Protein Science (1999), *8*:2072–2084. Cambridge University Press. Printed in the USA.

Barrel structures in proteins: Automatic identification and classification including a sequence analysis of TIM barrels

NOZOMI NAGANO,[1] E. GAIL HUTCHINSON,[1] AND JANET M. THORNTON[1,2]

[1]Biomolecular Structure and Modeling Group, Biochemistry & Molecular Biology Department, University College London, Gower Street, London WC1E 6BT, United Kingdom
[2]Crystallography Department, Birkbeck College, Malet Street, London WC1E 7HX, United Kingdom

(RECEIVED September 15, 1999; ACCEPTED August 13, 1999)

Abstract

Automated methods for identifying and characterizing regular β-barrels from coordinate data have been developed to analyze and classify various kinds of barrel structures based on geometric parameters such as the barrel strand number (n) and shear number (S). In total, we find 1,316 barrels in the January 1998 release of Protein Data Bank. Of 1,316 barrels, 1,277 barrels had an even shear number, corresponding to 50 nonhomologous families. The $(\beta\alpha)_8$ triose phosphate isomerase (TIM) barrel ($n = 8$, $S = 8$) fold has the largest number of apparently nonhomologous entries, 16, although the trypsin like antiparallel ($n = 6$, $S = 8$) barrels (representing only three families) are the most common with 527 barrels. Of all the protein families that exhibit barrel structures, 68% are found to be various kinds of enzymes, the remainder being binding proteins and transport membrane proteins. In addition, the layers of side chains, which form the cores of barrels with $S = n$ and $S = 2n$, are also analyzed. More sophisticated methods were developed for detecting TIM barrels specifically, including consideration of the amino acid propensities for the side chains that form the layers. We found that the residues on the outside of the eight stranded parallel β-barrel, buried by the α-helices, are much more hydrophobic than the residues inside the barrel.

Keywords: automatic detection; β-barrels; geometric parameters; protein structures; TIM barrels

Many proteins in the Brookhaven Protein Data Bank (PDB) (Bernstein et al., 1977) contain β-sheets, both parallel and antiparallel, which fold into a "closed" barrel. These barrel structures can be described using geometric parameters such as the number of strands and the so-called shear number, which is a measure of staggering of strands around the barrel (McLachlan, 1979; Lesk et al., 1989; Murzin et al., 1994a, 1994b). Previous analyses on relatively few examples of barrel structures were mainly performed by visual inspection (McLachlan, 1979; Lesk et al., 1989; Murzin et al., 1994a, 1994b). However, now there are more than 1,000 barrels in the PDB, so analyzing them all by visual inspection would be extremely difficult and time consuming.

For this reason, automated methods for identifying and characterizing β-barrels have been developed, based on geometric parameters. The layers of side chains, which form the core of the barrel, are also identified and more sophisticated methods for detecting and analyzing triose phosphate isomerase (TIM) barrels specifically were also developed.

Reprint requests to: Janet M. Thornton, Biomolecular Structure and Modeling Group, Biochemistry & Molecular Biology Department, University College London, Gower Street, London WC1E 6BT, United Kingdom; e-mail: thornton@biochem.ucl.ac.uk.

Definitions

Complete barrels and distorted barrels

Based on secondary structure information, barrel structures can be classified into two types: "complete barrels" and "distorted barrels." "Complete barrels" can be identified simply based on a complete ring of hydrogen bonds in the derived secondary structure information. Here, hydrogen bonds are calculated using SSTRUC, a local implementation (D.K. Smith, unpubl. data) of DSSP (Kabsch & Sander, 1983). In contrast, it is very difficult to detect "distorted barrels," because hydrogen bonds are not formed between some barrel strands, or some strands are not detected at all by the algorithm. In the present paper, only "complete barrels" are analyzed and classified.

Exterior and interior residues on the barrel strands

The direction of side chains of residues on the strands is calculated based on the direction of the vector from the Cα atom to the Cβ atom. For glycine, a dummy Cβ atom was generated. The barrel axis was calculated so that it goes through the center of the barrel, which is the average position of Cα atoms for the middle residue on each barrel strand, and the center of the barrel bottom, which is the average position of Cα atoms for the N-terminal

residue on each barrel strand. The angle made between the vector perpendicular to the barrel axis through the axis and the Cα atom and the vector from the Cα atom to the Cβ atom was calculated. Where the angle is between 0 and 90°, the residue is defined to be an "interior residue." Where the angle is between 90 and 180°, the residue is defined to be an "exterior residue." The "exterior residues" and "interior residues" are represented by ovals and rectangles, respectively, in Figure 1.

Barrel strand number n and the shear number S

There are two numbers used to describe the geometry of barrel structures. One is the number of strands making up the barrel structures (n); the other is the shear number (S), proposed by McLachlan (1979). This number is a measure of the extent to which the β-sheet is staggered. By rolling out the barrel, the hydrogen bonding pattern can be drawn (see Fig. 1). In the case of the left barrel of Figure 1, nine strands are shown, as the first strand is drawn on both edges of the sheet. The residues can be thought of as lying at the grid points of a virtual lattice. Starting from the first residue (Phe6) in strand 1, it is possible to move around the barrel until strand 1 is reached again. Here, the position of the first residue in the strand 1′ (Phe6) is now shifted by eight residues from its original position on strand 1 of the virtual lattice. The shear number for this barrel is eight. Since consecutive residues along a strand alternate between the two sides of the sheet, and corresponding residues on adjacent strands are on the same side of the sheet, the shear number should theoretically be an even integer.

Once calculated, the n and S values can be used to determine other geometrical parameters, such as strand tilt relative to the barrel axis and barrel radius (McLachlan, 1979; Lesk et al., 1989; Murzin et al., 1994a, 1994b).

An extended definition of layer structure

If S is equal to n or $2n$, particular sets of residues on the barrel lie in planes, perpendicular to the barrel axis, and form stacked "layers" (Lesk et al., 1989; Murzin et al., 1994a, 1994b) (see Fig. 1). However, in the case of $n < S < 2n$, such sets perpendicular to the barrel axis could not be identified.

Previously (Lesk et al., 1989), only interior residues involved in barrel packing were defined as "layers" of TIM barrel structures. However, exterior as well as interior residues lie in planes. Where $S = n$, a layer includes alternating interior and exterior residues. Where $S = 2n$, layers include only all interior or all exterior residues, and adjacent layers alternate. Here, "semi-complete layers," in which the direction of glycine is allowed to be flipped from interior to exterior or vice versa, are also detected.

Results and discussion

The classification of complete barrel structures based on their bifurcation

β-strands can be divided into three groups, based on the number of strands to which they are hydrogen bonded. "Usual strands," indicated by unlabeled boxes in Figure 2A, are hydrogen bonded to two adjacent strands. "Edge strands" (indicated by boxes labeled "E") are hydrogen bonded to only one adjacent β-strand, and "bifurcated strands" ("B") have more than two adjacent strands. Typical complete barrel structures are composed of only "usual

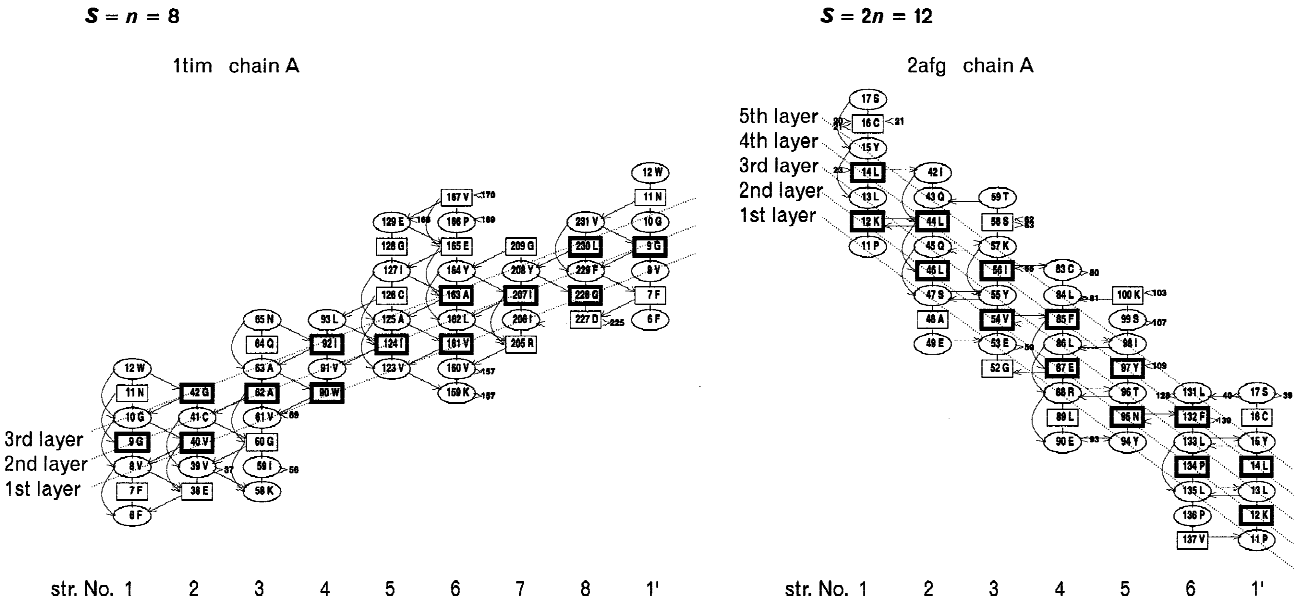

Fig. 1. Definition of the number of barrel strand n, shear number S, and the layers of barrel structures. The hydrogen bonding patterns obtained by rolling out barrel sheets are drawn using the HERA algorithm (Hutchinson & Thornton, 1990). The first strand (1) is repeated at the opposite end of the sheet (1′). Straight arrows represent main-chain hydrogen bonds. Ovals and rectangles represent exterior and interior residues, respectively. To calculate the shear number, start from the first residue in strand 1, move around the barrel until strand 1 is reached again. Here, the position of the first residue in strand 1′ is shifted by S residues from the original position on strand 1. The layers are represented by dotted lines. The interior residues on the layers, which are involved in the core of barrel, are indicated by the thick boxes. Chain A of 1tim [$(n,S) = (8,8)$] (triose phosphate isomerase; Banner et al., 1976) ($S = n$) (left) and chain A of 2afg [$(n,S) = (6,12)$] (acidic fibloblast growth factor; Blaber et al., 1996) ($S = 2n$) (right).

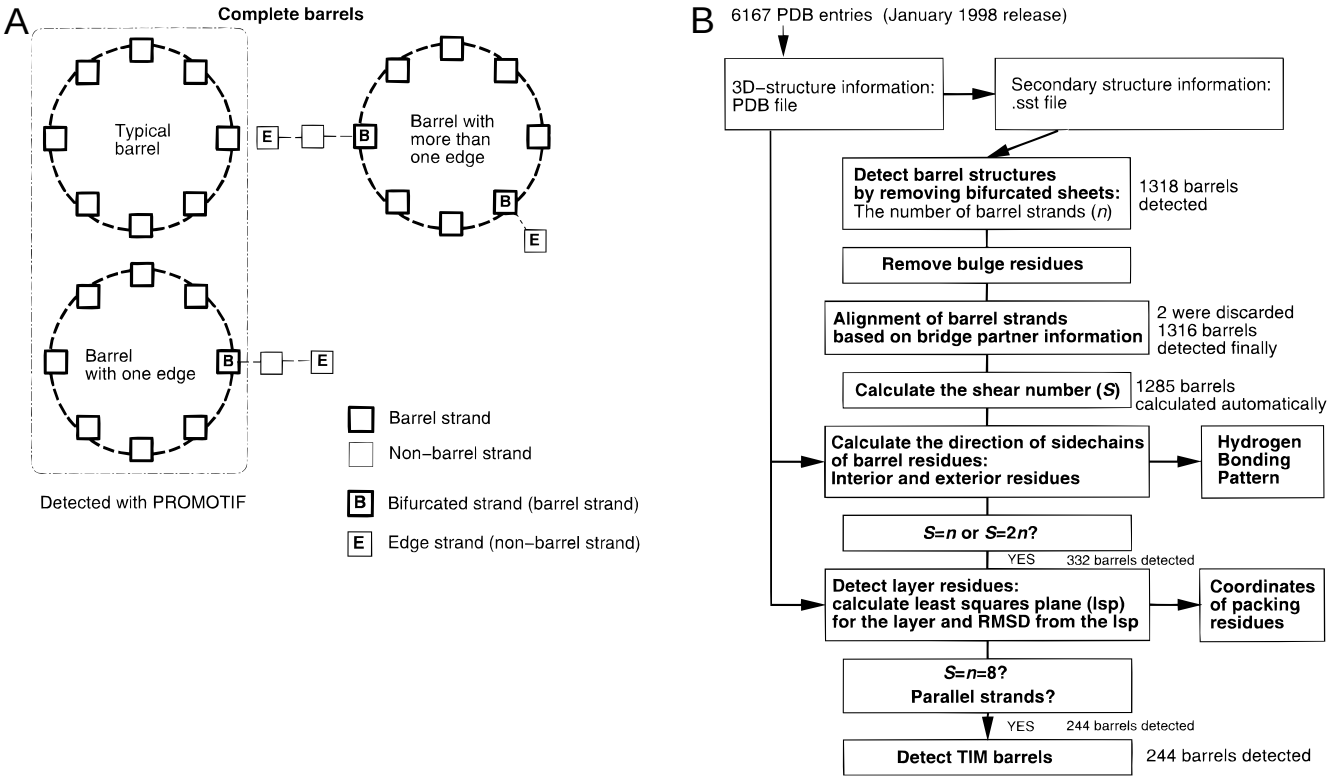

Fig. 2. A: The classification of "complete barrel" structures based on the bifurcation. "Edge strands," which are hydrogen bonded to only one adjacent β-strand, are indicated by boxes with letter "E," and their corresponding "bifurcated strands," which have more than two adjacent strands, are indicated by the box with letter "B." β-strands are classified into three types based on the number of adjacent strands to which the strands are hydrogen bonded. "Usual strands," which are hydrogen bonded to two adjacent strands, are indicated by boxes without any letter. "Typical complete barrel" structures are composed of only "barrel strands" indicated by thick boxes, all of which are "usual strands" with two adjacent strands. Barrel structures, which include "bifurcated strands," correspondingly have "edge strands," and they are classified based on the number of bifurcations. **B:** The flow diagram for detection of barrel structures in the present system.

strands." However, many barrel structures have at least one "bifurcated strand" and correspondingly an "edge strand."

The present procedure is shown in Figure 2B. When applied to all 6,167 PDB entries, excluding those with only Cα atoms and theoretical models, our refined procedure identifies 1,318 in total at this stage (see Table 1; Fig. 2B). Fifty two and three-tenths percent of these barrels were found to have bifurcated strands, and there were further classified according to the number of bifurcated strands (Table 1). Most have just one bifurcated strand (Table 1).

Calculation of shear number S

Shear numbers for all detected barrels were calculated (Table 2). The numbers were determined after removing bulge residues as detected by our algorithm (Chan et al., 1993). Thirty-three barrels had very complicated structures, and *S* and *n* had to be calculated manually by checking the hydrogen bonding patterns. Two of these 33 [thiol ester dehydrolase (PDB code, 1mka; Leesong et al., 1996), histocompatibility antigen (PDB code, 1roj; Rognan et al., 1994)] were finally eliminated as nonbarrel structures (Table 2). Seventeen barrels [e.g., cholera toxin B (PDB code, 1chp; Merritt et al., 1995), and enterotoxin (PDB code, 1tii; van den Akker et al., 1996)] had such complicated structures that their shear numbers could not be calculated (Table 2), although they do have complete rings of hydrogen bonds (data not shown). The shear numbers for the remaining 14 barrels [xylanase A (PDB code, 1clx; Harris et al., 1996), complex of viral proteins (PDB code, 1jxp; Y. Yan, S. Munshi, Y. Li, V. Sardana, J. Blue, B. Johns, J. Cole, C. Steinkueler, L. Tomei, R.D. Francesco, et al., in prep.), DNA-binding protein (PDB code, 1kaw; Raghunathan et al., 1997)] could only be calculated manually (Table 2).

Of 1,285 barrels calculated automatically, 1,277 had an even shear number, as predicted theoretically, while 8 barrels were found to have odd shear numbers. One example of a barrel with odd shear number is methylmalonyl-CoA mutase (PDB code, 1req; Mancia et al., 1996), a TIM barrel that has a shear number of 9. This value agrees with an independent study by Liu (1998). In this barrel, the hydrogen bonding between the 4th and 5th strands is so irregular that an odd shear number can occur (data not shown).

Most of our results are consistent with a previous study by Liu (1998), but for 20 of the 69 PDB files in his analysis, we obtained different shear numbers (data not shown). The shear numbers obtained by Liu are quite often odd integers. Our shear numbers were usually larger by one or two residues (data not shown). However, in four families, the differences were even greater [1prt (three), 1eag (four), 1efu (three), and 2mpr (three)] (1prt, pertussis toxin, Stein et al., 1994; 1eag, aspartic proteinase, Cutfield et al., 1995; 1efu, elongation factor, Kawashima et al., 1996; 2mpr, maltoporin,

Table 1. *Bifurcation of detected barrel structures*[a]

No. of bifurcated strands	No. of barrels	%	PROMOTIF	Examples[b]
0	629	47.7	✓	
1–6	689	52.3	✓	
1	513	38.9	✓	
2	149	11.3	X	
3	5	0.4	X	1ecp[*1]
4	0	0.0	—	
5	21	1.6	X	1chp[*2], 1gtp[*3]
6	1	0.1	X	1tdt[*4]

[a]Previously, 1,015 complete barrels were identified in the PDB entries using PROMOTIF. With the extended secondary structure definitions applied to all PDB files, we found an extra 303 barrels, giving 1,318 barrels in total.

[b]*1 = purine nucleoside phosphorylase; *2 = cholera toxin B; *3 = GTP cyclohydrolase I; *4 = N-succinyltransferase. (*The references are shown in each PDB file.)

Meyer et al., 1997). In particular, the first three of these contain complex bulges composed of three or four residues. The schematic diagram (Fig. 3) illustrates the calculation of two possible values for the shear number according to Liu's method. He defines shear numbers to be the change of residue numbers in a closed path. One path, avoiding all the bulges, is shown by the broken line yields a value $S_1 = 10$. The second path including the special bulge (Chan et al., 1993) is shown by the solid line and yields the value $S_2 = 6$. Liu quotes 6 as the minimal shear number of the barrel, while our method, based on the relative position of the residues on the staggered barrel, calculates 10. The latter is more consistent with the work of Murzin et al. (1994a), who argued that a barrel with $(n, S) = (6,6)$ would have too small a radius to accommodate the side chains. Considering that the observed average radius is 7.0 Å and the tilt of strands relative to the barrel axis for 1eag is 54°, this barrel obviously belongs to the (6,10) barrel group with theoretical

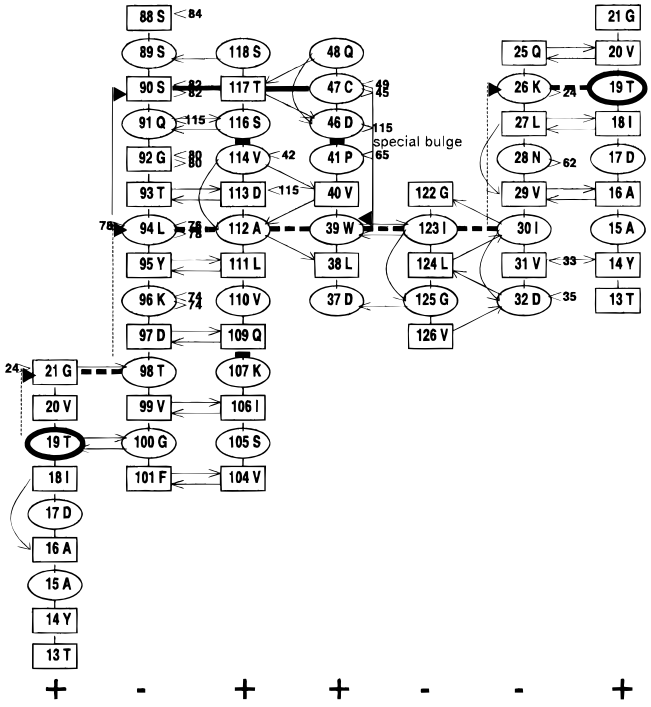

Fig. 3. Comparison of shear numbers for aspartic proteinase (PDB code 1eag; Cutfield et al., 1995) obtained in this work and by Liu (1998). Ovals and rectangles represent exterior and interior residues, respectively, as in Figure 1. The bulge residues are indicated by a bold short line. In his method, the shear numbers were defined as the change of residue numbers in a closed path. If all bulges are avoided starting from residue 19, indicated by bold oval, the shortest path could be the one through the residues as (21,98; 94,112; 112,39; 39,123; 123,30; 26,19), which is shown by the broken line. For this path, starting from the residue 19, considering the relative direction of each strand to that of the first strand, indicated by the signal of + and − below the hydrogen bonding pattern, the shear number could be calculated as $S_1 = (21–19) - (94–98) + (112–112) + (39–39) - (123–123) - (26–30) = 10$, which is consistent with our result. However, for the path that goes via the solid line and includes the special bulge from 42 to 45 (21, 98; 90, 117; 117, 47; 39, 123; 123, 30; 26, 19), the shear number is $S_2 = 2 - (-8) + 0 + (-8) - 0 - (-4) = 6$, which is as obtained by Liu.

Table 2. *Shear numbers of barrel structures*

Shear number	No. of barrels	Examples[a]
Calculated automatically	1,285	
Even numbers	1,277	
Odd numbers	8	1tox[*1]; $(n,S) = (5,11)$, 3fiv[*2]; $(n,S) = (6,9)$ 1req[*3]; $(n,S) = (8,9)$ (TIM barrel) 1ucw[*4]; $(n,S) = (8,7)$ (TIM barrel)
Not calculated automatically	33	
Calculated manually	14	1clx[*5]; $(n,S) = (8,8)$ (TIM barrel) 1jxp[*6]; $(n,S) = (6,8)$ 1kaw[*7]; $(n,S) = (5,10)$ (two barrels in one sheet)
Impossible to calculate shear number even manually	17	1chp[*8], 1tii[*9]
Identified as nonbarrels	2	1mka[*10]

[a]*1 = diphtheria toxin; *2 = acid proteinase; *3 = methylmalonyl-CoA mutase; *4 = transaldolase; *5 = xylanase A; *6 = viral proteins; *7 = DNA-binding protein; *8 = cholera toxin B; *9 = enterotoxin; *10 = thiol ester dehydrolase. (*The references are shown in each PDB file.)

radius, 7.0 Å, and tilt, 51°, rather than the (6,6) group with 5.0 Å and 37°.

Classification of complete barrels with even shear numbers based on (n,S)

Since the strand number n and shear number S are related to the geometry of the barrels, barrel structures can be classified according to these two numbers. Table 3A and Figure 4 show such a classification of all the barrels with even shear numbers.

The group with $n = 6$ and $S = 8$ is the most popular with 527 members. Most of these were members of the serine proteinase family. The second and third most common are the groups with $(n,S) = (6,10)$ and $(n,S) = (8,8)$, respectively. The group with $(n,S) = (8,8)$ corresponds to the TIM barrel family. The groups with either $n = 16$ or $n = 18$ contain membrane barrel proteins with very large theoretical radii, 15.5 and 17.2 Å, respectively, which allow molecules to be transported inside the barrel.

The shear numbers S for most of the barrel groups range from n to $2n$ as suggested by Murzin et al. (1994a). There are six exceptions, lipoamide dehydrogenase (PDB code, 1lpf and 1ojt; Mattevi et al., 1993; delaSierra et al., 1997) with $(n,S) = (5,12)$ and oxidoreductase (superoxide acceptor) (PDB code, 1sos and 3sod; Parge et al., 1992; McRee et al., 1990) with $(n,S) = (8,6)$.

Most of groups of barrels were antiparallel or mixed sheets. The only group with a parallel sheet is $(n,S) = (8,8)$.

In the present work, the number of "nonhomologous barrels" was also derived. Using the CATH nomenclature (Orengo et al., 1997), those domains with the same C, A, T, and H numbers are defined to be "homologous," while those with different numbers are defined to be "nonhomologous." For example, the group of barrel with $(n,S) = (5,8)$ has five "nonhomologous barrels" (Table 3A). This includes two mainly-β proteins with different folds (topologies) and three $\alpha\beta$ proteins, again with different topologies. In the Brookhaven PDB, some homologous families have many entries. For example, in the case of the barrels with $(n,S) = (6,8)$, although 527 barrels were detected, they include only three nonhomologous families (Table 3A). In contrast, the barrel group with $(n,S) = (8,8)$ is the most popular with 16 apparently nonhomologous families, although some of these might be distant relatives (Table 3A). All of them were $\alpha\beta$-barrels.

Functions of barrel structures

The function was also analyzed for each homologous barrel (Table 3B). Sixty-eight percent of them are enzymes, although some binding proteins and transporter protein, such as maltoporin and porin, are also found. In addition, nearly half of enzyme barrels are TIM barrels. Barrel enzymes perform many different functions and include primary EC numbers 1–6. Although most homologous barrel structures correspond to only one function, some have more than one function. Fifteen out of 16 TIM barrels are known to be

Table 3A. *Classification of barrel structures based on the number of strands (n) and the shear number (S) using 1,277 barrels with an even shear number*

No. of strands (n)	Shear number (S)	No. of barrels	No. of nonhomologous barrels	Relationships between n and S	Parallel, antiparallel, or mixed	Examples[a]
4	8	6	1	$S = 2n$	Antiparallel	1bnd[*1]
5	8	29	5	$n < S < 2n$	Mixed	1whi[*2], 1agn[*3]
5	10	55	3	$S = 2n$	Mixed	1asy[*4], 1enc[*5] (hydrolase, toxin)
5	12	3	1	$S > 2n$	Mixed	1lpf[*6], 1ojt[*6] ($\alpha\beta$ three-layer sandwich)
6	6	1	1	$S = n$	Antiparallel	1tdt[*7]
6	8	527	3	$n < S < 2n$	Antiparallel	1abi[*8], 1bbr[*8] (serine protease)
6	10	246	5	$n < S < 2n$	Mainly mixed	1cxs[*9], 1ttt[*10]
6	12	21	3	$S = 2n$	Mainly antiparallel	2afg[*11], 1ytf[*12] (β-trefoil or β-rolls)
7	10	14	2	$n < S < 2n$	Mixed	1aco[*13], 1pkm[*14]
8	6	3	1	$S < n$	Antiparallel	3sod[*15], 1sos[*15] (β-sandwich)
8	8	244	16	$S = n$	Mainly parallel	1tim[*16], 3enl[*17] (TIM barrels)
8	10	65	3	$n < S < 2n$	Antiparallel	1avd[*18], 1pts[*19]
8	12	22	1	$n < S < 2n$	Antiparallel	1brp[*20], 1epa[*21]
11	14	17	1	$n < S < 2n$	Mixed	1ema[*22], 1gfl[*22] (fluorescent protein)
14	14	1	1	$S = n$	Antiparallel	7ahl[*23]
16	20	3	1	$n < S < 2n$	Antiparallel	1prn[*24], 2por[*24] (membrane protein porin)
18	22	16	1	$n < S < 2n$	Antiparallel	1mpm[*25], 2mpr[*25] (membrane protein)
20	20	4	1	$S = n$	Antiparallel	1gtp[*26]
$n \le S \le 2n$		1,271	48			
$S < n$		3	1			
$S > 2n$		3	1			
Total		1,277	50			

[a] *1 = neurotrophic factor; *2 = ribosomal protein L14; *3 = σ alcohol dehydrogenase; *4 = aspartyl-tRNA synthetase; *5 = staphylococcal nuclease mutant; *6 = lipoamide dehydrogenase; *7 = transferase; *8 = thrombin; *9 = DMSO reductase; *10 = elongation factor; *11 = acidic fibloblast growth factor; *12 = TATA-box binding protein; *13 = aconitase; *14 = pyruvate kinase; *15 = oxidoreductase (superoxide acceptor); *16 = triose phosphate isomerase; *17 = enolase; *18 = avidin; *19 = streptavidin; *20 = retinol-binding protein; *21 = retinoic acid binding protein; *22 = green fluorescent protein; *23 = α-hemolysin; *24 = porin; *25 = maltoporin; *26 = GTP cyclohydrolase I. (*The references are shown in each PDB file.)

Table 3B. *Functions of barrel structures*[a]

n	*S*	No. of nonhomologous barrels	No. of enzymes	No. of nonenzymes	Functions
4	8	1	0	1	Growth factor
5	8	5	3	2	Major cold shock protein, Ribosomal protein L14, three enzymes: EC 3.6.1.1, EC 1.1.1.1, EC 2.4.2.1
5	10	3	2	1	Toxin, two enzymes: EC 6.1.1.6 or 6.1.1.12, EC 3.1.31.1
5	12	1	1	0	Enzyme: EC 1.8.1.4
6	6	1	1	0	Enzyme: EC 2.3.1.117
6	8	3	3	0	Three enzymes: EC 3.4.21.-, EC 3.6.1.34, EC 5.99.1.2
6	10	5	4	1	Elongation factor, four enzymes: dimethylsulfoxide reductase, EC 1.18.1.2, EC 3.2.1.4, EC 2.7.7.49 or 3.1.26.4 or 3.6.1.23 or 3.4.23.-
6	12	3	0	3	Acidic fibloblast growth factor or Interleukin-1, trypsin inhibitor, TATA-binding box
7	10	2	2	0	Two enzymes: EC 2.7.1.40, EC 4.2.1.3
8	6	1	1	0	Enzyme: EC 1.15.1.1
8	8	16^{*27}[b]	15^{*27}	1^{*27}	Seed protein[*27], 15 enzymes: EC 1.1.1.50, EC 1.14.14.3, EC 1.1.3.15 or 1.1.2.3, EC 2.2.1.2, EC 2.7.1.40, EC 2.7.9.1, EC 2.4.1.19 or 3.2.1.-, EC 3.2.1.52, EC 3.1.4.11, EC 4.1.1.39, EC 4.1.3.3 or 4.2.1.52 or 4.1.2.13, EC 4.2.1.11, EC 4.1.1.48 or 5.3.1.24, EC 5.1.1.1, EC 5.3.1.1
8	10	3	1	2	Chrysanthemi inhibitor, biotin-binding protein, enzyme EC 5.2.1.8
8	12	1	0	1	Retinol-binding protein, bilin-binding protein, or odorant-binding protein
11	14	1	0	1	Fluorescent protein
14	14	1	0	1	α-Hemolysin (cytolytic protein)
16	20	1	0	1	Integral membrane protein porin
18	22	1	0	1	Maltoporin (membrane protein)
20	20	1	1	0	Enzyme: EC 3.5.4.16
Total no.	50	34	16		

[a]The function was analyzed for each nonhomologous barrel.
[b]*27 = 15 out of 16 TIM barrels are known to be enzymes. The protein narbonin (Hennig et al., 1992) is found in plant seeds and may prove to have enzyme activities. (*The references are shown in each PDB file.)

enzymes. The remaining TIM barrel, narbonin (PDB code 1nar; Hennig et al., 1992), is found in plant seeds and may yet prove to have enzyme activities.

Analysis of layer structures

The detection of layer structure

As described above, if S is equal to n or $2n$, particular sets of residues are in planes, perpendicular to the barrel axis, which form stacked "layers." These layers are very easy to identify from the hydrogen bonding pattern as shown in Figure 1. In the present work, these layers are defined as a complete line of n residues forming part of the sheet (see Definitions section above) and are identified automatically. For 1tim (a parallel $(\beta\alpha)_8$ barrel) (triose phosphate isomerase; Banner et al., 1976), three layers were detected, while five layers were detected for 2afg (an antiparallel β-barrel) (acidic fibloblast growth factor; Blaber et al., 1996) (see Fig. 1). Where $S = n$, the layers comprise residues in adjacent strands with a stagger of one. In contrast, where $S = 2n$, layer residues occur on alternate strands with a stagger of two.

The interior residues on the layers, indicated by the thick boxes in Figure 1, are packed together in the barrel core.

The flatness of layer structure of barrels

The root-mean-square deviation (RMSD) from the least-squares plane through the Cα atoms of the layer residues for 1tim and 2afg are calculated (Table 4). In both structures, the layers are rather flat with an RMSD (on 8 atoms) = 0.36–0.47 Å for the three 1tim layers and RMSD (on 6 atoms) = 0.15–0.79 Å for the five layers in 2afg. In the highly twisted and sheared all-β barrels, the layers are not so distinct as in the regular $(\beta\alpha)_8$ barrel of TIM. In particular, the central external residue layer in 2afg (layer 3) is distorted, with alternate residue around the layer being slightly displaced up or down relative to the barrel axis (data not shown). The second and fourth layers, which form the core of the barrel, are however rather planar.

In addition, the average tilt relative to the barrel axis, as calculated from (n,S), is 37° for 1tim and 62° for 2afg. Since the strand tilt for 2afg is larger than in 1tim, the barrel structure of 2afg is squashed compared to 1tim. The shape of the 1tim barrel seems to be cylindrical, while that of 2afg is narrowed in the middle (data not shown).

The residues forming the core of barrel

In 1tim, the core is constructed from three layers, each with four residues from alternate strands, giving 12 residues in total (see Fig. 5A). In 2afg, the core is constructed from two layers, each of six residues from adjacent strands, again giving 12 residues in total (see Fig. 5A).

When the top view of all the layer residues in 1tim is observed, interior and exterior residues are obviously separate, and the interior residues are involved in the packing (see Fig. 5A). However, there are three glycines and two alanines among the packing residues, so the shape of the barrel is obviously distorted from circular. The exterior residues of TIM barrels are more hydrophobic

because they usually interact with the residues from helices outside the barrels.

In contrast, the shape of the barrel structure in 2afg is circular, and hydrophilic residues are distributed over the exterior residues (see Fig. 5B). In this protein, the barrel surface is exposed, which explains why hydrophilic residues occur. The interior residues are involved in the barrel packing as in 1tim.

The automatic detection of TIM barrel structures from their tertiary structure information

As described above, most of TIM barrels have shear numbers of 8, although there are a few exceptions such as 1ucw (transaldolase; Jia et al., 1997) and 1req (methylmalonyl-CoA mutase; Mancia et al., 1996), which have shear numbers of 7 and 9, respectively.

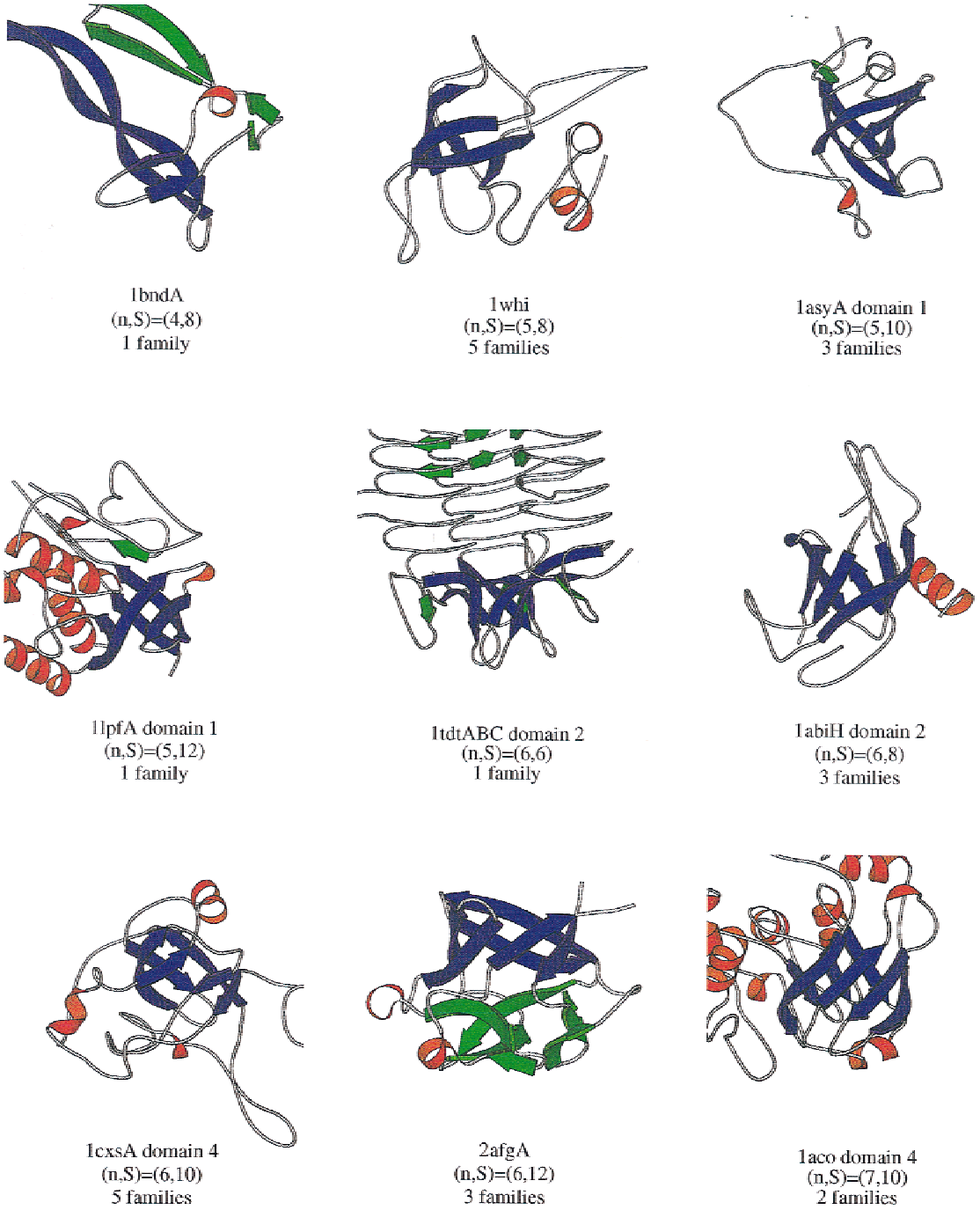

Fig. 4. MOLSCRIPT (Kraulis, 1991) representations of barrel structures for each (n,S) group. Barrel strands are colored blue, other strands green, helices red. PDB codes, chain name, domain number, and the number of nonhomologous families are also given. (*Figure continues on facing page.*)

In the present research, from the barrel group with $(n,S) = (8,8)$, all 244 barrels were found to be $(\beta\alpha)_8$ barrels. The TIM barrel is mainly composed of parallel strands, although sometimes two antiparallel β-strands are included. Therefore, we could identify $(\beta\alpha)_8$ structures by searching for barrels mainly composed of parallel strands with $(n,S) = (8,8)$.

In the most recent work on TIM barrels by Reardon and Farber (1995), 30 TIM barrels were analyzed in detail. Of the 30 barrels, 27 TIM barrels were categorized into seven families based mainly on their catalytic functions and sequence identity (Reardon & Farber, 1995). The present method was also applied to the 30 proteins. Out of the 30 barrels, two coordinates of the proteins, oligo-1,6-glucosidase and 2-dehydro-3-deoxyphosphogluconate aldolase, were not available, as the former was not deposited in the Brookhaven Protein Data Bank and the latter contained only those of Cα atoms. Of the remainder of 28 proteins, 10 proteins [trimethylamine dehydrogenase (PDB code, 2tmd; Bellamy et al., 1989), mandelate racemase (PDB code, 1mns; Neidhart et al., 1991), muconate cycloisomerase (PDB code, 1muc; Goldman et al., 1987), chloromuconate cycloisomerase (PDB code, 1chr; Hoier et al., 1994), xylose

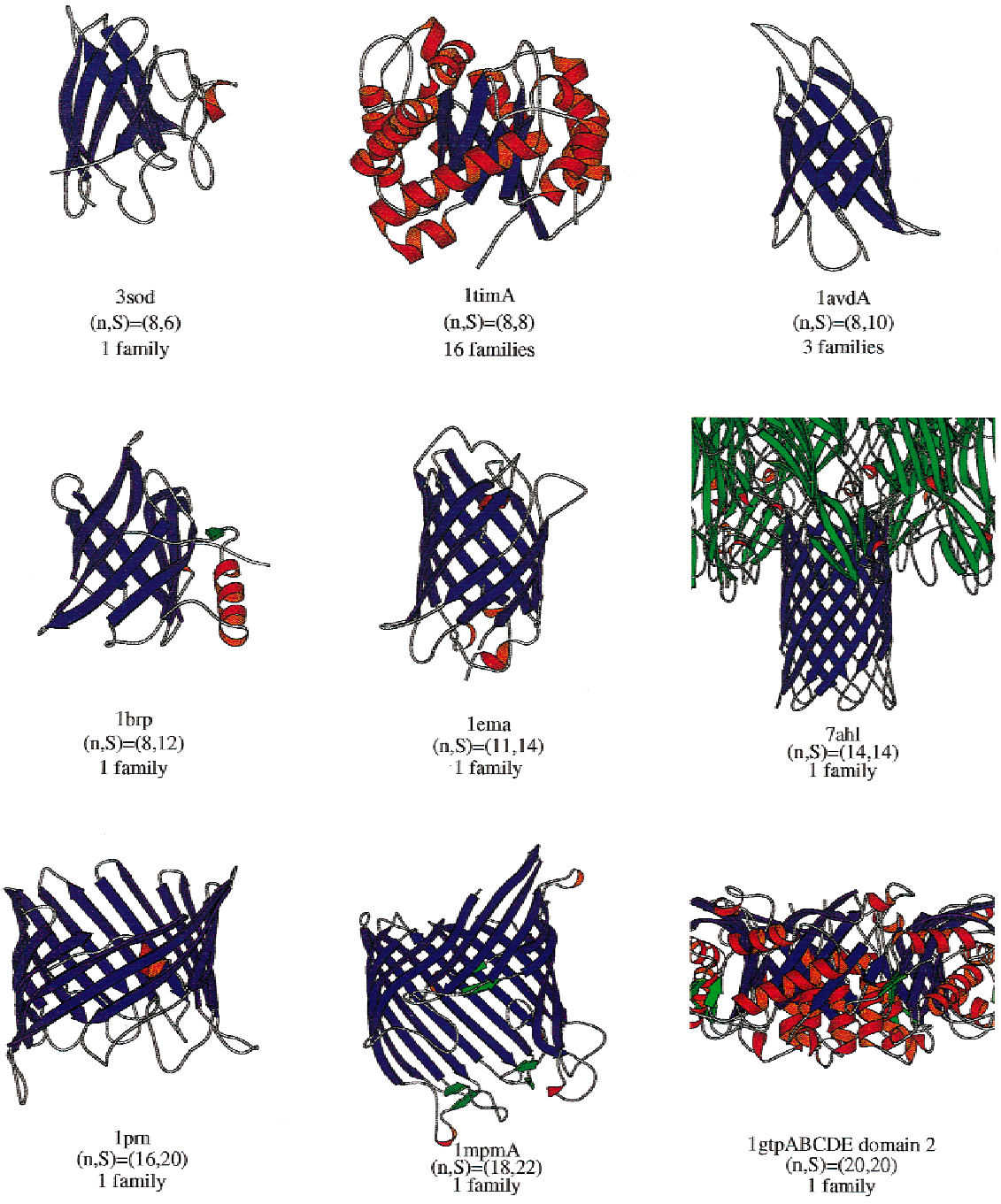

Fig. 4. *Continued.*

Table 4. *RMSD from the least-squares plane (lsp) through the Cα atoms of the layer residues*

PDB code[a]	Chain	Relationships between *n* and *S* residues	Observed radius (Å)	Tilt of strands (°)	Numbering of layers	No. of residues in the layer	RMSD from the lsp Interior residues	RMSD from the lsp exterior residues	RMSD from the lsp All residues
1tim*1	A	$S = n$	7.5	37	1	8	0.21	0.12	0.36
					2	8	0.31	0.54	0.47
					3	8	0.45	0.31	0.41
2afg*2	A	$S = 2n$	9.5	62	1	6	—	0.21	0.21
					2	6	0.27	—	0.27
					3	6	—	0.79	0.79
					4	6	0.15	—	0.15
					5	6	—	0.31	0.31

[a]*1 = triose phosphate isomerase; *2 = acidic fibloblast growth factor. (*The references are shown in each PDB file.)

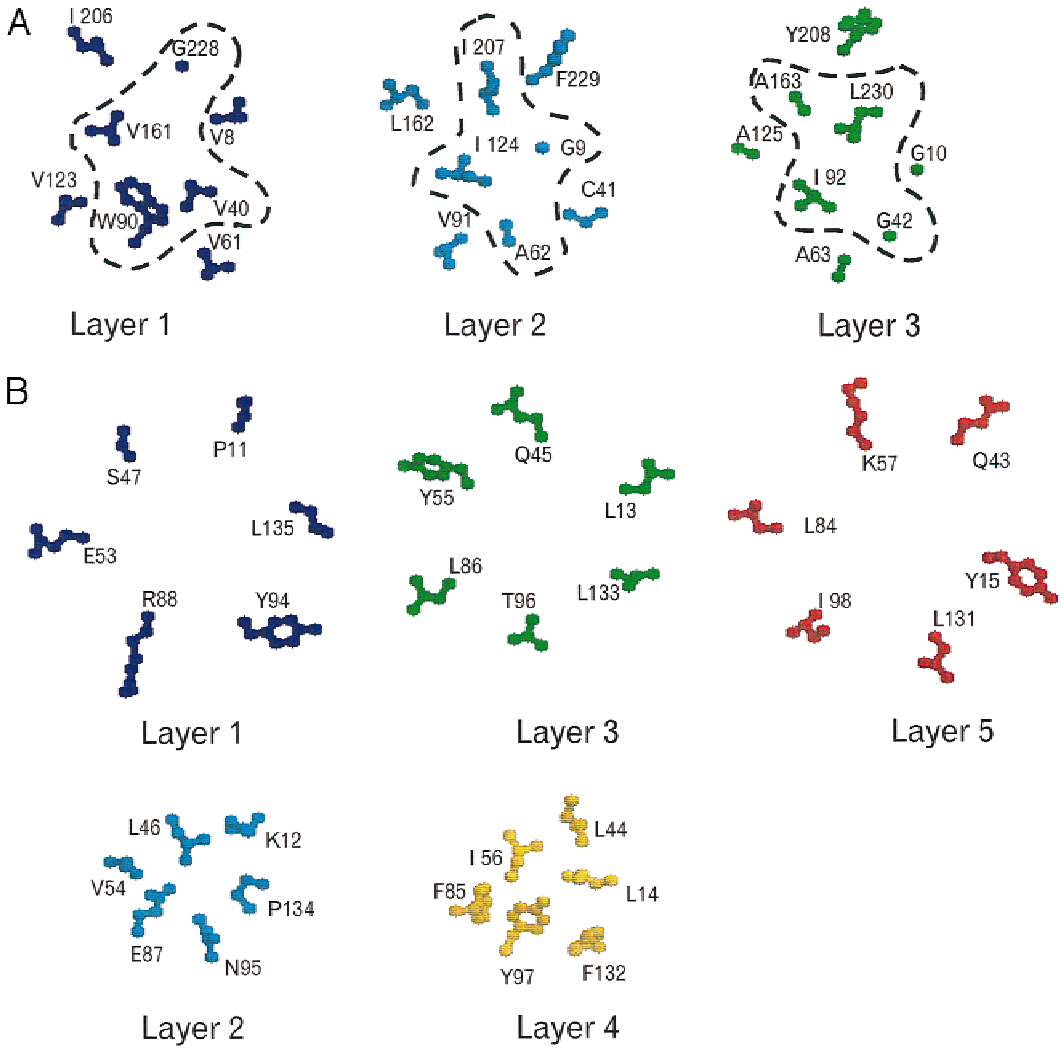

Fig. 5. Layer structures. The atoms for the first, second, third, fourth, and fifth layer are indicated by dark blue, light blue, green, yellow, and red, respectively. **A:** The top view of the side-chain atoms on the layers for chain A of 1tim. Each layer was drawn separately. The interior residues are surrounded by broken line. **B:** The top view of the side-chain atoms on the layers for chain A of 2afg. The side chains of layer 1, 3, and 5 are directing outward, while those of layer 2 and 4 are directing inward to contribute to the core of the barrel.

isomerase (PDB code, 1xis; Collyer et al., 1990), tryptophan synthase α-subunit (PDB code, 1wsy; Hyde et al., 1988), adenosine deaminase (PDB code, 2ada; Wilson et al., 1992), phosphotriesterase (PDB code, 1pta; Benning et al., 1994), urease (PDB code, 1kra; Jabri & Karplus, 1996), and old yellow enzyme (PDB code, 1oya; Fox & Karplus, 1994)], were regarded as "distorted" TIM barrels without any complete rings of hydrogen bonds in the present criteria. The rest of "complete" 18 TIM barrels were classified using our method differently from that by Reardon and Farber (1995). The families "A," "B," and "C" in their literature were divided into two, two, and three families, respectively, while three proteins [(1 $\rightarrow$ 3)-β-glucanase (PDB code, 1ghs; Varghese et al., 1994), (1 $\rightarrow$ 3, 1 $\rightarrow$ 4)-β-glucanase (PDB code, 1ghr; Varghese et al., 1994), endo-β-N-acetylglucosaminidase (PDB code, 2ebn; Roey et al., 1994)] were clustered together with the family of α-amylase (PDB code, 1amy; Brady et al., 1991), β-amylase (PDB code, 1btc; Mikami et al., 1992), and cyclodextrin glycosyltransferase (PDB code, 1cgt; Klein & Schulz, 1991) in the "C" family in the work of Reardon and Farber (1995). Eventually, the total number of complete TIM barrel families in the present criteria turned out to be nine in their list of the TIM barrels. Even if enolase family, which was not regarded as true TIM barrel family due to the antiparallel strands by Reardon and Farber (1995), was included in the group of TIM barrel families, only 10 TIM barrels were available at that point. For only these few years, six more families of TIM barrel have been reported.

In addition, based on the criterion of layers of TIM barrels, in which the interior and exterior residues occur alternately, the layers of TIM barrels were identified. As shown in Table 5, the number of layers ranges from 2 to 5 for all TIM barrels, though the cases of three and four layers were the most popular, with 110 and 97 examples, respectively. Five layers were identified for 1ctn (chitinase; Perrakis et al., 1994), in which the RMSD values from least-squares plane for the first and fifth layers were larger than those for the remainder.

In addition, the number of layers are crucially dependent on hydrogen bonding patterns, so related sequences can have a different number of layers. For example, 9rub (rubisco; Schneider et al., 1986) has chain A with four layers and chain B with three layers (Table 5).

Table 5. *The number of layers of TIM barrels*[a]

No. of layers	No. of barrels	No. of nonhomologous barrels	Examples[b]
1	0	0	
2	36	4	2amg[*1], 1fba[*2]
3	110	11	1tim[*3], 1gox[*4], 9rub[*5] chain B
4	97	9	9rub[*5] chain A, 1pkm[*6], 3tim[*3] chain A
5	1	1	1ctn[*7]
Total	244	25	

[a] From the barrel group with $(n,S) = (8,8)$, parallel barrels were detected, and the number of layers, in which the interior and exterior residues occur alternately, was also computed.

[b] *1 = G4-amylase; *2 = fructose-1,6-bisphosphate aldolase; *3 = triose phosphate isomerase; *4 = glycolate oxidase; *5 = ribulose-1,5-bisphosphate carboxylase-oxygenase; *6 = pyruvate kinase; *7 = chitinase. (*The references are shown in each PDB file.)

Amino acid propensities for layer residues on TIM barrels

At the next stage, from the list of α/β barrels based on the CATH database of domains, which are not overlapping with each other, such parallel β-barrels with $(n,S) = (8,8)$ were searched. In CATH, which has several levels of the classification such as homologous superfamily (H) and sequence family (S), α/β barrels are classified into the group with "CA number" (3.20). In the H-level, in which structures are grouped by their high structural similarity and similar function, there are 23 α/β barrel domains. In the S-level, in which structures are grouped by their sequence identity $>35\%$ but not identical to each other, there are 59 α/β barrel domains. From the H- and S-levels, 10 and 32 "complete TIM barrels," respectively, were detected with the present system.

Given these lists of complete TIM barrels, amino acid propensities of layer residues were analyzed. As shown in Table 6, from the lists of complete TIM barrels, 10 (H-level), and 32 (S-level) barrels had 33 and 95 layers, respectively. All the results from both levels for the amino acid propensity showed a similar tendency.

The amino acid propensities on inside and outside surfaces of barrels are very different (Table 6). The exterior residues are much more hydrophobic than interior residues. Charged residues are very uncommon on the exterior. In contrast, the interior residues that are involved in packing in the core of the barrel are more hydrophilic. In addition, many glycines are observed in the interior (Table 6).

Even though the number of hydrophobic residues are smaller inside the barrel than outside, the number of observed aromatic residues (F, Y, W, H) is slightly larger inside than outside (Table 6). Furthermore, there are very few proline residue inside the barrel (Table 6), except for one residue on the fourth layer of 1edg (cellulase; Ducros et al., 1995).

Discussion

In this paper, we have implemented an automated approach to barrel classification, based on the number of strands, the shear number (n,S) and the layer structure. Using this algorithm, we have performed our analysis of all the barrel structures in the Brookhaven Protein Data Bank. The method will only detect "complete" barrels with a continuous ring of hydrogen bonds, and so many barrels with irregularities are excluded from the analysis.

However, our current algorithm does include many barrels containing bifurcated strands, which were not identified by the original version of PROMOTIF (Hutchinson & Thornton, 1996). In a test case, 13% more barrels were identified (Table 1). More than half of the barrels identified include one or more bifurcated strands (Table 1).

The distribution of the complete barrel structures in the PDB is shown in Figure 6. Barrels have between 4 and 20 strands with clear preference for 5–8 strands. Six and 8 stranded barrels predominate in all PDB entries and among nonhomologous representatives. Shear numbers vary from 6 to 22, with a clear preference for $S = 8$ to 10, with $S = 8$ most popular.

The shear numbers S for most of the barrel groups range from n to $2n$. According to Murzin et al. (1994a), barrels, whose shear numbers are outside this range, would have greater constraints on their geometry. In the present work, this was indicated statistically. Only 4% of nonhomologous barrels were outside the range. There must be some other features stabilizing these barrels. In addition, Murzin et al. (1994a) suggested that there are 10 major barrel groups. Nine out of the 10 groups, with $(n,S) = (4,8)$, (5,8), (5,10),

Table 6. *Amino acid propensity of layer residues for TIM barrels*

	G	A	S	T	D	N	E	Q	K	R	H	P	L	I	M	V	F	Y	W	C	Total
Homologous families																					
10 complete TIM barrels (33 layers)																					
Internal	13	14	9	5	6	3	2	3	6	6	6	0	9	9	10	11	7	6	4	3	132
External	5	13	1	3	0	2	1	1	1	0	0	2	20	28	2	29	12	5	1	6	132
Total	18	27	10	8	6	5	3	4	7	6	6	2	29	37	12	40	19	11	5	9	264
Families with >35% sequence identity																					
32 complete TIM barrels (95 layers)																					
Internal	46	31	24	18	14	10	19	4	13	17	11	1	27	37	20	30	21	19	10	6	378
External	15	31	12	14	0	3	2	2	2	0	2	9	68	76	6	77	31	14	5	13	382
Total	61	62	36	32	14	13	21	6	15	17	13	10	95	113	26	107	52	33	15	19	760

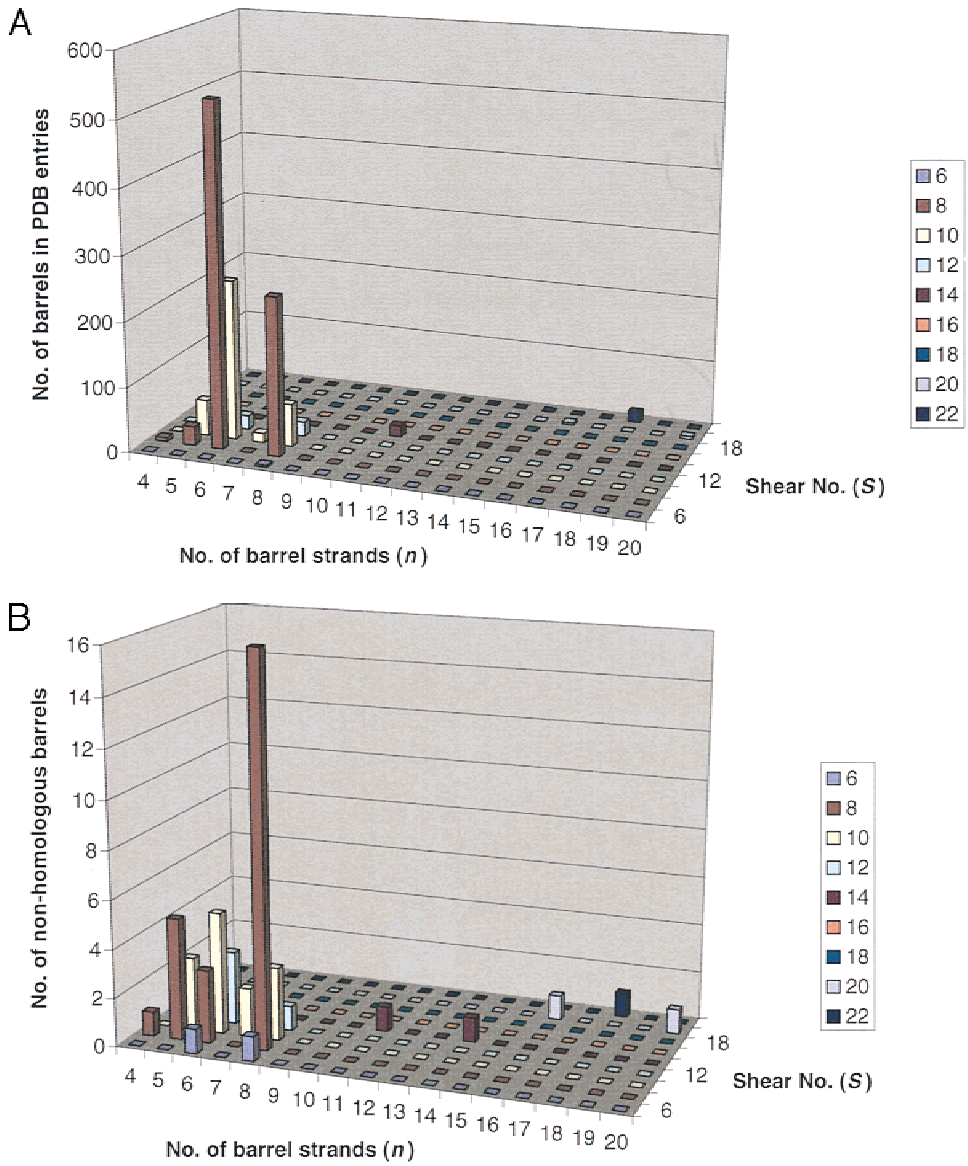

Fig. 6. The distribution of (n,S) values observed for complete barrels. **A:** The distribution for all PDB entries. **B:** The distribution for sequentially nonrelated entries.

(6,8), (6,10), (6,12), (7,10), (8,8), and (8,10), were found in the present work. $(n,S) = (7,8)$ was not identified because the barrel in cellobiohydrolase II (Koivula et al., 1996) is not complete, since hydrogen bonds do not occur between the first and seventh strands. Seven groups of the nine groups identified in this work had more than two nonhomologous barrels.

The functions of barrel structures are found to include predominantly various kinds of enzymes as well as binding proteins and transport membrane proteins. Although most of homologous barrel structures correspond to only one function, some correspond to more than one function.

In the present work, based on the number of strands, shear number, and layer structures, TIM barrels could be also analyzed. The vast majority of TIM barrel structures could be detected automatically from the barrel group with $(n,S) = (8,8)$ in more than 200 TIM barrels. There are only two exceptions with shear number ($S = 7$ or 9) (Table 2).

In the most recent work on analysis of TIM barrels (Reardon & Farber, 1995), only 30 TIM barrels were available to analyze. In the literature by Reardon and Farber (1995), 27 of 30 TIM barrels were categorized into seven families based mainly on their catalytic function and sequence identity. Since the present method of clustering proteins is different from their criteria, based exclusively on tertiary structure and sequence identity, the result of the family classification of the 28 TIM barrels, of which coordinates were available, were different from that by Reardon and Farber (1995). Although 10 of the 28 proteins were regarded as "distorted" TIM barrels without any complete rings of hydrogen bonds in the present criteria, the rest of 18 barrels were classified into 9 families using the present method. Even though enolase family, which was not regarded as true TIM barrel family by Reardon and Farber (1995), was included, only 10 TIM barrels had been reported at that point. For only a few years, the numbers of TIM barrels and of the families have been increased so rapidly (nearly 10 times of the barrel structures and 6 more families). Therefore, the present automated system for identifying one of dominant folds, TIM barrel, would be extremely essential.

The statistics of the number of layers could reflect the stable structure of the core of barrel. According to the previous literature on a limited number of examples (Lesk et al., 1989), the packing inside the sheet is limited to three layers, and the formation of a fourth layer is usually prevented by the protrusion of residues from the third layer. However, in the present work, four layers as well as three layers are found to be very common, which might reflect the stability of formation of four layers. Considering that 9rub (rubisco; Schneider et al., 1986) has chain A with four layers and chain B with three layers, the number of layers is very sensitive to the hydrogen bonding so that clearly related structures and sequences may apparently exhibit a different number of layers.

Based on the amino acid propensities, in TIM barrels, aromatic residues might be important in forming the core of barrel. In addition, proline residue might interrupt the formation of core of TIM barrels. The exterior hydrophobic residues interact with helices outside the barrels, while the interior hydrophilic residues seem to be related to the function of the proteins. The totally different amino acid propensities between inside and outside surfaces of the barrels might also be related to the curvature of β-sheet. Especially glycines in the interior might be necessary to allow the close packing and high curvature.

Material and methods

Automated methods for analysis and classification of "complete barrel" structures have been developed based on the HERA program (Hutchinson & Thornton, 1990), which draws schematic diagrams of hydrogen bonding pattern of β-sheets. This program is now part of the PROMOTIF suite of programs (Hutchinson & Thornton, 1996) for identifying and clarifying motifs in protein structures including β-sheets and β-bulges. This program uses the Brookhaven PDB (Bernstein et al., 1977) coordinate files (January 1998 release) and their secondary structure information, calculated using a local implementation, SSTRUC (D.K. Smith, unpubl. data) of the DSSP algorithm (Kabsch & Sander, 1983). To optimize the barrel assignments with SSTRUC, strands are extended by a residue at each end, where these form a single "sheet" hydrogen bond, but not the two required by DSSP.

Before calculating the shear number, the bulge residues that disrupt the regular hydrogen bonding patterns were removed using the algorithm developed by Chan et al. (1993). Alignments of barrel β-strands on the virtual lattice were performed based on the resulting hydrogen bonds.

Shear numbers were calculated from the resulting matrix of hydrogen bonding patterns. The shear numbers were calculated based on the relative position of same residue on the repeated edge strands of staggered barrel β-sheet as shown in Figure 1.

Two lists of protein domains were obtained based on the H and S levels of version 1.4 (13,338 domains) of the CATH classification (Orengo et al., 1997), in which structures are grouped into homologous families (H) and sequence families (S). The analysis of homologous proteins and their function was performed using algorithm developed by Martin et al. (1998).

Acknowledgments

This research was funded by the Naito Foundation in Japan and the Japan Society for the Promotion of Science. We acknowledge support for computing from the BBSRC. We thank Professor Kozo Nagano for useful discussions, particularly concerning TIM barrel structures.

References

Banner DW, Bloomer AC, Petsko GA, Phillips DC, Wilson IA. 1976. Atomic coordinates for triose phosphate isomerase. *Biochem Biophys Res Comm 72*:146–155.

Bellamy HD, Lim LW, Mathews FS. 1989. Studies of crystalline trimethylamine dehydrogenase in three oxidation states and in the presence of substrate and inhibitor. *J Biol Chem 264*:11887–11892.

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

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard, O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer based archival file for macromolecular structure. *J Mol Biol 112*:535–542.

Blaber M, Disalvo J, Thomas KA. 1996. X-ray crystal-structure of human acidic fibloblast growth factor. *Biochemistry 35*:2086–2094.

Brady RL, Brzozowski AM, Derewenda ZS, Dodson EJ, Dodson GG. 1991. Solution of the structure of *Aspergillus niger* acid α-amylase by combined molecular replacement and multiple isomorphous replacement methods. *Acta Crystallogr Sect D 47*:527–535.

Chan AWE, Hutchinson EG, Harris D, Thornton JM. 1993. Identification, classification, and analysis of beta-bulges in proteins. *Protein Sci 2*:1574–1590.

Collyer CA, Henrick K, Blow DM. 1990. Mechanism of aldose-ketose interconversion by D-xylose isomerase involving ring opening followed by a 1,2-hydride shift. *J Mol Biol 212*:211–235.

Cutfield SM, Dodson EJ, Anderson BF, Moody PC, Marchall CJ, Sullivan PA, Cutfield JF. 1995. The crystal structure of a major secreted aspartic proteinase from *Candida albicans* in complexes with two inhibitors. *Structure 3*:1261–1271.

delaSierra IL, Pernot L, Prange T, Saludjian P, Schiltz M, Fourme R, Pardon G. 1997. Molecular structure of the lipoamide dehydrogenase domain of a surface antigen from *Neisseria meningitidis*. *J Mol Biol 269*:129–141.

Ducros V, Czjzek M, Belaich A, Gaudin C, Fierobe HP, Belaich LP, Davies GJ, Haser R. 1995. Crystal structure of catalytic domain of a bacterial cellulase belonging to family-5. *Structure 3*:939–949.

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.

Goldman A, Ollis DL, Steitz TA. 1987. Crystal structure of muconate lactonizing enzyme at 3 Å resolution. *J Mol Biol 194*:143–153.

Harris GW, Jenkins JA, Connerton I, Pickersgill RW. 1996. Refined crystal structure of the catalytic domain of xylanase A from *Pseudomonas fluorescens* at 1.8 Å resolution. *Acta Crystallogr Sect D 52*:393–401.

Hennig M, Schlesier B, Dauter Z, Pfeiffer S, Betzel C, Hohne WE, Wilson KS. 1992. A TIM barrel protein without enzymatic activity? Crystal-structure of narbonin at 1.8 Å resolution. *FEBS Lett 306*:80–84.

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 Sect D 50*:75–84.

Hutchinson EG, Thornton JM. 1990. HERA—A program to draw schematic diagrams of protein secondary structure. *Proteins Struct Funct Genet 8*: 203–212.

Hutchinson EG, Thornton JM. 1996. PROMOTIF—A program to identify and analyze structural motifs in proteins. *Protein Sci 5*:212–220.

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.

Jabri E, Karplus PA. 1996. Structure of the *Krebsiella aerogenes* urease apoenzyme and two active-site mutants. *Biochemistry 35*:10616–10626.

Jia J, Schorken U, Lindqvist Y, Sprenger GA, Schneider G. 1997. Crystal structure of the reduced Schiff-base intermediate complex of transaldolase B from *Escherichia coli*: Mechanistic implications for class I aldolases. *Protein Sci 6*:119–124.

Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*:2577–2637.

Kawashima T, Berthetcolominas C, Wulff M, Cusack S, Leberman R. 1996. The structure of the *Escherichia coli* EF-Tu.EF-Ts complex at 2.5 Å resolution. *Nature 379*:511–518.

Klein C, Schulz GE. 1991. Structure of cyclodextrin glycosyltransferase refined at 2.0 Å resolution. *J Mol Biol 217*:737–750.

Koivula A, Reinikainen T, Ruohonen L, Valkeajarvi A, Claeyssens M, Teleman O, Kleywegt GJ, Szardenings M, Rouvinen J, Jones TA, Teeri TT. 1996. The active site of *Trichiderma reesei* cellobiohydrolase II: The role of tyrosine 169. *Protein Eng 9*:691–699.

Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Cryst 24*:946–950.

Leesong M, Henderson BS, Gilling JR, Schwab JM, Smith JL. 1996. Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: Two catalytic activities in one active site. *Structure 4*:253–264.

Lesk AM, Brändén C-I, 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.

Liu W. 1998. Shear numbers of protein β-barrels: Definition of refinements and statistics. *J Mol Biol 275*:541–545.

Mancia F, Keep NH, Nakagawa A, Leadlay PF, Mcsweeney S, Rasmussen B, Bosecke P, Diat O, Evans PR. 1996. How coenzyme B_{12} radicals are generated: The crystal structure of methylmalonyl-coenzyme A mutase at 2 Å resolution. *Structure 4*:339–350.

Martin ACR, Orengo CA, Hutchinson EG, Jones S, Karmirantzou M, Laskowski RA, Mitchell JBO, Taroni C, Thornton JM. 1998. Protein folds and functions. *Structure 6*:875–884.

Mattevi A, Obmolova G, Kalk KH, Vanberkel WJH, Hol WGL. 1993. 3-dimensional structure of lipoamide dehydrogenase from *Pseudomonas fluorescens* at 2.8 Å resolution—Analysis of redox and thermostability properties. *J Mol Biol 230*:1200–1215.

McLachlan AD. 1979. Gene duplications in the structural evolution of chymotrypsin. *J Mol Biol 128*:49–79.

McRee DE, Redford SM, Getzoff ED, Lepock JR, Hallewell RA, Tainer JA. 1990. Changes in crystallographic structure and thermostability of a Cu, Zn superoxide-dismutase mutant resulting from the removal of a buried cysteine. *J Mol Chem 265*:14234–14241.

Merritt EA, Sarfaty S, Chang TT, Palmer LM, Jobling MG, Holmes RK, Hol WGJ. 1995. Surprising leads for a cholera toxin receptor binding antagonist; crystallographic studies of CTB mutants. *Structure 3*:561–570.

Meyer JEW, Hofnung M, Schulz GE. 1997. Structure of maltoporin from *Salmonella typhimurium* ligated with a nitrophenyl-maltotrioside. *J Mol Biol 266*:761–775.

Mikami B, Sato M, Shibata T, Hirose M, Aibara S, Katsube Y, Morita Y. 1992. Three-dimensional structure of soybean β-amylase determined at 3.0 Å resolution: Preliminary chain tracing of the complex with α-cyclodextrin. *J Biochem 112*:541–546.

Murzin AG, Lesk AM, Chothia C. 1994a. Principles determining the structure of β-sheet barrels in proteins. I. A theoretical analysis. *J Mol Biol 236*:1369–1381.

Murzin AG, Lesk AM, Chothia C. 1994b. Principles determining the structure of β-sheet barrels in proteins. II. The observed structures. *J Mol Biol 236*:1382–1400.

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.

Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM. 1997. CATH—A hierarchic classification of protein domain structures. *Structure 5*:1093–1108.

Parge HE, Hallewell RA, Tainer JA. 1992. Atomic structures of wild-type and thermostable mutant recombinant human Cu, Zn superoxide-dismutase. *Proc Nat Acad Sci USA 89*:6109–6113.

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.

Raghunathan S, Ricard CS, Lohman TM, Waksman G. 1997. Crystal structure of the homo-tetrameric DNA binding domain of *Escherichia coli* single-stranded DNA-binding protein determined by multiwavelength X-ray diffraction on the selenomethionyl protein at 2.9 Å resolution. *Proc Nat Acad Sci USA 94*:6652–6657.

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

Roey PV, Rao V, Plummer TH, Tarentino AL. 1994. Crystal structure of endo-β-N-acetylglucosaminidase F1, an α/β-barrel enzyme adapted for a complex substrate. *Biochemistry 33*:13989–13996.

Rognan D, Scapozza L, Folkers G, Daser A. 1994. Molecular dynamics simulation of MHC-peptide complexes as a tool for predicting potential T cell epitopes. *Biochemistry 33:*11476–11485.

Schneider G, Lindqvist Y, Brändén C-I, Lorimer G. 1986. 3-dimensional structure of ribulose-1,5-bisphosphate carboxylase-oxygenase from *Rhodospirillum rubrum* at 2.9 Å resolution. *EMBO J 5*:3409–3415.

Stein PE, Boodhoo A, Armstrong GD, Cockle SA, Klein MH, Read RJ. 1994. The crystal structure of pertussis toxin. *Structure 2*:45–57.

van den Akker F, Sarfaty S, Twiddy EM, Connell TD, Holmes RK, Hol WG. 1996. Crystal structure of a new heat-labile enterotoxin, LT-IIb. *Structure 4*:665–678.

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 Nat Acad Sci USA 91*:2785–2789.

Wilson DK, Rudolph FB, Quiocho FA. 1992. Atomic structure of adenosine deaminase complexed with a transition-state analog: Understanding catalysis and immunodeficiency mutations. *Science 252*:1278–1284.