Characterizing the topology of protein beta-sheets by an axis.

Jean-Luc Jestin, Bernard Caudron

Characterizing the topology of protein beta-sheets by an axis.

Jean-Luc Jestin[1*], Bernard Caudron[2]

[1]Département de Virologie, Institut Pasteur ; E-mail : jjestin@pasteur.fr

[2]Centre d'Informatique pour la Biologie, Institut Pasteur ; E-mail : bernard.caudron@pasteur.fr

*Author to whom correspondence should be addressed ;

Tel : +33 1 4438 9496 ; Fax : +33 1 4568 8993

Abstract :

Beta-sheets and alpha-helices are the most frequent structural elements composing protein structures. Beta-sheets typically highlight complex curved and sequence-dependent surfaces of parallel and/or anti-parallel beta-strands. The topology of these curved surfaces are described in this work by a unique axis crossing all beta-sheet's strands. The distribution of the distances of the alpha carbons in each strand which are the closest to the axis is given. The frequency of the twenty amino acids along the axis is provided. Applications in the field of protein structure prediction are mentioned.

Keywords :

Protein structure ; polypeptide chain ; beta-strand ; secondary structure

1. Introduction

Protein beta-sheets composed of parallel or anti-parallel beta-strands are commonly represented as quasi-planar bidimensional surfaces (Pauling and Corey, 1951). Beta-strands are frequently six to eight amino acids long (Sternberg and Thornton, 1977a) and are juxtaposed, the side-chains being above and below the beta-sheet's surface, while the hydrogen bond network maintaining the common structure of the sheet is established between main chain chemical groups of the different alpha amino acids (Chothia, 1973) (Salemme, 1983) (Salemme and Weatherford, 1981b, a) (Rose et al., 2006). Two beta-strands occur in parallel or in anti-parallel depending on their orientation given by the position of the beta-strands' N- and C-termini. Parallel and anti-parallel strands differ by the mean values for the dihedral angles Phi and Psi which are respectively about -115° and +115° for parallel strands and about -145° and +145° for anti-parallel strands (Nesloney and Kelly, 1996). Parallel and anti-parallel strands differ also by their amino acid composition (Lifson and Sander, 1979) (Otaki et al., 2010). A rule predicting the anti-parallel character from the beta-strands' sequence was derived (Caudron and Jestin, 2012). Important rules defining the arrangement of strands within sheets were identified (Sternberg and Thornton, 1977a, c, b) (Richardson and Richardson, 2002). At a higher structural level, the packing of sheets and helices was analyzed (Chothia et al., 1977). Similarly, further rules providing three-dimensional structure information on beta-strands or on a beta-strand and a helix were recently identified depending on the length of the loops connecting the two secondary structural elements (Koga et al., 2012). Anti-parallel beta-strands show a great diversity of structures while parallel beta-strands are associated to less flexibility (Salemme, 1983).

Beta-sheets are well predicted from protein sequences (Rost and Sander, 1993) (Steward and Thornton, 2002) (Cheng and Baldi, 2005) (Zimmermann et al., 2007) (Zafer et al., 2011) (Subramani and Floudas, 2012). Cross-strand pairs of interacting amino acids were studied for beta-sheets (Von Heijne and Blomberg, 1977) (Lifson and Sander, 1979) (Wouters and Curmi, 1995) (Merkel et al., 1999) (Mandel-Gutfreund et al., 2001) (Zhang et al., 2010).

Observation of protein structures from the Protein Data Bank (PDB) (Abola et al., 1997) (Rose et al.) clearly shows that almost all protein beta-pleated sheets are far from planar surfaces, but instead are best described in three dimensions as curved bidimensional surfaces.

This observation was reviewed decades ago for beta-sheets composed of parallel as well as anti-parallel beta-strands (Salemme, 1983) (Koh et al., 2006). This curvature of beta-sheets is linked to the twist of beta-strands (Chou et al., 1982) which was characterized in particular by the angles Phi and Psi (Shamovsky et al., 2000).

Here, the curved bidimensional surface of protein beta-sheets is characterized by a sheet axis defined below.

2. Methods

The program Pdb22 available at the www address (http://mobyle.pasteur.fr/cgi-bin/portal.py#forms::pdb22) is a program written in perl. It uses as entry files, lists of PDB files (pdbxxxx.ent) corresponding to a protein or a protein domain structure which may be bound to other molecules and which is described as the first macromolecular chain. The program eliminates files associated to polypeptides of less than fifty amino acids from the lists as well as files containing overlapping strands, in which at least one amino acid is found in two different strands reported in the Sheet key of the PDB files. The list of non-redundant protein structures and their PDB files was set up using the program check.pl to remove highly homologous sequences such as those of antibodies. The Pdb22 output file (.xls) provides for each protein within the list its PDB name, the amino acid number and name in three-letter code, the start and the end of beta-strands indicated as amino acid numbers, the name of the sheet noted on the lines corresponding to amino acids found at the intersection of a beta-strand with the sheet axis and the distance D, which is calculated in Angströms and averaged per beta-strand for each sheet consisting of n strands using the following equation :

$$d = \frac{\sum_{i=2}^{n-1} \text{mindist}(i)}{n-2}$$

where mindist(i) is the minimal distance between an alpha carbon of strand i and the sheet axis. The distance d is estimated for each pair of amino acids defining an axis characterized by the atomic coordinates of one amino acid's alpha carbon in the first strand and another one in the sequence's last strand. The sheet axis is defined as the axis for which the distance d is minimal. For a sheet, the minimum of all distances d is noted D.

The probabilities to find an amino acid on a sheet axis were derived from the analysis of three lists of about 800 sheets. The errors on the probabilities given in the tables were calculated from the three distinct lists.

3. Results

To characterize the curved surface of beta-pleated sheets, we define here a sheet axis which is a straight line including at least one amino acid alpha carbon from each beta-strand of the sheet, and which is generally perpendicular to the sheets' beta-strands (Figure 1). The axis was chosen to cross the first and last strands at amino acids' alpha carbons. Figure 1 highlights for the two sheets A and B of a pyruvate phosphate dikinase domain structure the axes represented as circles including the set of amino acids which are on the axis or closest to the axis.

The mean distance of the closest alpha carbons to the sheets' axis was calculated from the atomic coordinates of protein structures reported in the Protein Databank (PDB) (Abola et al., 1997) (Rose et al.) (Laskowski, 2011). For a set of 1347 non-redundant domain structures, 2555 beta-sheets of three or more strands were analyzed to establish the distribution of the mean distances per strand to the sheet axis (Figure 2). Noticeably, more than 60% of the sheets have an average distance per strand of less than 1.5 Angströms.

The nature of the amino acids found at the intersection between the sheet axis and the beta-strands were further studied (Table 1). At these intersections, tryptophan (W) was found to be enriched by up to 37%, while asparagine (N) was counter-selected at those positions by 33%. Five out of six amino acids enriched at those intersections by 10% or more were hydrophobic amino acids: tryptophan, isoleucine, methionine, leucine and phenylalanine. The two most counterselected amino acids at those intersections were hydrophilic amino acids : asparagine, aspartic acid.

4. Discussion

Together with alpha-helices, beta-pleated sheets represent the most common structural components of proteins. The identification of an axis which minimizes the distance to alpha carbons of beta-strands provides a means to characterize the sheet's curved surface. If beta-sheets were planar, numerous sheet axes could be defined. The sheet's surface can be characterized by one axis crossing all beta-strands of the sheet and which is defined by a set of one alpha carbon per strand located on the axis or close to the axis (Figure 1).

Special cases of beta-sheets are beta-barrels, which are composed of beta-strands with no edge strand. The notion of an axis consisting of a straight line seems at first incompatible with the notion of barrel. For an SH3-type barrel of four beta-strands (NusG structure 2jvv.pdb) (Burmann et al., 2012), the length of the beta-strands is however such that a distance to the sheet axis (1.53 Angströms) could be computed. For a larger eight-stranded barrel (TRAP structure 4ae5.pdb), an average distance to the sheet axis of 4.50 Angströms per strand is significant (Henrick and Hirshberg, 2012).

Hydrophobic amino acids are located more frequently in the interior of beta-sheets (Sternberg and Thornton, 1977b) (Von Heijne and Blomberg, 1978). In this work it was observed that amino acids enriched significantly at the intersection of beta-strands and the corresponding sheet axis are hydrophobic amino acids. The enrichment or the counter-selection of specific amino acids at these intersections was found to be significant and larger than 30% : this property might be used in protein structure prediction approaches for the assessment of structure models' quality.

It is of interest to define an axis crossing all beta-strands within a beta-sheet: this axis was characterized by a distance found to be less than 2 Angströms per strand in average for three-quarters of the beta-sheets. By providing a new tool for the quality assessment of protein domain structure models, this simple computational approach may be of use in the field of protein structure prediction which is of increasing importance in structural genomics projects (Rost et al., 2002) (Fischer et al., 2003) (Moult et al., 2009) (Moult et al., 2011).

5. Acknowledgements

The authors thank M. Delarue for discussions and B. Néron for interfacing the program.

6. References

Abola, E.E., Sussman, J.L., Prilusky, J., & Manning, N.O., 1997. Protein Data Bank archives of three-dimensional macromolecular structures. Methods Enzymol. 277 556-71.

Burmann, B.M., Knauer, S.H., Sevostyanova, A., Schweimer, K., Mooney, R.A., Landick, R., Artsimovitch, I., & Roesch, P., 2012. An alpha helix to beta barrel domain switch transforms the transcription factor RfaH into a translation factor. Cell 150 (2), 291-303.

Caudron, B., & Jestin, J.L., 2012. Sequence criteria for the anti-parallel character of protein beta-strands. J. Theor. Biol. 315 146-149.

Cheng, J., & Baldi, P., 2005. Three-stage prediction of protein beta-sheets by neural networks, alignments and graph algorithms. Bioinformatics 21 i75-84.

Chothia, C., 1973. Conformation of twisted beta-pleated sheets in proteins. J. Mol. Biol. 75 (2), 295-302.

Chothia, C., Levitt, M., & Richardson, D., 1977. Structure of proteins: packing of alpha-helices and pleated sheets. Proc. Natl. Acad. Sci. USA 74 (10), 4130-4134.

Chou, K.C., Pottle, M., Nemethy, G., Ueda, Y., & Scheraga, H.A., 1982. Structure of beta-sheets- Origin of the right handed twist and of the increased stability of anti-parallel over parallel sheets. J. Mol. Biol. 162 (1), 89-112.

Fischer, D., Rychlewski, L., Dunbrack, R.L., Jr., Ortiz, A.R., & Elofsson, A., 2003. CAFASP3: the third critical assessment of fully automated structure prediction methods. Proteins 53 503-516.

Henrick, K., & Hirshberg, M., 2012. Structure of the signal transduction protein TRAP, target of RNAIII-activating protein. Acta Cryst. D Biol. Cryst. 68 744-750.

Koga, N., Tatsumi-Koga, R., Liu, G., Xiao, R., Acton, T.B., Montelione, G.T., & Baker, D., 2012. Principles for designing ideal protein structures. Nature 491 (7423), 222-227.

Koh, E., Kim, T., & Cho, H.S., 2006. Mean curvature as a major determinant of beta-sheet propensity. Bioinformatics 22 (3), 297-302.

Laskowski, R.A., 2011. Protein structure databases. Mol. Biotechnol. 48 (2), 183-198.

Lifson, S., & Sander, C., 1979. Antiparallel and parallel beta-strands differ in amino acid residue preferences. Nature 282 (5734), 109-111.

Lin, Y., Lusin, J.D., Ye, D., Dunaway-Mariano, D., & Ames, J.B., 2006. Examination of the structure, stability, and catalytic potential in the engineered phosphoryl carrier domain of pyruvate phosphate dikinase. Biochemistry 45 (6), 1702-1711.

Mandel-Gutfreund, Y., Zaremba, S.M., & Gregoret, L.M., 2001. Contributions of residue pairing to beta-sheet formation: conservation and covariation of amino acid residue pairs on antiparallel beta-strands. J. Mol. Biol. 305 (5), 1145-1159.

Merkel, J.S., Sturtevant, J.M., & Regan, L., 1999. Sidechain interactions in parallel beta sheets: the energetics of cross-strand pairings. Structure 7 (11), 1333-43.

Moult, J., Fidelis, K., Kryshtafovych, A., Rost, B., & Tramontano, A., 2009. Critical assessment of methods of protein structure prediction - Round VIII. Proteins 77 Suppl 9 1-4.

Moult, J., Fidelis, K., Kryshtafovych, A., & Tramontano, A., 2011. Critical assessment of methods of protein structure prediction (CASP)--round IX. Proteins 79 1-5.

Nesloney, C.L., & Kelly, J.W., 1996. Progress towards understanding beta-sheet structure. Bioorg. Med. Chem. 4 (6), 739-766.

Otaki, J.M., Tsutsumi, M., Gotoh, T., & Yamamoto, H., 2010. Secondary structure characterization based on amino acid composition and availability in proteins. J. Chem. Inf. Model. 50 (4), 690-700.

Pauling, L., & Corey, R.B., 1951. Configurations of Polypeptide Chains With Favored Orientations Around Single Bonds: Two New Pleated Sheets. Proc. Natl. Acad. Sci. USA 37 (11), 729-740.

Richardson, J.S., & Richardson, D.C., 2002. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. Proc. Natl. Acad. Sci. USA 99 (5), 2754-2759.

Rose, G.D., Fleming, P.J., Banavar, J.R., & Maritan, A., 2006. A backbone-based theory of protein folding. Proc. Natl. Acad. Sci. USA 103 (45), 16623-16633.

Rose, P.W., Beran, B., Bi, C.X., Bluhm, W.F., Dimitropoulos, D., Goodsell, D.S., Prlic, A., Quesada, M., Quinn, G.B., Westbrook, J.D., Young, J., Yukich, B., Zardecki, C., Berman, H.M., & Bourne, P.E., 2011. The RCSB Protein Data Bank. Nucleic Acids Research 39 D392-D401.

Rost, B., Honig, B., & Valencia, A., 2002. Bioinformatics in structural genomics. Bioinformatics 18 (7), 897-898.

Rost, B., & Sander, C., 1993. Prediction of protein secondary structure at better than 70% accuracy. J. Mol. Biol. 232 (2), 584-599.

Salemme, F.R., 1983. Structural properties of protein beta-sheets. Prog. Biophys. Mol. Biol. 42 (2-3), 95-133.

Salemme, F.R., & Weatherford, D.W., 1981a. Conformational and geometrical properties of beta-sheets in proteins. I. Parallel beta-sheets. J. Mol. Biol. 146 (1), 101-117.

Salemme, F.R., & Weatherford, D.W., 1981b. Conformational and geometrical properties of beta-sheets in proteins. II. Antiparallel and mixed beta-sheets. J. Mol. Biol. 146 (1), 119-141.

Shamovsky, I.L., Ross, G.M., & Riopelle, R.J., 2000. Theoretical studies on the origin of beta-sheet twisting. J. Phys. Chem. B 104 11296-11307.

Sternberg, M.J., & Thornton, J.M., 1977a. On the conformation of proteins: an analysis of beta-pleated sheets. J. Mol. Biol. 110 (2), 285-296.

Sternberg, M.J., & Thornton, J.M., 1977b. On the conformation of proteins: hydrophobic ordering of strands in beta-pleated sheets. J. Mol. Biol. 115 (1), 1-17.

Sternberg, M.J., & Thornton, J.M., 1977c. On the conformation of proteins: towards the prediction of strand arrangements in beta-pleated sheets. J. Mol. Biol. 113 (2), 401-418.

Steward, R.E., & Thornton, J.M., 2002. Prediction of strand pairing in antiparallel and parallel beta-sheets using information theory. Proteins 48 (2), 178-191.

Subramani, A., & Floudas, C.A., 2012. beta-sheet topology prediction with high precision and recall for beta and mixed alpha/beta proteins. PLoS One 7 (3), e32461.

Von Heijne, G., & Blomberg, C., 1977. The beta structure: inter-strand correlations. J. Mol. Biol. 117 (3), 821-824.

Von Heijne, G., & Blomberg, C., 1978. Some global beta-sheet characteristics. Biopolymers 17 (8), 2033-2037.

Wouters, M.A., & Curmi, P.M., 1995. An analysis of side chain interactions and pair correlations within antiparallel beta-sheets: the differences between backbone hydrogen-bonded and non-hydrogen-bonded residue pairs. Proteins 22 (2), 119-31.

Zafer, A., Yucel, A., & Hakan, E., 2011. Bayesian models and algorithms for protein beta-sheet prediction. IEEE ACM Trans. Comp. Biol. Bioinf. 8 395-409.

Zhang, N., Duan, G., Gao, S., Ruan, J., & Zhang, T., 2010. Prediction of the parallel/antiparallel orientation of beta-strands using amino acid pairing preferences and support vector machines. J. Theor. Biol. 263 (3), 360-368.

Zimmermann, O., Wang, L., & Hansmann, U.H., 2007. BETTY: prediction of beta-strand type from sequence. In Silico Biol. 7 (4-5), 535-542.

Table 1.
Amino acid composition of beta-sheets (S) and of intersections of beta-strands with sheet axes (I) in percentages. The probability to find an amino acid in (I) divided by the probability to find an amino acid in (S) is given for each amino acid in the third column by the ratio R. Errors given in parentheses derive from the analysis 2555 beta-sheets of 1347 proteins classified within three lists.

	S		I		R	
W	2.17	(0.12)	2.92	(0.25)	1.35	(0.19)
M	1.96	(0.06)	2.37	(0.11)	1.21	(0.09)
I	9.20	(0.15)	10.77	(0.38)	1.17	(0.06)
C	2.31	(0.14)	2.67	(0.14)	1.15	(0.13)
L	9.12	(0.32)	10.34	(0.24)	1.13	(0.07)
F	5.80	(0.15)	6.19	(0.21)	1.07	(0.06)
A	6.29	(0.14)	6.72	(0.3)	1.07	(0.07)
Y	5.49	(0.1)	5.84	(0.13)	1.06	(0.04)
V	12.81	(0.16)	13.46	(0.42)	1.05	(0.05)
E	4.22	(0.12)	4.09	(0.15)	0.97	(0.06)
R	3.94	(0.06)	3.78	(0.17)	0.96	(0.06)
H	2.13	(0.06)	1.99	(0.09)	0.94	(0.07)
Q	3.12	(0.1)	2.90	(0.28)	0.93	(0.12)
G	4.99	(0.14)	4.47	(0.57)	0.90	(0.14)
S	5.62	(0.19)	4.89	(0.15)	0.87	(0.06)
T	7.48	(0.04)	6.24	(0.38)	0.83	(0.06)
K	4.95	(0.09)	4.12	(0.26)	0.83	(0.07)
P	1.79	(0.07)	1.42	(0.21)	0.79	(0.15)
D	3.20	(0.1)	2.47	(0.06)	0.77	(0.04)
N	3.11	(0.05)	2.07	(0.07)	0.67	(0.03)

Figure Legends.

Figure 1.
Amino acids on the axis are shown for the *Clostridium symbiosum* pyruvate phosphate dikinase domain structure (2fm4.pdb) (Lin et al., 2006) whose ribbon is represented by links between adjacent alpha carbons from amino acids. The amino acids on the axis or closest to the axis are numbered 399, 467 and 426, 445 respectively in sheet A characterized by the averaged distance to the sheet axis D = 1.51 Angström per strand (Figure 1A) and 404, 506 and 497 in sheet B characterized by the distance D = 0.07 (Figure 1B).

Figure 2.

Distribution of the distance D between 0 and 9.5 Angströms for length intervals of 0.5 Angströms. The errors were found to be 0.019 for the highest probability of 0.318 for distances between 0.5 and 1.0 Angström (second bar), 0.006 for the probability 0.039 for distances between 2.5 and 3.0 Angströms (sixth bar) and between 0.001 and 0.005 for the remaining probabilities.

Figure 1.

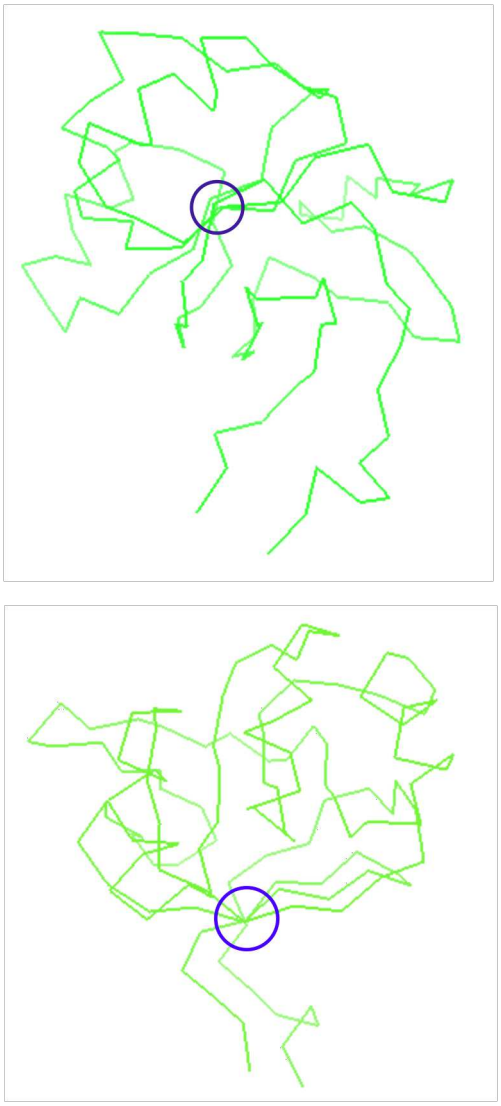
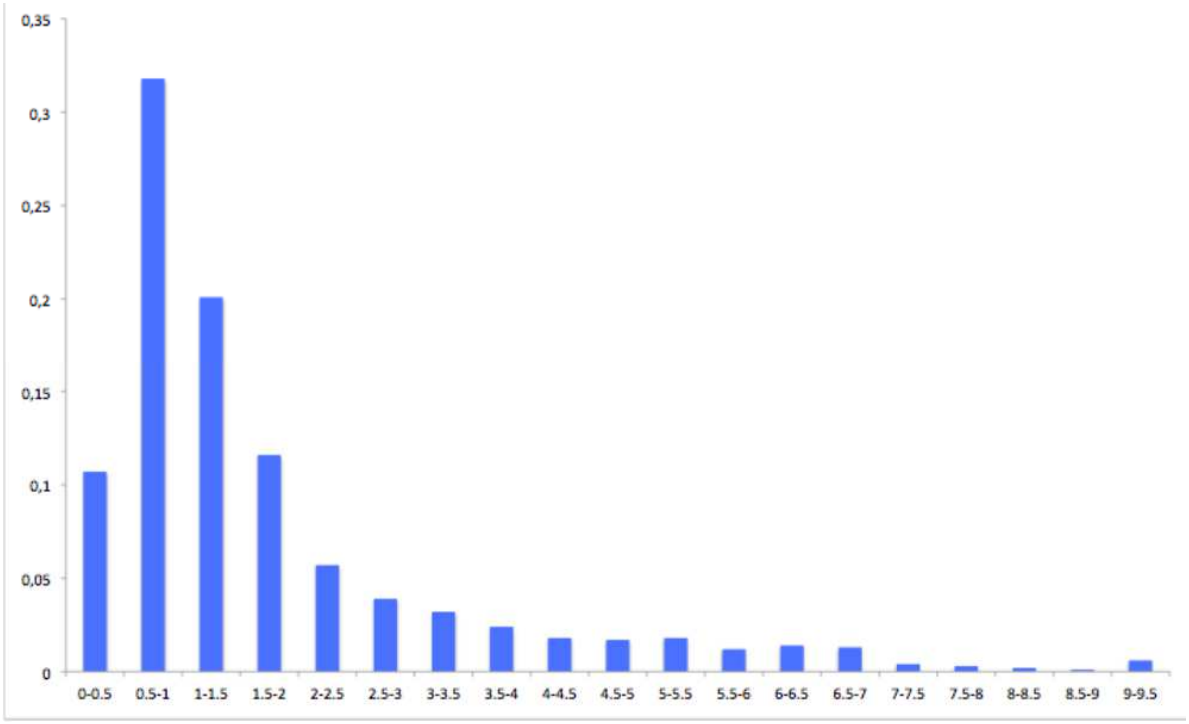


Figure 2.


0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
0-0.5
0.5-1
1-1.5
1.5-2
2-2.5
2.5-3
3-3.5
3.5-4
4-4.5
4.5-5
5-5.5
5.5-6
6-6.5
6.5-7
7-7.5
7.5-8
8-8.5
8.5-9
9-9.5