Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2008) 829–833
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Structural-symmetry-related sequence patterns of the proteins of beta-propeller family

Xiaochun Wang, Yanzhao Huang, Yi Xiao *

Biomolecular Physics and Modeling Group, Department of Physics, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China

Received 23 December 2006; received in revised form 29 April 2007; accepted 30 April 2007
Available online 3 May 2007

Abstract

It is still not very clear to what extent and how does the amino acid sequences of proteins determine their tertiary structures. In this paper, we report our investigations of the sequence–structure relations of the proteins in the beta-propeller fold family, which adopt highly symmetrical tertiary structures while their sequences appear ''random''. We analyzed the amino acid sequences by using a similarity matrix plus Pearson correlation method and found that the sequences can show the same symmetries as their tertiary structures only if we deduce the conditions of sequence similarity. This suggests that some key residues may play an important role in the formation of the tertiary structures of these proteins.

Keywords: Protein; Symmetry; Tertiary structure; Sequence; Beta-propeller family

The relationship between sequence and structure in the protein world has puzzled scientists for many years and a lot of efforts have been made to explore this problem [1–10]. In the present paper, we are interesting in understanding the sequence–structure relations of symmetric proteins. We shall use a novel method to explore this problem for one of typical protein families with symmetrical structures, i.e. β-propeller fold proteins.

The tertiary structures of β-propeller fold proteins show clear symmetries while they generally share no significant sequence similarity and frequently show no easily detectable sequence repeats [11,12]. Therefore, it is interesting to investigate whether the amino acid sequences contain the same symmetric patterns as their structures and to what extent the sequences determine the symmetries of their structures.

The structures of beta-propeller folds adopt a highly symmetric structure based on four to eight-fold repeats of a four-stranded antiparallel β-sheet motif [12–17]. Each of such motifs has been called a propeller blade and the whole structure an *n*-blade propeller. The motifs are connected by loop regions from β strand four in one motif to β strand 1 in the next motif. The twisted β-sheets pack face-to-face, radially arranged around a central tunnel. Overall, β-propeller domains are disc-shaped. We can take the structure of tetrameric neuraminidase heads as an example. Each of the four sub-units of the tetramer is folded into a single domain built up from six closely packed and similarly folded motifs. The motif is a simple up-and-down antiparallel β-sheet of four strands (Fig. 1a). The six motifs are arranged within each sub-unit with an approximate six-fold symmetry around an axis through the centre of the sub-unit (Fig. 1b). These six beta sheets are arranged like six blades of a propeller [15].

Five classes of β-propeller have been found, having a four-, five-, six-, seven-, or eight-fold pseudo-symmetry axis. From common view, these symmetries at structure level should be related to those at sequence level since the structures are determined by their sequences. However, although their modular building blocks all preserve the same basic conformation, they do not have similar sequences [16].

1. Methods

It is well known that if the primary sequences of two proteins have more than 25% identical amino acids, they likely have similar tertiary structures [18–20]. We expect the same rule may hold for the sub-structures of proteins, i.e. as long as two subsequences have certain number of identical amino acids, they likely have similar local tertiary structures. This suggests

* Corresponding author.
E-mail address: yxiao@mail.hust.edu.cn (Y. Xiao).

1093-3263/$ – see front matter
doi:10.1016/j.jmgm.2007.04.014

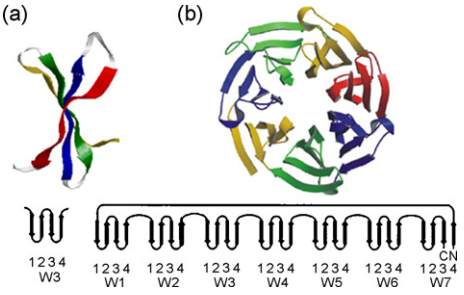

Fig. 1. Schematic and topological diagrams of the folding motif in neuraminidase from influenza virus. (a) The motif is built up from four antiparallel β strands joined by hairpin loops, an up-and-down open β sheet. (b) The sub-unit structure of the neuraminidase headpiece is built up from six similar, consecutive motifs [11].

that the amino acid sequences may reveal symmetric patterns as their structures if we lower the conditions for sequence similarity.

Based on this idea, we developed a method of analyzing the symmetries of the primary sequences. The detail of this method is as following:

Consider an arbitrary protein sequence $S = X_1X_2X_3\cdots X_N$, where N is the sequence length and x_i is one of the 20 amino acids. To take into account of the similarities of the 20 amino acids in physiochemical properties, we can divide them into different classes. For simplicity, we classify the 20 amino acids into two types according to their ability to form β-strands [10,21], i.e. strong β-strand formers (F, I, L, V, W, Y) and weaker β-strand formers (A, C, D, G, H, K, M, N, P, Q, R, S, T, E). It is easy to extend this method to use other schemes of classification. Then, we construct a set of all $(N - d + 1)$ possible segments of d consecutive symbols:

$$\begin{aligned} X_1(d) &= x_1x_2\ldots x_d \\ X_2(d) &= x_2x_3\ldots x_{d+1} \\ &\cdots\cdots \\ X_i(d) &= x_2x_3\ldots x_{d-1} \\ &\cdots\cdots \\ X_{N-d+1}(d) &= x_{N-d}x_{N-d+2}\ldots x_N \end{aligned} \tag{1}$$

where i denotes the location of the first residue of X_i in the sequence. Then, for each segment X_i (d) $(1 \le i \le N - d + 1)$, we find how many segments among the remaining ones are similar to it. We define that a segment $X_j(d)$ $(i \neq j,$ $1 \le j \le N - d + 1)$ is similar to $X_i(d)$ if the percent of their identical symbols is larger than s (the degree of similarity of two different segments).

In this way, for each segment $X_i(d)$, we can give a value $S_{d,i}$, which is the number of the segments similar to $X_i(d)$. Thus, for all the $(N - d + 1)$ segments $(X_1(d), X_2(d), \ldots, X_{N-d+1}(d))$, we obtain a set of the numbers $(S_{d,1}, S_{d,2}, \ldots, S_{d,N-d+1})$. The analysis of similar segments can be done for different segment lengths d. We can use $(S_{d,1}, S_{d,2}, \ldots, S_{d,N-d+1})$ for different d as rows to build a similarity matrix S for a protein sequence, i.e. the elements of S are $S_{d,i}$ (see Fig. 2).

To find the possible repetitions in the protein sequences, we further calculate the Pearson correlation coefficient r between sub-matrices of the same size in S (see Fig. 2). The sub-matrices have the same row size as S but smaller column size $(\le N/2)$. For example, if we consider sub-matrices with $N/2$ columns, we have sub-matrices S^1, $S^2, \ldots, S^{N/2+1}$ (see Fig. 2, where $N = 10$). These sub-matrices can be considered as a profile representation of the segments $(X_1(N/2), X_2(N/2), \ldots, X_{N/2+1}(N/2))$. If the protein sequence has a repetitive structure made of two modules of equal length (1 to $N/2$ and $(N/2 + 1)$ to N), the Pearson correlation coefficient between S^1 and $S^{N/2+1}$ should be a maximum. Any other two sub-matrices (e.g., 1 to $N/2$ and $N/4$ to $(3N/4 - 1)$) should give a much lower Pearson r. In a similar way, we can subdivide S into sub-matrices with the column size being N/n if we want to explore the n-fold repetition of protein structures. It is noted that the last sub-matrix, due to many zero elements (e.g., $S^{N/2+1}$ in Fig. 2) seem to be in an unequal position with other sub-matrices. In fact, for larger d, the corresponding elements of other sub-matrices are usually also zeros and so this problem will not affect our calculated results. The Pearson correlation analysis in our method incorporates information of repeats of different lengths and like a profile–profile comparison. This may also help us identifying repeats with complex structures, e.g. the repeats each of them is made of several non-consecutive subsequences.

$$\begin{array}{c} 5 \\ 4 \\ 3 \\ 2 \\ 1 \\ d\,/\,i \end{array}\!\!\begin{bmatrix} S_{5,1} & S_{5,2} & S_{5,3} & S_{5,4} & S_{5,5} & S_{5,6} & 0 & 0 & 0 & 0 \\ S_{4,1} & S_{4,2} & S_{4,3} & S_{4,4} & S_{4,5} & S_{4,6} & S_{4,7} & 0 & 0 & 0 \\ S_{3,1} & S_{3,2} & S_{3,3} & S_{3,4} & S_{3,5} & S_{3,6} & S_{3,7} & S_{3,8} & 0 & 0 \\ S_{2,1} & S_{2,2} & S_{2,3} & S_{2,4} & S_{2,5} & S_{2,6} & S_{2,7} & S_{2,8} & S_{2,9} & 0 \\ S_{1,1} & S_{1,2} & S_{1,3} & S_{1,4} & S_{1,5} & S_{1,6} & S_{1,7} & S_{1,8} & S_{1,9} & S_{1,10} \end{bmatrix}\!\!\begin{array}{c} \\ \\ \\ \\ \\ \\ 1 \end{array}\!\!\begin{array}{cccccccccc} & & & & & & & & & \\ & & & & & & & & & \\ & & & & & & & & & \\ & & & & & & & & & \\ & & & & & & & & & \\ & & & & & & & & & \\ 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{array}$$

$$S^1 = \begin{bmatrix} S_{5,1} & S_{5,2} & S_{5,3} & S_{5,4} & S_{5,5} \\ S_{4,1} & S_{4,2} & S_{4,3} & S_{4,4} & S_{4,5} \\ S_{3,1} & S_{3,2} & S_{3,3} & S_{3,4} & S_{3,5} \\ S_{2,1} & S_{2,2} & S_{2,3} & S_{2,4} & S_{2,5} \\ S_{1,1} & S_{1,2} & S_{1,3} & S_{1,4} & S_{1,5} \end{bmatrix}\!\!\begin{array}{c} \\ \\ \\ \\ \\ i \end{array}\!\!\begin{array}{ccccc} & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ 1 & 2 & 3 & 4 & 5 \end{array}\quad S^2 = \begin{bmatrix} S_{5,2} & S_{5,3} & S_{5,4} & S_{5,5} & S_{5,6} \\ S_{4,2} & S_{4,3} & S_{4,4} & S_{4,5} & S_{4,6} \\ S_{3,2} & S_{3,3} & S_{3,4} & S_{3,5} & S_{3,6} \\ S_{2,2} & S_{2,3} & S_{2,4} & S_{2,5} & S_{2,6} \\ S_{1,2} & S_{1,3} & S_{1,4} & S_{1,5} & S_{1,6} \end{bmatrix}\!\!\begin{array}{c} \\ \\ \\ \\ \\ i \end{array}\!\!\begin{array}{ccccc} & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ 2 & 3 & 4 & 5 & 6 \end{array}\quad \cdots S^6 = \begin{bmatrix} S_{5,6} & 0 & 0 & 0 & 0 \\ S_{4,6} & S_{4,7} & 0 & 0 & 0 \\ S_{3,6} & S_{3,7} & S_{3,8} & 0 & 0 \\ S_{2,6} & S_{2,7} & S_{2,8} & S_{2,9} & 0 \\ S_{1,6} & S_{1,7} & S_{1,8} & S_{1,9} & S_{1,10} \end{bmatrix}\!\!\begin{array}{c} \\ \\ \\ \\ \\ i \end{array}\!\!\begin{array}{ccccc} & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ 6 & 7 & 8 & 9 & 10 \end{array}$$

Fig. 2. The description of the construction of the similarity matrix S (top) and its sub-matrices $S^{(i)}$ (bottom) of column size $N/2$ for an amino acid sequence of length $N = 10$, where d is the segment length, i is the location of the first amino acid residue of the segment and $S^{(i)}$ is a sub-matrices with its columns beginning from the ith column (amino acid residue) of S.

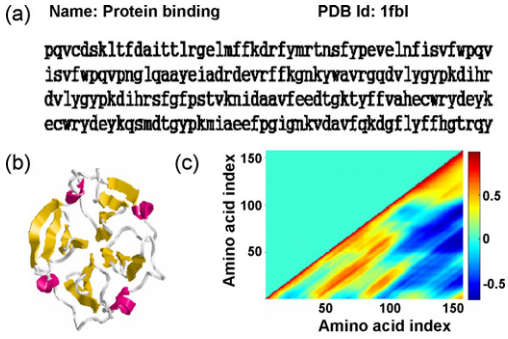

Fig. 3. The C-terminal domain of the Protein Metalloprotease (PDB id:1fbl). (a) The primary sequence; (b) cartoon representation of the tertiary structure; (c) pseudo-color plot of the Pearson's correlation coefficients between different sub-matrices of the similarity matrix. Both the horizontal and vertical axes denote the amino acid index in the sequence. The magnitude of the correlation coefficients is indicated by the colorbar.

The Pearson correlation coefficients r between two sub-matrices is defined as:

$$r(j,k) = \frac{\sum_m \sum_n (S^j_{mn} - \bar{S}^j)(S^k_{mn} - \bar{S}^k)}{\sqrt{(\sum_m \sum_n (S^j_{mn} - \bar{S}^j)^2)(\sum_m \sum_n (S^k_{mn} - \bar{S}^k)^2)}} \quad (2)$$

where $r(j, k)$ is the value of the correlation coefficient between the sub-matrix S^j and S^k, m and n are, respectively, the row and the column indices of the elements in S^j and S^k, and $\bar{S}^j$ and $\bar{S}^k$ are the average value of all elements in S^j and S^k. The Pearson correlation coefficients between all possible pairs of sub-matrices construct a matrix r, which can be visualized by using pseudo-color plot (Fig. 3c). If the protein structure has recurring substructures, the Pearson correlation coefficient between the corresponding sub-matrices should be very high and this can be obviously seen from the pseudo-color plot.

2. Results and discussions

We shall take the C-terminal domain (275–466) of the Metalloprotease (PDB ID:1fbl) as an example to show the possible symmetric pattern in its primary sequence (Fig. 3). The C-terminal domain of 1fbl is a member of the beta-propeller four-fold and has a four-fold symmetric structure (Fig. 3b), but its primary sequence appears irregular without clear symmetric signals (Fig. 3a). Since this is a protein with four-fold symmetry, we calculated the Pearson's correlation coefficients between every sub-matrices of column size (N/4) for the similarity matrix S. The pseudo-color plot of the Pearson's correlation coefficients between different sub-matrices of the similarity matrix is shown in (Fig. 3c). Since the plot is symmetric about the main diagonal, only the part below is shown. In the plot, the color at the point (i, j) denotes the value of the Pearson's correlation coefficient between the sub-matrices S^i and S^j, i.e. the similarity between the two subsequences of length 4/N beginning at the ith and jth residues, respectively. There are four red bands paralleled with the main diagonal. They represent larger correlation coefficients between these sub-matrices. The main diagonal band

Table 1
The calculated results of repeat number at sequence level for 57 representative proteins of four-blade to eight-blade beta propeller families

Fold class	PDB id	Repeat number		s
		Structure level	Sequence level	
Four-bladed beta-propeller	1hxn	4	4	0.40
	1itvA	4	4	0.30
	1qhuA	4	4	0.32
	1pex	4	4	0.28
	1fbl	4	4	0.40
	1gen	4	4	0.25
Five-bladed beta-propeller	1tl2	5	5	0.60
	1uypA	5	5	0.65
	1gyhA	5	5	0.50
	1oygA	5	5	0.50
	1s18A	5	5	0.50
Six-bladed beta-propeller	1kit	6	6	0.60
	1f8eA	6	6	0.50
	1ijqA	6	6	0.60
	1ofzA	6	6	0.70
	1suuA	6	6	0.30
	1npeA	6	6	0.30
	1ms9A	6	6	0.60
	1cruA	6	6	0.55
	1h6lA	6	6	0.60
	1crzA	6	6	0.35
	3sil	6	6	0.60
	1e1aA	*6*	*6*	*0.60*
	4aahA	*6*	*6*	*0.60*
	1k32A	*6*	*6*	*0.55*
	1e8uA	*6*	*6*	*0.55*
	1q7fA	6	6	0.52
	1eur	6	6	0.70
	1v3eA	6	6	0.60
	2sli	6	6	0.60
	1v04A	6	6	0.50
Seven-bladed beta-propeller	1ri6A	7	7	0.60
	1l0qA	7	7	0.50
	1c9lA	7	7	0.60
	1nr0A	7	7	0.30
	1jtdB	7	7	0.60
	1tbgA	7	7	0.55
	1erjA	7	7	0.60
	1gxrA	7	7	0.60
	1k8kC	7	7	0.50
	1nexB	7	7	0.50
	1p22A	7	7	0.55
	1gof	7	7	0.55
	1a12A	7	7	0.60
	1jjuB	7	7	0.60
	1jjuH	7	7	0.70
	1m1xA	7	7	0.70
	1qfmA	7	7	0.60
	1k32A	*7*	*7*	*0.60*
	1jofA	*7*	*7*	*0.55*
	2bbkH	7	7	0.55
	1qniA	*7*	*7*	*0.55*
	1olzA	7	7	0.70
Eight-bladed beta-propeller	4aahA	8	8	0.60
	1flgA	*8*	*8*	*0.60*
	1kb0A	8	8	0.70
	1gq1A	8	8	0.60
	1pfqA	*8*	*8*	*0.50*

represents self-similarity of the sub-matrices and so the correlation coefficient is 1. The three non-diagonal bands indicate there exist four internal repeats. The repeats begin at the residue 1, 50, 86 and 150, respectively and the Pearson's correlation coefficients of the last three repeats with the first one are 0.64, 0.53, and 0.54, respectively. This means that the sequences of these four parts are similar, i.e., has a four-fold symmetric pattern. It is noted that the locations of the four sequence repeats are in good agreement with those of the structural repeats. These results indicate that the sequence of the second domain of the protein metalloprotease can indeed reveal the same symmetry as its structure if we lower the condition for sequence similarity. This suggests that the structure symmetry may be determined by part of amino acids in the sequence.

It is easy to extend the analysis above to the amino acid sequences of all other propeller fold proteins. Table 1 gives the calculated results of repeat number at sequence level for 57 representative proteins. These proteins are selected from the four to eight bladed beta-propeller folds in SCOP and cover all families in these five folds [22]. Furthermore, among them the identical amino acids between any two sequences are less than 30%. Therefore, these proteins can be taken as the representatives of five types of beta-propellers. In table, we also list the value of s at which the plot of Pearson correlation coefficients show clear symmetric patterns. Fig. 4 shows the structures and pseudo-color plot of the Pearson's correlation coefficients between different sub-matrices of the similarity matrix of some propeller fold proteins. It is surprising to find that almost all of them reveal the same pseudo-symmetries as their structures. However, the sequence symmetries can emerge only for very lower s. For examples, $s = 0.3$–0.4 for most four-blade β-propeller folds and $s = 0.5$–0.7 for most other larger β-propeller folds. In fact, it is also possible to see the sequence symmetric patterns of four-blade β-propeller folds when $s = 0.5$–0.6 but they are not very clear. We investigated the similarities between any two random binary sequences with lengths of 30–100 units and found they have about 65% identical units on average. This implies that the symmetric sequence patterns obtained here are not the results of the similarities of random sequence segments. Furthermore, since the symmetric patterns can emerge only for very lower s, even we have considered the physiochemical similarities between amino acids; it suggests that these symmetric patterns may be determined by a small part of amino acids in their sequences, i.e. a set of key amino acids. This is possible. For example, the protein domain G_β from transducin (PDB id:1tbg) is a propeller protein composed of seven blades, or called WD-repeats [15,23]. It is known that sequence similarities between its blades are much weak and there are only five residues that are almost totally invariant in each blade. These five structurally conserved residues connect the outer strand of each blade to the inner three strands of the next blade and are certainly considered as the key residues critical for the structural stability of each blade in the G_β protein [15]. In Ref. [23], we have investigated the key amino acids in a β-propeller protein by using an energy-based contact method. Therefore, in the next step, it is interesting to apply this method and others to identify the possible key amino acids of the whole β-propeller family.

From Table 1 it is found that our method can show clear sequence symmetries for 49 among 57 or 86% representative β-propeller proteins. The other eight proteins (marked as *italic* in Table 1) also show certain sequence symmetries but it is difficult to determine their degrees, even deducing the condition of sequence similarity further. This may due to the complex

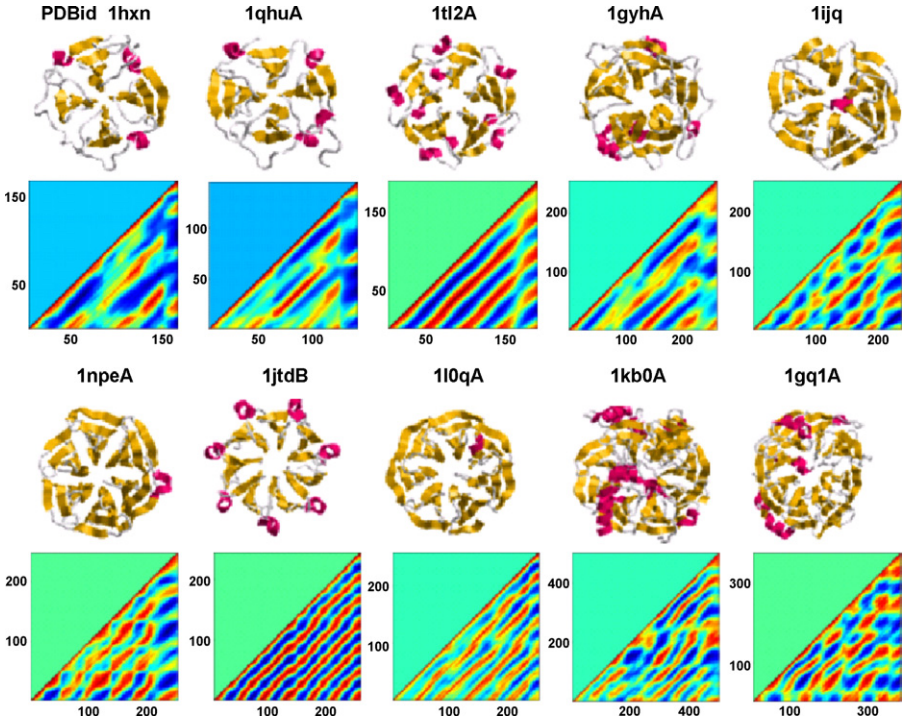

Fig. 4. The structures and pseudo-color plot of the Pearson's correlation coefficients between different sub-matrices of the similarity matrix of propeller fold proteins.

structures of these proteins or due to that we only considered the β-strand formation preference of amino acid propensity, although it is a reflection of the combined effect of some physiochemical properties of amino acids. In fact, the formations of protein tertiary structures need many aspects such as hydrophobic effect, polarity, electronic property and so on.

It is also noted that we have tried different physiochemical parameters of amino acids for their classification, e.g. hydrophobicity, charge, volume, beta-preference and so on, but the beta-preference gives the best results, i.e. the number of the proteins that our method can reveal the sequence symmetry is the largest by using beta-preference. It is known that the hydrophobic propensity plays a key role in the protein structures. We think our classification scheme may be a balance of the propensities of both beta-preference and hydrophobicity because five of the stronger beta-formers are also hydrophobic. Furthermore, in most cases, a secondary structure prediction will be difficult to suggest the presence of repeat beta units because structural variations associated with the β-propeller architecture, corresponding to the number of β-strands in the blade, number of amino acid residues in equivalent β-strands in the different blades, the occurrence and locations of β-helices [24,25].

In summary, our results demonstrate that most sequences of the beta-propeller fold proteins have hidden symmetric patterns in their amino acid sequences that coincide with their tertiary structure. This suggests that the formations of the symmetric tertiary structures of these protein domains may be the results of their sequence symmetries. In other words, the symmetries of these tertiary structures may be encoded by the sequences, specially a set of key amino acids. These results are in agreement with the theory that modern proteins evolved by gene duplications and fusions [26] and helpful to understand the sequence–structure relationship of proteins. On the other hand, our method may provides a qualitative approach to search the internal symmetries of proteins and can be used to predict whether the structure of a protein have symmetry directly from its sequence, and thus reduce the range of possible structures predicted by the conventional 3-D structure prediction method.

Acknowledgement

This work is supported by the NSFC under Grant no. 30525037 and 30470412 and the Foundation of the Ministry of Education of China.

References

[1] C.B. Anfinsen, Principles that govern the folding of protein chains, Science 181 (1973) 223–227.
[2] U. Mayor, N.R. Guydosh, C.M. Johnson, J.G. Grossmann, S. Sato, G.S. Jas, S.M.V. Freund, D.O.V. Alonsok, V. Daggettk, A.R. Fersht, The complete folding pathway of a protein from nanoseconds to microseconds, Nature 421 (2003) 863–867.
[3] C.M. Dobson, M. Karplus, The fundamentals of protein folding: bringing together theory and experiment, Curr. Opin. Struct. Biol. 9 (1999) 92–101.
[4] A. Sali, E. Shakhnovich, M. Karplus, How does a protein fold? Nature 369 (1994) 248–251.
[5] A. Heger, L. Holm, Rapid automatic detection and alignment of repeats in protein sequences, Prot. Struct. Funct. Genet. 41 (2000) 224–237.
[6] R. Szklarczyk, J. Heringa, Tracking repeats using significance and transitivity, Bioinformatics 20 (Suppl. 1) (2004) i311–i317.
[7] M.A. Andrade, C.P. Ponting, T.J. Gibson, P. Bork, Homology-based method for identification of protein repeats using statistical significance estimates, J. Mol. Biol. 298 (2000) 521–537.
[8] J. Heringa, The evolution and recognition of protein sequence repeats, Comput. Chem. 18 (1994) 233–243.
[9] M. Pellegrini, E.M. Marcotte, T.O. Yeates, A fast algorithm for genome-wide analysis of proteins with repeated sequences, Prot. Struct. Funct. Genet. 35 (1999) 440–446.
[10] R. Xu, Y. Xiao, A common sequence-associated physiochemical feature for proteins of beta-trefoil family, Comput. Biol. Chem. 29 (2005) 79–82.
[11] W.R. Taylor, J. Heringa, F. Baud, T.P.A Flores, Fourier analysis of symmetry in protein structure, Protein Eng. 15 (2002) 79–89.
[12] V. Fülöp, D.T. Jones, Beta Propellers: structural rigidity and functional diversity, Curr. Opin. Struct. Biol. 9 (1999) 715–721.
[13] T. Pons, R. Gómez, G. Chinea, A. Valencia, Beta- propellers: associated functions and their role in human diseases, Curr. Med. Chem. 10 (2003) 505–524.
[14] A.G. Murzin, Structural principles for the propeller assembly of β-sheets: the preference for seven-fold symmetry, Prot. Struct. Funct. Genet. 14 (1992) 191–201.
[15] C. Branden, J. Tooze, Introduction to Protein Structure, 2nd ed., Garland Publishing, New York, 1999.
[16] M. Paoli, Protein folds propelled by diversity, Prog. Biophys. Mol. Biol. 76 (1/2) (2001) 103–130.
[17] J.N. Varghese, W.G. Laver, P.M. Colman, Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 A resolution, Nature 303 (1983) 35–40.
[18] S.B. Needleman, C.D. Wunsch, A general method applicable to the search for similarities in the amino acid sequences of two proteins, J. Mol. Biol. 48 (1970) 443–453.
[19] T.F. Smith, M.S. Waterman, Identification of common molecular subsequences, J. Mol. Biol. 147 (1981) 195–197.
[20] D.W. Mount, Bioinformatics: Sequences and Genome Analysis, Cold Spring Harbor Laboratory Press, New York, 2001, 428.
[21] P.Y. Chou, G.D. Fasman, Prediction of the secondary structure of proteins from their amino acid sequence, Adv. Enzymol. Relat. Areas Mol. Biol. 47 (1978) 45–47.
[22] T.J. Hubbard, B. Ailey, S.E. Brenner, A.G. Murzin, C. Chothia, SCOP: a structural classification of proteins database, Nucl. Acids Res. 27 (1999) 254–256.
[23] C. Chen, L. Lin, Y. Xiao, Identification of key residues in proteins by using their physical characters, Phys. Rev. 73 (2006), 041926-1-041926-7.
[24] K. Guruprasad, P. Dhamayanthi, Structural plasticity associated with the β-propeller architecture, Inter. J. Biol. Macromol. 34 (2004) 55–61.
[25] A. Kotu, K. Guruprasad, The automatic detection of known β-propeller structural motifs from protein tertiary structure, Inter. J. Biol. Macromol. 36 (2005) 176–183.
[26] C. Chothia, J. Gough, C. Vogel, S.A. Teichmann, Evolution of the protein repertoire, Science 300 (2003) 1701–1703.