop (14 residues) that connects β -strands 5 and 6. The three boxes between β -strand 8 and helix 8 represent the additional helix 8’ (upper box) and the two loops that connect this helix with adjacent secondary structural elements.

60; Nagata et al., 1989). One oddity in PRAI from *E. coli* is the replacement of helix 5 by a peptide segment in a partially extended conformation.

In Figure 2, the lengths of the strands, helices, and loops are compared schematically using the assignments of secondary structure determined with the program DSSP (Kabsch & Sander, 1983). In general, the lengths of the strands of the central β -barrels agree well, whereas somewhat more variability is observed in the lengths of the helices, in particular helices 6 and 8. All loops at the N-terminal side of the central β -barrels are less than eight residues long. Some loops as long as 26 residues are found at the C-terminal side of the central β -barrels.

The active sites of all β/α -barrel enzymes known so far are located at the C-termini of the central β -strands and involve residues from the loops connecting the β -strands and α -helices (Farber & Petsko, 1990). The greater variation in the lengths of some C-terminal loops is assumed to be a consequence of the different catalytic functions of the respective β/α -barrel enzymes. Interestingly, the loop following β -strand 6 is longer than 10 residues in all three enzymes and appears to be very flexible in the crystal structures of PRAI and TRPS(α). We postulate that this loop is involved in binding the corresponding substrates in the active sites by undergoing conformational rearrangement upon substrate binding.

Common Phosphate Binding Site. Two bound phosphate ions have been detected in the high-resolution structure of PRAI:IGPS. They are located in the active sites in equivalent positions at the C-terminal side of the two β/α -barrels (Priestle et al., 1987) and are enclosed between two neighboring loops in a strikingly similar way (Figure 3). These phosphate ions are hydrogen bonded (i) to the peptidyl amido groups of both the N-terminal residues of the short helix 8’ and the preceding loop and (ii) to the peptidyl amido groups of the first residues of the loop that connects β -strand 7 with helix 7. Helix 8’ comprises residues 237–240 in IGPS and residues 428–430 in PRAI, where it has the conformation of a short 3_{10} -helix. In the TRPS(α) crystal structure, the phosphate moiety of

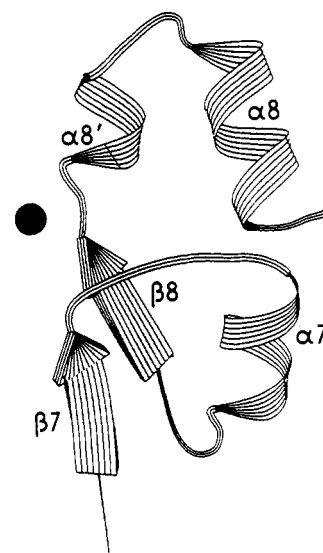


FIGURE 3: Ribbon drawing (Priestle, 1988) of the part of the β/α -barrel structure that is involved in the binding of the phosphate group. The phosphate position is indicated by the black circle. The view direction is the same as in the bottom panel of Figure 1.

indolepropanol phosphate (IPP), a competitive inhibitor, is bound at the equivalent position (Hyde et al., 1988). Helix 8’ comprises residues 235–242 in TRPS(α) and thus is somewhat longer than the equivalent helices in IGPS and PRAI. The superposition of the residues involved in phosphate binding in these three enzymes is shown in Figure 1 (bottom).

In all three enzymes the phosphate is held by main-chain NH groups in equivalent positions of the supersecondary as well as of the tertiary structure. The binding is strengthened and the phosphate charge is partially compensated by the macrodipole moment originating from the short helix 8’ that is present in all three β/α -barrel structures. In PRAI and TRPS(α) only neutral side chains contribute to binding of the phosphate ion whereas in IGPS one phosphate charge is compensated by the positive charge of the ϵ -amino group of

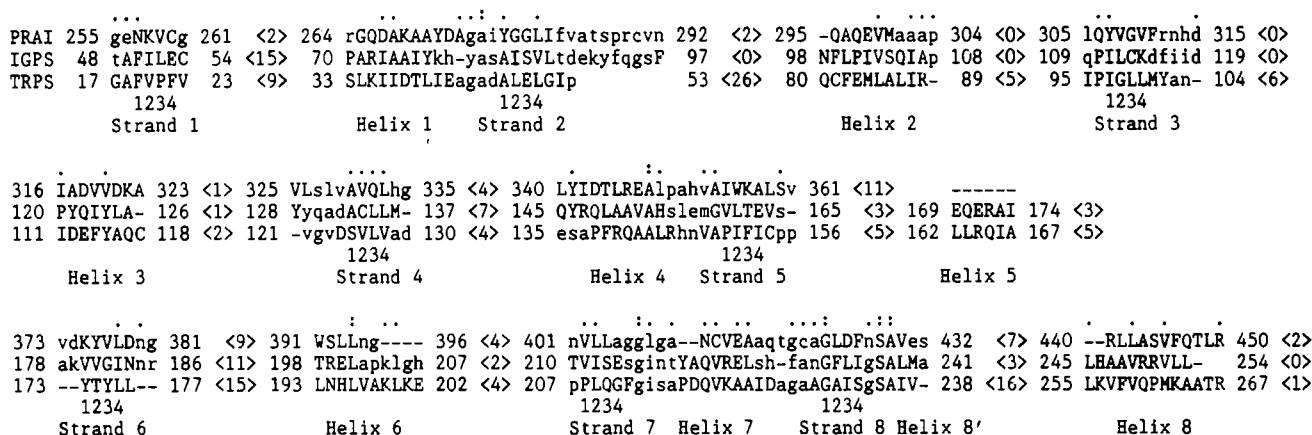


FIGURE 4: Amino acid sequence alignment of PRAI (top), IGPS (middle), and TRPS(α) (bottom) from *E. coli* based on structural superposition. The sequence numbers of the first and the last residue in each segment are indicated, and the number of intervening residues in each enzyme is shown as $\langle n \rangle$. Residues that have been assigned to β -strands and α -helices are indicated with capital letters. The four residues in each of the eight strands that define the four central residue layers of the central β -barrel are indicated with 1234 below. Sequence identities found in any two enzymes are indicated with a single dot above, and sequence identities common to all three enzymes are indicated with a colon above. Further statistics from this alignment are presented in Table I.

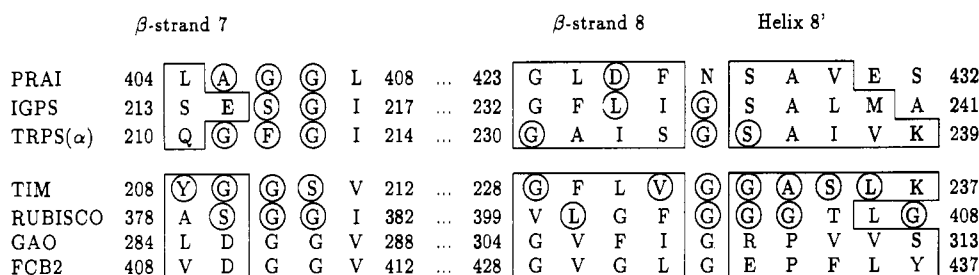


FIGURE 5: Alignment of the amino acid sequences of PRAI from *E. coli* (Horowitz et al., 1983), IGPS from *E. coli* (Horowitz et al., 1983), TRPS(α) (Nichols & Yanofsky, 1979), TIM from *E. coli* (Pichersky et al., 1984), RUBISCO from spinach (Zurawski et al., 1986), GAO from spinach (Volokita & Somerville, 1987), and FCB2 from *S. cerevisiae* (Guiard, 1985) in the regions involved in phosphate group binding. Residues belonging to secondary structural elements [PRAI, IGPS, TRPS(α), and GAO, determined with the program DSSP (Kabsch & Sander, 1983); TIM (Lolis et al., 1990); RUBISCO (Schneider et al., 1990); and FCB2 (Xia & Mathews, 1990)] are framed. The first and last residue of each section are numbered. Residues conserved among all known sequences of a particular enzyme [PRAI and IGPS (Niermann & Kirschner, 1990), TRPS(α) (Crawford et al., 1987), TIM (Lolis et al., 1990), and RUBISCO (Schneider et al., 1990)] are encircled.

Lys55 which is hydrogen bonded to it. The rms deviations between the phosphate binding sites are about 1 Å, and the positions of the phosphorus atoms of both the phosphate ion (PRAI, IGPS) and the inhibitor IPP [TRPS(α)] differ by only about 0.7 Å (Table I).

Inorganic phosphate is a weak inhibitor of PRAI and IGPS, and it has been postulated that it binds to the site occupied by the phosphate moieties of the substrates (Bisswanger et al., 1979). Therefore, it is reasonable to suppose that the common phosphate site is part of the substrate binding site in all three enzymes. This notion is confirmed by preliminary results from the crystal structure of the isolated IGPS domain complexed with the substrate analogue *N*-(5'-phosphoribitol-1-yl)-anthranilate that has been partially refined to a current *R*-factor of 23% at 2.17-Å resolution (Wilmanns et al., 1990; M. Wilmanns, unpublished results). The phosphate moiety of this inhibitor corresponds to the position of the phosphate ion in the active site of the IGPS domain in the crystal structure of the bienzyme PRAI:IGPS.

Comparison of Aligned Sequences. The superposition of the three-dimensional structures permits the assignment of structurally equivalent residues between the three enzymes and a comparison of the corresponding amino acids. To improve the alignment further, residues 340–351 of PRAI were shifted by one residue relative to the IGPS and TRPS(α) sequences. The resulting amino acid sequence alignment is shown in Figure 4 and the relevant statistics are summarized in Table II. It is seen that roughly 50–75% of the residues in each enzyme are structurally equivalent. The number of amino acid

Table I: Superposition of Phosphate Binding Sites

pairs of superimposed structures	rms deviation ^a (Å)	difference in phosphorus atom position ^b (Å)
PRAI/IGPS	0.83	0.69
IGPS/TRPS(α)	1.19	0.62
PRAI/TRPS(α)	1.31	0.81
PRAI/GAO	1.27	1.07
IGPS/GAO	0.87	1.07
TRPS(α)/GAO	1.01	1.18

^a Rms distances for the main-chain atoms of the 15 residues involved in the phosphate group binding site superposition [structure alignment in Figure 1 (bottom), sequence alignment in Figure 5]. ^b Differences between the positions of the phosphorus atoms of the phosphate ions (PRAI, IGPS) or groups [of IPP in TRPS(α) and of FMN in GAO].

sequence identities in the corresponding regions of pairs of superimposed enzymes ranges from 13.6% to 17.8%. Thus, only low levels of amino acid sequence identity are found, even in the structurally equivalent regions.

Significant similarity between the amino acid sequences of these three enzymes is observed in a segment of about 40 residues that includes β -strand 7, helix 7, β -strand 8, helix 8', and helix 8 (abbreviated $\beta 7\alpha 7\beta 8\alpha 8'\alpha 8$). One glycine in the loop following β -strand 7 [G407 in PRAI, G216 in IGPS, and G213 in TRPS(α)] is the only invariant residue among all known sequences of these three β/α -barrel enzymes (Crawford et al., 1987; Niermann & Kirschner, 1991). The sequences of the segments involved in phosphate binding in the *E. coli* enzymes are aligned in Figure 5. It is evident that this region

Table II: Statistical Results of Sequence Alignments Based on Structural Superposition

	PRAI/ IGPS	IGPS/ TRPS(α)	PRAI/ TRPS(α)
no. of equivalent residues ^a	147	146	139
rms deviation for paired C α atoms	2.26	2.42	2.81
equivalent residues per enzyme ^b (%)	75/58	57/54	71/52
identical amino acids ^c (%)	20 (13.6)	26 (17.8)	22 (15.8)
identical DNA base pairs ^d (%)	156 (35.4)	148 (33.8)	131 (31.4)

^aNumber of structurally equivalent residues between each pair of enzymes. Total number of residues: PRAI, 197; IGPS, 255; TRPS(α), 269. ^bPercentage of residues per enzyme that have structurally equivalent positions. ^cNumber (percentage) of identical amino acids at structurally equivalent positions. ^dNumber (percentage) of identical nucleotides at structurally equivalent positions.

of the three enzymes is conserved both in sequence and in three-dimensional structure.

In order to further examine the relationship between these enzymes, we calculated the nucleotide sequence similarities of all possible pairs *trpF/trpC*, *trpC/trpA*, and *trpF/trpA* by using 7 *trpF* genes coding for PRAI, 9 *trpC* genes coding for IGPS, and 11 *trpA* genes coding for TRPS(α). The DNA sequence alignments were again limited to the structurally equivalent segments of the corresponding enzyme structures. The bias toward fortuitously high sequence identities should be reduced in this kind of alignment for the following reasons. First, no a priori sequence information was included in the structural superposition that provides the register for the amino acid and the nucleotide sequence alignments. Secondly, the redundancy of the genetic code weakens the bias introduced in the amino acid sequence alignment by the preference of certain residue types for particular secondary structural elements (Keim et al., 1981).

The average fractions of identical base pairs are as follows: 31.1% (based on 63 comparisons) for *trpF/trpC*, 34.3% (based on 99 comparisons) for *trpC/trpA*, and 32.1% (based on 77 comparisons) for *trpF/trpA*. The similarities of the three pairs of the *E. coli* nucleotide sequences are close to the average values (Table II). However, pairs of nucleotide sequences, in particular *trpC/trpA*, are considerably more closely related in some organisms. The *trpC* nucleotide sequence from *H. volcanii* shares more than 40% identical bases with 6 out of 11 *trpA* nucleotide sequences used (*P. aeruginosa*, 47.2%; *T. thermophilus*, 46.3%; *H. volcanii*, 42.9%; *C. crescentus*, 42.6%; *P. putida*, 42.4%; *K. aerogenes*, 40.1%). Base-pair identities higher than 40% are also observed for 3 out of 11 *trpA* nucleotide sequences (*P. putida*, 41.9%; *T. thermophilus*, 41.7%; *P. aeruginosa*, 40.8%) compared to the *trpC* nucleotide sequence from *P. putida*. Moreover, comparing the *trpC* and *trpF* nucleotide sequences, two pairs of sequences are found with base identities higher than 40%: 40.8% for the pair *C. crescentus/K. aerogenes* and 40.0% for the pair *A. niger/P. aeruginosa*.

How significant are these nucleotide sequence similarities? To assume a completely random distribution of the four nucleotides (25%) is incorrect, because (i) these are used unequally in the three codon positions, (ii) unequal numbers of codons for different amino acids are available, and (iii) a preferred subset of codons is used by several organisms (Staden & McLachlan, 1982; Staden, 1990). Furthermore, structurally related amino acids are often encoded by codons that are related at two of three codon positions. In order to estimate

base-pair similarities that code for residues in similar structural environments (Keim et al., 1981), the following two tests were carried out. (i) Pairs of DNA sequences, correctly aligned by structural superposition, were shifted relative to each other by -2, -1, +1, and +2 codons. The average percentages of identical bases for the three pairs *trpF/trpC*, *trpC/trpA*, and *trpF/trpA* varied between 26% and 29%. (ii) The second test was restricted to 3 × 8 regions of the nucleotide sequences that code for the central parts of the eight β -strands (each four residues long) forming the central β -barrels of the *trpF*, *trpC*, and *trpA* gene products. These nucleotide sequence segments were compared in all possible combinations, providing alignments of sequences that code for residues in the quasi-identical structural environment of the central β -barrel. If the amino acid environment were to exert a significant bias on the choice of codons, this bias would become most clearly apparent in the comparison of these central β -strands. The resulting base identities are 30.9% for *trpF/trpC*, 34.9% for *trpC/trpA*, and 30.7% for *trpF/trpA*. The averaged identities of the corresponding but randomized nucleotide sequence regions are 28.9%, 30.7%, and 29.1%, respectively. Codon by codon comparisons indicate that the first two positions of each codon are in general considerably more closely related than the third positions, which are about 25% identical.

The nucleotide sequence similarities found do not indicate significant overall sequence homology between these three enzymes, although a tendency to slightly higher similarities is observed for correctly aligned DNA sequences than for randomized pairs of nucleotide sequences. This trend is most obvious for the comparison of the *trpC/trpA* nucleotide sequences. It should be noted that the sequence comparison results are presumably slightly weakened because the superposition of related tertiary structures might not always provide the correct alignment of the corresponding sequences. For instance, a helix from the outer helical ring of a particular β/α -barrel might easily be rotated around its axis, thus inducing local errors in structural superpositions (Castagnoli et al., 1989).

Considerably higher similarities are sometimes found in nucleotide sequence comparisons between the three pairs of some specific organisms, in particular *P. putida* and *H. volcanii*. However, it remains unclear whether these higher similarities are due to a possibly higher "random" level of codon usage in these organisms or a consequence of closer evolutionary relationship between the corresponding genes.

Comparison of Cyclically Permuted β/α -Barrels. It has been suggested that, because structural rearrangements of β/α -barrel enzymes could have occurred during evolution, related portions of different β/α -barrel enzymes may be located at different positions in the barrel (Lebioda et al., 1982; Goldman et al., 1987; Lebioda & Stec, 1988; Farber et al., 1987; Rey et al., 1988). Some β/α -barrel structures have been analyzed by fitting cross sections of the central β -barrels to ellipsoids and comparing their major and minor axes (Goldman et al., 1987; Lebioda & Stec, 1988; Farber & Petsko, 1990). Two cyclically permuted variants of PRAI from yeast have stability and catalytic activity comparable to those of the native enzyme (Luger et al., 1989), indicating that the structural integrity of this β/α -barrel enzyme is not disturbed by fusion of the original N- and C-termini and cleavage of a surface loop.

To test the above ideas with the three enzymes under investigation, the β/α -barrel structures of PRAI, IGPS, and TRPS(α) were cyclically permuted by rotating the central β -barrels about the barrel axis relative to each other by successive shifts of one β -strand. Subsequently, they were su-

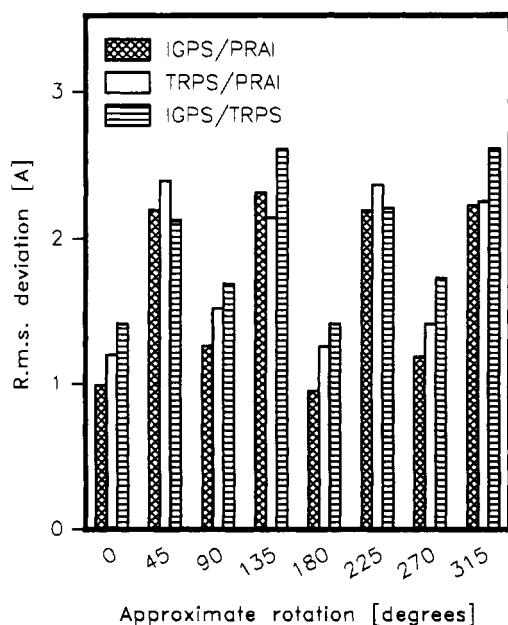


FIGURE 6: The rms deviations (in angstroms) of global superpositions using the 32 C_{α} atoms of the β -strand residues that constitute the central four layers of the β/α -barrel. The rotations (0° , 45° , ..., 315°) about the barrel axis correspond to shifts of an integral number (0, 1, ..., 7) of β -strands.

perimposed in pairs and compared. The α -helices and loop regions were not included in this process, because their superposition is much less satisfactory. Figure 6 shows that the rms deviations in C_{α} positions for all even-numbered shifts, corresponding to relative rotations of 0° , 90° , 180° , and 270° , are about half as large as for the odd-numbered ones. This finding indicates that the originally proposed 8-fold symmetry of parallel β/α -barrels (Lebioda et al., 1982; Goldman et al., 1987; Lebioda & Stec, 1988) must be reduced to 4-fold symmetry if one takes into account the nature of the interstrand hydrogen-bond network and the side-chain orientations of the central β -strand residues (Lesk et al., 1989). Side chains of residues of adjacent β -strands within one particular β -barrel layer alternately point toward the inside and the outside of the β -barrel.

Interestingly, direct superposition of these three β/α -barrel structures (0° rotation) does not lead to a distinct minimum rms value for any pair of these β/α -barrel structures (see Figure 6). The rms deviations for the even-numbered shifts are mainly determined by the side-chain packing within the central layers of the β -barrel (Lasters et al., 1988; Lesk et al., 1989). Thus, analysis of the eight possible superpositions of cyclically permuted β/α -barrel structures, corresponding to rotation steps of 45° , is not a method that can be used to establish evolutionary relationships for β/α -barrel enzymes.

DISCUSSION

Evolutionary Implications. The discovery of an equivalent phosphate binding site in the β/α -barrel structures of PRAI, IGPS, and TRPS(α) catalyzing three sequential steps in tryptophan biosynthesis suggests that these three enzymes are related by divergent evolution. Earlier work demonstrated that two enzymes catalyzing two consecutive steps in methionine biosynthesis, namely cystathionine γ -synthase and β -cystathionase, are also related by divergent evolution, since they show extensive sequence homology (Belfaiza et al., 1986).

N. H. Horowitz postulated 45 years ago that enzymes in biosynthetic pathways may have evolved in a retrograde fashion by a series of gene duplication events with subsequent

independent development (Horowitz, 1945, 1965). The results presented here are partially consistent with this hypothesis. The amino acid sequence and the three-dimensional structure of the β -subunit of tryptophan synthase are completely unrelated to PRAI, IGPS, and TRPS(α) (Hyde et al., 1988). Our findings therefore do not fit the original hypothesis of Horowitz, which states that retrograde evolution began with the ultimate enzyme of a biosynthesis pathway. Recent sequence comparisons suggest that the gene for the β -subunit of tryptophan synthase may be related to those of other pyridoxal phosphate dependent enzymes such as threonine synthase (Parsot, 1987; Bork & Rohde, 1990). This similarity is consistent with Jensen's alternative hypothesis (Jensen, 1976) that divergent evolution proceeded from primitive enzymes with broad substrate specificity. It amounts to only a minor modification of the original Horowitz hypothesis if retrograde evolution began with the penultimate enzyme of the tryptophan biosynthesis pathway. By contrast, TRPS(α) might have diverged according to Jensen's hypothesis from an enzyme belonging to the central glycolytic pathway, for example, triosephosphate isomerase (Banner et al., 1975; Lolis et al., 1990).

Comparison to Other β/α -Barrel Structures. Binding of phosphate to residues of the loop between β -strand 7 and helix 7 and to the N-terminal residues of helix 8' of the β/α -barrel also occurs in five other parallel β/α -barrel structures: TIM (Banner et al., 1975; Lolis et al., 1990), the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (RUBISCO; Chapman et al., 1987; Andersson et al., 1989; Knight et al., 1990) and the FMN-binding domains of glycolate oxidase (GAO; Lindquist & Bränden, 1985; Lindquist, 1989), trimethylamine dehydrogenase (TMDH; Lim et al., 1986), and flavocytochrome b_2 (FCB2; Xia et al., 1987; Xia & Mathews, 1990).

All these enzymes display some local amino acid sequence similarity within the loop between β -strand 7 and α -helix 7 as well as in the section comprising β -strand 8 and the additional helix 8' (Figure 5). For convenience only a single sequence of each enzyme is included. However, all known sequences were considered in the following comparison. The alignment shows that the two glycines at the N-terminus of the first loop (G405 and G406 of the PRAI sequence) are conserved in at least five out of seven enzymes. The residue that corresponds to L408 of the PRAI sequence is always a hydrophobic residue of medium size (valine, leucine, isoleucine). The only exception found is in TIM from *S. cerevisiae* (alanine). The first residue of β -strand 8 (G423 of the PRAI sequence) is a glycine in all sequences except for RUBISCO. However, this residue is not strictly conserved in PRAI and IGPS. Since G423 is located at the N-terminal side of the central parallel β -barrel, no obvious structural requirement accounts for the invariance of this glycine. In all crystal structures the "loop" between β -strand 8 and the subsequent additional helix 8' is only one residue long (Figure 2). In all sequences except PRAI this loop residue (N427 of the PRAI sequence) is a glycine. In PRAI, N427 has an unusual, energetically unfavorable main-chain conformation ($\phi = -139^{\circ}$, $\psi = -99^{\circ}$), usually occupied only by glycine. Since this residue determines the angle between the β -strand 8 and the additional helix 8', we believe that the structure of the phosphate binding site requires that this residue be a glycine.

In four of these eight β/α -barrel structures the loop between β -strand 6 and α -helix 6 is extraordinarily long and highly flexible in the crystal structures of the unliganded enzymes:

in PRAI (M. Wilmanns, unpublished results), in TRPS(α) (Hyde et al., 1988), in TIM (Lolis et al., 1990), and in RUBISCO (H. Schreuder, P. M. G. Curmi, D. Cascio, and D. Eisenberg, personal communication). Structural comparisons of the nonliganded and liganded structures of TIM (Lolis and Petsko, 1990) and RUBISCO (H. Schreuder, P. M. G. Curmi, D. Cascio, and D. Eisenberg, personal communication) as well as biochemical characterization of a deletion mutant of TIM (Pompliano et al., 1990) demonstrate that this loop moves considerably toward the active site in these enzymes upon binding of a substrate analogue. These data imply that in at least these four β/α -barrel enzymes [PRAI, TRPS(α), TIM, and RUBISCO] the loop between β -strand 6 and helix 6 fulfils an equivalent function during catalysis, closing down upon the bound substrate.

A comparison of PRAI, IGPS, and TRPS(α) at atomic resolution with β/α -barrel structures of different function is currently only possible for GAO from spinach, for which the high-resolution structure is deposited in the Protein Data Bank, Brookhaven National Laboratory (Lindquist, 1989; Lindquist & Brändén, 1989). We have superimposed the binding site of the phosphate moiety of FMN in GAO onto the phosphate binding sites of the three β/α -barrel enzymes involved in tryptophan biosynthesis. Figure 1 (bottom) demonstrates that the four phosphate binding sites are very similar. Table I shows that the position of the phosphorus atom of FMN deviates by only about 1 Å from the phosphorus atoms of the phosphate ions in PRAI and IGPS and of IPP in TRPS(α).

The observation that this common phosphate binding site occurs in at least five other β/α -barrel enzymes implies that also they may have descended from a common ancestor. According to Farber and Petsko (1990), the currently known β/α -barrel enzymes fall into two groups: the first would contain GAO, FCB2, and TMDH, which are all part of larger multidomain enzymes and bind FMN as a common cofactor. The second group would include TIM, PRAI, IGPS, and TRPS(α), all single-domain enzymes that are specific for phosphorylated substrates. It is probably not fortuitous that the product of the TRPS(α) reaction, D-glyceraldehyde 3-phosphate, is identical with the substrate of TIM. RUBISCO is a somewhat special case, since it consists of three domains and its substrate, ribulose biphosphate, contains two phosphate moieties. One of these is bound to the common phosphate binding site (Andersson et al., 1989).

In addition to TIM there are three other β/α -barrel enzymes from the glycolytic pathway of which the three-dimensional structures have been determined: pyruvate kinase (Muirhead et al., 1986), enolase (Lebioda et al., 1989; Stec & Lebioda, 1990), and fructosebiphosphate aldolase (FALD; Sygusch et al., 1987). The first two of these structures do not possess the common phosphate binding site. FALD, which is also a single-domain enzyme, has an additional helix 8' that is preceded by a short loop of two residues (Sygusch et al., 1987). An inhibitor locus, described for the low-resolution structure of FALD (Sygusch et al., 1985), is located near the N-terminus of this additional helix 8' (Sygusch et al., 1987). Therefore, FALD is a further candidate for having the common phosphate binding site.

The central β -barrel of the β/α -barrel enzymes is assumed to be a rather stable structural core in proteins due to an extensive and uninterrupted hydrogen bonding network between the individual β -strands. Therefore, a process of convergent evolution toward this stable fold cannot be excluded a priori. However, two cases have recently been reported where strong evidence for divergent evolution of functionally

distinct β/α -barrel enzymes was discovered. Mandelate racemase and muconate lactonizing enzyme, both found as octamers with 422-point symmetry in their respective crystal forms, share 26% sequence identity. About 60% of the residues of the corresponding β/α -barrel domains can be superimposed with an rms distance of 1.3 Å. Their close structural relation is highlighted by a common binding site for a divalent cation (Neidhart et al., 1990). Second, GAO and FCB2, both members of the group of β/α -barrel enzymes that contain an equivalent phosphate binding site, share 37% sequence identity. The structures of the flavin-binding domains of both enzymes superimpose with an rms distance of 0.93 Å. The cofactor FMN is located in an almost identical position in GAO and FCB2 (Lindquist et al., 1991).

The observation of a common phosphate binding site in at least eight out of 18 β/α -barrel enzymes suggests that all these enzymes are related by divergent evolution, although significant sequence similarity is limited to the phosphate binding sites. This notion is further supported by statistical considerations. If one takes the eight loops of the C-terminal side of the β -barrel as a priori equivalent, there are eight positions where this type of phosphate binding site could be located. The probability for such a phosphate binding site to be located between the same two adjacent loops of eight β/α -barrel structures would be $1/8^8$, which corresponds to about 1:16 000 000. Further analysis, based on three-dimensional structures refined at high resolution and on sequences of these enzymes, is needed to ascertain whether the evolutionary relationship, demonstrated here for the three enzymes of the tryptophan biosynthesis pathway, extends to further β/α -barrel proteins.

ACKNOWLEDGMENTS

We are grateful to C. Henn for creating Figure 1. C.C.H. and D.R.D. thank Drs. Gary Felsenfeld, Edith Miles, and Mark Boguski for their comments and advice. The critical comments of a reviewer have led to considerable improvement of the sequence comparisons and are highly appreciated.

REFERENCES

- Andersson, I., Knight, S., Schneider, G., Lindquist, Y., Lindquist, T., Brändén, C. I., & Lorimer, G. (1989) *Nature* 337, 229–234.
- Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D., & Waley, S. G. (1975) *Nature* 225, 609–614.
- Belfaiza, J., Parsot, C., Martel, A., de la Tour, C. B., Margarita, D., Cohen, G. N., & Saint-Girons, I. (1986) *Proc. Natl. Acad. Sci. U.S.A.* 83, 867–871.
- Bisswanger, H., Kirschner, K., Cohn, W., Hager, V., & Hansson, E. (1979) *Biochemistry* 18, 5946–5953.
- Bork, P., & Rohde, K. (1990) *Biochem. Biophys. Res. Commun.* 171, 1319–1325.
- Castagnoli, L., Scarpa, M., Kokkinidis, M., Banner, D. W., Tsernoglou, D., & Cesareni, G. (1989) *EMBO J.* 8, 621–629.
- Cederlund, E., Lindquist, Y., Söderlund, G., Brändén, C. I., & Jörnvall, H. (1988) *Eur. J. Biochem.* 173, 523–530.
- Chapman, M. S., Suh, S. W., Cascio, D., Smith, W. W., & Eisenberg, D. (1987) *Nature* 329, 354–356.
- Chothia, C. (1988) *Nature* 333, 598–599.
- Chou, P. Y., & Fasman, G. D. (1974) *Biochemistry* 13, 211–222.
- Crawford, I. P., Niermann, T., & Kirschner, K. (1987) *Proteins: Struct., Funct., Genet.* 1, 118–129.

- Farber, G. K., & Petsko, G. A. (1990) *Trends Biochem. Sci.* 15, 228-234.
- Farber, G. K., Petsko, G. A., & Ringe, D. (1987) *Protein Eng.* 1, 459-466.
- Finkelstein, A. V., & Ptitsyn, O. B. (1988) *Prog. Biophys. Mol. Biol.* 50, 171-190.
- Garnier, J., Osguthorp, D. J., & Robson, B. (1978) *J. Mol. Biol.* 120, 97-120.
- Goldman, A., Ollis, D. L., & Steitz, T. A. (1987) *J. Mol. Biol.* 194, 143-153.
- Guiard, B. (1985) *EMBO J.* 4, 3265-3272.
- Horowitz, H., van Arsdel, J., & Platt, T. (1983) *J. Mol. Biol.* 169, 775-797.
- Horowitz, N. H. (1945) *Proc. Natl. Acad. Sci. U.S.A.* 31, 153-156.
- Horowitz, N. H. (1965) in *Evolving Genes and Proteins* (Bryson, V., & Vogel, H. J., Eds.) pp 15-26, Academic, New York.
- Hyde, C. C., Ahmed, S. A., Padlan, E. A., Miles, E. W., & Davies, D. R. (1988) *J. Biol. Chem.* 263, 17857-17871.
- Jensen, R. A. (1976) *Annu. Rev. Microbiol.* 30, 409-425.
- Jones, T. A. (1978) *J. Appl. Crystallogr.* 11, 268-272.
- Kabsch, W., & Sander, C. (1983) *Biopolymers* 22, 2577-2637.
- Keim, P., Heinrikson, R. L., & Fitch, W. M. (1981) *J. Mol. Biol.* 151, 179-197.
- Knight, S., Andersson, I., & Brändén, C. I. (1990) *J. Mol. Biol.* 215, 113-160.
- Lasters, I., Wodak, S., Alard, P., & van Cutsem, E. (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85, 3338-3342.
- Lebioda, L., & Stec, B. (1988) *Nature* 333, 683-686.
- Lebioda, L., Hatada, M. H., Tulinsky, A., & Mavridis, I. M. (1982) *J. Mol. Biol.* 162, 445-458.
- Lebioda, L., Stec, B., & Brewer, J. M. (1989) *J. Biol. Chem.* 264, 3685-3693.
- Lesk, A. M., Brändén, C. I., & Chothia, C. (1989) *Proteins Struct., Funct., Genet.* 5, 139-148.
- Levitt, M., & Chothia, C. (1976) *Nature* 261, 552-558.
- Lim, L. W., Shamala, N., Mathews, F. S., Steenkamp, D. J., Hamlin, R., & Xuong, N. h. (1986) *J. Biol. Chem.* 261, 15140-15146.
- Lindquist, Y. (1989) *J. Mol. Biol.* 209, 151-166.
- Lindquist, Y., & Brändén, C. I. (1985) *Proc. Natl. Acad. Sci. U.S.A.* 82, 6855-6859.
- Lindquist, Y., & Brändén, C. I. (1989) *J. Biol. Chem.* 264, 3624-3628.
- Lindquist, Y., Brändén, C. I., Mathews, F. S., & Lederer, F. (1991) *J. Biol. Chem.* 266, 3198-3207.
- Lolis, E., & Petsko, G. A. (1990) *Biochemistry* 29, 6619-6625.
- Lolis, E., Alber, T., Davenport, R. C., Rose, D., Hartman, F. C., & Petsko, G. A. (1990) *Biochemistry* 29, 6609-6618.
- Luger, K., Hommel, U., Herold, M., Hofsteenge, J., & Kirschner, K. (1989) *Science* 243, 206-210.
- Muirhead, H., Clayden, D. A., Barford, D., Lorimer, C. G., Fothergill-Gilmore, L. A., Schiltz, E., & Schmitt, W. (1986) *EMBO J.* 5, 475-481.
- Nagata, S., Hyde, C. C., & Miles, E. W. (1989) *J. Biol. Chem.* 264, 6288-6296.
- Neidhart, D. J., Kenyon, G. L., Gerlt, J. A., & Petsko, G. A. (1990) *Nature* 347, 692-694.
- Nichols, B. P., & Yanofsky, C. (1979) *Proc. Natl. Acad. Sci. U.S.A.* 76, 5244-5248.
- Niermann, T., & Kirschner, K. (1991) *Protein Eng.* 4, 359-370.
- Parsot, C. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 5207-5210.
- Pichersky, E., Gottlieb, L. D., & Hess, J. F. (1984) *Mol. Gen. Genet.* 195, 314-320.
- Pompliano, D. L., Peyman, A., & Knowles, J. R. (1990) *Biochemistry* 29, 3186-3194.
- Priestle, J. P. (1988) *J. Appl. Crystallogr.* 21, 531-534.
- Priestle, J. P., Grütter, M. G., White, J. L., Vincent, M. G., Kania, M., Wilson, E., Jardetzky, T. S., Kirschner, K., & Jansonius, J. N. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 5690-5694.
- Rey, F., Jenkins, J., Janin, J., Lasters, I., Alard, P., Claessens, M., Matthysens, G., & Wodak, S. (1988) *Proteins: Struct., Funct., Genet.* 4, 165-172.
- Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. C., & Jones, T. A. (1990) *Science* 249, 380-386.
- Satow, Y., Cohen, G. H., Padlan, E. A., & Davies, D. R. (1986) *J. Mol. Biol.* 190, 593-604.
- Schneider, G., Knight, S., Andersson, I., Brändén, C. I., Lindquist, Y., & Lundquist, T. (1990) *EMBO J.* 9, 2045-2050.
- Staden, R. (1990) *Methods. Enzymol.* 183, 163-180.
- Staden, R., & McLachlan, A. D. (1982) *Nucleic Acids Res.* 10, 141-156.
- Stec, B., & Lebioda, L. (1990) *J. Mol. Biol.* 211, 235-248.
- Sygusch, J., Boulet, H., & Beaudry, D. (1985) *J. Biol. Chem.* 260, 15286-15290.
- Sygusch, J., Beaudry, D., & Allaire, M. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 7846-7850.
- Taylor, W. R. (1986) *J. Theor. Biol.* 119, 205-218.
- Volokita, M., & Somerville, C. R. (1987) *J. Biol. Chem.* 262, 15825-15828.
- Wilmanns, M., Schlagenhauf, E., Fol, B., & Jansonius, J. N. (1990) *Protein Eng.* 3, 173-180.
- Xia, Z. X., & Mathews, F. S. (1990) *J. Mol. Biol.* 212, 837-863.
- Xia, Z. X., Shamala, N., Bethge, P. H., Lim, L. W., Bellamy, H. D., Xuong, N. h., Lederer, F., & Mathews, F. S. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 2629-2633.
- Zurawski, G., Whitfield, P. R., & Bottomley, W. (1986) *Nucleic Acids Res.* 14, 3975.