Journal of Theoretical and Computational Chemistry
Vol. 3, No. 4 (2004) 521–526

World Scientific
www.worldscientific.com

TWISTING: A QUANTITY DESCRIBING THE TOPOLOGICAL RELATION OF TWO PEPTIDE FRAGMENTS IN PROTEIN STRUCTURES

ZONG-HAO ZENG

Institute of Biophysics, Chinese Academy of Sciences
15 Datun Road, Chaoyang District, Beijing 100101, China

Received 7 July 2004
Accepted 9 October 2004

A method for analyzing the topological relation of any two peptide fragments in a protein structure is given by making analogy of the two fragments with a braid. Knotted relations are analyzed by calculating polynomials for the two fragments separately and for the two as a whole, respectively. Twisting is calculated from the writhe number, which is a measure to the internal relative rotation during formation of the protein structure. Statistical analysis on the internal rotations of all known structures found that the relative rotation like a one-dimensional random walk with decreasing step lengths.

Keywords: Protein structure; protein topology; protein folding; knot and braid.

1. Introduction

Proteins are linear chain molecules folded into compact forms when in their native states. The folding may be sequential or nonsequential.[1] Hyperhelices, either α or β, are examples of sequential folding, whereas the immunology globulin and "Jelly role"[2] are typical types of nonsequential folding. More extreme cases of non-sequentially folded structure are knotted structures.[3] If a peptide is cut into two pieces, the two fragments can easily be separated apart in case of sequential folding, while they could not in case of nonsequential folding. Sequentiality is a topological relation between two peptide fragments and it describes the complexity of folding. In this paper, a simple method that conveniently evaluates sequentiality is proposed, while taking into account more general topological relations between any two peptide fragments in a protein.

The theory of knot and braid is employed in the method. Usually, to see whether a line or belt is knotted, one needs to hold and take apart the two ends and then see what happen to the line or belt. Ambiguities exist, when any one of the two ends is too close to, or even inside, the rest of the line or belt. For proteins, instead of

stretching, contracting can be used while fixing the two ends.[3] This avoids ambiguous situations produced by stretching, on the understanding that the ways of contracting the peptide are always the same. In the case of two peptide fragments, a similar method can be employed by taking the two fragments as a two-string braid. As a result, the topological relations of the two fragments can easily be analyzed.

The method is tested on 3729 protein structures selected out from the Protein Data Bank (PDB)[4] using the standard of less than 40% sequence homology and without chain interruptions. Statistical analysis is also performed on this dataset.

The Two-String Braid

One need not to cut the peptide at any place, but at places where the way of folding changes. Secondary structures are essentially straight. A program is designed to automatically divide a polypeptide into straight segments (SSs). Each of these SSs may cover one or more secondary structural segments. Cut a polypeptide chain at places outside these SSs, then, each of the resulting two peptide fragments will consist of one or more contiguous SSs. For any two such peptide fragments, there are two vectors. Each of them begins from the C-end of one fragment and towards the N-end of another fragment. With the one most near the N- and C-end of the whole polypeptide chain as the bottom vector, the other vector can be called as the top vector. Attach each vector a plane perpendicular to the common perpendicular of the two vectors (Fig. 1). Contract the two peptide fragments in a way similar to that given by Ref. 3 and at the same time avoid any parts of the contracting peptide cutting each other. In this way, a setting resemble a two-string braid[5] is obtained.

Generally the setting is not standard according to the definition given by Murasugi and Kurpita.[5] Not only are the two vectors not parallel, but mainly

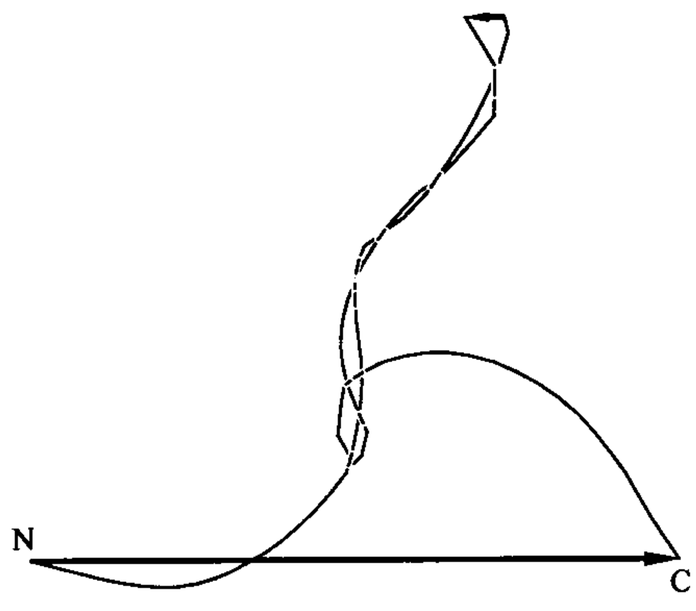

Fig. 1. A non-standard two-string braid. The two arrows show the top and the bottom vectors. The two peptide fragments are contracted and are negatively twisted. N and C denote the N- and C-end of the polypeptide respectively. This is a real example from the structure with PDB code 2nap.

because, in cases, one may even not be able to contract the two strings (the two polypeptide fragments) to let them getting inside the two parallel planes. This may affect the analysis on knots and tangles, but not on twisting. A further work is to project the two strings to a plane parallel to the common perpendicular, and then calculate nodes, i.e. crossing points of the projections. Another purpose of contraction is to decrease the number of nodes to accelerate further calculations in the following steps. The contraction itself is also CPU-time-consuming. Therefore, after the number of nodes is reduced to a given value, the contraction is stopped.

2. Checking for Knotted Structures

It has been known that some of the protein structures are knotted.[3] Knots are checked for the two fragments separately and for the two fragments as a whole. For this purpose, the N- and C-end of the two fragments are taken to be the same point to make them into two closed lines, respectively. By including the top and the bottom vectors, the two peptide fragments is also embedded into a closed line. Then the Kauffman polynomials, or Jones polynomials, are calculated for each of them. The Kauffman polynomials are calculated by the formulae[6]

$$K(a) = (-a)^{-3w} \sum a^{\alpha(s)-\beta(s)}(-a^2 - a^{-2})^{\gamma(s)-1}, \tag{1}$$

where the summation runs over all states s. There are two different states, A and B, for a node after destroying it. If there are n nodes, then the total number of states is 2^n. In Eq. (1), w is the writhe number of the closed line, $\alpha(s)$ and $\beta(s)$ are number of nodes in states A and B, respectively, and $\gamma(s)$ is the number of circles obtained after destroying all the nodes. For more details, see Prasolov and Sossinsky.[6] The Jones polynomials are obtained from the Kauffman polynomials by the transformation $a = q^{-1/4}$ as

$$J(q) = K(q^{-1/4}). \tag{2}$$

Only two kinds of knotted or tangled structures[3] were found in PDB. They are displayed in Figs. 2(a) and 2(b), and have Kauffman polynomial

$$a^{-4} + a^{-12} - a^{-16} \quad \text{and} \quad a^8 - a^4 + 1 - a^{-4} + a^{-8},$$

respectively. The example in Fig. 2(c) is a slippy knot. If we check the whole polypeptide chain, no knot will be found.

3. Twisting

For nonsequential folding, in most of the cases the two peptide fragments are simply twisted. Excluding all the self-crossings in the projection, which are produced by one fragment with itself, then the twisting, φ, is calculated from the writhe number as

$$\varphi = -180^* \text{writhe} + \theta, \tag{3}$$

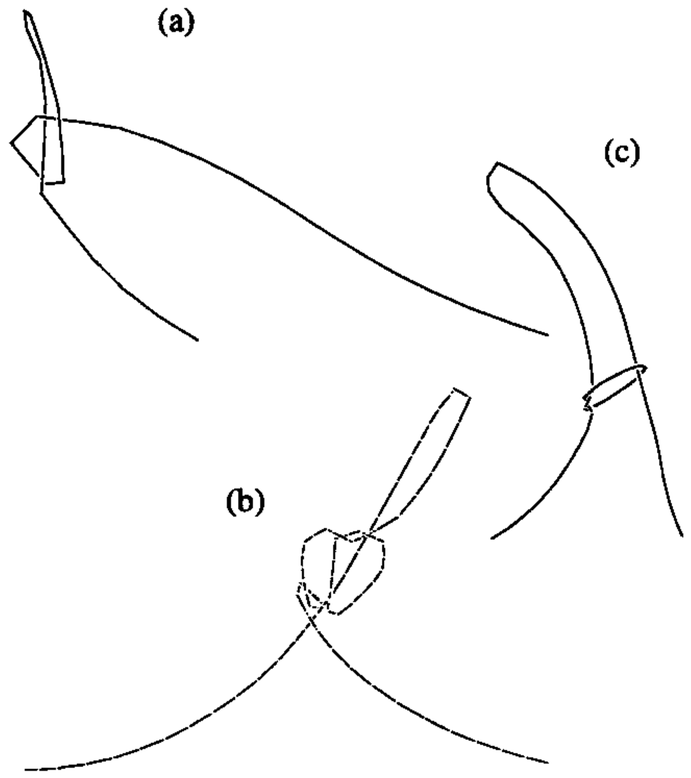

Fig. 2. Examples of knotted structures found in PDB: (a), 1 mxi; (b), 1 qmg; (c), 1 ed8. The two peptide fragments were contracted. More situations are those where one fragment is knotted while the other is not. Example (c) is a slippy knot, where one fragment is knotted, but the whole chain is not.

when the projection plane is perpendicular to the bottom vector, and where θ is the angle made by the top and the bottom vector. The negative sign is to make the positive and negative values consistent with the right and left rotations, respectively.

Twisting is a measure to the internal relative rotation between two peptide fragments in a folding. To check the statistical behaviour of this relative rotation, define φ_1 of a polypeptide chain as the largest value of twisting calculated by cutting a chain at various places, and then average φ_1 on chains that have the same number of SSs to get the average values $\langle\varphi_1\rangle$, respectively for positive and negative values of φ_1. Finally, $\langle\varphi_1\rangle$ is fitted to the following formula

$$\langle\varphi_1\rangle = 2\mathrm{arcctg}(\alpha(n-1)^{1/3} + \beta)(n-1)^{1/2} \tag{4}$$

where n is the number of SSs, and α and β are two parameters. This formula describes the average length performed by a one-dimensional random walk with decreasing step lengths. Like a fishing line wound on a reel, the line will get distant to the reel's axis, and the angle span will decrease for a fixed length of the line. This is why the step length decreases. Parameters obtained by fitting the calculated data are listed in Table 1. Figure 3 shows the calculated $\langle\varphi_1\rangle$ and the fitting lines.

Table 1. Parameters of fitting relative rotation angles to the formula of one-dimensional random walk with decreasing step lengths.

Parameters	α	β
Positive twisting	0.2455	0.2658
Negative twisting*	0.1274	0.8524

*Values corresponding to $-\varphi$.

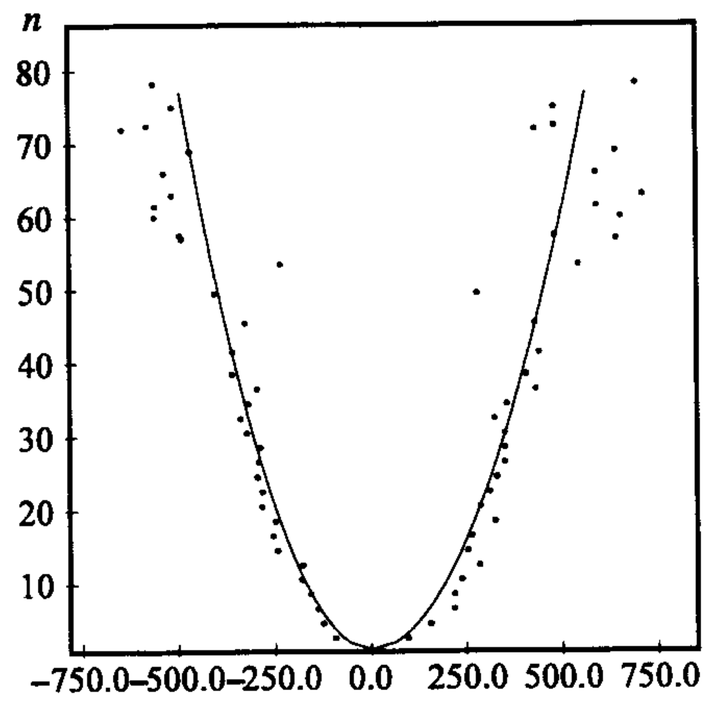

Fig. 3. Relation between $\langle\varphi_1\rangle$ and the number of SSs, n.

4. Discussions and Conclusions

If only the main chain conformation is taken into account, the conformation of a polypeptide chain is completely determined by the Ramachandran angles, ϕ and ψ. These are rotations around single bonds. In this detailed view, the microscopic forces dominate. These microscopic forces restrict the possible types of secondary structures and ϕ and ψ have their favored values.[2] While in a coarser view, at the levels beyond an elementary structural fragment, the randomness of thermal motions appears. If we look into a pool of proteins, i.e. a dilute solution of polypeptide chains with various foldable sequences, the various structural types displayed by all the proteins can be regarded as a reflection and a fixation of the thermal motions. Although each sequence has its own structure, the establishment of this structure needs thermodynamic motions. Each structure fixes a specific moment of motion in a condensed form, just like a snapshot records a moment of a race into a film. When all the snapshots are analyzed, we find that the relative rotation between different fractions of the chain showing random characters.

A structure formed by internal relative rotation can only defold through internal relative counter rotation; therefore the twisting restricts the possible ways of folding and unfolding. The elegant statistical rule extracted from the structural data reveals that the concept of twisting in protein structures is useful to understand some aspects of the thermodynamic principles governing the formation of protein structures and therefore useful in classifying protein structures in a way as consistent to folding as possible.

Acknowledgement

Grant sponsor: National Natural Science Foundation of China; Grant number 30080004.

References

1. Tsai C-J, Maizel JV Jr., Nussinov R, Distinguishing between sequential and nonsequentially folded proteins: Implications for folding and misfolding, *Protein Sci* **8**:1591–1604, 1999.
2. Scheef ED, Fink JL, Fundamentals of Protein Structure, in Bourne PE, Weissig H (eds.), *Structural Bioinformatics*, Wiley–Liss, NJ, 2003.
3. Taylor WR, A deeply knotted protein structure and how it might fold, *Nature* **406**: 916–919, 2000.
4. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE, The Protein data bank, *Nucleic Acids Res* **28**:235–242, 2000.
5. Murasugi K, Kurpita BI, *A Study of Braids*, Kluwer Academic Publishers, Dordrecht, 1999.
6. Prasolov VV, Sossinsky AB, *Knots, Links, Braids and 3-Manifolds*, American Mathematical Society, Providence, Rhode Island, 1997.