Is Protein Folding Still a Challenge?

http://www.jbsdonline.com

Mittal et al. in their paper entitled “A stoichiometry driven universal spatial organization of backbones of folded proteins: Are there Chargaff’s rules for protein folding?” have asked whether there is a unifying theme or concept underlying the magnificent diversity of folded protein structures (1). They have observed that preferential interactions between amino acids do not drive protein folding, contrary to all prevailing views. This view is so prevailing that preferential interactions between amino acids are the basis for introducing knowledge-based potentials, which in turn provides the underpinning for present day 3D protein structure prediction by modeling and simulation (2-5 and references therein).

Every student of protein biochemistry is instilled with the idea that the protein folding process is complex and modulated by several factors. Levanthal’s paradox (6) is often stated to emphasize the complexity of protein folding. Delineating the pathway of protein folding and identification of folding intermediates has been and still is a favourite research topic. Structural propensities of amino acids to occur in a specific conformation such as $\alpha$-helix, $\beta$-sheet or $\beta$-turns were proposed based on analysis of protein structures in the crystalline state way back in the 1970’s by Chou and Fasman (7, 8). Prediction of protein folding was attempted based on structural propensities of amino acids and probability values (8). Extensive analysis of protein crystal structures, have provided insights into interaction between aromatic amino acids (9), disulfide conformation (10) and occurrence of $\beta$-hairpin structures (11, 12).

Mittal et al. have analyzed 3718 protein X-ray structures and delineated interacting partners of amino acids based on C$\alpha$ distances (excluding those adjacent to each other connected by a peptide bond). Their reasoning is that if the two methyl groups in Leu and Ile are proximal to each other and interact via hydrophobic interactions, this would be reflected in the C$\alpha$ distances between them. Their analysis also suggests that any one of the 20 coded amino acids could interact with any of the other 19 coded amino acids. Also, interactions between cationic or anionic amino acids are not precluded. Their intriguing observation is that, in the crystalline state, the interaction between amino acids is independent of the nature of the amino acid. The frequency of occurrence rather than chemical nature determines the interaction between amino acids. Their conclusion is that folding of a protein is simply dictated by frequency of occurrence of amino acids and not the sequence.

An important conclusion by Mittal et al. is that for a folded protein, stoichiometric ratios of individual amino acids should have values indicated in Table I. The implication of this is that it would not be possible for a protein to fold if the percentage of Ala deviates from 7.8+/−3.4 or that of His deviates from 2.3+/−1.4. This information would be of immense help to protein engineers in determining whether a protein encoded by an open reading frame can fold or possibly exist.

R. Nagaraj

Centre for Cellular and Molecular Biology, CSIR, Uppal road, Hyderabad 500 007, India

Corresponding Author:
R. Nagaraj
Phone: +91-40-2716022
Fax: +91-40-27160591/031
E-mail: nraj@ccmb.res.in
Mittal et al. have proposed in their paper that the most important parameters for protein folding are (i) exclusion of water and (ii) shape characteristics of individual amino acids along the sequence that would minimize surface-to-volume ratio. However, they summarize that protein folding is primarily dictated by frequencies of occurrence of amino acids. This appears to be at variance with their proposal in (ii).

Chargaff’s rule stems from the association of Adenine (A) with Thymine (T) and Guanine (G) with Cytosine (C) via hydrogen bonding. A similar association of amino acids does not emerge from the analysis by Mittal et al. In fact, their analysis stresses the absence of any preferential relationship between amino acids. Hence, the basis for stochiometry-driven spatially organized double helical structure of DNA does not apply to organization of amino acids in proteins. The authors are using the term Chargaff’s rule to emphasize the stoichiometric relationship that exist among the amino acids in Table I, this stoichiometry being the underpinning of protein folding. From species to species, source to source, the proportions of A, G, C and T in DNA may vary widely, but universally, the molar ratios of A and T and that of G and C in DNA were not far from unity. In the same manner, from protein to protein, the proportions of amino acids may vary, but in folded globular proteins, the stoichiometry of the amino acids should not be far from what is in Table I in Mittal et al. (1).

The process of protein folding in solution can be modulated by several factors such as pH, temperature, salt concentration and the presence of organic solvents. The study by Mittal et al. gives a fascinating and new insight into organization of amino acids in folded proteins in the crystalline state. However, knowledge of only amino acid composition may not provide insights into how the folded conformation is achieved, particularly in solution. Protein folding continues to be a challenge and a problem yet to be solved.

References