Protein Science (1996), 5:1136–1143. Cambridge University Press. Printed in the USA.

Invariant glycines and prolines flanking in loops the strand β2 of various $(\alpha/\beta)_8$-barrel enzymes: A hidden homology?

ŠTEFAN JANEČEK

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

(RECEIVED December 11, 1995; ACCEPTED March 5, 1996)

Abstract

The question of parallel $(\alpha/\beta)_8$-barrel fold evolution remains unclear, owing mainly to the lack of sequence homology throughout the amino acid sequences of $(\alpha/\beta)_8$-barrel enzymes. The "classical" approaches used in the search for homologies among $(\alpha/\beta)_8$-barrels (e.g., production of structurally based alignments) have yielded alignments perfect from the structural point of view, but the approaches have been unable to reveal the homologies. These are proposed to be "hidden" in $(\alpha/\beta)_8$-barrel enzymes. The term "hidden homology" means that the alignment of sequence stretches proposed to be homologous need not be structurally fully satisfactory. This is due to the very long evolutionary history of all $(\alpha/\beta)_8$-barrels. This work identifies so-called hidden homology around the strand β2 that is flanked by loops containing invariant glycines and prolines in 17 different $(\alpha/\beta)_8$-barrel enzymes, i.e., roughly in half of all currently known $(\alpha/\beta)_8$-barrel proteins. The search was based on the idea that a conserved sequence region of an $(\alpha/\beta)_8$-barrel enzyme should be more or less conserved also in the equivalent part of the structure of the other enzymes with this folding motif, given their mutual evolutionary relatedness. For this purpose, the sequence region around the well-conserved second β-strand of α-amylase flanked by the invariant glycine and proline (56_GFTAIWITP, *Aspergillus oryzae* α-amylase numbering), was used as the sequence–structural template. The proposal that the second β-strand of $(\alpha/\beta)_8$-barrel fold is important from the evolutionary point of view is strongly supported by the increasing trend of the observed β2-strand structural similarity for the pairs of $(\alpha/\beta)_8$-barrel enzymes: α-amylase and the α-subunit of tryptophan synthase, α-amylase and mandelate racemase, and α-amylase and cyclodextrin glycosyltransferase. This trend is also in agreement with the existing evolutionary division of the entire family of $(\alpha/\beta)_8$-barrel proteins.

Keywords: α-amylase; $(\alpha/\beta)_8$-barrel enzymes; conserved β2-strands; evolutionary relationships; hidden homology

$(\alpha/\beta)_8$-Barrel structural motif formed by the inner barrel composed from eight parallel β-strands surrounded by eight α-helices is one of the most frequently occurring folding motifs in proteins (Brändén, 1991; Orengo et al., 1994). The family of proteins containing this fold comprises more than 30 different enzyme specificities that belong to all commissions of enzyme nomenclature except ligases (for a recent review, see Reardon & Farber, 1995).

This structural motif was first found in the structure of chicken muscle triosephosphate isomerase (Banner et al., 1975). Since then, the question of evolution of this fold has remained in the center of general attention. Unfortunately, the answer is not wholly clear yet. In fact, three different forms of evolution of $(\alpha/\beta)_8$-barrels are possible: (1) divergent evolution; (2) convergent evolution; and (3) exon shuffling (Farber, 1993). Combinations of these possibilities should also be taken into account (Doolittle, 1994). Nevertheless, the strongest argument exists for divergence of many members of the family, i.e., conserving the location of the active sites of all currently known $(\alpha/\beta)_8$-barrel enzymes at the C-termini of β-strands forming the inner β-barrel (Farber & Petsko, 1990; Brändén, 1991; Farber, 1993; Orengo et al., 1994; Reardon & Farber, 1995). On the other hand, the apparent lack of sequence homologies throughout the family argues against a divergent way of evolution. This fact, together

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

Abbreviations: AAMY, α-amylase; ADA, adenosine deaminase; BGLA, (1→3,1→4)-β-glucanase; CBH, cellobiohydrolase II; CGT, cyclodextrin glycosyltransferase; EGLA, endoglucanase; ENOL, enolase; IGPS, indole-3-glycerolphosphate synthase; MR, mandelate racemase; OGLU, oligo-1,6-glucosidase; OYE, old yellow enzyme; PDB, Protein Data Bank; PRAI, *N*-(5′-phosphoribosyl) anthranilate isomerase; RBCO, ribulose-1,5-bisphosphate carboxylase/oxygenase; RMSD, RMS deviation; TADH, trimethylamine dehydrogenase; TSA, α-subunit of tryptophan synthase; URE, urease; XYI, xylose isomerase.

with the structural principles determining the packing of the sheet structure into a stable $(\alpha/\beta)_8$-barrel, offer the opportunity of the convergence of all $(\alpha/\beta)_8$-barrel proteins (Lesk et al., 1989). However, it has been indicated recently (Raine et al., 1994) that simple genetic alterations may be responsible for switching the nature of chain packing observed in $(\alpha/\beta)_8$-barrels.

Limited sequence similarities can only be found between several pairs or groups of $(\alpha/\beta)_8$-barrel enzymes that constitute so-called subfamilies (Farber & Petsko, 1990), e.g., α-amylase (AAMY), oligo-1,6-glucosidase (OGLU), and cyclodextrin glycosyltransferase (CGT) (MacGregor & Svensson, 1989; Kizaki et al., 1993); mandelate racemase (MR), muconate lactonizing enzyme, and chloromuconate cycloisomerase (Neidhart et al., 1990; Hoier et al., 1994); glycolate oxidase, flavocytochrome b_2, trimethylamine dehydrogenase (TADH), and old yellow enzyme (OYE) (Lindqvist et al., 1991; Fox & Karplus, 1994; Scrutton, 1994); aldose reductase and aldehyde reductase (El-Kabbani et al., 1994); and *N*-acetylneuraminate lyase and dihydrodipicolinate synthase (Izard et al., 1994; Mirwaldt et al., 1995). Recently, a large group of $(\alpha/\beta)_8$-barrel enzymes covering several families of glycosyl hydrolases (for the classification of glycohydrolases, see Henrissat, 1991; Henrissat & Bairoch, 1993) that contain their catalytic glutamates on strands β4 and β7, has been revealed (Henrissat et al., 1995; Jenkins et al., 1995). However, no clear sequence evidence has been offered up to now that would suggest homology among the substantial part of the seemingly unrelated $(\alpha/\beta)_8$-barrels except for the finding of two sequence regions spanning the phosphate binding site in eight $(\alpha/\beta)_8$-barrels (Wilmanns et al., 1991; Bork et al., 1995). It is worth mentioning, however, that motif searches with the key sites of the common phosphate binding site match a region in the vitamin B12 binding site of adenosylcobalamin-dependent mutase that belongs to another protein fold (Rossman fold). Therefore, this sequence motif is rather weak at discriminating the $(\alpha/\beta)_8$-barrel fold. Interestingly, the conserved residues of both sites occur in the same secondary structural elements (A. Bateman, pers. comm.).

Perhaps the $(\alpha/\beta)_8$-barrel enzymes have diverged so far that no detectable sequence homologies appear among them (Lesk & Boswell, 1992). It seems easier to predict an $(\alpha/\beta)_8$-barrel correctly from the sequence (Crawford et al., 1987; Jespersen et al., 1993; Wilmanns & Eisenberg, 1993; Niermann & Kirschner, 1995) than to trace the homologies among the $(\alpha/\beta)_8$-barrels of known structure. A few years ago, an idea was presented (Janeček, 1993) that a conserved sequence region of one $(\alpha/\beta)_8$-barrel enzyme should be more or less conserved in the structures of the other $(\alpha/\beta)_8$-barrels due to their evolutionary relatedness. This proposal has been supported recently by the observation of functionally essential, invariant glutamates located near the C-terminus of strand β5 in various $(\alpha/\beta)_8$-barrel enzymes (Janeček & Baláž, 1995), the β5-strand being well-known as a conserved sequence region in AAMYs and related starch-hydrolases adopting the $(\alpha/\beta)_8$-barrel fold (for this purpose, see Jespersen et al., 1993; Janeček, 1994a).

Here, a report on another sequence-structural feature is presented that may join a substantial group of enzymes from the $(\alpha/\beta)_8$-barrel family. The strand β2 of AAMY (56_GFTAI WITP, *Aspergillus oryzae* AAMY numbering, β2-strand underlined) flanked in loops by glycine and proline residues (Matsuura et al., 1984) and found to be well conserved in AAMY (Janeček, 1994b) was identified to have its sequence-structural equivalents in more than 15 other $(\alpha/\beta)_8$-barrel enzymes. These findings are in agreement with the recent proposal (Janeček, 1995) that sequence homologies in $(\alpha/\beta)_8$-barrels exist as so-called hidden homologies in which the homologous amino acid residues from the primordial barrel(s) might have adopted different roles in function and/or different positions in structure.

Background

Published amino acid sequences and three-dimensional structures of $(\alpha/\beta)_8$-barrel enzymes were explored in the search for a common sequence-structural feature. The search was based on the previously reported idea that a conserved sequence region of an $(\alpha/\beta)_8$-barrel enzyme should be more or less conserved also in the equivalent part of structure of the other $(\alpha/\beta)_8$-barrel enzymes owing to their mutual evolutionary relatedness (Janeček, 1993). In this context, AAMY was taken as the template $(\alpha/\beta)_8$-barrel enzyme because: (1) it is considered to be the representative of a large enzyme group of starch hydrolases adopting the $(\alpha/\beta)_8$-barrel fold; (2) the evolutionary relationships of these homologous enzymes are well recognized; and (3) the sequences of AAMYs exhibit a very low degree of similarity (about 10% in general), but strong conservation is seen around its β-strands (MacGregor & Svensson, 1989; Janeček & Baláž, 1993; Jespersen et al., 1993; Janeček, 1994a, 1994b; Davies & Henrissat, 1995). The stretch around the strand β2 of *A. oryzae* AAMY (belonging to the best conserved β-strands) that is flanked in loops by glycine and proline (Gly 56 and Pro 64, respectively) was used as the template segment in the search for similarities among the other $(\alpha/\beta)_8$-barrel enzymes.

The strand β2 has been identified recently to be conserved not only in AAMYs, but also in the sequences of the other members of the AAMY family, such as CGT, OGLU, pullulanase, isoamylase, etc. (MacGregor & Svensson, 1989; Jespersen et al., 1993; Janeček, 1994a; Janeček et al., 1995). In all these enzymes, this strand is flanked at the N and C terminus by glycine and proline, respectively (Table 1), that are invariant in each enzyme specificity. The length of the peptide from Gly to Pro is not, however, constant in all cases, and this allows one to discriminate between the closely related specificities, such as AAMY and CGT (Janeček, 1994a; Janeček et al., 1995).

Structural information concerning the $(\alpha/\beta)_8$-barrel enzymes that contain their strand β2 flanked in loops by glycine and proline was extracted from the literature and from the Protein Data Bank (PDB) files (files containing the coordinates of C_α-atoms only were not used). The following enzymes were investigated: TADH from methylotropic bacterium W3A1 (Barber et al., 1992) (PDB entry, 2TMD); OYE from *Saccharomyces cerevisiae* (Fox & Karplus, 1994); CGT from *Bacillus circulans* (Lawson et al., 1994) (1CDG); AAMY from *A. oryzae* (Matsuura et al., 1984) (2TAA); endoglucanase (EGLA) from *Clostridium thermocellum* (Dominguez et al., 1995); OGLU from *B. cereus* (Kizaki et al., 1993); barley (1 → 3,1 → 4)-β-glucanase (BGLA) (Varghese et al., 1994); cellobiohydrolase II (CBH) from *Trichoderma reesei* (Rouvinen et al., 1990); urease (URE) from *Klebsiella aerogenes* (Jabri et al., 1995); mouse adenosine deaminase (ADA) (Wilson et al., 1991); ribulose-1,5-bisphosphate carboxylase/oxygenase (RBCO) from *Rhodospirillum rubrum* (Schneider et al., 1990) (5RUB); indole-3-glycerolphosphate synthase (IGPS) from *Escherichia coli* (Wilmanns et al., 1992) (1PII); enolase (ENOL) from *S. cerevisiae* (Lebioda et al., 1989)

Table 1. *Sequence stretches around the strand β2 of (α/β)$_8$-barrel enzymes from the α-amylase family*

EC	Enzyme	Sequence[a]	Ac. no.[b]
3.2.1.1	α-Amylase	56_GFTAIWITP	P10529
3.2.1.10	Oligo-1,6-glucosidase	44_GIDVIWLSP	P21332
3.2.1.20	α-Glucosidase	60_GMDGVWLSP	P13080
3.2.1.41	Pullulanase	458_GVTHVELLP	P07206
3.2.1.1/41	α-Amylase-pullulanase	435_GISVIYLNP	P16950
3.2.1.54	Cyclomaltodextrinase	197_GVNALYFNP	X62576
3.2.1.60	Maltotetraohydrolase	50_GFSAIWMPVP	P22963
3.2.1.68	Isoamylase	277_GVTAVEFLP	P10342
3.2.1.70	Dextran glucosidase	44_GVMAIWLSP	M30944
3.2.1.93	Trehalose-6-phosphate hydrolase	46_GVDAIWLTP	P28904
3.2.1.98	Maltohexaohydrolase	38_GITAVWIPP	P19571
3.2.1.116	Maltotriohydrolase	35_GVSAIWIPQP	D26510
3.2.1.133	Maltogenic amylase	65_GVTTIWLSP	P19531
3.2.1.135	Neopullulanase	186_GVTALYFTP	D13178
3.2.1.-	Maltopentaohydrolase	31_GFAAVQISP	D10769
2.4.1.18	Glycogen branching enzyme	314_GYSHIELLP	P16954
2.4.1.19	Cyclodextrin glycosyltransferase	70_GVTAIWISQP	X78145
2.4.1.25	Glycogen debranching enzyme	177_GYMNIHFTP	L10605

[a] The start of the segment in the amino acid sequence is given. The invariant glycine and proline residues are in bold.
[b] Accession numbers from SwissProt and GenBank Sequence Databases. All SwissProt database numbers start with P.

(7ENL); α-subunit of tryptophan synthase (TSA) from *Salmonella typhimurium* (Hyde et al., 1988) (1WSY); MR from *Pseudomonas putida* (Neidhart et al., 1991) (2MNR); xylose isomerase (XYI) from *Streptomyces olivochromogenes* (Farber et al., 1989) (3XIA); and *N*-(5′-phosphoribosyl)anthranilate isomerase (PRAI) from *E. coli* (Wilmanns et al., 1992) (1PII).

For initial structural comparison of the β2-strand segments, the program HYPERCHEM (Autodesk, Inc.) was used. To compare the stretches structurally in details, the CCP4 software (CCP4, 1994) was used. The β2-strands of AAMY and CGT, AAMY and MR, and AAMY and TSA were superimposed using least-squares minimization, the values of RMS and maximum deviations between corresponding C_α atoms being calculated using LSQAB program (CCP4, 1994). For plotting figures of the overlapped protein segments, the program PLUTO (CCP4, 1994) was used.

Results and discussion

Theoretical support for the proposed homology

The question of the evolution of (α/β)$_8$-barrel enzymes is hazy mainly due to the general lack of sequence homologies or, at least, similarities. There is no easy way to trace the homologies (similarities) as the structurally satisfactory alignments (see, e.g., Pickett et al., 1992; Sergeev & Lee, 1994). The necessity of a slightly unusual and different approach in the search for sequence regions important from the evolutionary point of view has been indicated recently (Janeček, 1995). It takes into account the very long evolutionary history of the family of (α/β)$_8$-barrels. The result of this divergent process could be the adoption of different structural positions and functional roles in the structures of the present-day (α/β)$_8$-barrel enzymes by the amino acid residues that were homologous in a primordial barrel. Therefore, the aligned homologies (similarities) need not be fully structurally satisfactory. The description of the invariant glutamates near the C-terminus of the fifth β-strand that might have adopted different functional roles in more than 10 various (α/β)$_8$-barrel enzymes has been manifested already (Janeček & Baláž, 1995). The aligned sequence stretches presented here (Table 2) are proposed to represent the other evolutionarily conserved segment of the (α/β)$_8$-barrel ancestor's polypeptide chain preceded and followed by a glycine and a proline, respectively, the prolines being considered as adopting different structural positions in the investigated (α/β)$_8$-barrel enzymes. Note that a sequence stretch similar to the template β2-strand of AAMY has appeared also in another region of an (α/β)$_8$-barrel protein. This interesting observation of the similarity between the second β-strand of AAMY and the strand β4 of glycolate oxidase, flavocytochrome b_2, and TSA has been discussed recently (Janeček, 1995). The fact that TSA contains both β-strands (β2 and β4) flanked in loops by Gly and Pro (Hyde et al., 1988; Janeček, 1995; cf. Table 2) should be taken into account.

The identification of the eventual homology comprising the β2-strands is based on the simple idea (Janeček, 1993) that highly conserved sequence regions of a particular (α/β)$_8$-barrel enzyme should be more or less conserved in the sequences of the other evolutionarily related enzymes. The second β-strand of the barrel of AAMY and related starch hydrolases (Table 1) belongs, indeed, to the best conserved regions of this subfamily of (α/β)$_8$-barrel enzymes. It is, in fact, easily localized in the sequence due to characteristic length of the peptide stretch (7 or 8 residues) between the invariant glycines and prolines. There are only a few cases of a glycine substitution by an alanine when comparing more than 100 different sequences of these starch hydrolases and related enzymes (Š. Janeček, unpubl. results).

For these reasons, the results summarized in Table 2 should be of interest. Remarkably, more than 15 different (α/β)$_8$-barrel enzymes have their strand β2 flanked in loops by a glycine and a proline. And many other (α/β)$_8$-barrels contain either a glycine in the loop connecting the β2-strand to the preceding helix α1 or a proline in the loop connecting the β2-strand to the fol-

Table 2. *Sequence stretches around the strand β2 of different $(\alpha/\beta)_8$-barrel enzymes*

EC[a]	Enzyme[b]	Sequence[c]	Length
1.5.99.7	TADH	53_GWAALNTEYCSINP	14
1.6.99.1	OYE	65_GTMIITEGAFISP	13
2.4.1.19	CGT	70_GVTAIWISQP	10
3.2.1.1	AAMY	56_GFTAIWITP	9
3.2.1.4	EGLA	41_GFDHVRLPFDYP	12
3.2.1.10	OGLU	44_GIDVIWLSP	9
3.2.1.73	BGLA	26_GIKSMRLYAP	10
3.2.1.91	CBH[d]	160_GNYAGQFVVYDLP	13
3.5.1.5	URE	150_GVTTMVGGGTGP	12
3.5.4.4	ADA[e]	94_GVVYVEVRYSP	11
4.1.1.39	RBCO	186_GGDFIKNDEP	10
4.1.1.48	IGPS[f]	92_GASAISVLTEP	11
4.2.1.11	ENOL[g]	162_GALALQEFMIAP	12
4.2.1.20	TSA	44_GADALELGVP	10
5.1.2.2	MR	159_GFRAVKTKIGTP	12
5.3.1.5	XYI	46_GAHGVTFHDDDLIP	14
5.3.1.16	PRAI	275_GAIYGGLIFVATSP	14

[a] Enzymes are ordered according to their EC numbers.

[b] Full names of the enzymes and their sources can be found in the Abbreviations and Background sections, respectively.

[c] The start of the segment in the amino acid sequence is given. The invariant glycine and proline residues are in bold. Residues forming the strand β2 are underlined.

[d] A barrel equivalent to the fold of CBH II is present in the structure of endocellulase E2 (EC 3.2.1.4) from *Thermomonospora fusca* (Spezio et al., 1993) with the sequence around the strand β2 65_GKIPILVV YNAP.

[e] The exact position of the second β-strand in the ADA sequence was not available either in the literature (Wilson et al., 1991) or from the PDB file (entry 1ADA).

[f] The β2-strand region of *E. coli* IGPS is not flanked by Gly and Pro (Wilmanns et al., 1992). Here the equivalent stretch of *S. cerevisiae* IGPS is shown, the structure of which has not been solved yet.

[g] The strand β2 of ENOL is oriented in an antiparallel fashion to the rest of the barrel strands (Lebioda et al., 1989).

lowing helix α2 (Table 3). There are only several $(\alpha/\beta)_8$-barrel enzymes that do not have the sequence either starting with Gly or terminating with Pro at (or around) their second β-strand. These are flavocytochrome b_2 (Xia & Mathews, 1990), luciferase (e.g., Fisher et al., 1995), xylanase (e.g., Harris et al., 1994), chitinase (e.g., Perrakis et al., 1994), endo-β-*N*-acetylglucosaminidase (Van Roey et al., 1994), and fructose-1,6-bisphosphate aldolase (e.g., Sygusch et al., 1987).

The length of the segment between the glycine and the proline varies from 7 to 12 residues. The fact that the regions are not structurally fully satisfactory could be explained by the long evolutionary history of the barrels during which the primordial sequence homologies were destroyed by the number of insertions, deletions, and substitutions resulting in adoption of different structural positions by the previously equivalent proline residues. In these terms, the absence of a sequential homology throughout the stretches presented in Table 2 should not be surprising. Nevertheless, there are only hydrophobic amino acid residues in the fifth position of all these enzymes except for CBH and PRAI (both have glycines), this hydrophobic place being preceded by alanines in TADH, CGT, AAMY, CBH, IGPS, ENOL, TSA, and MR. It is possible to align the stretches listed in Table 2 manually in order to start with glycines and to end with prolines, but the presented segments are too short for doing it with justification.

Structural support for the proposed homology

If the structural similarity between the β2-strands exists, it cannot comprise all the enzymes listed in Table 2. This is due to the fact that the enzymes belong to different families of all $(\alpha/\beta)_8$-barrels (see the division of this enzyme family given by Farber & Petsko, 1990). Importantly, in agreement with this division, a clear trend can be seen in the structural comparison of the β2-strands, i.e., the similarity is higher for the enzymes that are more closely related and is lower for the enzymes that are more distantly related (Figs. 1, 2, 3). This is supported namely by the statistical data (Table 4) that clearly reflect the facts that: (1) AAMY is in one close subfamily with CGT (Fig. 1; RMSD 0.695 Å); (2) AAMY is in one broader family with MR (Fig. 2; RMSD 1.150 Å); and (3) AAMY is in different family than TSA (Fig. 3; RMSD 1.484 Å).

As far as the β2-strand segments of the other enzymes from Table 2 are concerned, mostly they do not fit each other in such a way as AAMY, CGT, MR, and TSA do. It should be pointed out, however, this was not expected at all because the structurally based approaches (e.g., Pickett et al., 1992; Sergeev & Lee, 1994), producing the sequence alignments that are satisfactory from the structural point of view, also did not detect the homologies. The fact that there is an analogical situation with the β5-strand segments comprising the invariant and functionally essential glutamate residues in a similar set of $(\alpha/\beta)_8$-barrels (Janeček & Baláž, 1995) also supports the idea that the sequence homologies important from the ancient point of view exhibit in the present-day $(\alpha/\beta)_8$-barrel enzymes some destruction of the secondary structure equivalence that was probably present in the primordial barrel(s). And finally, this is consistent with the proposal that there could be a family of enzymes whose members have diverged so much that no detectable sequence homologies are left (Lesk & Boswell, 1992).

Conclusion

It is postulated that the sequence similarities or, perhaps, homologies among different $(\alpha/\beta)_8$-barrel enzymes that would be structurally fully satisfactory cannot be expected. This is supported by the fact that the structurally based approaches (Pickett et al., 1992; Sergeev & Lee, 1994) offer satisfactory alignments, but they do not necessarily bring the homologies. It seems possible, however, that sequence homologies do occur in $(\alpha/\beta)_8$-barrel enzymes (Table 2), but, due to their very long evolutionary history and distantly related functions, the homologous amino acid residues from the primordial barrel structure adopted different functional roles (Janeček & Baláž, 1995) and/or different positions in the structure (the prolines in Figs. 1, 2, 3). The homologies are therefore *hidden* in $(\alpha/\beta)_8$-barrel enzymes.

The β2-strand segment presented here could constitute a hidden homology in $(\alpha/\beta)_8$-barrels because: (1) it was revealed by using the conserved sequence region of one of them (AAMY) as structural template; (2) it is characterized by two invariant amino acid residues (Gly and Pro); (3) its length varies from 9 to 14 residues (e.g., insertions and deletions); (4) the similarity between these segments is higher for more closely related

Table 3. *Sequence stretches around the strand β2 of different $(\alpha/\beta b)_8$-barrel enzymes with conserved either Gly or Pro amino acid residue*

EC	Enzyme	Sequence[a]	Reference
Gly			
1.1.1.2	Aldehyde reductase	40_GYRHIDCAAIY	El-Kabbani et al. (1994)
1.1.1.21	Aldose reductase	38_GYRHIDCAH	Harrison et al. (1994)
1.1.1.50	3α-Hydroxysteroid dehydrogenase[b]	45_GFRHFDSAY	Hoog et al. (1994)
1.1.3.15	Glycolate oxidase	100_GTIMTLSS	Lindqvist (1989)
2.7.1.40	Pyruvate kinase	67_GMNVARLNFSH	Muirhead et al. (1986)
3.1.8.1	Phosphotriesterase	65_GVRTIVDVS	Benning et al. (1994)
3.2.1.2	β-Amylase	46_GVDGVMVDVWWG	Mikami et al. (1993)
3.2.1.39	(1→3)-β-Glucanase	26_GINGMRIYF	Varghese et al. (1994)
3.2.1.85	6-Phospho-β-galactosidase	67_GVNGIRISIAWS	Wiesmann et al. (1995)
4.1.2.21	2-Keto-3-deoxy-6-phosphogluconate aldolase	51_GIRTLEVTLRSQ	Mavridis et al. (1982)
4.1.3.3	*N*-Acetylneuraminate lyase	38_GIDGLYVGG	Izard et al. (1994)
4.2.1.52	Dihydrodipicolinate synthase	35_GTSAIVSVG	Mirwaldt et al. (1995)
Pro			
3.2.1.21	β-Glucosidase	84_DMNLDAYRFSISW**P**	Barrett et al. (1995)
3.2.1.23	β-Galactosidase	384_FNAVRCSHY**P**	Jacobson et al. (1994)
5.3.1.1	Triosephosphate isomerase	35_NVEVVIC**PP**	Lolis et al. (1990)
5.5.1.1	Muconate cycloisomerase	162_RHRVFKLKIGANP	Helin et al. (1995)
5.5.1.7	Chloromuconate cycloisomerase	159_HNRFKVKLGFRS**P**	Hoier et al. (1994)

[a] Residues forming the β2-strand are underlined.
[b] The exact position of the second β-strand was not available.

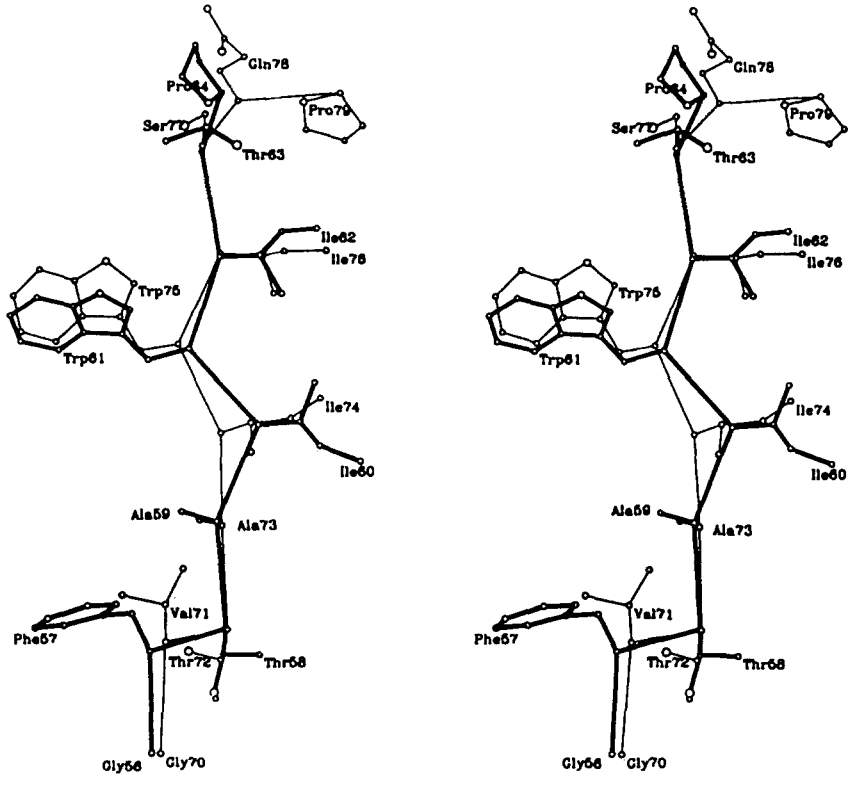

Fig. 1. Overlay of β2-strands of closely related $(\alpha/\beta)_8$-barrel enzymes. The β2-strand of cyclodextrin glycosyltransferase from *B. circulans* strain 251 (thin lines) is overlapped on the β2-strand of α-amylase from *A. oryzae* (thick lines).

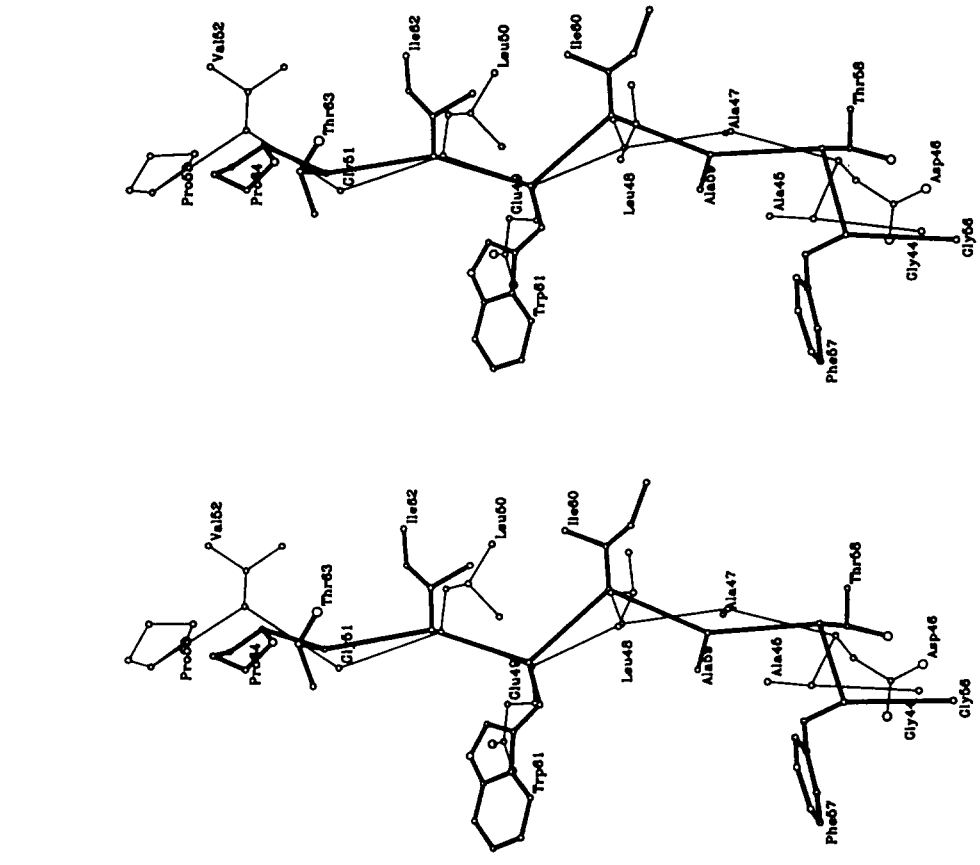

Fig. 3. Overlay of β2-strands of distantly related $(\alpha/\beta)_8$-barrel enzymes. The β2-strand of tryptophan synthase (α-subunit) from *S. typhimurium* (thin lines) is overlapped on the β2-strand of α-amylase from *A. oryzae* (thick lines).

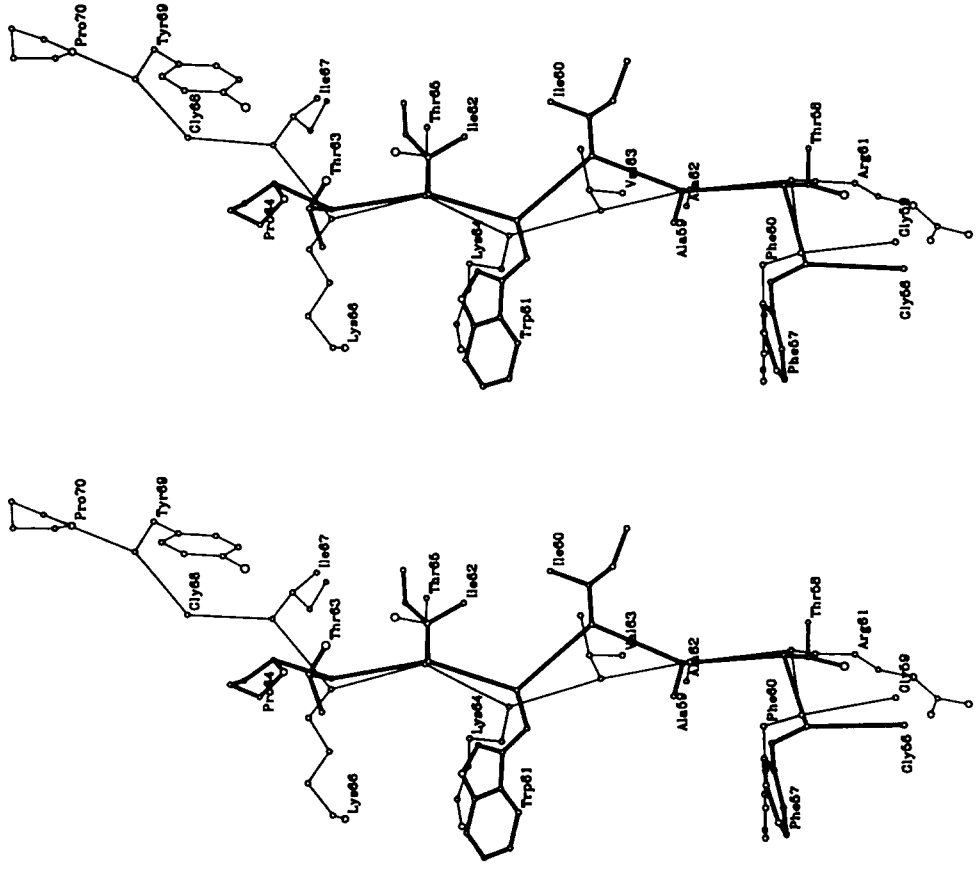

Fig. 2. Overlay of β2-strands of related $(\alpha/\beta)_8$-barrel enzymes. The β2-strand of mandelate racemase from *P. putida* (thin lines) is overlapped on the β2-strand of α-amylase from *A. oryzae* (thick lines).

Table 4. *Statistical values from the superpositions of β2-strands*

Enzyme pair	RMSD (Å)	Maximum deviation (Å)
AAMY & CGT	0.695	1.368
AAMY & MR	1.150	2.061
AAMY & TSA	1.484	2.850

$(\alpha/\beta)_8$-barrel enzymes and is lower for more distantly related ones (Figs. 1, 2, 3; Table 4); and (5) it could be traced, in part, also in the other $(\alpha/\beta)_8$-barrel enzymes (Table 3) in which either the glycine or the proline preceding and succeeding the β2-strand, respectively, is conserved. It is worth mentioning that the visual inspection of many amino acid sequences of several $(\alpha/\beta)_8$-barrel enzymes (e.g., pyruvate kinase [Muirhead et al., 1986], (1 → 3)-β-glucanase [Varghese et al., 1994], triosephosphate isomerase [Lolis et al., 1990]) indicates that the glycine or proline that are in the center of attention of this study are replaced by alanine or serine. Note that the Gly(Pro) → Ala and Gly(Pro) → Ser mutations are controlled by substitutions of only one base.

It seems that the segment **G**-X7 to 12-**P**, comprising the second β-strand of the $(\alpha/\beta)_8$-barrel fold, is a better marker than the catalytic residues that are, in general, localized at different β-strands (see, e.g., Farber & Petsko, 1990; Brändén, 1991). This is consistent with the postulate of divergent evolution that the three-dimensional structure of a protein evolves most slowly, that protein sequence evolves more quickly, and that the chemical mechanism of an enzyme evolves most rapidly. This has been demonstrated recently on the eventual evolution of sugar isomerases (Banerjee et al., 1995). It might be concluded that the regions of hidden homology in $(\alpha/\beta)_8$-barrel proteins should evolve more slowly than the rest of the amino acid sequence (G. Farber, pers. comm.).

Acknowledgments

This work was promoted in part by support from The Slovak Literary Fund. I am grateful to Dr. Štefan Baláž (Department of Biochemical Technology, Slovak Technical University, Bratislava) for the opportunity to use the HYPERCHEM program; Dr. Jozef Ševčík (Institute of Molecular Biology, Slovak Academy of Sciences, Bratislava) for generating the protein structure overlaps using the CCP4 software; Dr. Alex Bateman (MRC Laboratory of Molecular Biology, Cambridge) for providing unpublished results; Dr. Bernard Henrissat (CERMAV, CNRS, Grenoble) for reading the manuscript and fruitful discussion; and Dr. Gregory K. Farber (Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park) for helpful suggestions.

References

Banerjee S, Anderson F, Farber GK. 1995. The evolution of sugar isomerases. *Protein Eng 8*. Forthcoming.

Banner DW, Bloomer AC, Petsko GA, Phillips DC, Pogson CI, Wilson IA, Corran PH, Furth AJ, Milman JD, Offord RE, Priddle JD, Waley SG. 1975. Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 Å resolution using amino acid sequence data. *Nature 255*:609-614.

Barber MJ, Neame PJ, Lim LW, White S, Mathews FS. 1992. Correlation of X-ray deduced and experimental amino acid sequences of trimethylamine dehydrogenase. *J Biol Chem 267*:6611-6619.

Barrett T, Suresh CG, Tolley SP, Dodson EJ, Hughes MA. 1995. The crystal structure of a cyanogenic β-glucosidase from white clover, a family 1 glycosyl hydrolase. *Structure 3*:951-960.

Benning MM, Kuo JM, Raushel FM, Holden HM. 1994. Three-dimensional structure of phosphotriesterase: An enzyme capable of detoxifying nerve agents. *Biochemistry 33*:15001-15007.

Bork P, Gellerich J, Groth H, Hooft R, Martin F. 1995. Divergent evolution of a β/α barrel subclass: Detection of numerous phosphate-binding sites by motif search. *Protein Sci 4*:268-274.

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

CCP4. 1994. Collaborative Computational Project Number 4. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D 50*: 760-763.

Crawford I, Niermann T, Kirschner K. 1987. Prediction of secondary structure by evolutionary comparison: Application to the α subunit of tryptophan synthase. *Proteins Struct Funct Genet 2*:118-129.

Davies G, Henrissat B. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure 3*:853-859.

Dominguez R, Souchon H, Spinelli S, Dauter Z, Wilson KS, Chauvaux S, Beguin P, Alzari PM. 1995. A common protein fold and similar active site in two distinct families of β-glycanases. *Nature Struct Biol 2*:569-576.

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

El-Kabbani O, Green NC, Lin G, Carson M, Narayana SVL, Moore KM, Flynn TG, DeLucas LJ. 1994. Structures of human and porcine aldehyde reductase: An enzyme implicated in diabetic complications. *Acta Crystallogr D 50*:859-868.

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

Farber GK, Glasfeld A, Tiraby G, Ringe D, Petsko GA. 1989. Crystallographic studies of the mechanism of xylose isomerase. *Biochemistry 28*:7289-7297.

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

Fisher AJ, Raushel FM, Baldwin TO, Rayment I. 1995. Three-dimensional structure of bacterial luciferase from *Vibrio harveyi* at 2.4 Å resolution. *Biochemistry 34*:6581-6586.

Fox KM, Karplus PA. 1994. Old yellow enzyme at 2 Å resolution: Overall structure, ligand binding, and comparison with related flavoproteins. *Structure 2*:1089-1105.

Harris GW, Jenkins JA, Connerton I, Cummings N, Lo Leggio L, 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.

Harrison DH, Bohren KM, Ringe D, Petsko GA, Gabbay KH. 1994. An anion binding site in human aldose reductase: Mechanistic implications for the binding of citrate, cacodylate, and glucose 6-phosphate. *Biochemistry 33*:2011-2020.

Helin S, Kahn PC, Guha BL, Mallows DG, Goldman A. 1995. The refined X-ray structure of muconate lactonizing enzyme from *Pseudomonas putida* PRS2000 at 1.85 Å resolution. *J Mol Biol 254*:918-941.

Henrissat B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J 280*:309-316.

Henrissat B, Bairoch A. 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J 293*: 781-788.

Henrissat B, Callebaut I, Fabrega S, Lehn P, Mornon JP, Davies G. 1995. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc Natl Acad Sci USA 92*: 7090-7094.

Hoier H, Schlömann M, Hammer A, Glusker JP, Carrell HL, Goldman A, Stezowski JJ, Heinemann U. 1994. Crystal structure of chloromuconate cycloisomerase from *Alcaligenes eutrophus* JMP134 (pJP4) at 3 Å resolution. *Acta Crystallogr D 50*:75-84.

Hoog SS, Pawlowski JE, Alzari PM, Penning TM, Lewis M. 1994. Three-dimensional structure of rat liver 3α-hydroxysteroid/dihydrodiol dehydrogenase: A member of the aldo-keto reductase superfamily. *Proc Natl Acad Sci USA 91*:2517-2521.

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.

Jabri E, Carr MB, Hausinger RP, Karplus PA. 1995. The crystal structure of urease from *Klebsiella aerogenes*. *Science 268*:998–1004.

Jacobson RH, Zhang XJ, DuBose RF, Matthews BW. 1994. Three-dimensional structure of β-galactosidase from *E. coli*. *Nature 369*:761–766.

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

Janeček Š. 1995. Similarity of different β-strands flanked in loops by glycines and prolines from distinct $(\alpha/\beta)_8$-barrel enzymes: Chance or a homology? *Protein Sci 4*:1239–1242.

Janeček Š, Baláž Š. 1993. Evolution of parallel β/α-barrel enzyme family lightened by structural data on starch-processing enzymes. *J Protein Chem 12*:509–514.

Janeček Š, Baláž Š. 1995. Functionally essential, invariant glutamate near the C-terminus of strand β5 in various $(\alpha/\beta)_8$-barrel enzymes as a possible indicator of their evolutionary relatedness. *Protein Eng 8*:809–813.

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.

Jenkins J, Lo Leggio L, Harris G, Pickersgill R. 1995. β-Glucosidase, β-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8-fold β/α architecture and with two conserved glutamates near the carboxy-terminal ends of β-strands four and seven. *FEBS Lett 362*:281–285.

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, Brewer JM. 1989. The structure of yeast enolase at 2.25-Å resolution. An 8-fold $\beta+\alpha$-barrel with a novel $\beta\beta\alpha\alpha(\beta\alpha)_6$ topology. *J Biol Chem 264*:3685–3693.

Lesk AM, Boswell DR. 1992. Does protein structure determine amino acid sequence? *BioEssays 14*:407–410.

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.

Lolis E, Alber T, Davenport RC, Rose D, Hartman FC, Petsko GA. 1990. Structure of yeast triosephosphate isomerase at 1.9-Å resolution. *Biochemistry 29*:6609–6618.

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.

Mavridis IM, Hatada MH, Tulinsky A, Lebioda L. 1982. Structure of 2-keto-3-deoxy-6-phosphogluconate aldolase at 2.8 Å resolution. *J Mol Biol 162*:419–444.

Mikami B, Hehre EJ, Sato M, Katsube Y, Hirose M, Morita Y, Sacchettini JC. 1993. The 2.0-Å resolution structure of soybean β-amylase complexed with α-cyclodextrin. *Biochemistry 32*:6836–6845.

Mirwaldt C, Korndörfer I, Huber R. 1995. The crystal structure of dihydrodipicolinate synthase from *Escherichia coli* at 2.5 Å resolution. *J Mol Biol 246*:227–239.

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.

Neidhart DJ, Howell PL, Petsko GA, Powers VM, Li R, Kenyon GL, Gerlt JA. 1991. Mechanism of the reaction catalyzed by mandelate racemase. 2. Crystal structure of mandelate racemase at 2.5-Å resolution: Identification of the active site and possible catalytic residues. *Biochemistry 30*:9264–9273.

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

Niermann T, Kirschner K. 1995. The predicted secondary structure of the G-type glutamine amidotransferase is compatible with TIM-barrel topology. *Protein Eng 8*:535–542.

Orengo CA, Jones DT, Thornton JM. 1994. Protein superfamilies and domain superfolds. *Nature 372*:631–634.

Perrakis A, Tews I, Dauter Z, Oppenheim AB, Chet I, Wilson KS, Vorgias CE. 1994. Crystal structure of a bacterial chitinase at 2.3 Å resolution. *Structure 2*:1169–1180.

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.

Raine ARC, Scrutton NS, Mathews FS. 1994. On the evolution of alternate core packing in eightfold β/α-barrels. *Protein Sci 3*:1889–1892.

Reardon D, Farber GK. 1995. The structure and evolution of α/β barrel proteins. *FASEB J 9*:497–503.

Rouvinen J, Bergfors T, Teeri T, Knowles JKC, Jones TA. 1990. Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science 249*:380–386.

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.

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

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

Spezio M, Wilson DB, Karplus PA. 1993. Crystal structure of the catalytic domain of a thermophilic endocellulase. *Biochemistry 32*:9906–9916.

Sygusch J, Beaudry D, Allaire M. 1987. Molecular architecture of rabbit skeletal muscle aldolase at 2.7-Å resolution. *Proc Natl Acad Sci USA 84*: 7846–7850.

Van Roey P, Rao V, Plummer TH Jr, Tarentino AL. 1994. Crystal structure of endo-β-*N*-acetylglucosaminidase F_1, an α/β-barrel enzyme adapted for a complex substrate. *Biochemistry 33*:13989–13996.

Varghese JN, Garrett TPJ, Colman PM, Chen L, Hoj PB, Fincher GB. 1994. Three-dimensional structures of two plant β-glucan endohydrolases with distinct substrate specificities. *Proc Natl Acad Sci USA 91*:2785–2789.

Wiesmann C, Beste G, Hengstenberg W, Schulz GE. 1995. The three-dimensional structure of 6-phospho-β-galactosidase from *Lactococcus lactis*. *Structure 3*:961–968.

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, Nierman 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.

Wilson DK, Rudolph FB, Quiocho FA. 1991. Atomic structure of adenosine deaminase complexed with a transition-state analog: Understanding catalysis and immunodeficiency mutations. *Science 252*:1278–1284.

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