J. Mol. Biol. (1994) 236, 1382–1400

Principles Determining the Structure of β-Sheet Barrels in Proteins

II. The Observed Structures

Alexey G. Murzin[1,2]†, Arthur M. Lesk[3] and Cyrus Chothia[1,2]

[1]*MRC Laboratory of Molecular Biology*
[2]*Cambridge Centre for Protein Engineering and*
[3]*Department of Haematology, University of Cambridge Clinical School*
Hills Road, Cambridge CB2 2QH, U.K.

In the accompanying paper we derived a set of principles that, we argue, govern the structure of β-sheet barrels. Barrel structures are classified in terms of two integral parameters: the number of strands in the β-sheet, n, and a measure of the stagger in the β-sheet, S. We derived a set of equations that show how the (n, S) values of a barrel structure determine the arrangement of its strands; its general shape; the twist and coiling of the β-sheet, and the arrangement of residues in the barrel interior. This work suggested that there are ten different combinations of n and S that form barrels with good β-sheet geometries and interiors close packed by β-sheet residues.

In this paper we demonstrate the validity of these principles. We analyse in detail the observed structures of 39 different β-sheet barrels. These structures include representatives of all the different barrel structures currently known and for which atomic co-ordinates are available.

We show that the observed arrangement of the strands, and the extent of the twist and coiling of the β-sheets, are very close to those calculated from the (n, S) values for the barrel. Of the 39 structures, 34 have one of the ten (n, S) values that we expect to form barrels with good β-sheet geometries and interiors close packed by β-sheet residues. The other five have one of two (n, S) values that give good β-sheet geometries but radii so large the β-sheet residues leave cavities at the centre of the barrels. In at least four of these the cavities have a functional role.

Keywords: analysis of observed protein folds; β-sheet structure; barrel distortions; residue packing; ligand binding sites

1. Introduction

In the accompanying paper (Murzin *et al.*, 1994; referred to here as paper I) we presented a set of principles that, we argued, govern the barrel structures formed by β-sheets in proteins. In this paper we describe the main structural features of all the observed β-sheet barrels that have distinctly different amino acid sequences and for which atomic co-ordinates are available. This work demonstrates the validity of the principles described in paper I.

2. Structures and Co-ordinates

The structures discussed in this paper are listed in Table 1. With two sets of exceptions, part 1A of this list contains all the barrel structures for which atomic co-ordinates were available to us‡. The first set of exceptions involves large families of proteins whose members have clear sequence homology: these are represented by a few distantly related members. The second set of exceptions involve the α/β-barrel and β-trefoil proteins. Though most of the proteins within these two large groups have no significant sequence similarities, they do have very similar structures (Lesk *et al.*, 1989; Wilmanns *et al.*,

† On leave from the Institute of Mathematical Problems of Biology, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia.

‡ Not considered here are the sandwich structures, formed by the face to face packing of two β-sheets, for which the term "barrel structure" has been used incorrectly. (The twisted β-sheets in sandwich structures give them the shape of twisted rectangular prisms and these are sometimes confused with the shapes formed by β-sheet barrels.) We also do not consider structures in which single β-sheets fold upon themselves but have no significant contacts between the two edge strands.

0022-2836/94/101382-19 $08.00/0

Table 1
Proteins containing β-sheet barrels

Protein, protein family or common fold	PDB file entry	Reference	Barrel location in the structure†
A. *Closed barrels*			
Ovomucoid domain III	3OVO	Musil *et al.* (1991)	Oligomer of N-terminal regions
Alcohol dehydrogenase	5ADH	Eklund *et al.* (1976)	Barrel 2‡: β-sheet S3
Inorganic pyrophosphatase	1PYP	Arutiunian *et al.* (1981)	β-Sheet BAR
Purine nucleotide phosphorylase	1ULB	Ealick *et al.* (1991)	β-Sheets A + B
Major cold shock protein	1CSP§	Schindelin *et al.* (1993)	β-Sheet
Rhinovirus 14 coat protein	4RHV	Arnold & Rossmann (1990)	Oligomer of N-terminal regions
Staphylococcal nuclease	1SNC	Loll & Lattman, (1989)	β-Sheets S1 + S2
Verotoxin-1 B subunit	1BOV§	Stein *et al.* (1992)	β-Sheets B1A+B2A
Asp-tRNA synthetase	DRS§	Cavarelli *et al.* (1993)	β-Sheet of N-terminal domain
Serine proteases			
γ-chymotrypsin	2GCH	Cohen *et al.* (1981)	Barrel 1‡: β-sheet 1; barrel 2‡: β-sheet 2
α-lytic protease	2ALP	Fujinaga *et al.* (1985)	Barrels 1 and 2‡: β-sheets of both domains
Sindbis virus capsid protein	2SNV	Tong *et al.* (1993)	β-Sheet A2
Di-ubiquitin	1AAR	Cook *et al.* (1992)	Dimer of subunit β-sheets (A + B)
Reductase Family			
Ferredoxin reductase	1FNR	Karplus *et al.* (1991)	β-Sheet S1
Phthalate dioxygenase reductase	2PIA	Correll *et al.* (1992)	β-Sheet BB1
Acid proteases			
HIV protease	3HVP	Wlodawer *et al.* (1989)	β-Sheet COR
penicillopepsin	3APP	James & Sielecki (1983)	Barrel 1‡: β-sheets II + VI
Elongation factor Tu	ETU§	Kjeldgaard & Nyborg (1992)	Barrel 1 and 2‡: β-sheets of domains 2 and 3
Gln-tRNA synthetase	1GSG	Rould *et al.* (1989)	Barrel 1 and 2‡: β-sheets of the last 2 domains
β-Trefoil fold			
Erythrina trypsin inhibitor	1TIE	Onesti *et al.* (1991)	β-Sheet B1
interleukin-1β	1I1B	Finzel *et al.* (1989)	β-Sheet BRL
fibroblast growth factor	1BAR§	Zhu *et al.* (1991)	"Barrel" part of fold
Cellobiohydrolase	3CBH§	Rouvinen *et al.* (1990)	β-Sheet
Aconitase	5ACN	Robbins & Stout (1989)	β-Sheets S5 + S6
α/β Barrels			
glycolate oxidase	1GOX	Lindqvist (1989)	β-Sheet BAR
RuBisCO	5RUB	Schneider *et al.* (1990)	β-Sheet ACT
TIM (chicken)	1TIM§	Banner *et al.* (1975)	β-Sheet A
Catalase	8CAT	Fita & Rossmann (1985)	β-Sheet S1A
Streptavidin	STR§	Hendrickson *et al.* (1989)	β-Sheet
Cyclophilin A	1CPL§	Ke *et al.* (1991)	β-Sheet
Lipocalin binding proteins			
retinol-binding protein	1RBP§	Cowan *et al.* (1990)	β-Sheet A
bilin-binding protein	1BBP	Huber *et al.* (1987)	β-Sheets S1A + S2A
major urinary protein	1MUP	Bocskei *et al.* (1992)	β-Sheet
E2 DNA-binding protein	1BOP§	Hegde *et al.* (1992)	Dimer of subunit β-sheets
Porin	2POR§	Weiss & Schulz (1992)	β-Sheet S1
B. *Partly open barrels*			
Alcohol dehydrogenase	8ADH	Eklund *et al.* (1976)	Barrel 1‡: β-sheets S1 + S2
SH3-homology domain	1SHG§	Musacchio *et al.* (1992)	β-Sheet
Photoreaction centre			
H-chain	1PRC	Deisenhofer *et al.* (1985)	β-Sheet of C-terminal domain
Heat-labile enterotoxin			
B-Subunit (monomer)	1LTT§	Sixma *et al.* (1992)	β-Sheet(s) of one subunit
Acid proteinases			
Penicillopepsin (C-domain)	3APP	James & Sielecki (1983)	β-Sheets III + VII

Structures are listed here in the same order as they appear in Table 3.
† When assigned a PDB file (as of October 1993), names of β-sheets that make the barrel are given.
‡ Barrel number in the structures with more than one barrel; it is added to the 3-letter protein name in the text and the other Tables.
§ Co-ordinate files were given to us by their authors; those without the first digit have not been deposited with PDB before this publication.

1991; Murzin *et al.*, 1992) and we have taken, for each group, three examples to be adequate representatives.

The set of proteins considered here contains 39 different barrel structures; four of the 35 different proteins contain two barrel structures (see Table 1A).

We list in Table 1B proteins that contain what we call partly open barrel structures. These are β-sheet barrels that are not fully closed because they contain one strand which has its N-terminal region hydrogen bonded to one neighbour, its C-terminal region to the other neighbour and these two regions separated by two or more residues in a non-β conformation. As will be shown at the end of the paper, the geometries of the partly open barrels are, to a close approximation, the same as the related fully closed structures.

Note that the barrel structures in these proteins are found in very different structural contexts. In

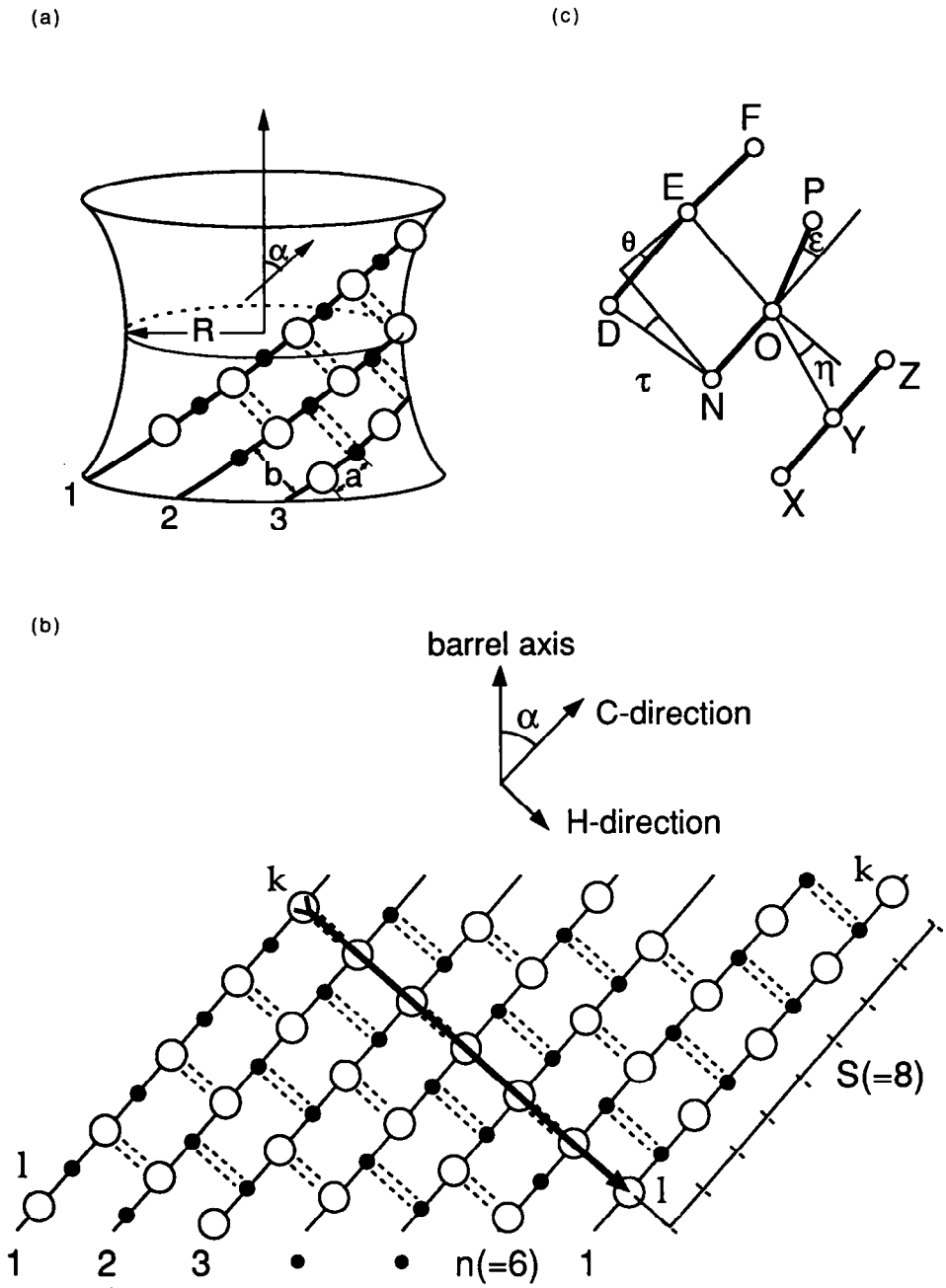

Figure 1. The structural features of β-sheet barrels and their symbols. (a) Part of the β-sheet that forms a barrel structure is shown with open circles for residues that point out of the barrel and filled circles for residues that point in. The general shape of the barrel structure is indicated by the line drawing. On this part of the Figure we mark the mean radius of the barrel, R; the mean slope of the strands to the barrel axis, α; the C^{α} to C^{α} distance along the strands, a (3·3 Å), and the interstrand distance, b (4·4 Å). (b) A plan of the β-sheet that forms the barrel structure produced by cutting along strand 1 and unrolling it onto a flat surface. The first strand is shown twice to illustrate how its hydrogen bonding to the second and last strands closes the barrel structure. On this part of the Figure we mark the number of strands in the β-sheet, n (6 here) and the measure of its stagger: the shear number S. S can be determined by starting from residue k in strand 1, move around the barrel, in a direction perpendicular to the direction of the strand, until strand 1 is reached again. Because of the stagger the point of return will not be residue k but one displaced from it, l. The shear number is $|l-k|$ (in this example $S = 8$). Note that along a strand consecutive residues alternate in the direction of their hydrogen bonding; this means that S must be an even integer. (c) We show a small section of the β-sheet: residues, D, E, F in strand 1, N, O, P in strand 2 and X, Y, Z in strand 3. The small open circles represent the position of these residues when projected on to the surface of the barrel. On this part of the Figure we show the angles used to characterize the twisting and coiling of β-sheets: the mean twist of the β-sheet about an axis perpendicular to the strand direction, θ; the mean twist of the β-sheet about an axis parallel to the strand direction, τ (it is related to the twist angle θ by the equation $\tau = (a/b)\theta$); the mean coiling of the β-sheet along the strands, ε, and the mean coiling of the β-sheet along a line perpendicular to the strands, η.

some they form almost the whole protein; in others only a small part. They include barrels formed by both parallel and anti-parallel β-sheets. Four barrels are formed by the association of subunits.

All calculations in this paper were carried out by the methods described either here or in our previous publications (Lesk *et al.*, 1989; Murzin *et al.*, 1992). We determined the residues that form β-sheet barrels by calculating from atomic co-ordinates of each protein the main-chain torsion angles and hydrogen bonds.

3. Summary of the Principles Governing β-Barrel Structures

In the accompanying paper I, we presented on the basis of our own work and that of previous authors, particularly McLachlan (1979), a set of principles for the structure of β-barrels. These principles are in two parts. The first part is a set of equations that relate the main structural characteristics of barrels (Table 2A). The second part is a geometrical description of those barrel structures that are most likely to occur in protein structures (Table 2B).

The main structural characteristics of barrel structures are defined in Figure 1 and the relationships between them given by the equations (1) to (5) in Table 2A. These relationships arise from the general stereochemical properties of β-sheets and the specific constraints arising from the formation of closed structures in which the first strand hydrogen bonds to the last.

The structural features of a barrel are determined mainly by the number of strands that form the β-sheet, n, and its shear number, S. Other features can be calculated from n and S using equations (1) to (6). This means that structures with the same n, S have very similar structures and that n, S values can be used to classify barrel structures.

The equations, by themselves, define a wide variety of structures. The structures actually found in proteins fit the equations and also usually satisfy the requirement that the residues of the β-sheet that point into the interior are close packed. (Exceptions can occur if proteins have ligand binding sites in their interior, or if they form pores.) In Table 2B we describe the ten barrel structures that, we argued, fit the equations and whose interiors can be close packed by the inward pointing residues of the β-sheet (see paper I).

As was discussed in paper I, the residue packing in barrels with $S = n$ or $S = 2n$ is different from that in barrels with S between n and $2n$. We refer to the former as "true" barrels and the latter as "orthogonal" barrels.

The analysis of the observed structures reported here shows that all the proteins in Table 1 have geometries that are well described by equations (1) to (6). In addition we find that all but five of the proteins belong to one of the ideal classes listed in Table 2B. The exceptions are proteins that accommodate other molecules or a segment of polypeptide in their interior. This allows them to have geometries that go beyond the limits imposed by the close packing of β-sheet residues alone.

Table 2

Relationships between the structural characteristics of β-sheet barrels and ideal geometries for β-barrels

A. Equations relating the structural characteristics of β-sheet barrels

$$\tan\alpha = Sa/nb \quad (1)$$

$$R = [(Sa)^2 + (nb)^2]^{1/2}/[2n\sin(\pi/n)] \quad (2)$$

$$n\theta + S\varepsilon = 2\pi\sin\alpha \quad (3)$$

$$S(a/b)\theta + n\eta = 2\pi\cos\alpha \quad (4)$$

$$(\theta-\theta^{o})^2 + \varepsilon^2 + \eta^2 = \text{constrained minimum}, \quad (5)$$

where n is the number of strands in the sheet; S is the shear number (see Figure 1); α is the mean slope of the strands (Figure 1); R is the mean radius of the barrel; a is the C^{α} to C^{α} distance along the strands, 3·3 Å; b is the interstrand distance, 4·4 Å; θ is the mean twist of the sheet; ε is the mean coiling of the strands; η is the mean coiling of the β-sheet about a direction perpendicular to the strands.

	Class		Ideal geometries				
B. *Ideal geometries for β-barrels*							
	n	S	R(Å)	$\alpha(0)$	$\theta(0)$	$\varepsilon(0)$	$\eta(0)$
True barrels	4	8	5·6	56	32	21	1
	5	10	6·7	56	27	16	−1
	6	12	7·9	56	24	13	−2
	8	8	7·2	37	29	−3	15
Orthogonal barrels	5	8	5·8	50	34	13	5
	6	8	6·2	45	33	7	10
	6	10	7·0	51	29	11	2
	7	8	6·7	41	32	2	12
	7	10	7·4	47	29	6	4
	8	10	7·9	43	28	2	7

4. Strands, Shear Number, Mean Radius and Mean Strand Tilt in the Barrel Structures

The theoretical analysis of barrel structures that can be formed by β-sheets (McLachlan, 1979; paper I) indicated that in such structures:

$$\tan \alpha = Sa/nb \quad (1)$$

$$R = [(Sa)^2 + (nb)^2]^{1/2}/[2n \sin (\pi/n)], \quad (2)$$

where n is the number of strands in the β-sheet; S, the shear number, is a measure of the extent to which the strands in the β-sheet are staggered (Figure 1); α is the mean slope of the strands to the barrel axis (Figure 1); R is the mean radius of the barrel; a is the C^α to C^α distance along the strands (3·3 Å) and b is the interstrand distance (4·4 Å). We measured the values of n, S, R and α in the observed barrel proteins to determine how well these equations described the relations between these quantities.

(a) *Determination of the strand number shear number, strand tilt and mean radius*

Strands: The determination of the number of strands in a sheet, n, is straightforward and unambiguous given the convention that any β-strand, even when it is interrupted by a β-bulge (Richardson *et al.*, 1978), is considered to be a single strand.

Shear number: To determine the shear number in each barrel structure, the main-chain hydrogen bonds were calculated and the plan of each β-sheet drawn. Residues bulging out from the strand were not included. Shear numbers were then read from the sheet plan, see Figure 1.

Strand tilt: For each structure, the tilt of the strands to the barrel axis was determined by first transforming the co-ordinates by a rotation and translation that put the barrel axis along the z-axis and its waist in the xy plane. Each strand of the barrel was approximated by a least-squares line through the main-chain N, C^α and C atoms. The dihedral angles between these lines and z-axis were then calculated and averaged for all strands in the barrel to give the mean tilt of the strands, α.

Mean radius: The mean radius of the each barrel, R, was determined by fitting a least-squares cylinder, the axis of which is fixed in the z direction, to the straight line segments fitted to the strands.

The value for the mean barrel radius is sensitive to the choice of the length of the strands used in the calculation. On one hand, barrels usually widen at both ends (see paper I) which means that the use of longer strands may give a larger mean radius. On the other hand, the approximation by straight lines of strands that are coiled effectively reduces the calculated mean radius. To avoid these complications the cylinder was fitted to the three residues in each strand that are closest to the waist of the barrel.

The barrels were classified by their n, S values. In Table 3 we list for each structure the observed values of the strand and shear numbers, the mean radius and the mean tilt of the strands. We also list the "ideal" values of the radius and tilt calculated from n, S by equations (1) and (2).

(b) *The fit of the observed strand numbers, shear numbers, strand tilts and barrel radii to the theoretical relationships*

(i) *Strand and shear numbers*

The number of strands forming the closed barrel structures, n, ranges between 4 and 16. Their shear numbers, S, are 8, 10, 12 or 20. (Recall that the shear number has to be an even integer, see paper I.) For each structure the values of n and S are given in Table 2. Of the 39 observed combinations of n and S, 34 are within the range the theoretical analysis had indicated as being favourable. The five exceptions have larger n, S values: 8, 12 or 16, 20. As already mentioned, these structures can be formed because ligands, water molecules or external loops of polypeptide contribute to the close packing of the barrel interior.

(ii) *Mean tilt of the strands to the barrel axis*

In Table 3 we list both the observed mean tilt angles, α, and those calculated from the equation:

$$\tan \alpha = Sa/nb. \quad (1)$$

The individual results (Table 3) show close agreement between the theoretical values and the observed: most differ by no more than 4°, and the average difference is 2·5°.

(iii) *The mean radius of the barrels*

The observed values of the mean barrel radii are given in Table 3 together with the theoretical values calculated from the equation:

$$R = [(Sa)^2 + (nb)^2]^{1/2}/[2n \sin (\pi/n)]. \quad (2)$$

The agreement of the observed and theoretical values is very close: no difference is greater than 0·4 Å and the average difference is 0·12 Å.

A simpler form of the equation that is symmetrical with respect to n and S:

$$R_0 = [(Sa)^2 + (nb)^2]^{1/2}/2\pi$$

gives radii that are smaller than those observed by 0·4 Å on average. A discrepancy of this size was anticipated for this equation (see paper I).

5. Twist and Coiling of the Strands in the Barrels

To form a barrel structure β-sheets have to twist and coil. The theoretical analysis (paper I) suggested that if θ is the mean twist of the sheet, ε is the mean coiling of the strands and η is the mean coiling of the β-sheet about a direction perpendicular to the strands; the twisting and coiling

Table 3

β-Sheet barrels: ideal and observed geometries

Barrel class		Ideal geometries					Observed geometries				
n	S	R(Å)	α(°)	θ(°)	ε(°)	Protein	R(Å)	α(°)	θ(°)	ε(°)	A.R.
4	8	5·6	56	32	21	OVO	5·5	50	30	21	1·05
5	8	5·8	50	34	13	ADH2	5·8	43	44	3	1·10
						PYP	5·8	49	36	11	1·25
						ULB	5·7	49	32	14	1·25
						CSP	6·1	51	38	10	1·15
5	10	6·7	56	27	16	RHV	6·6	59	43	10	1·00
						SNC	7·1	57	37	12	1·15
						BOV	6·9	51	33	14	1·45
						DRS	6·9	55	28	16	1·20
6	8	6·2	45	33	7	GCH1	6·4	42	31	9	1·40
						GCH2	6·2	42	36	4	1·30
						ALP1	6·2	44	34	8	1·10
						ALP2	6·2	43	30	10	1·25
						SNV	6·2	47	33	4	1·25
						AAR	6·2	42	47	−7	1·30
6	10	7·0	51	29	11	FNR	6·9	44	28	10	1·35
						PIA	7·0	49	32	10	1·45
						APP1	6·9	44	29	8	1·40
						HVP	7·1	47	34	6	1·30
						ETU1	6·9	46	31	8	1·30
						ETU2	7·1	49	29	11	1·15
						GSC1	7·1	45	33	8	1·10
						GSC2	6·9	49	28	12	1·25
6	12	7·9	56	24	13	TIE	8·1	59	44	2	1·05
						I1B	8·0	55	38	5	1·00
						BAR	7·9	57	45	1	1·15
7	8	6·7	41	32	2	CHB	6·7	42	33	1	1·10
7	10	7·4	47	29	6	ACN	7·4	46	30	6	1·45
8	8	7·2	37	29	−3	GOX	7·3	36	24	3	1·05
						RUB	7·3	35	30	−3	1·10
						TIM	7·0	36	24	3	1·55
8	10	7·9	43	28	2	CAT	7·7	42	24	7	1·25
						STR	7·7	40	21	9	1·20
						CPL	7·8	43	27	5	1·20
8	12	8·7	46	27	3	RBP	8·6	46	22	6	1·35
						BBP	8·5	46	23	5	1·20
						MUP	8·6	47	24	8	1·25
						BOP	8·6	42	22	7	1·35
16	20	15·5	43	17	−1	POR	15·8	42	13	2	1·15

n, S, R, α, θ, and ε are defined in Table 2. A.R., the axial ratio of the barrel cross-section, is defined in section 6(a). Full names of the proteins are given in Table 1 where they are listed in the same order.

Ideal geometries were calculated from n, S values using the equations in Table 2A. Observed geometries were calculated using the atomic co-ordinates using the procedures described in sections 4, 5 and 6.

angles in a barrel structure are related to each other, and to the other geometrical features, by the equations:

$$n\theta + S\varepsilon = 2\pi \sin \alpha \quad (3)$$

$$S(a/b)\theta + n\eta = 2\pi \cos \alpha. \quad (4)$$

It also indicated that for given values of n and S there are ideal values of θ, ε, and η (Table 2B) and that departures from these ideal values should be correlated.

(a) *Determination of the twist and coiling of the strands and β-sheets*

Twist: To determine the twist between adjacent strands, each strand was divided first into overlapping segments four residues long. (Residues in β-bulges were not included.) The twist angle between two strands was then calculated by taking the average value of the angles between the segments that are in register in the two strands. The values for each pair of strands were then averaged to give the mean twist of the β-sheet, θ.

C-Coiling: The same four residue segments were used to determine the coiling of the strands. For each strand the angles between all its constituent segments were calculated. These angles were then averaged to give a value for the mean strand coiling ε. In the averaging, the angles were given weights corresponding to the number of residues by which the segments were shifted relative to each other: 1 for the angles between segments shifted by one

Table 4

Correlation of the differences between the ideal and the observed values of the twist and coiling angles (°)

Barrel class		Ideal values for twist and coiling				Observed values for twist and coiling			Correlation of the differences between ideal and observed values	
n	S	θ	ε	η	Protein	θ	ε	η	$\Delta\theta+\Delta\varepsilon$	$\Delta\theta+\Delta\eta$
7	10	29	6	4	ACN	30	6	4	−1	−1
8	8	29	−3	15	GOX	24	3	19	+1	−1
8	8	29	−3	15	RUB	29	−3	17	−3	+2
8	10	28	2	7	CAT	24	7	10	+1	−1
8	10	28	2	7	STR	21	9	15	0	+1
8	12	27	3	−1	RBP	22	6	9	−2	+5
8	12	27	3	−1	BBP	23	5	5	−3	+2
16	20	17	−1	0	POR	13	2	5	−1	+1

residue, 2 for two residue shifts and 3 for three residue shifts. The sign of each angle is that of the triple product $(\mathbf{s}_1 \times \mathbf{s}_2)\cdot\mathbf{z}$ where $\mathbf{s}_1$ and $\mathbf{s}_2$ are unit vectors determining the two segments and $\mathbf{z}$ is the unit vector in the z direction. Long strands were truncated to seven residues with the middle one being the closest to the waist of the barrel.

(Note that this method does not distinguish between the cases where the coiling is smoothly spread over several residues and those where it arises from a sharp bend produced by a single residue.)

H-Coiling: The coiling of the β-sheets in a direction perpendicular to that of the strands, that is, the direction of the hydrogen bonds, was determined by a method similar to that for determining the coiling along the strands. The peptides along lines perpendicular to the strand direction were divided into sets each containing an overlapping set of three peptides. A least-squares line was fitted to each set. The angles between all constituent sets were calculated and averaged with the weighting scheme described above to give the mean value for the H-coiling, η.

This method of calculating η requires at least four peptide groups in each H line; a condition found only in certain of the larger barrel structures. This means that it was possible to calculate η values for only eight of the barrel structures discussed here (Table 4).

(b) *The observed values of the twist and coiling of the strands and the relations among these quantities and to other aspects of the structures*

(i) *The twist of the β-sheets*

The average value of the twist angle, θ, is 32° and its standard deviation is 8°. This means the twist of the β-sheets in barrel structures is greater than is generally found in the two other major classes of structures in which β-sheets are found. In the α/β structures, where α-helices pack on one or both sides of a β-sheet, the mean twist angle is about 19° (Janin & Chothia, 1980). In the aligned class, where β-sheets pack face to face, the mean twist angle is 17° (Chothia & Janin, 1981).

For individual strands, the extent of the twist is related to the number of β-sheet hydrogen bonds they form. In the majority of cases the strands in the barrel structures are linked to each other by three to six hydrogen bonds. The distribution of twist angles and the number of interstrand hydrogen bonds is shown in Figure 2(a). For the set with four or more hydrogen bonds the twist angles tend to be smaller and the distribution of angles more compact. This is because very large twists prevent the formation of more than a few hydrogen bonds. Strands that have twists $\theta > 45°$ form on average just two hydrogen bonds to each other. Thus barrels with high mean twist angles have β-sheets with several strands linked by only a few hydrogen bonds; for example, *Erythrina* trypsin inhibitor with $n = 6$ and a mean θ of 44° has a β-sheet in which three pairs of strands have only two hydrogen bonds linking them (Onesti *et al.*, 1991; Murzin *et al.*, 1992).

(ii) *C-Coiling of the β-sheet strands*

The different structures have mean strand coiling angles, ε, between −7° and 21° (Table 3). Most values, however, are between 1° and 14° and the average value is 7°.

Strong coiling is produced by values of the main-chain torsion angles (ϕ, ψ) that produce, alternately, small and large values of $|\phi-\psi|$ (see paper I). For such strands there is a strong tendency for (1) residues pointing into the barrel to be small and to have (ϕ, ψ) angles in the upper left corner of the β-region (and hence large $|\phi-\psi|$ values) and (2) residues pointing out of the barrel to be large with (ϕ, ψ) near the centre of the β-region (and hence small values of $|\phi-\psi|$). This suggests that the coiling is sequence-dependent. However, though this pattern of residues may facilitate coiling, it is not a necessity as there are a few cases of strands that have strong regular coiling and large residues pointing in and out of the barrel.

Strong coiling of a strand tends to reduce the number of β-sheet hydrogen bonds that it makes. In

θ

n_H	0°	4°	8°	12°	16°	20°	24°	28°	32°	36°	40°	44°	48°	52°	56°	60°	Total
1											1		1	1	1		4
2			1		1	1			1	1		4	5	3		3	20
3			1			3	1	1	6	3	4	3	1	1	2		26
4			2		2	6	4	12	11	5	7	2					51
5		1			1	9	9	10	8	7	1						46
6			1	3		4	5	4	9	3	1	1					31
>6		3	5	7	9	9	10	9	3	4	1						66
Total	0	4	10	10	13	32	35	36	38	23	15	10	7	5	3	3	244

(a)

ε

n_P	0°	3°	6°	9°	12°	15°	18°	21°	24°	27°	30°	Total
0		1	1	2	5	2	1	1		1		14
1		2	4			6	3					15
2	2	9	14	16	15	3	5	4				68
3	2	8	12	9	4	6	2	2		1		46
4	6	9	9	6	4	5						39
5	3	6	8	5		1						23
>5	7	10	13	4	1	4						39
Total	20	45	61	42	29	27	11	7	0	2	0	244

(b)

Figure 2. Interstrand hydrogen bonds and strand geometry. (a) For each pair of adjacent strands, we show the joint distribution of their twist angles, θ, and the number of interstrand hydrogen bonds, n_H. (b) For all individual strands, we show the joint distribution of their coiling angles ε and the number of *peptides* that make hydrogen bonds to *both* adjacent neighbours, n_P.

Figure 2(b) we plot the coiling angle, ε, of each strand against the number of peptides that make hydrogen bonds to both neighbouring strands, n_p. The larger the value of n_p the smaller the mean value of ε. Inspection of the few strongly coiled strands with $n_p > 3$ shows that most of these hydrogen bond to strands with bulges or some other irregularity.

In several cases, large values of ε are produced by a sharp bend at a single site in a strand. Residues at such sites are invariably glycines in extended conformations with ϕ, ψ values outside the normally allowed β-sheet region. These "Gly-kinks" allow the regular hydrogen bonding to be maintained on both sides of the kinked strand; unlike β-bulges. The coiling in most of the strands for which $\varepsilon \geq 15°$ and $n_p > 4$ is produced by Gly-kinks.

(iii) *H-Coiling of the β-sheets*

The individual 8 barrel structures for which H-coiling of the β-sheets can be calculated have mean coiling angles, η, between 5° and 19° (Table 4). The average value for all the structures is 11°.

(iv) *The observed and ideal values for the twist and coiling angles*

The theoretical analysis presented in paper I suggested that, if barrels with given values of n and S have ideal isotropic stressed β-sheets, the ideal values of the twist angle, θ, and coiling angles, ε and η, can be calculated by minimizing the function:

$$f = (\theta - \theta_0)^2 + \varepsilon^2 + \eta^2, \quad (5)$$

subject to the conditions imposed by equations (3) and (4). The ideal values of θ, ε and η for different barrel structures are given in Table 2.

The analysis also predicts that for individual structures the differences between the ideal and observed values of the twist and coiling angles will be nearly complementary so that:

$$\Delta\theta + \Delta\varepsilon \approx 0$$

and

$$\Delta\theta + \Delta\eta \approx 0.$$

Comparison of the observed and ideal values of θ and ε (Table 3) shows that in most structures they are very close. For 31 of the structures, $\Delta\theta$ and $\Delta\varepsilon$

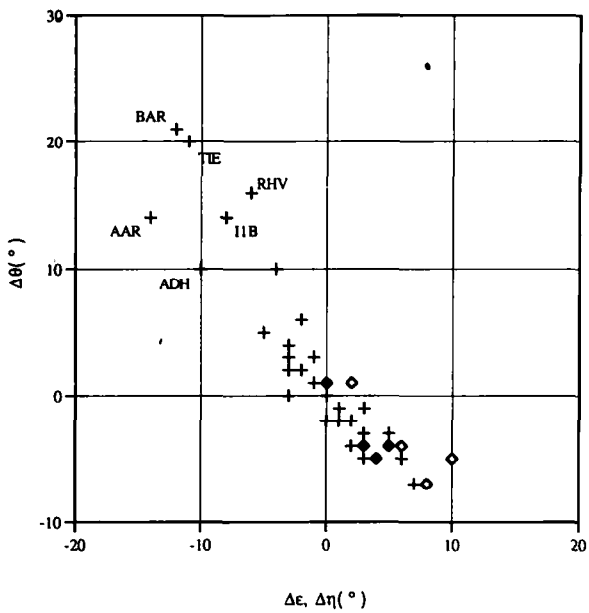

Figure 3. The differences between the ideal and observed values of the twist and coiling angles are small and correlated. We plot the difference between the observed and calculated values of the twist ($\Delta\theta$) and coiling angles ($\Delta\varepsilon$ and for those structures where there is data, $\Delta\eta$) (see Tables 3 and 4). Points relating values of $\Delta\theta$ and $\Delta\varepsilon$ are indicated by + and those relating values of $\Delta\theta$ and $\Delta\eta$ are indicated by ◇. Most of the differences are small and the few larger differences are discussed in the text, see section 5(b)(iv). The theoretical analysis predicted that for individual structures the differences between the ideal and observed values of the twist and coiling angles will be nearly complementary so that:

$$\Delta\theta + \Delta\varepsilon \approx 0$$

and

$$\Delta\theta + \Delta\eta \approx 0$$

as is generally observed.

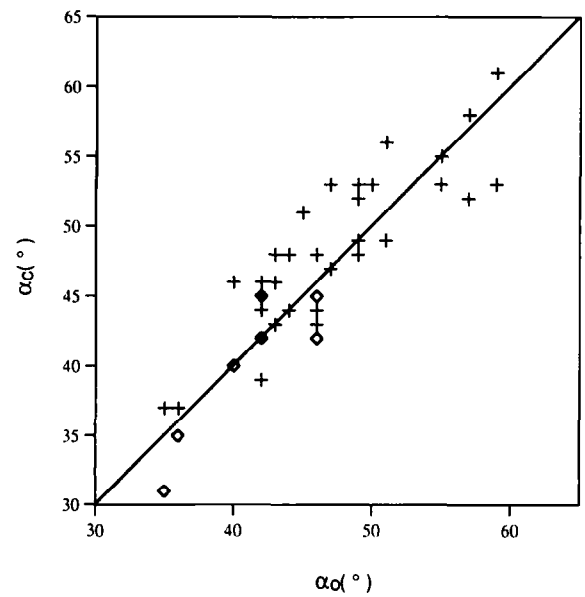

Figure 4. The observed and calculated values of the tilt angle α. The theoretical analysis showed that the tilt angle α is related to the other geometrical features by the equations:

$$n\theta + S\varepsilon = 2\pi \sin\alpha$$

and

$$S(a/b)\theta + n\eta = 2\pi \cos\alpha.$$

Here we show the observed values α_0 plotted against those calculated from the observed values of the other geometrical features, α_c. The α_c calculated from the first equation are indicated by + and those calculated from the second equation are indicated by ◇.

are no more than 7° and average 3° and 2·5°, respectively. The largest deviations are found in the class of structures for which $n = 6$ and $S = 12$. These have $\Delta\theta = 14°$ to 21° and $\Delta\varepsilon = -8°$ to $-12°$ (Table 3). Inspection of these structures shows that at one end of the barrel the strands twist and coil away to become parts of different β-sheet structures (see Murzin *et al.*, 1992).

For each structure the values of $\Delta\theta$ and $\Delta\varepsilon$ are plotted in Figure 3. This clearly shows that the differences are correlated and tend to cancel; for 30 of the structures the value of $|\Delta\theta + \Delta\varepsilon|$ is 2° or less. Large values are found only in the $n = 6$, $S = 12$ structures discussed in the last paragraph.

The values observed for the coiling of the β-sheet perpendicular to the strand direction, η, exceed the ideal values by up to 10°; the average difference is 5°. As with ε, the differences in η are correlated with departures of the twist angle from the ideal values so that the values of $\Delta\theta + \Delta\eta$ are less than 2° (Table 4).

(v) *The relation between the twist and coiling and the tilt of the strands to the barrel axis*

The mean twist of the β-sheet, θ, the mean coiling of its strands, ε, and the tilt of the strands to the barrel axis, α, should be linked by the approximate equation:

$$n\theta + S\varepsilon = 2\pi \sin\alpha. \quad (3)$$

To determine how well this relationship applies to real structures, we calculated, for each barrel, values of α using the observed values of n, θ, S and ε. In Figure 4 these calculated values of α are plotted against the observed values. This Figure shows clearly that the two sets of values are very similar: the average difference between them is less than 3°.

A similar equation should relate θ and η to α:

$$S(a/b)\theta + n\eta = 2\pi \cos\alpha. \quad (4)$$

For the eight structures for which we are able to calculate a value of η, we used equation (4) to determine α. The results are given in Figure 4. These α values differ from the observed values by between 0° and 4° and the average difference is less than 2°.

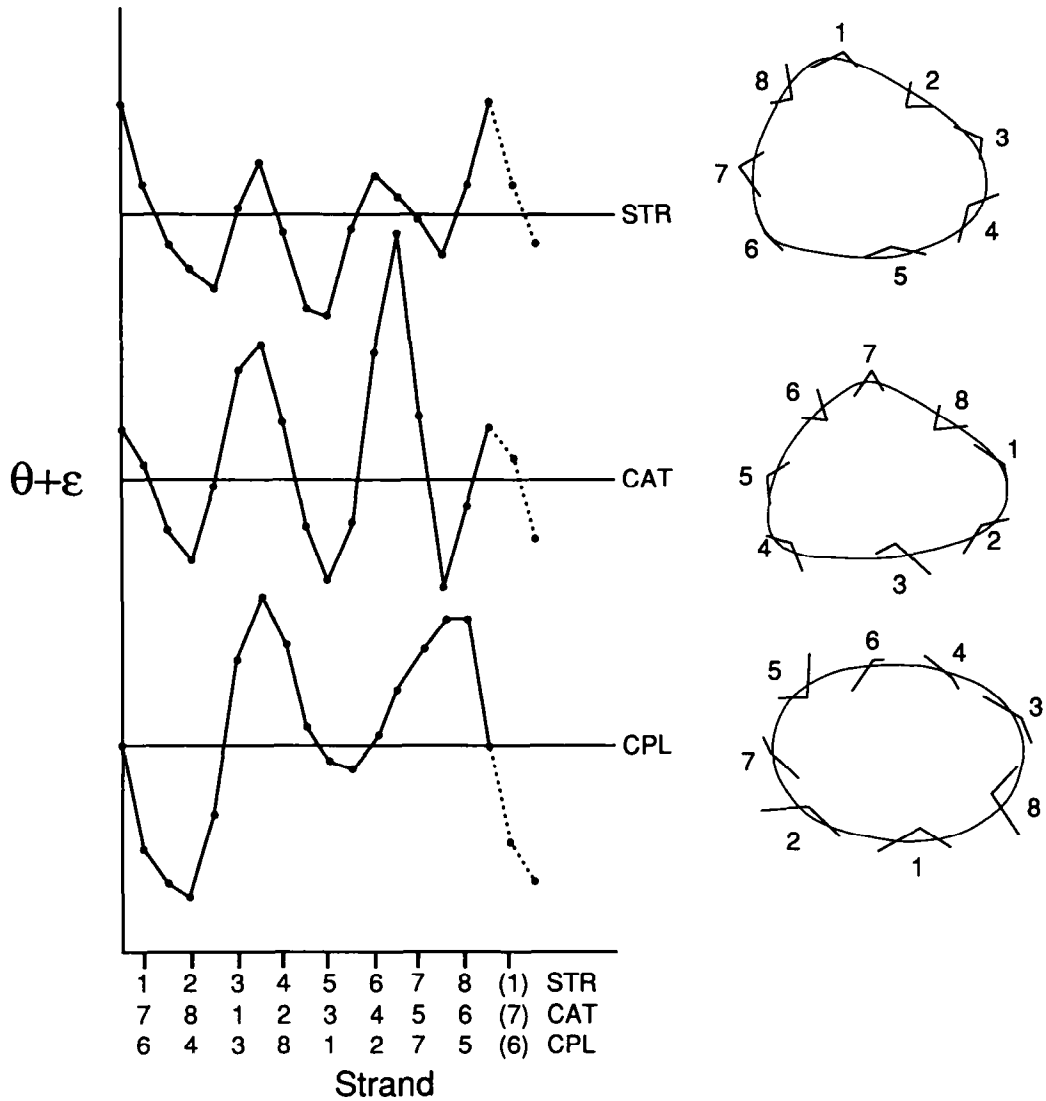

Figure 5. The effect of local variations in the twist angle, θ, and coiling angle ε on the shape of barrel cross-sections. Here, for 3 barrels with $n = 8$, $S = 10$: streptavidin (STR), catalase (CAT) and cyclophilin (CPL), we give on the left of the Figure the value of $\theta+\varepsilon$ for each individual strand and interstrand region. For the individual strands, the value of $\theta+\varepsilon$ is the sum of the mean value of ε and of the average of the θ values given the strand's 2 neighbours. For the regions between 2 strands the value of $\theta+\varepsilon$ is the sum of their twist angle, θ, and of the mean value of their coiling angles ε. On the right, we use the positions of the C^{α} of 3 residues from the central region of each strand to illustrate the shape of the equatorial cross-sections in the 3 barrels. The order of the strands counted clockwise from the top of each diagram corresponds to the order of the strands in the plots. Comparisons of the left and right sides of the Figure shows how "corners" and "flat sides" in the distorted barrels correspond to maxima and minima in the function $G(\theta,\varepsilon) = \theta+\varepsilon$.

6. Barrel Deformations

The previous sections have shown that the equations in Table 2 accurately relate and describe the principal structural features of β-barrels. Detailed inspection of the structures shows, however, that they have local departures from ideal geometry most of which are not reflected in the mean values of the structural features. They arise from the changes in β-sheet conformation that facilitate the packing of residues in the barrel interior, as will be discussed in section 7. Their most striking effect is on the shape of the equatorial cross-section of the barrels. In this section we first describe the extent to which the β-barrel structures treated here are deformed and then discuss how the deformations are facilitated by modifications in the conformation and hydrogen bonding of the barrel β-sheet.

(a) *The axial ratios of the cross-sections of β-barrel structures*

Inspection of the barrel structures shows that most of them are distorted so that their cross-sections are not circular. Usually they are elliptical; though in a few of the larger structures they are more like a rounded triangle (Figure 5). A general measure of the extent of this sort of distortion in the barrel structures is given by fitting an ellipse to the

cross-sections of the barrels and determining the axial ratio: the ratio of the major and minor axes of the ellipse.

A set of residues that defines the "waist" of each barrel was chosen and the molecule oriented so that the z axis was perpendicular to a least-squares plane through the C^{α} atoms of these residues. For all waist residues i the atoms $C^{\alpha i-1}$, C^{i-1}, N^{i}, $C^{\alpha i}$, C^{i}, N^{i+1}, $C^{\alpha i+1}$ were projected onto the xy plane and a least-squares ellipse fitted to this set of points. We give, in Table 3, the axial ratio of the barrel cross-section in each structure. The values vary between 1·00 and 1·55.

Most of the true barrel structures, those where $S = n$ or $S = 2n$, have little or no distortion (Table 3). Of the 11 true barrel structures discussed here, nine have axial ratios between 1·00 and 1·20 and a mean value of 1·10. However, the two exceptions, TIM and BOV, are among the most distorted of all the barrels with values of 1·55 and 1·45, respectively.

Distortions are much more common in the orthogonal barrels, those where $n < S < 2n$. They have axial ratios of between 1·10 and 1·45 with a mean value of 1·25 (Table 3). This difference in the behaviour of the two sets of barrels is discussed below in the section on residue packing.

(b) *Changes in shape through variations in twist and coiling*

For the barrels to have elliptical or triangular cross-sections, rather than circular, the local curvature of the β-sheet has to be variable. In paper I we showed that the radius of curvature at each point on the equatorial cross-section is roughly proportional to the sum of the local values of the twist and coiling angles, $\theta+\varepsilon$. This means a representation of the shape of the equatorial cross-section should be given by a plot of the sum of $\theta+\varepsilon$ *versus* the strand number. Regions with high values of $\theta+\varepsilon$ will have high curvature and correspond to "corners"; those with low values will have low curvature and correspond to flatter "sides" (Figure 5).

As described above most β-barrels have elliptical cross-sections with two corners and two sides as in, for example, cyclophilin. In a few cases they have three corners giving a triangular cross-section, as occurs in catalase. In Figure 5 we plot the values of $\theta+\varepsilon$ *versus* strand number for these two proteins.

In cyclophilin one corner is located between strands 3 and 8 and the second in strand 7 (Figure 5). In each of these strands strong local coiling is produced by β-bulges. In catalase (Figure 5) one corner is between strands 1 and 2 and is produced by a Gly-kink in strand 1; a second corner is between strands 4 and 5 which are strongly twisted with $\theta = 49°$, and the third corner is between strands 6 and 7 and results from the combination of moderate twist and coiling in both strands.

(c) *Hydrogen bonding in corner strands*

Strong coiling or twisting limits the extent of the hydrogen bonding possible between a strand and one or both of its neighbours. This means that corner strands usually have disruptions in the patterns of hydrogen bonds. In most cases this takes one of two forms: either an unusually small number of hydrogen bonds is formed by the corner strand to one of the neighbouring strands; or the first part of the strand hydrogen bonds to one neighbour and the second part hydrogen bonds to the other.

In most barrels, corners are not equally affected by the distortions of the structure: there is a strong tendency for the defects to accumulate in just one corner where the hydrogen bonding pattern is disrupted. Thus, in the examples just discussed, strands 4 and 5 in catalase have only two hydrogen bonds between them, and the two halves of strand 3 in cyclophilin hydrogen bond to different neighbours. The most extreme cases are found in cellobiohydrolase (CBH), cold shock protein (CSP) and the reductases (FNR and PIA) where the barrels are closed by only a single hydrogen bond from the end residue of a short strand.

(d) *Unrelated proteins in the same (n, S) class may have similar distortions and related proteins can have different distortions*

In the previous section we described how the catalase barrel has three corners so that the cross-section of the barrel is triangular in shape. A very similar distortion is found in a barrel that belongs to the same class ($n = 8$, $S = 10$) as catalase: streptavidin. Although their cross-sections are very similar in appearance the distortions occur in different positions. In catalase the corners are found at strands 1-2, 4-5 and 6-7; in streptavidin they occur at strands 3-4, 6 and 8-1.

The close similarity of the distortions in the two proteins is demonstrated by a least-squares fit of their atomic co-ordinates. If strands 1, 2, 3...., 8 of streptavidin are superposed on strands 7, 8, 1,...6 of catalase, respectively, the r.m.s. difference in the position of the main-chain atoms of 41 β-sheet residues is 1·26 Å.

Conversely, proteins that are related but whose sequences have diverged greatly can have quite different distortions. For example, the distantly related retinol and bilin binding proteins have, respectively, elliptical and triangular cross-sections.

7. Structural Comparisons of β-Sheet Barrels with the Same Number of Strands

Similarities in the structures of certain β-sheet barrels have been noted previously (Ke, 1992; Flower, 1993). Now the n, S values of an ideal barrel determine its radius, the slope of the strands relative to the barrel axis and other geometrical features. This means that ideal barrels with the

same n, S values will have the same structure. In practice, however, the general shape of most barrels will be deformed so as not to have a circular cross-section, as was discussed in the last section. In addition, the residues at the ends of strands often have local differences in the extent of their twist and coiling. To determine the effects of these features we calculated the differences in the atomic co-ordinates of some barrels that have the same values of n, S.

The structures of all examples of eight-stranded barrels were compared: that is the four with $n = 8$, $S = 12$, the three with $n = 8$, $S = 10$, and the three with $n = 8$, $S = 12$ (Table 3). For barrels with the same n, S values, the β-sheets were superposed and the r.m.s. difference in position of C^{α} atoms of the common β-sheet regions calculated. We also calculated the r.m.s. differences for all main-chain atoms. We could do this because all the structures that we use here with $n = 8$ and $S = 10$ and 12 have anti-parallel β-sheets and all those with $n = 8$ $S = 8$ have parallel β-sheets. These differences are the same as, or only marginally larger than those for the C^{α} atoms (Table 5A).

These calculations show that for groups of proteins with the same n, S values the r.m.s. differences in the co-ordinates of their main-chain atoms are between 1·3 and 2·3 Å (Table 5). The larger differences are mainly the result of a strand accommodating a β-bulge in one structure but not the other. Removing residues at the ends of strands from the comparisons reduces the size of the differences: if we take for each strand just the three residues that form the waist of the barrel (see section 6(a) and Figure 5) the r.m.s. differences are mostly in the range 1·0 to 1·2 Å.

The results in Table 5A show that barrels with the same n, S values have very similar structures. This is in spite of some of the barrels having very different structural features. In several cases the strands are connected quite differently. The glycolate oxidase (GOX) barrel is almost circular in cross-section (the axial ratio is 1·05) but that in triose phosphate isomerase (TIM) has the largest axial ratio of the structures considered here (1·55): the r.m.s. difference of the main-chain atoms of the 36 residues that form their common β-sheet is 1·5 Å. The retinol binding protein (RBP) barrel is formed by a large single β-sheet; that in E2 DNA-binding protein (BOP) is formed at the interface between two subunits by two four stranded β-sheets that have α-helices packed on their surfaces: the r.m.s. difference for the 49 residues that form their common β-sheet is 1·8 Å.

We also compared structures that have the same value for n but different values for S. For a given n, different S values produce different radii and strand tilts (Table 3). If Δ_0 is the mean r.m.s. difference between barrels with the same S values, the r.m.s. difference between barrels with different S values, Δ_S can be estimated from the equation:

$$\Delta_S^2 = \Delta_0^2 + (\Delta R_S)^2 + (l\pi\Delta\alpha_S/720)^2,$$

where ΔR_S and $\Delta\alpha_S$ are the differences in the radius and the tilt, respectively; and l is the mean length of a strand, typically $l = 18$ Å (five residues). This means:

$$\Delta_S^2 = \Delta_0^2 + (\Delta R_S)^2 + (0{\cdot}08\Delta\alpha_S)^2.$$

From the values of R and α in Table 3 we would expect that the contributions of ΔR_S and $\Delta\alpha_S$ terms will be comparable with Δ_0 only when S values differ by 4 or more.

Examination of the observed values (Table 5B) shows that this is indeed the case. When the S values differ by 2 (i.e. 8 *versus* 10 and 10 *versus* 12) the r.m.s. differences are the same as or slightly larger than the values found for structures with the same S value; but when the S values differ by 4, the r.m.s. differences are roughly twice as great.

Table 5

Structural comparisons of eight stranded β-sheet barrels

A. *Barrels with the same* n *and* S *values*

		RBP	BBP	MUP	BOP
	RBP	—	1·4 (1·5)	1·6 (1·7)	1·8 (1·8)
$n=8$	BBP	45	—	2·1 (2·1)	2·1 (2·2)
$S=12$	MUP	44	44	—	2·2 (2·3)
	BOP	49	45	44	—

		CAT	CPL	STR
	CAT	—	1·7 (1·8)	1·5 (1·6)
$n=8$, $S=10$	CPL	40	—	1·6 (1·8)
	STR	48	40	—

		GOX	RUB	TIM
	GOX	—	1·2 (1·3)	1·5 (1·5)
$n=8$, $S=8$	RUB	37	—	1·5 (1·6)
	TIM	36	39	—

B. *Barrels with the same* n *but different* S *values*

		GOX	STR	BOP
$n=8$, $S=8$	GOX	—	1·9 (—)	3·7 (—)
$n=8$, $S=10$	STR	41	—	2·2 (2·2)
$n=8$, $S=12$	BOP	34	41	—

Full protein names are given in Table 1. In the comparison tables we give for each pair of proteins 2 or 3 numbers. On the lower left we give the number of residues in the β-sheet common to both proteins. On the top right we give, in Å, the r.m.s. difference in position of the C^{α} atoms of these residues and, in parenthesis, the difference for all main-chain atoms where possible.

8. Residue Packing in the Barrel Interiors

Although the structural details of the packing within any particular barrel structure are complicated and probably unique, some generalizations can be made. These concern (1) the arrangement of the residues in barrel interiors, (2) their role in barrel distortions and (3) the relation between the barrel radius and the mean volume of the interior residues. The residue arrangements in true barrels,

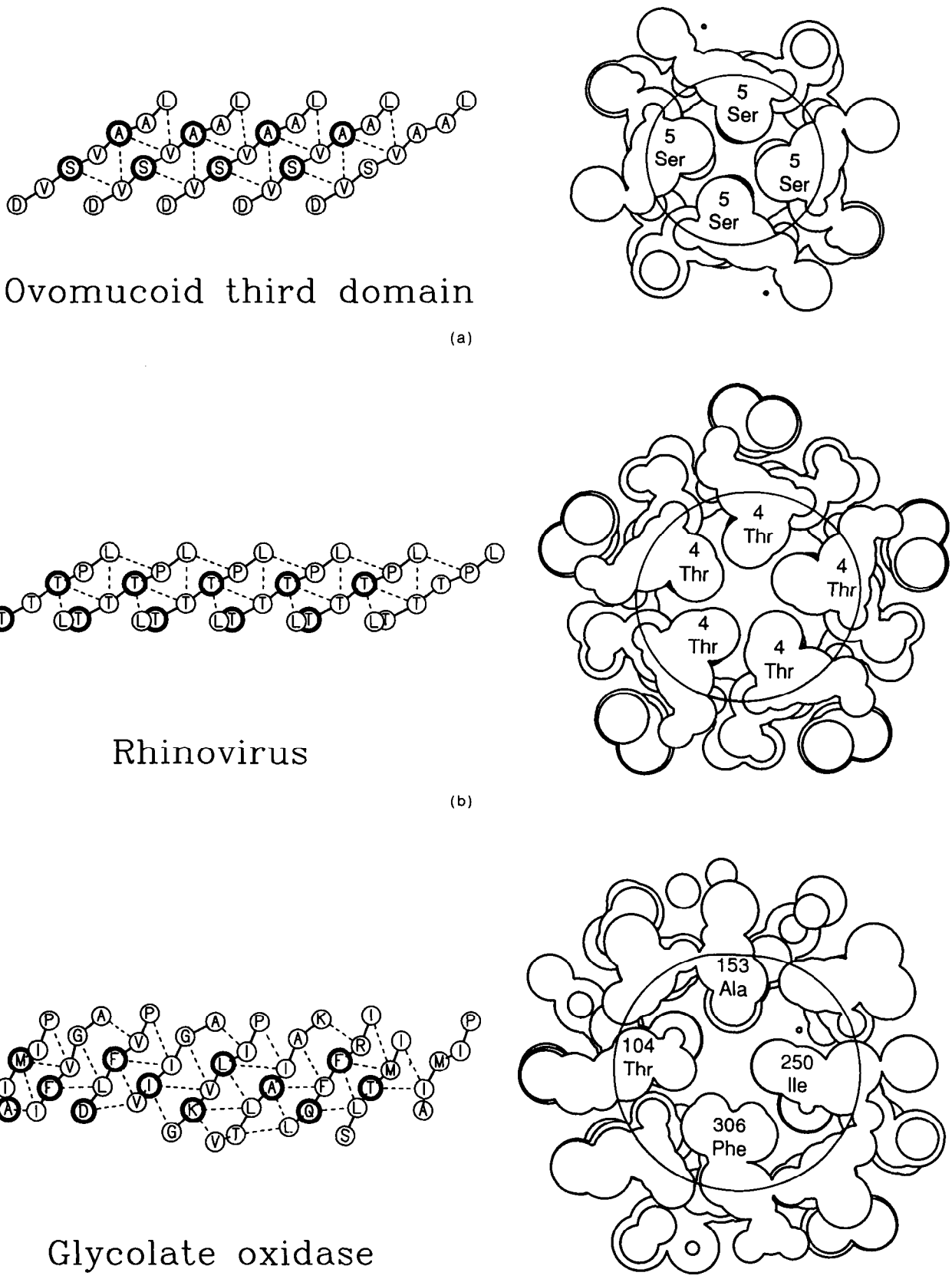

Figure 6. Residue packing in true barrels. On the left, we show plans of the β-sheets that form barrel structures in (a) ovomucoid (b) rhinovirus and (c) glycolate oxidase. On the right, we use serial sections cut perpendicular to the barrel axes to show the residue packing at the centre. In these structures where $S = 2n$, in the first 2 cases and $S = n$, in the third case, the residues in the interior pack in clear layers.

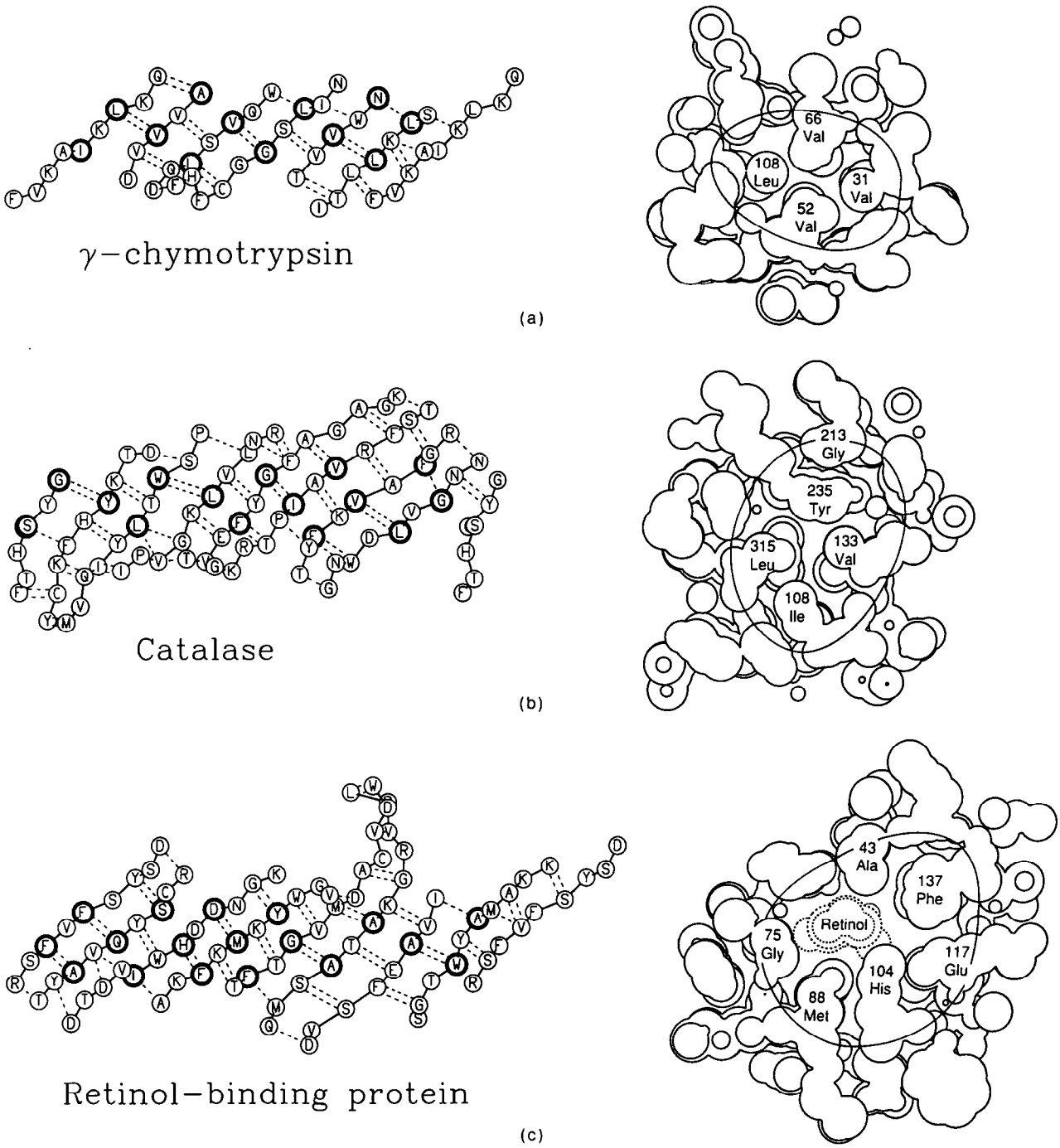

Figure 7. Residue packing in orthogonal barrels. On the left, we show plans of the β-sheets that form barrel structures in (a) chymotrypsin, (b) catalase and (c) retinol binding protein. On the right, we use serial sections cut perpendicular to the barrel axes to show the residue packing at the centre. In chymotrypsin and catalase the barrels are distorted to bring residues on opposite faces close together. In retinol binding protein the n, S values of the barrel are 8, 12 which produce a barrel with a large radius and a binding pocket for retinol.

where $S = n$ or $S = 2n$, are simpler than, and distinct from, those in orthogonal barrels, where S has a value between n and $2n$. We discuss separately these two groups of barrels.

(a) *Residue packing in true barrels*

In true barrel structures, the tilt of the strands places the interior residues in layers normal to the barrel axis (Figure 6).

When $S = 2n$ and $\alpha \approx 56°$, each strand contributes one residue to each layer. This means that for $n = 4$, 5 or 6 there are 4, 5 or 6 residues in each layer. An example of each of these types of packing is illustrated in Figure 6. When $S = n$ and $\alpha \approx 36°$, the strands are tilted so that alternate strands contribute residues to alternate layers. In practice only $n = S = 8$ is found and in these structures each layer has four residues. An example of this type of packing is shown in Figure 6.

In section 6 we saw that most true barrels have little distortion. This is because if the interior residues are similar in size the symmetric packing arrangement gives undistorted barrels. In some cases (OVO and RHV) the symmetry is exact because different subunits provide strands of the same sequence to give barrels with the rotational symmetry of the oligomers (Musil *et al.*, 1991; Arnold & Rossmann, 1990). In the β-trefoil proteins the large diameter of the barrel is filled by large residues packing in an arrangement that has pseudo three-fold symmetry (Murzin *et al.*, 1992).

Barrel distortions have the effect of moving opposing faces relative to one another and so allow heterogeneous sets of side-chains to close pack. TIM has the most distorted of all the barrels discussed here. Inspection of its structure shows that it has a large cavity in the centre of the barrel next to a Gly residue. The large distortion in the barrel brings the side with the Gly residue closer to that on the opposite side to reduce the size of the cavity. (An inspection of the TIM sequences from different species shows that residues around this cavity are mostly conserved (unpublished results). This suggests that the cavity and the concomitant distortion are important for the structure of the active site.)

(b) *Residue packing in orthogonal barrels*

When $n < S < 2n$ the tilt of the strands is such that residues in the interior do not form layers that have any simple relation to the symmetry of the barrel. The rows of residues that form the layers in the true barrel structures are, in the orthogonal barrels, at an angle to the plane perpendicular to the barrel axis (Figure 7).

This irregular arrangement of residues, when combined with the normal heterogeneous nature of side-chains, means that there is a strong tendency for orthogonal barrels to be distorted; as was described in section 6. The most common distortion gives the barrel an elliptical cross section. This has the effect of moving the two long "sides" closer together and large residues on these sides come into contact. The patterns of these contacts are discussed in the caption to Figure 7. The "corners" tend to be occupied by small residues and the residues making contacts between the sides tend to form irregular layers that fill the interior.

(c) *Relation between the mean volume of interior residues and the barrel radius*

In paper I it was proposed that if the barrel was filled just by the interior β-sheet residues the mean radius of the barrel should be related to the mean volume of the residues by the equation:

$$V - V_0 = abR, \qquad (6)$$

where $a = 3{\cdot}3$ Å and $b = 4{\cdot}4$ Å (see Table 2) and V_0 is the volume of the part of the interior residue's main-chain that lies outside the perimeter of the barrel. To analyse the accuracy of this equation we determined, for each barrel structure, the mean volumes of interior residues and plotted the values against those of the mean radii (Figure 8).

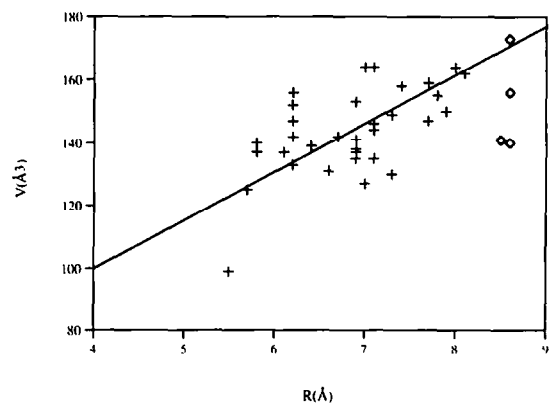

Figure 8. A plot of the weighted mean volume of residues buried in barrel interiors, V, against the barrel radius, R (see section 7(c)). The values for the 10 barrels classes in which the interior is close packed by β-sheet residues are shown by + symbols. The values for the barrels with $n = 8$, $S = 12$ are shown by ◇ symbols. The line on the graph corresponds to the equation $V = abR + V_0$, where $ab = 15$ Å^2 and $V = 40$ Å^3 (see section 7(c)).

Mean residue volumes were calculated using the residues that formed the three central layers of the barrels. There is some tendency in the outer layers for large residues to protrude from the barrel or, if some interior residues are small, for external residues to pack into the barrel. In calculating the mean volumes, therefore, we gave residues on the central layer (i.e. those near the equatorial plane) twice the weight of those in the outer layers.

The plot of the mean volume of the interior residues *versus* radius is shown in Figure 8. The correlation coefficient between V and R is 0·6. Thus there is only a weak relation between V and R. Inspection of the structures shows that, on a more detailed level, those lying above the theoretical line in Figure 8, i.e. those with "excess" volume have residues that protrude from the barrel. In nearly all cases, those that lie below the line will either have external residues fill the space or have it filled by distortions of the barrel.

9. Barrels With Large Radii

Five of the barrel structures have (n, S) values that are not one of the ten combinations we described as consistent with structures having good β-sheet geometries and interiors closed packed by β-sheet residues. Four of these, RBP, BBP, MUP and BOP have $n = 8$ and $S = 12$. The fifth, POR, has $n = 16$ and $S = 20$.

The structures with $n = 8$, S = 12 have radii of 8·5 to 8·6 Å. This means that the β-sheet residues cannot fill the barrel and must leave a channel at the centre. Three, RBP, BBP and MUP are

members of the lipocalin family of transport proteins. Their binding sites are at one end of the channel; the other is filled by large side-chains from a segment of polypeptide that folds across the ends of their barrels. In the apo form of RBP water molecules occupy the space that is occupied by retinol in the holo form (Zanotti *et al.*, 1993). A similar set of water molecules is seen in the axial channel of the BOP barrel. This barrel is formed by the association of two subunits of which the barrel is only a small part (Hegde *et al.*, 1992).

The fifth structure is the largest barrel known at present, porin, which has $n = 16$ and $S = 20$ and a radius of 15·8 Å. This integral membrane protein forms pores (hence its name). The pore structure is formed mainly by a large loop of polypeptide that fills much of the space left at the centre by the β-sheet residues.

For all five of these large barrels the values R, α, θ, ε and η are very close to those calculated from their (n, S) values (Tables 3 and 4). We showed in paper I that, if the packing of the internal β-sheet residues is not taken into consideration, the value of S that gives the least stressed barrel is related to n and is an even number close to that given by the equation:

$$S = n + 4{\cdot}2.$$

These five structures in which the large radii reduce the extent of the interactions between the internal β-sheet residues have (n, S) values that fit this equation: 8, 12 and 16, 20.

10. Chain Topology in Barrel Structures

(a) *Strand connections*

A variety of studies, reviewed in paper I, have shown that the topological regularities found in protein structures can be expressed in terms of three rules that hold in almost all cases:

(1) pieces of secondary structure that are adjacent in the sequence are also often in contact in three dimensions;

(2) the connections in β-X-β units (where the βs are parallel strands in the same β-sheet, though not necessarily adjacent, and X is an α helix, a strand in a different β-sheet or an extended piece of polypeptide) are right-handed; and

(3) the connections between secondary structures neither cross each other nor make knots in the chain.

Inspection of Table 6, in which we describe the frequency of different types of interstrand connections occurring in the 39 barrel structures, shows that the three rules accurately describe what occurs. More than 80% of the connections are between adjacent strands (rule 1). All but one of the connections between parallel strands are right-handed (rule 2).

Table 6

Chain topology in the barrel structures

Type of interstrand connection	Number of cases
A. *Interstrand connections*	
Turns and loops between:	
1. Adjacent antiparallel strands	133
2. Non-adjacent antiparallel (across the ends of the barrels)	27
3. Non-adjacent antiparallel (across the outside of the barrels)	5
4. Adjacent parallel strands (right handed connections)	32
5. Adjacent parallel strands (left handed connections)	0
6. Non-adjacent parallel strands (right handed connections)	2
7. Non-adjacent parallel strands (left handed connections)	1
Total number of connections:	202

Barrel class n	Barrel class S	Number of structures	Number of different topologies
B. *Number of different chain topologies known at present for each class of barrel*			
4	8	1	1
5	8	4	3
5	10	4	2
6	8	T+1	2
6	10	A+6	4
6	12	11	1
7	8	1	1
7	10	1	1
8	8	>23	2
8	10	3	2
8	12	4	2
16	20	1	1

T: trypsin family of proteases.
A: acid protease family.

(b) *Chain topologies in each* n, S *class*

In Table 6B we list the number of different topologies known at present for the structures in each n, S class. The number of observed topologies is only a small fraction of the possible number. All structures in the 6, 12 class have the topology of the β-trefoil fold (Murzin *et al.*, 1992). Of the 23 or so structures in the 8, 8 class, all but one structure has the same α/β barrel fold with an 8-fold alternation of strands and helices: $(\beta\alpha)_8$. The exception is enolase which begins with two strands and then two helices: $\beta_2\alpha_2(\beta\alpha)_6$ (Lebioda *et al.*, 1989). In the 5, 10 class all members have the topology of the OB fold (Murzin, 1993) except for the RHV barrel formed around a 5-fold axis in rhinovirus by the five N-terminal peptides.

The tendency of certain topologies to cluster in particular n, S classes probably arises from biological reasons; for example, very distant evolutionary relationships or the natural selection of topologies that have functional advantages.

11. Barrels Formed by Partly Open β-Sheets

As mentioned in the Introduction, certain barrel structures are not fully closed because they have one strand which has its N-terminal region hydrogen bonded to one neighbour, its C-terminal region to the other neighbour and these two regions are separated by a linking segment of two or more residues. Proteins that contain such structures are listed in Table 1B.

Though the linking segment in the strand means that such barrels are not quite as constrained as those formed by completely closed β-sheets, their structures are described to a good approximation by the same principles.

(a) *The conformations of the linking segments*

The linking segment in LTT consists of a two-residue bulge. In ADH1, SHG and PRC the link contains four residues that form one turn of a 3_{10} helix. In APP2 the link contains 19 residues some of which form a hairpin in another β-sheet. These linking regions produce large sharp kinks, up to about 90°, between the two parts of the strand that form the barrel structure. The linking regions themselves make few contacts with the barrel interior.

(b) *The geometry of the partly open barrels*

To compare the partly open barrels with the fully closed we need to determine effective numbers for the strand count and the shear, n^* and S^*. The strand with the long link is counted as one strand in calculating the effective number of strands, n^*. The effective shear number, S^*, can be calculated in the normal manner if the linking region is simply deleted from the H-bond plan of the β-sheet. The values of n^*, S^* for the partly open structures are given in Table 7.

The observed values of the tilt angles, α, and their mean radii, R, were determined by the same procedures as for the ordinary barrels, and are given in Table 5. The tilt angles are the same as, or slightly larger than, the ideal values for fully closed barrels. The values for the radii are about 1 Å greater.

Table 7
Geometry of partly open barrels

Barrel class		Ideal values for fully closed barrels		Observed values for partly open barrels		
n^*	S^*	R(Å)	α(°)	Protein	R(Å)	α(°)
4	8	5·6	56	ADH1	6·5	58
4	8	5·6	56	SHG	6·4	56
5	8	5·8	50	PRC	6·3	50
5	10	6·7	56	LTT	7·5	60
6	10	7·6	51	APP2	8·0	55

The partly open barrels also have larger mean residue volumes for the residues in the barrel interior. Indeed, the linking region, increased radius and increased residue size are all interdependent features of partly open barrel structures.

APP2, the second domain of the acid proteases, has a structure closely related to the fully closed first domain, APP1, and there is good evidence that they resulted from gene duplication (Tang *et al.*, 1978). This clearly implies that partly open structures can evolve from fully closed, and *vice versa*, by the co-ordinated evolution of the size of their interior residues and linking regions.

12. Conclusion: A Limited Repertoire for the Barrel Structures Formed by β-Sheets

β-Sheet barrels can be classified by two integral parameters, n and S (McLachlan, 1979). In paper I we argued that the mean values of the main geometrical features of barrels can be accurately calculated from the n, S values and this means that barrels in the same class have very similar structures; their main differences being strand direction and connectivity and the extent of distortions from ideal geometries. We argued further that structures in which β-sheets have low energy conformations, and whose interiors are filled by β-sheet residues, are likely to belong to one of ten (n, S) classes. Good β-sheet geometries can also occur in certain barrels too large to be filled by the β-sheet residues.

In this paper we have described an analysis of all the observed barrel structures that have distinctly different amino acid sequences and for which atomic co-ordinates are available. The results reported here clearly show that the principles presented in paper I, and summarized here in Table 2, do accurately describe the main features of the β-sheet barrels. They all have geometries whose features can be calculated from their (n, S) values. Of the 39 structures treated here, 34 fall into one of the ten expected (n, S) classes. The other five have good β-sheet geometries and cavities at their centre. In

four of these the cavity has been shown to play a functional role.

The work for this paper was greatly helped by generous gifts of atomic co-ordinates prior to their publication and we thank Drs R. S. Hegde, U. Heinemann, W. Hendrickson, W. G. J. Hol, T. A. Jones, H. Ke, M. Kjeldgaard, D. Moras, J. Nyborg, R.J. Read, B. Rees, D. C. Rees, M. Saraste, G. E. Schulz, P. E. Stein, P. B. Sigler and T. K. Sixma.

We acknowledge the support of a long term EMBO fellowship (A.G.M.) and from the Kay Kendall Foundation (A.M.L).

References

Arnold, E. & Rossmann, M. G. (1990). Analysis of the structure of a common cold virus, human rhinovirus 14, refined at a resolution of 3 Å. *J. Mol. Biol.* **211**, 763–801.

Arutiunian, E. G., Terzian, S. S., Voronova, A. A., Kuranova, I. P., Smirnova, E. A., Vainstein, B. K., Hohne, W. E. & Hansen, G. (1981). X-ray diffraction study of inorganic pyrophosphatase from baker's yeast at the 3 Å resolution (in Russian). *Dokl. Akad. Nauk SSSR*, **258**, 1481–1484.

Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D. & Waley, S. G. (1975). Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2·5 Å resolution using amino acid sequence data *Nature (London)*, **255**, 609–614.

Bocskei, Z., Groom, C. R., Flower, D. R., Wright, C. E., Phillips, S. E. V., Cavaggioni, A., Findlay, J. B. & North, A. C. T. (1992). Pheromone binding to two rodent urinary proteins revealed by X-ray crystallography. *Nature (London)*, **360**, 186–188.

Cavarelli, J., Rees, B., Ruff, M., Thierry, J. C. & Moras, D. (1993). Yeast $tRNA^{Asp}$ recognition by its cognate class II aminoacyl-tRNA synthetase. *Nature (London)*, **361**, 181–184.

Chothia, C. & Janin, J. (1981). Relative orientation of close packed β-sheets in proteins. *Proc. Nat. Acad. Sci., U.S.A.* **74**, 4130–4134.

Cohen, G. H., Silverton, E. W. & Davies, D. R. (1981). Refined crystal structure of γ-chymotrypsin at 1·9 Å resolution. *J. Mol. Biol.* **148**, 449–479.

Cook, W. J., Jeffrey, L. C., Carson, M., Chen, Z. & Pickart, C. M. (1992). Structure of a diubiquitin conjugate and a model for interaction with ubiquitin conjugating enzyme (E2). *J. Biol. Chem.* **267**, 16467–16471.

Correl, C. C., Batie, C. J., Ballou, D. P. & Ludwig, M. L. (1992). Phthalate dioxygenase reductase: a modular structure for electron transfer from pyridine nucleotides to [2Fe-2S]. *Science*, **258**, 1604–1610.

Cowan, S. W., Newcomer, M. E. & Jones, T. A. (1990). Crystallographic refinement of human serum retinol binding protein at 2 Å resolution. *Proteins*, **8**, 44–61.

Deisenhofer, J., Epp, O., Miki, K., Huber, H. & Michel, H. (1985). Structure of the protein subunits in the photosynthetic reaction centre of *Rhodopseudomonas viridis* at 3 Å resolution. *Nature (London)*, **318**, 618–624.

Ealick, S. E., Babu, Y. S., Bugg, C. E., Erion, M. D., Guida, W. C., Montgomery, J. A. & Secrist, J. A., III (1991). Application of crystallographic and modeling methods in the design of purine nucleoside phosphorylase inhibitors. *Proc. Nat. Acad. Sci., U.S.A.* **88**, 11540–11544.

Eklund, H., Nordstrom, B., Zeppezauer, E., Soderlund, G., Ohlsson, I., Boiwe, T., Sonderberg, B.-O., Tapia, O. & Brändén, C.-I. (1976). Three-dimensional structure of horse liver alcohol dehydrogenase at 2·4 Å resolution. *J. Mol. Biol.* **102**, 27–59.

Finzel, B. C., Clancy, L. L., Holland, D. R., Muchmore, S. W., Watenpaugh, K. D. & Einspahr, H. M. (1989). The crystal structure of recombinant human interleukin-1β at 2·0 Å resolution. *J. Mol. Biol.* **209**, 779–791.

Fita, I. & Rossmann, M. G. (1985). The NADPH binding site on beef liver catalase. *Proc. Nat. Acad. Sci., U.S.A.* **82**, 1604–1608.

Flower, D. R. (1993). Srructural relationship of streptavidin to the calycin protein superfamily. *FEBS Letters*, **333**, 99–102.

Fujinaga, M., Delbaere, L. T. J., Brayer, G. D. & James, M. N. G. (1985). Refined structure of α-lytic protease at 1·7 Å resolution: analysis of hydrogen bonding and solvent structure. *J. Mol. Biol.* **184**, 479–502.

Hegde, R. S., Grossman, S. R., Laimins, L. A. & Sigler, P. B. (1992). Crystal structure at 1·7 Å of bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature (London)*, **359**, 505–512.

Hendrickson W. A., Pahler, A., Smith, J. L., Satow, Y., Merritt, E. A. & Phizackerley, R. P. (1989). Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation. *Proc. Nat. Acad. Sci., U.S.A.* **86**, 2190–2194.

Huber, R., Schneider, M., Mayr, I., Mueller, R., Deutzmann, R., Suter, F., Zuber, H., Falk, H. & Kayser, H. (1987). Molecular structure of the bilin binding protein (BBP) from *Pieris brassicae* after refinement at 2·0 Å resolution. *J. Mol. Biol.* **198**, 499–513.

James, M. N. G. & Sielecki, A. R. (1983). Structure and refinement of penicillopepsin at 1·8 Å resolution. *J. Mol. Biol.* **163**, 299–361.

Janin, J. & Chothia, C. (1980). Packing of α-helices onto β-pleated sheets and the anatomy of α/β proteins. *J. Mol. Biol.* **143**, 95–128.

Karplus, P. A., Daniels, M. J. & Herriott, J. R. (1991). Atomic structure of ferredoxin-$NADP^+$ reductase: prototype for a structurally novel flavoenzyme family. *Science*, **251**, 60–66.

Ke, H. (1992). Similarities and differences between human cyclophilin A and other β-barrel structures. *J. Mol. Biol.* **228**, 539–550.

Ke. H., Zydowsky, L. D., Liu, J. & Walsh, C. T. (1991). Crystal structure of recombinant human T cell cyclophilin A at 2·5 Å resolution. *Proc. Nat. Acad. Sci., U.S.A.* **88**, 9483–9487.

Kjeldgaard, M. & Nyborg, J. (1992). Refined structure of elongation factor EF-Tu from *Escherichia coli*. *J. Mol. Biol.* **223**, 721–742.

Lebioda, L., Stec, B. & Brewer, J. M. (1989). The structure of yeast enolase at 2·25 Å resolution. An eightfold β/α barrel with a novel $\beta\beta\alpha\alpha(\beta\alpha)_6$ topology. *J. Biol. Chem.* **264**, 3685–3693.

Lesk, A., Brändén, C. -I. & Chothia, C. (1989). Structural principles of α/β barrel proteins: the packing of the interior of the sheet. *Proteins*, **5**, 139–148.

Lindqvist, Y. (1989). Refined structure of spinach glycolate oxidase at 2 Å resolution. *J. Mol. Biol.* **209**, 151–166.

Loll, P. J. & Lattman, E. E. (1989). The crystal structure of the ternary complex of staphylococcal nuclease, Ca^{2+}, and the inhibitor pdTp, refined at 1·65 Å. *Proteins*, **5**, 183–201.

McLachlan, A. D. (1979). Gene duplications in the structural evolution of chymotrypsin. *J. Mol. Biol.* **128**, 49–79.

Murzin, A. G. (1993). OB(oligonucleotide/oligosaccharide binding)-fold. *EMBO J.* **12** 861–867.

Murzin, A. G., Lesk, A. M. & Chothia, C. (1992). β-trefoil fold. *J. Mol. Biol.* **223**, 531–543.

Murzin, A. G., Lesk, A. M. & Chothia, C. (1994). Principles determining the structure of β-barrels in proteins. I A theoretical analysis. *J. Mol. Biol.* **236**, 1369–1381.

Musacchio, A., Noble, M., Pauptit, R., Wierenga, R. & Saraste, M. (1992). Crystal structure of a Src-homology 3 (SH-3) domain. *Nature (London)*, **359**, pp. 851–855.

Musil, D., Bode, W., Huber, R., Laskowski, M., Jr, Lin, T.-Y. & Ardelt, W. (1991). Refined X-ray crystal structures of the reactive site modified ovomucoid inhibitor third domains from silver pheasant (OMSVP3*) and from Japanese quail (OMJPQ3*). *J. Mol. Biol.* **220**, 739–755.

Onesti, S., Brick, P. & Blow, D. M. (1991). Crystal structure of a Kunitz-type trypsin inhibitor from *Erythrina caffra* seeds. *J. Mol. Biol.* **217**, 153–176.

Richardson, J. S., Getzoff, E. D. & Richardson, D. C. (1978). The β-bulge: a common small unit of nonrepetitive protein structure. *Proc. Nat. Acad. Sci., U.S.A.* **75**, 2574–2578.

Robbins, A. H. & Stout, C. D. (1989). The structure of aconitase. *Proteins*, **5**, 289–312.

Rould, M. A., Perona, J. J., Soell, D. & Steitz, T. A. (1989). Structure of *E. coli* glutaminyl-tRNA synthetase complexed with $tRNA^{Gln}$ and ATP at 2·8 Å resolution. *Science*, **246**, 1135–1142.

Rouvinen, J, Bergfors, T., Teeri, T., Knowles, J. K. C. & Jones, T. A. (1990). Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science*, **249**, 380–386.

Schindelin, H., Marahiel, M. A. & Heinmann, U. (1993). Universal nucleic acid-binding domain revealed by crystal structure of the *B. subtilis* major cold-shock protein. *Nature (London)*, **364**, 164–168.

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.

Sixma, T. K., Pronk, S. E., Kalk, K. H., van Zanten, B. A. M., Berghuis, A. M. & Hol, W. G. J. (1992). Lactose binding to heat-labile enterotoxin revealed by X-ray crystallography. *Nature (London)*, **355**, 561–564.

Stein, P. E., Boodhoo, A., Tyrrell, G. J., Brunton, J. L. & Read, R. J. (1992). Crystal structure of the cell-binding B oligomer of verotoxin-1 from *E. coli*. *Nature (London)*, **355**, 748–750.

Tang, J., James, M. N. G., Hsu, J. N., Jenkins, J. A. & Blundell, T. L. (1978). Structural evidence for gene duplication in the evolution of the acid proteases. *Nature (London)*, **271**, 618–621.

Tong, L., Wengler, G. & Rossmann, M. G. (1993). The refined structure of sindbis virus core protein in comparison with other chymotrypsin-like serine proteinase structures. *J. Mol. Biol.* **230**, 228–247.

Weiss, M. S. & Schulz, G. E. (1992). Structure of porin refined at 1·8 Å resolution. *J. Mol. Biol.* **227**, 493–509.

Wilmanns, M., Hyde, C. C., Davies, D. R., Kirschner, K. & Jansonius, J. N. (1991). Structural conservation in parallel β/α barrel enzymes that catalyse three sequential reactions in the pathway of tryptophan biosynthesis. *Biochemistry*, **30**, 9161–9169.

Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J. & Kent, S. B. H. (1989). Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science*, **245**, 616–621.

Zanotti, G., Ottonello, S., Berni, R. & Monaco, H. L. (1993). Crystal structure of the trigonal form of human plasma retinol-binding protein at 2·5 Å resolution. *J. Mol. Biol.* **230**, 613–624.

Zhu, X., Komiya, H., Chirino, A., Faham, S., Fox, G. M., Arakawa, T., Hsu, B. T. & Rees, D. C. (1991). Three-dimensional structure of acidic and basic fibroblast growth factors. *Science*, **251**, 90–93.

Edited by F. Cohen

(Received 1 September 1993; accepted 8 December 1993)