J. Mol. Biol. (1998) **275**, 541–545

JMB

COMMUNICATION

Shear Numbers of Protein β-Barrels: Definition Refinements and Statistics

Wei-min Liu

Department of Computer and Information Science, Indiana University Purdue University Indianapolis, 723 W. Michigan Street, Indianapolis IN 46202-5132, USA

The original definition of shear number for a β-barrel is not unique if it contains one or more uneven β-bulges. We define the shear number of a β-barrel as the minimal change of residue numbers in the backbone direction for all closed paths on the β-barrel. We also discuss how to overcome some computational difficulties. It is pointed out that some closed β-sheets should not be considered as β-barrels. The pertinent statistics obtained from a representative list of the Protein Data Bank entries are summarized. All β-barrels have positive shear numbers, i.e. they are right-twisted. The shear numbers of most β-barrels are even, but exceptions do exist. The sizes of β-ladders in a β-barrel vary significantly. Most β-barrels contain uneven β-bulges, which may have important biological functions.

Keywords: β-barrel; β-bulge; closed β-sheet; shear number

Protein β-barrels play important functional roles. They can contain active sites of enzymes (Reardon & Farber 1995; Eads *et al*., 1997), bind DNA (Lodi *et al*., 1995) and RNA (Bycroft *et al*., 1997), provide membrane transportation channels (Schirmer, 1995; Montal, 1996) and contribute to electron tunneling (Langen *et al*., 1995). The structure of a β-barrel is characterized by the barrel size (the number of β-strands in the barrel), and its shear number (McLachlan, 1979; Chou *et al*., 1990; Murzin *et al*., 1994a,b).

Refinements of definition

The shear number of a β-barrel is defined by McLachlan (1979) as the change of residue numbers on a β-strand when a point moves in the left hydrogen bond direction back to the same β-strand. Murzin *et al*. (1994a,b) use the absolute value of this number. The benefit of the latter definition is that it is independent of the moving direction and can make the computation simpler. Because β-barrels are right-twisted (Chothia, 1973; Chou *et al*., 1983), these two definitions are consistent. However, we do not use the absolute value in definition, and let statistics show that all β-barrels have positive shear numbers.

Figure 1 is an outside view of six strands in a regular β-barrel cut along a strand, which is drawn twice. We can define either direction of the barrel axis as positive (the barrel axis will be discussed later). Murzin *et al*. (1994a,b) call the hydrogen bond direction the H-direction, and the main chain direction the C-direction. We define the positive C-direction as the C-direction that forms an acute angle with the positive direction of the barrel axis, and the positive H-direction as the right-handed direction with respect to the positive direction of the barrel axis. If a point moves left along the hydrogen bonds (i.e. in the negative H-direction) from the residue A on the strand D_1 back to the same strand at the residue B, the increase of residue number (from residue A' to residue B in the C-direction) is eight, so the shear number of this barrel is eight. To count the increase of residue numbers automatically, we introduce the sign, s_j, of the strand D_j. The sign s_j is 1 if the positive C-direction is from the N terminus to the C terminus on the strand D_j, and is -1 otherwise. The increase of residue number is $s_1(n_B - n_A)$, where n_A and n_B are, respectively, the residue numbers of A and B in the Protein Data Bank (PDB) file. Because residues A and B are in the same β-strand, there is no chain break between them. The benefit of this counting method is that the result is independent of the choice of positive direction of the barrel axis, which is especially important for automatic computation.

The path AB makes a full turn from the strand D_1 back to itself, but this path is not fully closed.

0022–2836/98/040541–05 $25.00/0/mb971501

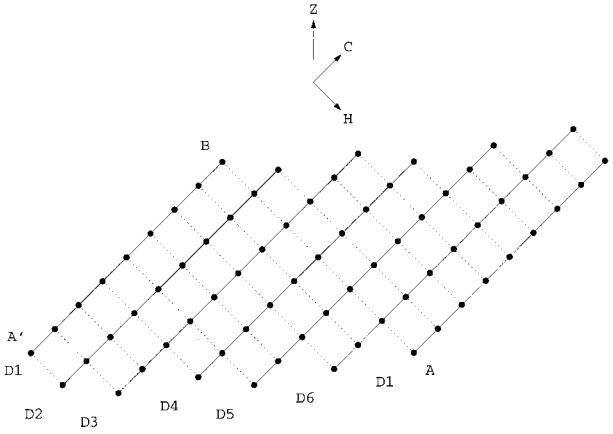

Figure 1. An outside view of a β-barrel. One can choose either direction of the barrel axis as positive and it is denoted by the vector *Z*. The positive C-direction, denoted by the vector *C*, has an acute angle with the vector *Z*. The positive H-direction, denoted by the vector *H*, is right-hand coupled with the positive direction of the barrel axis. The shear number of a β-barrel is the difference between residue numbers *B* and *A*′ multiplied by the sign s_1 of the strand D_1, when one makes a full turn travel *AB* in the negative H-direction from strand D_1 back to itself. *A*′ and *A* represent the same residue on strand D_1. The shear number is also the change of residue numbers in the C-direction on the closed path *A*′*BA*, which is in the positive H-direction. This is a hypothetic β-barrel. Real β-barrels are not so regular.

Instead, we can use a closed path *A*′*BA* moving in the positive H-direction, and up or down in the C-direction on the barrel. The shear number can be defined as the change of residue numbers in the C-direction of the closed path. With a fully closed path, we eliminate the need for a starting strand in the definition. This number does not depend on the path, as long as the β-barrel does not contain any uneven β-bulge. Richardson *et al.* (1978) define a β-bulge as a region between two consecutive β-type hydrogen bonds that includes two residues on one strand opposite a single residue on the other strand. This definition is extended to allow a regular β-sheet to be disrupted by at most one extra residue on one strand and at most four extra residues on the other strand (Kabsch & Sander, 1983; Chan *et al.*, 1993). If the numbers of residues in two strands of a β-bulge are different, we call it an uneven β-bulge. We denote a β-bulge with the symbol *a*:*b*, where *a* and *b* are the numbers of extra residues.

A problem in the definition of shear number is that when there are one or more uneven β-bulges in a β-barrel, the change of residue numbers in the C-direction of a closed path is not unique. In other words, it is path-dependent. We call these different numbers the shear numbers of the corresponding closed paths, and use their minimum to define the shear number of a β-barrel. In the meantime, it also makes sense to take the maximal possible deviation into account. Mathematically speaking, the shear number of a β-barrel is:

$$S = S_{\min} = \min(S_1, S_2, \ldots, S_p) \quad (1)$$

where S_1, S_2, . . . , S_p are shear numbers of all closed paths on the β-barrel. Let:

$$S_{\max} = \max(S_1, S_2, \ldots, S_p) \quad (2)$$

The maximal deviation of shear number is defined as:

$$\Delta S = S_{\max} - S_{\min} \quad (3)$$

which is always non-negative. When there is no uneven β-bulge, $\Delta S = 0$.

On many real β-barrels, one cannot make a full turn travel without jumps on intermediate strands. Therefore, we need to count the change of residue numbers on a strand for a closed path. Consider a β-barrel with *n* strands D_1, D_2, . . . , D_n. Assume that they are in the positive H-direction. For convenience, the pseudobond between two C^α atoms in a regular β-structure is called a β-bridge, and repeating β-bridges between two β-strands form a β-ladder (Kabsch & Sander, 1983). We call the number of β-bridges in a β-ladder the ladder size. A closed path on an *n*-strand barrel can be denoted by a sequence of left and right residue numbers on β-bridges: (L_1, R_1; L_2, R_2; . . . ; L_n, R_n), where L_j and R_j are, respectively, the left and right residue numbers of a β-bridge between the strands D_j and D_{j+1}. The strand D_{n+1} is defined to be the strand D_1. The change, S_i, of residue numbers in the C-direction of closed path *i* is the algebraic sum of the changes on all its strands:

$$S_i = \sum_{j=1}^{n} S_{i,j} \quad (4)$$

where $S_{i,j} = s_j\ (L_j - R_{j-1})$ is the change of residue numbers on the strand D_j of closed path *i*; R_0 is equal to R_n; s_j is the sign of the strand D_j.

However, for some data in the PDB, the directions of β-strands vary significantly, and it is difficult to assign signs to all strands so that they are consistent with the parallelity or antiparallelity of β-strands. Therefore, we can choose the strand having the smallest angle with the barrel axis as the first strand, determine its sign s_{j+1}, and assign the signs of the remaining strands on the basis of parallelity or antiparallelity of β-strands: $s_{j+1} = s_j$, if the strands D_j and D_{j+1} are parallel; and $s_{j+1} = -s_j$, if they are antiparallel. With this method, we can assign strand signs consistently for all closed β-sheets with an even number of antiparallel β-ladders.

To determine the barrel axis, we can first calculate the centroid of C^α atoms in every β-ladder, then use the centroid of these β-ladder centroids as the centroid of the β-barrel. The barrel axis can be considered as a straight line passing through the β-barrel centroid and in the normal direction of a plane with the least-squares error for the β-ladder centroids. In practice, the average normal direction

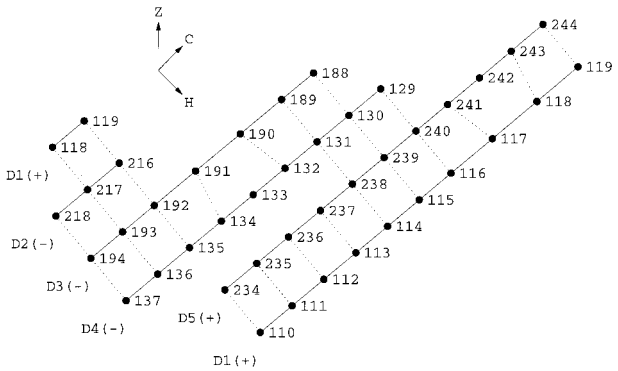

Figure 2. The β-barrel of a pentosyltransferase (PDB code 1PBN). Positive or negative signs are assigned to all strands for counting changes of residue numbers on them. Different closed paths may have different shear numbers and their minimum is defined as the shear number of the β-barrel.

determined by the barrel centroid and two adjacent ladder centroids works as well.

The β-barrel of a pentosyltransferase (PDB code 1PBN) is shown in Figure 2. Let closed path 1 be (118, 217; 216, 192; 194, 137; 130, 239; 244, 119). These numbers are the PDB residue numbers. The positive direction of the barrel axis is chosen upward, and the signs of the strands are $s_1 = s_5 = 1$, $s_2 = s_3 = s_4 = -1$, and the change of residue numbers in closed path 1 is:

$$S_1 = (118 - 119) - (216 - 217) - (194 - 192) - (130 - 137) + (244 - 239) = 10$$

Let closed path 2 be (118, 217; 216, 192; 188, 130; 130, 239; 240, 116):

$$S_2 = (118 - 116) - (216 - 217) - (188 - 192) - (130 - 130) + (240 - 239) = 8$$

which is the minimal shear number, S_{min}, for all closed paths and is defined to be the shear number of the β-barrel. Since $S_{max} = S_1$, and the maximal deviation of shear number $\Delta S = 10 - 8 = 2$.

To compute the shear number of a β-barrel, we do not have to compute the shear numbers of all possible closed paths (the total number of them is huge). We can take advantage of the fact that the shear number is invariant for the closed paths whose difference does not include any uneven β-bulge. Thus, we need only to calculate the shear number of one closed path and use the information about uneven bulges to obtain the shear number of the β-barrel.

Pertinent statistics

The statistical information of existing structures is important for structural classification, and critical for protein modeling and design. We use the DSSP program (Define Secondary Structure of Proteins, Kabsch & Sander, 1983) to determine the secondary structures of all protein entries in a representative list of 998 PDB chains selected in May 1997 by PDB-SELECT, with less that 35% sequence identity (Hobohm *et al.*, 1992; Hobohm & Sander, 1994). The total number of PDB entries is 928. We consider β-strands with lengths at least two; that is, their residues have the secondary structure summary E in the DSSP files. Both intrachain and interchain β-bridges are taken into account. A specially developed computer program is used to form a graph of the β-structure. The nodes in the graph represent β-strands and the edges represent the β-ladders between two β-strands. We remove the nodes and edges that are not in any cycle, as well as those connecting two cycles.

Not all closed β-sheets have the shape of β-barrels. If two adjacent β-ladders in a closed β-sheet do not share a common residue, and if there is at least an extra residue on the common strand between the two ladders, we do not count it in our statistics. The examples include the three-stranded structure in the β-trefoil (1TIE, 1WBA, Murzin *et al.*, 1992), and the 20-strand structure in the cholera toxin B pentamer (1LTS and 1TII). These closed β-sheets include an odd number of anti-parallel ladders and are discussed by Liu (1998) from the surface topologic viewpoint. Here, we remark that for these structures it is impossible to assign strand signs that are consistent with the parallelity and antiparallelity of ladders and to keep the invariance of shear number for closed paths whose difference does not include an uneven β-bulge. Figure 3 shows the topologic connection of three antiparallel β-ladders in a β-trefoil. Two ladders are separated by an extra residue in their common strand, which is hydrogen bonded to another strand not shown here. For any assignment of signs to the three strands, there must be at least two strands sharing the same sign. When two closed paths differ in the β-bridges between these two strands, the changes of residue numbers in the two paths will be different, even though there is no uneven β-bulge.

We group β-barrels according to their sizes (n) shear numbers (S) and deviations of shear number(ΔS). We count the frequency of occurrence, ladder sizes (maximal, minimal and mean values) and largest β-bulges. The largest β-bulge in a class is the bulge with the largest size difference in two strands; when this difference is the same, we take the one with maximal size. The statistics are summarized in Table 1.

From Table 1, we see that all β-barrels have positive shear numbers, i.e. they are right-twisted. Also note that the shear numbers for most β-barrels are even because of the zigzags of β-strands (Murzin *et al.*, 1994a,b). However, there exist exceptional cases whose shear numbers are odd.

We observe that ladder sizes vary significantly in a β-barrel. The average ladder sizes have some relationship to barrel sizes and shear numbers. For

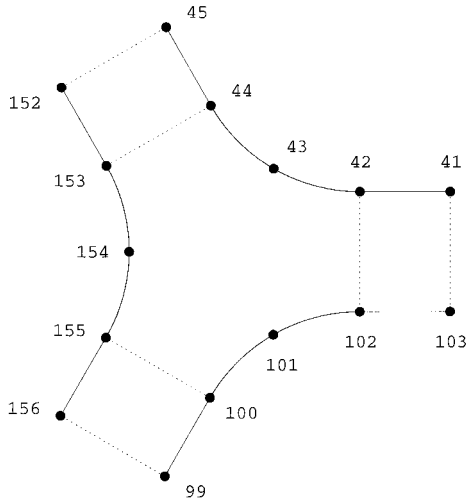

Figure 3. A structural motif in β-trefoil. The real structure is non-planar. Three antiparallel β-strands form three β-ladders. Two ladders are separated by an extra residue in their common β-strand. It is impossibe to assign signs to the β-strands so that the change of residue numbers remains constant when no uneven β-bulge is involved. The residue numbers are in the PDB file of a trypsin inhibitor (1TIE). The hydrogen bonds between residues 41 and 103, and between 99 and 156 are not perfect.

example, β-barrels of sizes larger than eight tend to have larger average ladder sizes; β-barrels of size eight have larger ladder sizes if the shear numbers are ten or more. We also notice that large uneven bulges 4:1 and 3:0 appear only in β-barrels of size five or six.

Possible impact

The shear number is an essential feature of a β-barrel. It has been used in modeling transmembrane β-barrels (Sansom & Kerr, 1995). From the statistics in Table 1, we can clearly see that most β-barrels include uneven β-bulges ($\Delta S > 0$). Therefore, we should take β-bulges into account in modeling β-barrels.

Richardson *et al.* (1978) propose that β-bulges can compensate for the effects of a single-residue insertion or deletion within a β-structure and provide the strong local twist required for forming β-barrels. Recent studies show that they can also play important biological roles.

Phosphoinositide-specific phospholipase C isozymes (PI-PLC) are involved in most signal transduction cascades and are crucial in the control of cellular responses, including cell growth, proliferation, contraction, excitation and secretion. Their catalytic domain contains a TIM-barrel (triosephosphate isomerase-like barrel). Residue Lys438 of the PDB entry 2ISD is in an uneven β-bulge. It forms a salt-bridge to the 4-phosphoryl group. The mutation of Lys438 leads to loss of phosphatidylinositol-4,5-bisphosphate hydrolysis (Simões *et al.*, 1995). Moreover, the catalytic residue His311 is on the tip of another β-buldge (Essen *et al.*, 1996).

Table 1. Statistics of β-barrel shear numbers

n	S	ΔS	Freq.	Ladder size Max	Ladder size Min	Ladder size Mean	Largest bulge	Examples (PDB codes)
4	7	1	2	11	1	4.25	1:0	1AOC
5	7	⩾2	3	10	2	6.00	3:0	1PRT 1WHI
5	8	⩾2	11	10	1	4.78	2:1	1ECP 1INO 1MJC 1PBN
5	10	1	1	6	3	4.40	1:0	1SNC
5	10	⩾2	15	11	1	5.72	3:0	1LYL 1PRT 1TSS
6	6	⩾2	1	11	2	4.83	4:1	1EAG
6	7	⩾2	6	12	2	5.45	4:1	1ECL 1EFU 2SGA 4SGB
6	8	1	5	7	2	4.00	1:0	1DSU 3RP2 1TON
6	8	⩾2	28	8	1	4.11	3:1	1BMF 1DSU 1HCG 1UCY
6	9	1	1	11	2	4.00	1:0	1SME
6	9	⩾2	4	12	1	4.42	2:1	2ER7 1LYB 1MPP
6	10	0	2	9	1	3.50	0:0	1EXS 1FMB
6	10	1	4	8	1	4.00	2:1	2ENG 1FIV 2HPE
6	12	0	4	7	1	4.04	0:0	1I1B 1ILR 1WBA
6	12	1	1	6	1	3.50	1:0	1TIE
7	10	⩾2	1	6	2	4.00	2:1	8ACN
8	6	⩾2	3	7	2	3.71	1:0	1CEO 1ECE
8	7	⩾2	5	5	2	3.72	2:1	1CBG 1DIK 1EDG 1PBG
8	8	0	12	6	1	3.41	0:0	1AMY 1BPL 1IGS 5TIM
8	8	1	18	7	1	3.54	2:1	2AMG 1BYB 1CTN 1SFT
8	8	⩾2	12	8	1	3.96	2:1	1CLX 2EBN 1FBA 1XYZ
8	9	1	2	5	2	3.31	2:1	1REQ
8	10	0	2	10	2	5.88	0:0	1AVD
8	10	⩾2	5	10	2	5.95	3:1	2CAE 1HBQ 1SMP 1SRI
8	11	⩾2	6	10	2	6.23	2:1	1BBP 1OBP
8	12	⩾2	2	10	2	6.31	1:0	1EPA
11	14	0	1	11	5	8.36	0:0	1EMA
16	20	1	2	13	5	8.72	1:0	1PRN 2POR
18	19	⩾2	3	16	6	10.50	2:0	2MPR
20	20	⩾2	4	10	4	7.25	2:1	1GTP

The enzyme 2,3-dihydroxybiphenyl 1,2-dioxygenase plays an important role in degrading aromatic toxin. It has an uneven β-bulge including His209 and His210 in the PDB entry 1HAN. These two residues together with His194 and His195 contribute to the environment of the active site Fe atom, and are conserved (Han *et al.*, 1995). The β-sheet of 1HAN is not fully closed according to the DSSP program, but is approximately closed.

In short, further investigations of shear numbers, uneven β-bulges, as well as distribution of ladder sizes of β-barrels in the perspective of their biological functions are helpful for protein analysis and design.

Acknowledgments

The author thanks the two anonymous referees for their critical comments and constructive suggestions. He is grateful to the NIH/NCI for the computer time provided by the NCI's Frederick Biomedical Supercomputing Center.

References

Bycroft, M., Hubbard, T. J., Proctor, M., Freund, S. M. & Murzin, A. G. (1997). The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold. *Cell,* **88**, 235–242.

Chan, A. W. E., Hutchinson, E. G., Harris, D. & Thornton, J. M. (1993). Identification, classification, and analysis of β-bulges in proteins. *Protein Sci.* **2**, 1574–1590.

Chothia, C. (1973). Conformation of twisted β-sheets in proteins. *J. Mol. Biol.* **75**, 295–302.

Chou, K. C., Nemethy, C. & Scheraga, H. A. (1983). Role of interchain interactions in the stabilization of the right-handed twist of β-sheets. *J. Mol. Biol.* **168**, 389–407.

Chou, K. C., Carlacci, L. & Maggiora, G. G. (1990). Conformational and geometrical properties of idealized β-barrels in proteins. *J. Mol. Biol.* **213**, 315–326.

Eads, J. C., Ozturk, D., Wexler, T. B., Grubmeyer, C. & Sacchettini, J. C. (1997). A new function for a common fold: the crystal structure of quinolinic acid phosphoribosyltransferase. *Structure,* **5**, 47–58.

Essen, L.-O., Perisic, O., Cheung, R., Katan, M. & Williams, R. L. (1996). Crystal structure of a mammalian phosphoinositide-specific phospholipase Cδ. *Nature,* **380**, 595–602.

Han, S., Eltis, L. D., Timmis, K. N., Muchmore, S. W. & Bolin, J. T. (1995). Crystal structure of the biphenyl-cleaving extradiol dioxygenase from a PDB-degrading pseudomonad. *Science,* **270**, 976–980.

Hobohm, U. & Sander, C. (1994). Enlarged representative set of protein structures. *Protein Sci.* **3**, 522–524.

Hobohm, U., Scharf, M., Schneider, R. & Sander, C. (1992). Selection of a representative set of structures from the Brookhaven Protein Data Bank. *Protein Sci.* **1**, 409–417.

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

Langen, R., Chang, I. J., Germanas, J. P., Richards, J. H., Winkler, J. R. & Gray, H. B. (1995). Electron tunneling in proteins: coupling through a β strand. *Science,* **268**, 1733–1735.

Liu, W.-M. (1998). Is there a Möbius band in closed protein β sheets? *Protein Eng.* In the press.

Lodi, P. J., Ernst, J. A., Kuszewski, J., Hickman, A. B., Engleman, A., Craigie, R., Clore, G. M. & Gronenborn, A. M. (1995). Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry,* **34**, 9826–9833.

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

Montal, M. (1996). Protein folds in channel structure. *Curr. Opin. Struct. Biol.* **6**, 499–510.

Murzin, A. G., Lesk, A. M. & Chothia, C. (1992). β-Trefoil fold: patterns of structure and sequence in the Kunitz inhibitors interleukins-1β and 1α and fibroblast growth factors. *J. Mol. Biol.* **223**, 531–543.

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

Murzin, A. G., Lesk, A. M. & Chothia, C. (1994b). Principles determining the structure of β-sheet barrels in proteins, II. *J. Mol. Biol.* **236**, 1382–1400.

Reardon, D. & Farber, G. K. (1995). The structure and evolution of alpha/beta barrel proteins. *FASEB J.* **9**, 497–503.

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

Sansom, M. S. & Kerr, I. D. (1995). Transbilayer pores formed by β-barrels: molecular modeling of pore structures and properties. *Biophys. J.* **69**, 1334–1343.

Schirmer, T., Keller, T. A., Wang, Y. F. & Rosenbusch, J. P. (1995). Structural basis for sugar translocation through maltoporin channels at 3.1 Å resolution. *Science,* **267**, 512–514.

Simões, A. P., Camps, M., Schnabel, P. & Gierschik, P. (1995). Mutational analysis of a putative polyphosphoinositide binding site in phospholipase C-β2. *FEBS Letters,* **365**, 155–158.

Edited by J. M. Thornton

(Received 24 July 1997; received in revised form 15 October 1997; accepted 27 October 1997)