Structure Entropy of Peptide Fragments

Jenn-Kang Hwang[1] and Chen Hsiung Chan[2]

[1]Department of Biological Science & Technology and Institute of Bioinformatics, National Chiao Tung University, Hsin Chu, Taiwan, R.O.C. [2]Department of Life Sciences, National Tsing Hua University, Hsin Chu, Taiwan, R.O.C.
E-mail: jkhwang@cc.nctu.edu.tw

Introduction

To predict 3D protein structure directly from its primary sequence is one of the most challenging problems in biochemistry. It is generally believed that the sequence of a protein can determines its 3D structure (Fig. 1) [Anfinsen, 1973], but the mechanism is still not well understood despite great advances in recent years. At present, "homology modeling" is still the most reliable and routinely used procedures to predict 3D structure directly from protein sequence, if the structure of its homologue is available. "Homology modeling" is based on the observation that proteins with 25% or greater sequence homology are generally of similar structural folds [Guex and Peitsch, 1999]. Of course, exceptions do occur. In 1994, Rose and Creamer [Rose & Creamer, 1994] offered an award of 1,000 dollars out of their own pocket to whoever can transform a protein to a radically different conformation, while at the same time alter no more than 50% of its original amino acid composition. In 1997, Regan's group [Dala et al,1997] succeeded in transforming a protein of an all $\beta-$sheet conformation from Streptococcus bacteria to an all $\alpha-$helix protein under the "50% rule". This is no small feat, since their work shows that proteins with 50% homology can have completely different conformations. Another example is prion (pronounced as [`pree on], not [`prI on]), which causes "mad cow disease" [Prusiner, 1997]. Prion can adopt two different conformations - one form is predominantly -helical and does not cause disease, while the other is mostly $\tilde{\beta}$ sheet and does [Prusiner, 1998]. At present, it is still not clear how such transformation takes place. There are also examples that identical sequences will adopt different conformations, depending on the solvent environment. For example, trifluroethanol (TFE) is a solvent known to induce helical structure from proteins that will otherwise have different conformations in aqueous solution. Obviously, the conformation of a protein is determined by two factors: the intrinsic structural propensity encoded by its sequence and the interactions from the external environment such as the surrounding amino residues or solvent

molecules. The statement that "sequence determines its 3D structure" is an overly simplified one, which overlooks the importance of the effects from the external environment.

Structural information of short peptide fragments

In the case of peptide segments of lengths ranging from 4~25 amino acids, it frequently happens that an identical sequence can adopt different conformations in different proteins [Kabsch and Sander, 1984; Argos, 1987]. We observed that, while some sequences seems to prefer only one or two conformations, some other sequences do not seem to have any conformational preferences in terms of secondary structure. Examples are shown in Figure 2. It would be very interesting if we can quantify such structural information contents of short sequence fragments. This should be useful in, for example, *de novo* protein design, since such knowledge will help predict whether a new sequence is a good, foldable one before experiment. Recent experiments in protein folding show the existence of "hot spot regions" that are important to protein folding using ϕ-values analysis [Fersht et al., 1992; Itzhaki et al., 1995]. Plaxco et al have been shown that these hot spots (residues with high -values) are not necessarily correlated with sequence conservation (Plaxco et al, 2000), and they suggested that these regions are correlated with the "structural conserved regions". Obviously, the knowledge of information of structural content should be useful in the understanding of protein folding. Kabsch and Sander (1984) have reported that pentapeptides, a peptide composed of five amino acid residues, have different structures in different proteins. It has been shown that, while the structures of most pentapeptides depend on the context of its protein environments, 20% of pentapeptides have consistent structures in different proteins [Argos, 1987].

Structural entropy of peptide fragments

We have recently developed an approach, the Structural Entropy of Peptide Sequence method (SEPS) to evaluate the structural information contents of peptide fragments. Using this approach, we can evaluate the entropy of structural physical properties of peptides. These structural physical properties can be the secondary structural elements, solvent accessible surfaces or other 3D geometrical properties. The general formulation is as follows:

$$s^x = -\sum_i p_i^x \ln p_i^x \qquad (1)$$

where s^x is the entropy associated with the peptide sequence x of a given length L, and p_i^x is the probability that the peptide x is in the state of structural elements i. Using this equation, we can assign each peptide fragment of a given sequence a value of entropy of the designated physical properties. The length of the peptide fragment is arbitrary, ranging from 1 to 16 amino residues. In the case of single amino acid, if we use the secondary structure assignment to encode the peptide fragments, then Eq. (1) will give the secondary structure propensity of each amino acid.

Structural entropy in protein folding

One of the applications of SEPS is in the study of protein folding. We have recently applied this method to Cardiotoxin III (CTX III). CTX III is a small, all β-sheet protein with a length of 60 amino acids (Fig. 2). The β-sheets of CTX III can be classified as two groups: the two-stranded group composed of S1 and S2 connected by loop I, and the three-stranded group composed of S3, S4 and S5 connected through loop II and loop III. These three loops are considered functionally important in the toxicity of cardiotoxin [Gatineau et al., 1990; Marchot et al. 1988]. There are four structure-stabilizing disulfide bridges in CTX III. The melting temperature of CTX III is higher than 90ºC, making CTX III an extremely thermally stable protein [Jayaraman et al., 1996]. Folding studies on CTX III [Sivaraman et al., 1998] have shown that the two-stranded sheet unfolds faster than the three-stranded sheet.

We used secondary structural elements to encode the peptide sequences. A widow of a peptide fragment is used to evaluate the entropy of the given residue, which is usually chosen as the central residue of the window. We used a window of size of 5 residues because pentapeptides are the minimal number of residues that can cover most secondary structural elements such as α-helices, β-strands or turns. The plot of structural entropy vs. sequence of CTX III is shown in figure 4. The strands S1 and S2 are of the highest entropy among these five β-strands and it happens that these two strands are among the first to unfold in folding experiment [Sivaraman et al., 1998]. Hydrogen exchange experiment shows that S4 and S5 are among the first structural elements to form; the low entropy of S4 and S5 in figure 5 reflects this observation. We also notice that the entropy of the loop LX is lower than not only the rest of the loops and turns but also than strands S1 and S2. This is interesting since in general β-strands usually has lower entropy than the loop structures. Hence, the loop LX could play a crucial role in the folding process, possibly positioning S5 around S3, though this prediction still remains to be verified by experiment.

Conclusion

Though protein sequence usually uniquely defines its 3D structure, this observation in general does not apply to the conformations of peptide fragments. Some peptide segments in proteins will adopt only a few conformations, while others will adopt quite a few conformations in different proteins. At present, there is not yet an efficient, fast theoretical approach available to identify regions important to protein folding. SEPS should be useful in the study of protein folding, as shown in the case of the snake toxin protein.

References

Anfinsen, C. B., (1973) Principles that govern the folding of protein chains. *Science*, **181**, 223-230.

Argos, P. (1987) Analysis of sequence-similar pentapeptides in unrelated protein tertiary structures. *J. Mol. Biol.*, **197**, 331-348.

Dalal, S., Balasubramanian, S., Regan L. (1997) Protein alchemy: changing beta-sheet into alpha-helix. *Nat Struct Biol.* **4**, 548-52.

Fersht, A. R., Matouschek, A. and Serrano, L. (1992) The folding of an enzyme I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.*, **224**, 771-782.

Gatineau, E., Takechi, M., Bouet, F., Mansuelle, P., Rochat, H., Harvey, A. L., Montenay-Garestier, T., Menez, A. (1990) Delineation of the functional site of a snake venom cardiotoxin: preparation, structure, and function of monoacetylated derivatives. *Biochemistry*, **29**, 6480-6489.

Guex, N. and Peitsch, M. C. (1999) Molecular modelling of proteins. *Immunology News*, **6**,132-134.

Itzhaki, L. S., Otzen, D. E. and Fersht, A. R. (1995) The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.*, **254**, 260-288.

Jayaraman, G., Kumar, T. K., Sivaraman, T., Chang, D. K., Lin, W. Y. and Yu, C. (1996) Thermal denaturation of an all beta-sheet protein--identification of a stable partially structured intermediate at high temperature. *Int. J. Biol. Macromol.*, **18**, 303-306.

Kabsch, W. and Sander, C. (1984) On the use of sequence homologies to predict protein structure: Identical pentapeptides can have completely different conformations. *Proc. Natl. Acad. Sci. USA*, **81**, 1075-1078.

Marchot, P., Bougis, P. E., Ceard, B., Van Rietschoten, J., Rochat, H. (1988) Localization of the toxic site of Naja mossambica cardiotoxins: small synthetic peptides express an in vivo lethality. *Biochem. Biophys. Res. Commun.*, **153**, 642-647.

Plaxco, K. W., Larson, S., Ruczinski, I., Riddle, D. S., Thayer, E. C., Buchwitz, B., Davidson, A. R. and Baker, D. (2000) Evolutionary conservation in protein folding kinetics. *J. Mol. Biol.*, **298**, 303-312.

Prusiner, S. B. (1997) Prion Diseases and the BSE Crisis. *Science*, **278**, 245-251.

Prusiner, S. B. (1998) Prions. *Proc. Natl. Acad. Sci. USA*, **95**, 13363-13383.

Rose, G. D. and Creamer, T. P. (1994) Proteins: *Struct. Funct. Genet.* **19**, 1-3.

Sivaraman, T., Kumar, T. K. S., Chang, D. K., Lin, W. Y. and Yu, C. (1998) Events in the kinetic folding pathway of a small, all b-sheet protein. *Biochemistry*, **273**, 10181-10189.

Figure Captions

Figure 1. Protein sequence in general uniquely defines its 3D structure.

Figure 2. The identical sequence fragment IKMFIKN will assume a α-sheet conformation, shown in color, in (a) 1PGS, while an β-sheet conformation, shown in color, in (b) 1DL3. These structures were drawn by RasMol.

Figure 3. The 3D structure of CTX III in ribbon representation. This picture was generated using SwissPdbViewer and Pov-Ray.

Figure 4. The structural entropy vs. sequence of CTX III. The label Lx, Tx correspond to Loop x and Turn x in Fig. 3, respectively.

Figure 1

TTCCPSIVARSNFNVCRLPGTPEAICATYTGCIIIPGATCPGDYAN
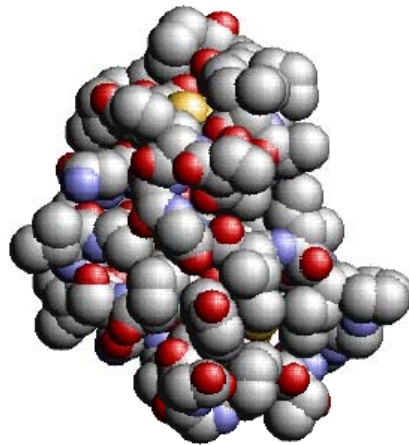
Protein sequence uniquely
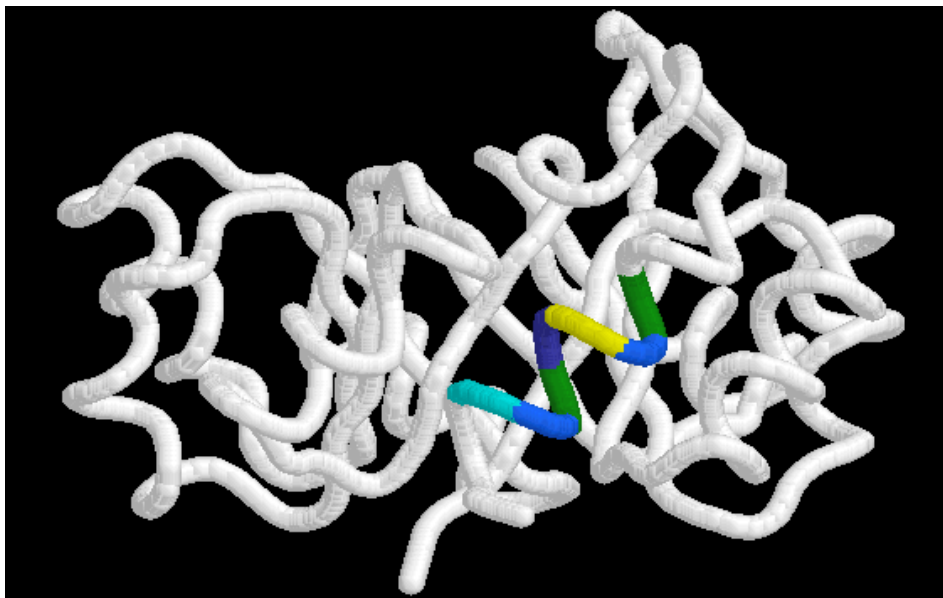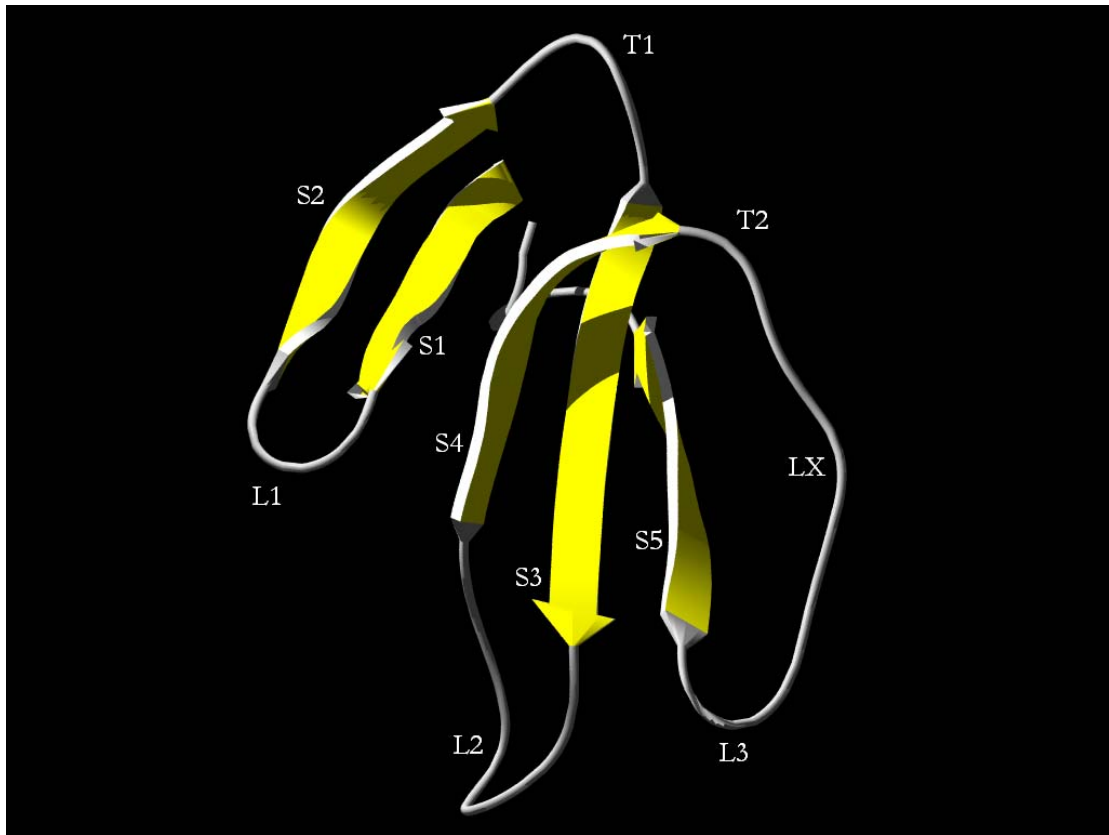defines its 3D structure.

Figure 2a

Figure 2b

Figure 3

Figure 4