Protein Science (1995), *4*:1239–1242. Cambridge University Press. Printed in the USA.

FOR THE RECORD

Similarity of different β-strands flanked in loops by glycines and prolines from distinct $(\alpha/\beta)_8$-barrel enzymes: Chance or a homology?

ŠTEFAN JANEČEK

Institute of Ecobiology, Slovak Academy of Sciences, Štefánikova 3, SK-81434 Bratislava, Slovakia

(RECEIVED March 8, 1995; ACCEPTED March 21, 1995)

Abstract: Many $(\alpha/\beta)_8$-barrel enzymes contain their conserved sequence regions at or around the β-strand segments that are often preceded and succeeded by glycines and prolines, respectively. α-Amylase is one of these enzymes. Its sequences exhibit a very low degree of similarity, but strong conservation is seen around its β-strands. These conserved regions were used in the search for similarities with β-strands of other $(\alpha/\beta)_8$-barrel enzymes. The analysis revealed an interesting similarity between the segment around the β2-strand of α-amylase and the one around the β4-strand of glycolate oxidase that are flanked in loops by glycines and prolines. The similarity can be further extended on other members of the α-amylase and glycolate oxidase subfamilies, i.e., cyclodextrin glycosyltransferase and oligo-1,6-glucosidase, and flavocytochrome b_2, respectively. Moreover, the α-subunit of tryptophan synthase, the $(\alpha/\beta)_8$-barrel enzyme belonging to the other subfamily of $(\alpha/\beta)_8$-barrels, has both investigated strands, β2 and β4, similar to β2 of α-amylase and β4 of glycolate oxidase. The possibilities of whether this similarity exists only by chance or is a consequence of some processes during the evolution of $(\alpha/\beta)_8$-barrel proteins are briefly discussed.

Keywords: α-amylase; $(\alpha/\beta)_8$-barrel enzymes; conserved β-strands; evolutionary relatedness; glycolate oxidase; tryptophan synthase

The question of parallel $(\alpha/\beta)_8$-barrel fold evolution is still unclear. Divergent evolution (Farber & Petsko, 1990) is favored mainly by the active site conservation of different $(\alpha/\beta)_8$-barrel enzymes at the COOH-terminal end of the inner β-barrel sheet regardless of the catalytic function of the enzymes. On the other hand, a general lack of sequence homology among the individual $(\alpha\text{-}\beta)_8$-barrel proteins supports the concept of their convergency to a symmetric and stable fold (Lesk et al., 1989). The third possibility that should be taken into account is the exon theory of genes, which is anchored in the idea that proteins evolved by a combination of exons coding for small structural or functional units (Gilbert & Glynias, 1993). It seems acceptable that during the evolution of $(\alpha/\beta)_8$-barrels all the three possibilities or their combinations have been used (Brändén, 1991; Farber, 1993; Doolittle, 1994). In these terms the search for sequence homologies or, at least, similarities among these enzymes is of special importance. Several groups, such as pyruvate kinase and enolase (Lebioda & Stec, 1988), α-amylase (AMY), cyclodextrin glycosyltransferase (CGT), and oligo-1,6-glucosidase (OGL) (MacGregor & Svensson, 1989; Watanabe et al., 1990), mandelate racemase and muconate lactonizing enzyme (Neidhart et al., 1990), glycolate oxidase (GOX) and flavocytochrome b_2 (FCB) (Lindqvist et al., 1991), and indole-3-glycerolphosphate synthase, *N*-(5′-phosphoribosyl)-anthranilate isomerase, and the α-subunit of tryptophan synthase (TSA) (Wilmanns et al., 1991), have been found to possess divergency-favoring common sequence-structural features. These groups represent, in fact, more or less independent subfamilies of the entire present-day family of $(\alpha/\beta)_8$-barrel proteins as divided by Farber and Petsko (1990). Some features joining the subfamilies of $(\alpha/\beta)_8$-barrels have been already recognized: e.g., (1) two sequence regions spanning their phosphate binding sites (Wilmanns et al., 1991); (2) the sequence motif based on their structurally derived alignment (Pickett et al., 1992); or (3) five sequence stretches corresponding to the five conserved regions of AMY (Janeček, 1993). Nevertheless, until now no clear sequence-structural evidence has been offered to demonstrate the existence of evolutionary relatedness of seemingly unrelated $(\alpha/\beta)_8$-barrel subfamilies.

As already known, β-strands of $(\alpha/\beta)_8$-barrels are better conserved than the α-helices that are more variable in length and sequence (Lesk et al., 1989). Therefore an attempt was made to investigate the sequences of β-strands of various $(\alpha/\beta)_8$-barrel enzymes for whether some mutual homology could be found

Reprint requests to: Štefan Janeček, Institute of Ecobiology, Slovak Academy of Sciences, Stefanikova 3, SK-81434 Bratislava, Slovakia; e-mail: sjanecek@ue.savba.sk.

Abbreviations: AMY, α-amylase; ANL, *N*-acetylneuraminate lyase; CGT, cyclodextrin glycosyltransferase; FCB, flavocytochrome b_2; GOX, glycolate oxidase; OGL, oligo-1,6-glucosidase; PDB, Protein Data Bank; TSA, α-subunit of tryptophan synthase.

among them. The detailed results of this analysis will appear elsewhere. Here, the interesting observation of remarkable mutual similarity between the strands β2 of AMY and CGT, β4 of GOX and FCB, and both β2 and β4 of TSA as being well flanked by glycines and prolines in loops is described and discussed from the evolutionary point of view.

AMY was used as the template $(\alpha/\beta)_8$-barrel enzyme (Kinemage 1) in this study because (1) AMY is the leading member of a large subfamily of divergently evolved $(\alpha/\beta)_8$-barrel starch hydrolases and related enzymes with published evolutionary trees, and (2) the amino acid sequences of AMY possess very low degrees of similarity (around 10% in general), but, on the other hand, their β-strands are strongly conserved (Nakajima et al., 1986; MacGregor & Svensson, 1989; Jespersen et al., 1993; Janeček, 1994a, 1994b; Svensson, 1994). Of the β-strands (β2, β3, β4, β5, β7) of AMY that are best conserved, the stretch around the β2-strand flanked in loops by glycine and proline residues was chosen as the template segment in the search for similarities among the other $(\alpha/\beta)_8$-barrels. The importance of this segment is double: (1) it belongs to one of the sequence similarities in different AMYs and CGTs (MacGregor & Svensson, 1989; Janeček, 1994a), and (2) it simultaneously constitutes one of the characteristic sequential differences allowing discrimination of CGTs from AMYs (Jespersen et al., 1993; Janeček, 1994b; Janeček et al., 1995). Structural information concerning the enzymes from the AMY subfamily (AMY, CGT, OGL) was extracted from the literature (Matsuura et al., 1984; Kizaki et al., 1993; Lawson et al., 1994) and from Protein Data Bank files: 6TAA for AMY from *Aspergillus oryzae* and 1CDG for CGT from *Bacillus circulans* strain 251. The atomic coordinates of OGL from *Bacillus cereus* (Kizaki et al., 1993) were not available from the PDB. Structural details of the flavin mononucleotide-dependent enzymes (GOX and FCB) were also found in published papers (Lindqvist, 1989; Xia & Mathews, 1990; Lindqvist et al., 1991). Their PDB files used were 1GOX for spinach GOX and 1FCB for FCB from *Saccharomyces cerevisiae*. Similarly, the data on TSA from *Salmonella typhimurium* were taken from the literature (Hyde et al., 1988) and from the relevant PDB file: 1WSY.

The amino acid region around the β2-strand of AMY (Table 1) is, in fact, the sequence region between the invariant residues G56 and P64 (*A. oryzae* AMY numbering) that has been recognized as a sequence similarity of AMYs and CGTs (Janeček, 1994b) and as a feature that simultaneously allows discrimination of CGTs from AMYs by the specific peptide length from glycine to proline (Jespersen et al., 1993; Janeček et al., 1995). The equivalent sequence stretch of the third $(\alpha/\beta)_8$-barrel enzyme from the AMY subfamily, OGL, is also added in Table 1. Interestingly, the other pair of already known homologous $(\alpha/\beta)_8$-barrel enzymes, GOX and FCB, have amino acid sequences around their β4-strands similar to those around the β2-strands of the enzymes from the AMY subfamily (Table 1). The member of the third $(\alpha/\beta)_8$-barrel subfamily, TSA, may join by its two relevant β-strands (β2 and β4) AMY, CGT, and OGL on the one side and GOX and FCB on the other side (cf. Table 1).

In all presented enzymes (Kinemages 2–5), this stretch is characterized by two amino acid residues, a glycine at the start located in a turn preceding the β-strand and a proline at the end positioned 8–11 residues from the glycine. Most of the variable residues represent conservative substitutions, such as valine, leucine, isoleucine and phenylalanine (the second, fifth, and seventh positions in the stretches shown in Table 1). Secondary structure (β-strands) is also well conserved.

Table 1. *Similarity between the β2- and β4-strands of distinct groups of $(\alpha/\beta)_8$-barrel enzymes*

Enzyme	E.C.	Sequence[a]	Structure	Numbering	Length
AMY	3.2.1.1	GFTAIWITP	β2	56–64	9
CGT	2.4.1.19	GVTAIWISQP	β2	70–79	10
OGL	3.2.1.10	GIDVIWLSP	β2	44–52	9
TSA	4.2.1.20	GADALELGVP	β2	44–53	10
TSA	4.2.1.20	GVDSVLVADVP	β4	122–132	11
GOX	1.1.3.15	GFKAIALTVDTP	β4	148–159	12
FCB	1.1.2.3	GVKALFVTVDAP	β4	273–284	12

[a] The β-strands are underlined; the glycines and prolines at the starts and the ends of segments, respectively, are signified by bold letters.

The close homology between AMY and CGT on the one side and GOX and FCB on the other side has already been manifested (MacGregor & Svensson, 1989; Lindqvist et al., 1991). To demonstrate close similarity between the β2-strand of starch hydrolases and related enzymes from the AMY subfamily and the β4-strand of the flavin mononucleotide-dependent enzymes, the structures of investigated segments were superimposed (Fig. 1; Kinemages 2–5). Although the overlaps indicate a high level of mutual similarity, they do not answer the question of why this similarity occurs. Basically, two principally different explanations are possible: chance or a homology.

A chance likeness of the β-strand segments shown in Table 1 is the simplest explanation. Indeed, other $(\alpha\beta)_8$-barrel enzymes also have their β-strands flanked in loops of glycines and prolines, such as ribulose-1,5-bisphosphate carboxylase/oxygenase (Schneider et al., 1990) and *N*-acetyl-neuraminate lyase (ANL) (Izard et al., 1994), and there are several others in which glycines and prolines flanking the strands in loops are hardly traced or the distance between the two amino acid residues is too long, e.g., pyruvate kinase (Muirhead et al., 1986) and xylanase (Harris et al., 1994). The presence of the two β-strands (β2 and β4) in TSA (Table 1) could be used as an argument against the eventuality that the studied similarity exists only by chance. There is a very high degree of mutual sequence similarity (four identical residues and four conservative substitutions by introducing a gap between G51 and V52 in the β2-strand, i.e., 73%, as well as with the β2-strand of AMY, CGT, and OGL and the β4-strand of GOX and FCB. All of these segments are very similar also from the structural point of view like those that are shown overlapped in Figure 1 and Kinemages 2–5. As far as the orientation of terminating prolines is concerned, they are not equally distant from the invariant glycines and therefore they are not structurally equivalent.

The second argument against a chance likeness is the fact that there are about 10 other $(\alpha/\beta)_8$-barrel enzymes that have just the β2-strand flanked in loops by glycines and prolines (Š. Janeček, unpubl.). And finally, the segment around the β4-strand of the recently recognized $(\alpha/\beta)_8$-barrel enzyme, ANL (Izard et al., 1994), which belongs neither to the AMY subfamily nor to the flavin mononucleotide-dependent enzymes, has the sequence GFDAVSAVTP (β4-strand underlined), which is re-

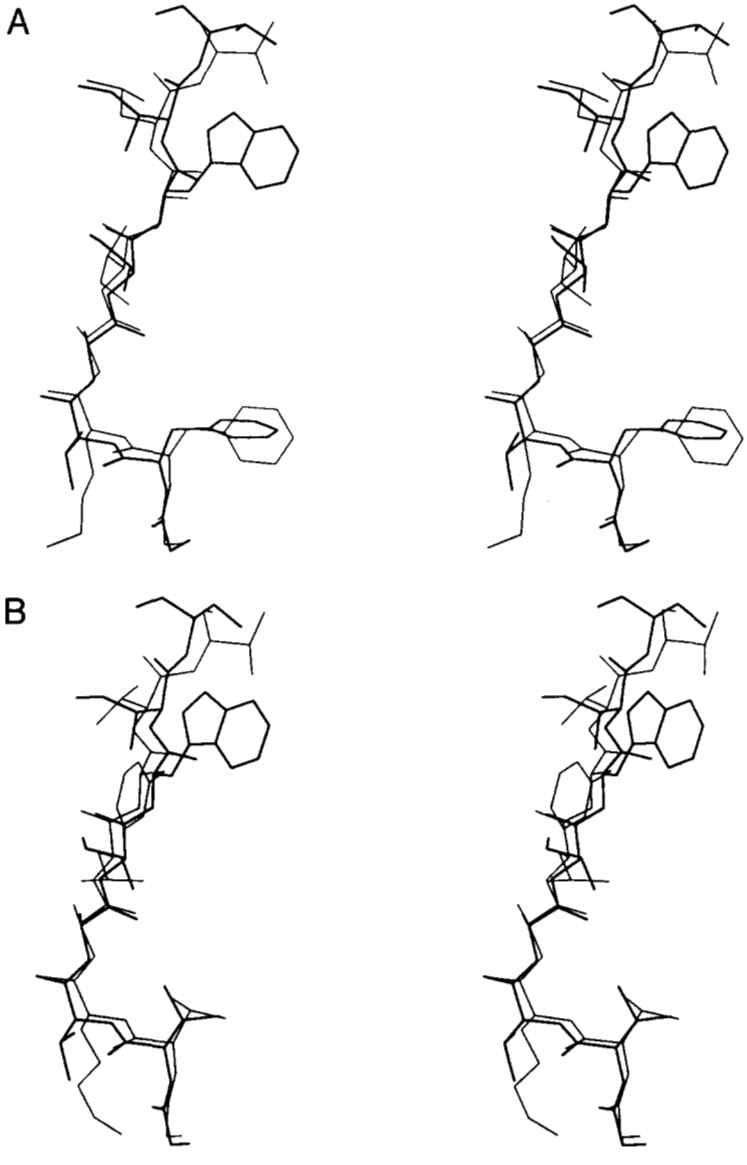

Fig. 1. Stereo views of β-strand segments of nonhomologous $(\alpha/\beta)_8$-barrel enzymes. **A:** The β4-strand of GOX from spinach (thin lines; sequence: 148-GFKAIALT) is overlapped on the β2-strand of AMY from *A. oryzae* (thick lines; sequence: 56-GFTAIWIT). **B:** The β4-strand of FCB from *S. cerevisiae* (thin lines; sequence: 273-GVKALFVT) is overlapped on the β2-strand of CGT from *B. circulans* (thick lines; sequence: 70-GVTAIWIS). The overlaps were based on the superposition of the backbone atoms, i.e., N, C_α, C_O (for eight amino acid residues, i.e., 24 atom pairs). The RMS deviations were calculated with the program ALCHEMY (Tripos Associates, Inc.) and the overlaps were displayed with the program HYPERCHEM (Autodesk, Inc.). The RMS deviations are 0.506 Å for segment overlap of GOX on AMY and 0.766 Å for segment overlap of FCB on CGT. When comparing the mutually exchanged pairs, i.e., AMY with FCB and CGT with GOX, these values are 0.706 Å and 0.541 Å, respectively.

markably similar to those listed in Table 1. Moreover, two other proteins, dihydrodipicolinate synthase and MosA (an enzyme implicated in rhizopine synthesis), which have been suggested to share a similar structure of ANL (Izard et al., 1994), also have stretches starting with glycine and ending with proline residues, GIVGCLTVTP and GADGVLIVSP, respectively, in the region equivalent to its β4-strand. Because the atomic coordinates of ANL are not yet available from PDB, the relevant segments cannot be compared structurally. Nevertheless, in the light of comparison of AMY, CGT, TSA, GOX, and FCB, the segment structural similarity of ANL with the rest of enzymes should be forthcoming.

Although the eventuality of a chance similarity seems to be less probable, the above-presented arguments do not address the nature of the evolutionary relatedness of these enzymes caused by a possible homology between their β-strands. The aim of this contribution was, however, neither to align the β-strands of $(\alpha/\beta)_8$-barrel enzymes (for this, see, e.g., Pickett et al., 1992; Taylor et al., 1994; Sergeev & Lee, 1994) nor to specify the types of evolutionary relationships among them (for this, see, e.g., Farber & Petsko, 1990; Brändén, 1991; Farber, 1993). The main aim was to demonstrate the observation of sequence-structural similarity (although limited) between seemingly unrelated subfamilies of $(\alpha/\beta)_8$-barrel enzymes. In these terms the possibility of shifting the main elements of secondary structure (β-strands) relative to each other while retaining their fold in divergently evolved $(\alpha/\beta)_8$-barrel enzymes (Chothia, 1988) should also be taken into account. It seems possible that sequence similarities do occur in $(\alpha/\beta)_8$-barrels, but due to their very long evolutionary history and distantly related functions the homologous amino acid residues have adopted different structural and/or functional roles. Paradoxically, to be successful in the search for homology among different $(\alpha/\beta)_8$-barrel proteins, one probably cannot produce sequence alignments that are as strictly structurally satisfactory as is necessary when working with clearly homologous sequences.

Acknowledgment

I am grateful to Dr. Štefan Baláž (Slovak Technical University, Bratislava) for providing facilities necessary for visualizing and overlapping protein structures.

References

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

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

Doolittle RF. 1994. Convergent evolution: The need to be explicit. *Trends Biochem Sci 19*:15–18.

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.

Gilbert W, Glynias M. 1993. On the ancient nature of introns. *Gene 135*: 137–144.

Harris GW, Jenkins JA, Connerton I, Cummings N, Leggio LL, Scott M, Hazlewood GP, Laurie JI, Gilbert HJ, Pickersgill RW. 1994. Structure of the catalytic core of the family F xylanase from *Pseudomonas fluorescens* and identification of the xylopentaose-binding sites. *Structure 2*:1107–1116.

Hyde CC, Ahmed SA, Padlan EA, Miles EW, Davies DR. 1988. Three-dimensional structure of the tryptophan synthase $\alpha_2\beta_2$ multienzyme complex from *Salmonella typhimurium*. *J Biol Chem 263*:17857–17871.

Izard T, Lawrence MC, Malby RL, Lilley GG, Colman PM. 1994. The three-dimensional structure of *N*-acetylneuraminate lyase from *Escherichia coli*. *Structure 2*:361–369.

Janeček Š. 1993. Sequence similarities in $(\alpha/\beta)_8$-barrel enzymes revealed by conserved regions of α-amylase. *FEBS Lett 316*:23–26.

Janeček Š. 1994a. Sequence similarities and evolutionary relationships of microbial, plant and animal α-amylases. *Eur J Biochem 224*:519–524.

Janeček Š. 1994b. Parallel β/α-barrels of α-amylase, cyclodextrin glycosyltransferase and oligo-1,6-glucosidase versus the barrel of β-amylase: Evolutionary distance is a reflection of unrelated sequences. *FEBS Lett 353*: 119–123.

Janeček Š, MacGregor EA, Svensson B. 1995. Characteristic differences in the primary structure allow discrimination of cyclodextrin glucanotransferases from α-amylases. *Biochem J 305*:685–686.

Jespersen HM, MacGregor EA, Henrissat B, Sierks MR, Svensson B. 1993. Starch- and glycogen-debranching and branching enzymes: Prediction of structural features of the catalytic $(\beta/\alpha)_8$-barrel domain and evolutionary relationship to other amylolytic enzymes. *J Protein Chem 12*: 791–805.

Kizaki H, Hata Y, Watanabe K, Katsube Y, Suzuki Y. 1993. Polypeptide folding of *Bacillus cereus* ATCC7064 oligo-1,6-glucosidase revealed by 3.0 Å resolution X-ray analysis. *J Biochem 113*:646–649.

Lawson CL, van Montfort R, Strokopytov B, Rozeboom HJ, Kalk KH, de Vries GE, Penninga D, Dijkhuizen L, Dijkstra BW. 1994. Nucleotide sequence and X-ray structure of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 in a maltose-dependent crystal form. *J Mol Biol 236*:590–600.

Lebioda L, Stec B. 1988. Crystal structure of enolase indicates that enolase and pyruvate kinase evolved from a common ancestor. *Nature 333*:683–686.

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

Lindqvist Y. 1989. Refined structure of spinach glycolate oxidase at 2 Å 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.

MacGregor EA, Svensson B. 1989. A super-secondary structure predicted to be common to several α-1,4-D-glucan-cleaving enzymes. *Biochem J 259*:145–152.

Matsuura Y, Kusunoki M, Harada W, Kakudo M. 1984. Structure and possible catalytic residues of Taka amylase. *J Biochem 95*:697–702.

Muirhead H, Clayden DA, Barford D, Lorimer CG, Fothergill-Gilmore LA, Schiltz E, Schmitt W. 1986. The structure of cat muscle pyruvate kinase. *EMBO J 5*:475–481.

Nakajima R, Imanaka T, Aiba S. 1986. Comparison of amino acid sequences of eleven different α-amylases. *Appl Microbiol Biotechnol 23*:355–360.

Neidhart DJ, Kenyon KL, Gerlt JA, Petsko GA. 1990. Mandelate racemase and muconate lactonizing enzyme are mechanistically distinct and structurally homologous. *Nature 347*:692–694.

Pickett SD, Saqi MAS, Sternberg MJE. 1992. Evaluation of the sequence template method for the protein structure prediction. Discrimination of the $(\beta/\alpha)_8$-barrel fold. *J Mol Biol 228*:170–187.

Schneider G, Lindqvist Y, Lundqvist T. 1990. Crystallographic refinement and structure of ribulose-1,5-bisphosphate carboxylase from *Rhodospirillum rubrum* at 1.7 Å resolution. *J Mol Biol 211*:989–1008.

Sergeev Y, Lee B. 1994. Alignment of β-barrels in $(\beta/\alpha)_8$ proteins using hydrogen-bonding pattern. *J Mol Biol 244*:168–182.

Svensson B. 1994. Protein engineering in the α-amylase family: Catalytic mechanism, substrate specificity, and stability. *Plant Mol Biol 25*:141–157.

Taylor WR, Flores TP, Orengo CA. 1994. Multiple protein structure alignment. *Protein Sci 3*:1858–1870.

Watanabe K, Kitamura K, Iha H, Suzuki Y. 1990. Primary structure of the oligo-1,6-glucosidase of *Bacillus cereus* ATCC7064 deduced from the nucleotide sequence of the gene. *Eur J Biochem 192*:609–620.

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.

Xia ZX, Mathews FS. 1990. Molecular structure of flavocytochrome b_2 at 2.4 Å resolution. *J Mol Biol 212*:837–863.