UNIVERSALITIES IN PROTEIN TERTIARY STRUCTURES: SOME NEW CONCEPTS

B.JAYARAM[1,2*], ADITYA MITTAL[1*], AVINASH MISHRA[1], CHANCHAL ACHARYA[1] AND GARIMA KHANDELWAL[2]

[1]*Kusuma School of Biological Sciences, Indian Institute of Technology Delhi Hauz Khas, New Delhi 110016, India.*

[2]*Department of Chemistry & Supercomputing Facility for Bioinformatics & Computational Biology, Indian Institute of Technology Delhi Hauz Khas, New Delhi 110016, India.*

[*]*E-mail: bjayaram@chemistry.iitd.ac.in; amittal@bioschool.iitd.ac.in*

We discuss the recent extraction of signatures of stoichiometry driven universal spatial organization of backbones of folded proteins regardless of their size, shape/structure and function. We present further evidence for secularity of amino acids in protein structures from the perspectives of surface area and energy. While conceptual fragmentation to gain insights into the diversity of protein structures appears to be a popular approach, we believe that the secrets to solving the protein folding problem lie in appreciating concepts that are universally applicable.

1. Introduction

Historically, the ideas of Pauling[1-3] and Ramachandran[4] proposed in 1950s and 1960s established universality among secondary structures in proteins. Pauling's work led to our understanding that proteins, irrespective of their structure and function, are made up of regular secondary structural elements called alpha helices and/or beta sheets and irregular regions connecting these called loops with Ramachandran's work providing a raison d'etre in terms of a stereochemical interpretation for these. Each secondary structural element (alpha helix or beta sheet), is characterized by a well-defined allowed region in the dihedral angle (ϕ, ψ) space of the backbones of proteins. Not surprisingly, even the so called "irregular regions", i.e. the loops, assume either helical or sheet like dihedral values[5]. A protein consisting of n peptide linkages shows up as n points in the 2D- (ϕ, ψ) Ramachandran plot, exhibiting a clustering of points as per its secondary structural composition. Structural studies via crystallography and NMR (RCSB[6]) have verified the hypotheses of Pauling and Ramachandran time and again.

How to extend these ideas of commonalities among proteins to tertiary structures remains a pending question however. A specification of the 2n Ramachandran angles (n ϕs and n ψs) leads to a coarse-grained description of the tertiary structure of a protein. The overall conformation of a protein thus corresponds to a single point in this 2n dimensional Hyper-Ramachandran plot. Of course, one could add more dimensions to account for the degrees of freedom associated with the side chains. If free energy is added to this 2n-dimensional surface, native structure of the protein, according to Anfinsen[7], corresponds to the bottom most point or the global minimum in free energy on this (2n+1) dimensional surface. Thus ensued several proposals on energy landscapes which overall conform to the concept of minimum free energy for the native structure[8-10]. Several other physico-chemical parameters based on size, shape, area, energy (intrinsic as well as transfer) have been investigated[11-14] but few universal ideas applicable to all proteins have emerged. This has led to extensive classifications of both secondary and tertiary structures of proteins such as various flavors of helices, turns, super-secondary structural motifs, folds etc.[15-19]. The corollary of all such classifications which chronicle the architectural splendor of proteins is to give up on universality.

We revisit proteins from a new perspective here and show evidence for very compelling clues to the existence of some universal principles, not on the folding pathways but, on the organization of protein tertiary structures. We hope that the perspective presented here paves the way for re-embracing unifying principles of protein structures rather than develop numerous subtle fragmentations.

2. Stoichiometry

It has been demonstrated recently[20-24] that amino acid space of proteins is not infinite rather proteins have well defined stoichiometries with bounds set on amino acid compositions (Table 1) as seen from the sequence data available in Swissprot/Uniprot[25]. The compositions are non-random and the deviations from the averages (called the margin of life[20]) account for the diversity of proteins.

It is logical to expect that these stoichiometries are the essence of protein size, shape, structure and function. In fact, a sequence analysis reveals the requirement of all 20 amino acids in all known protein sequences, as shown in Figure 1. This, coupled with the complete absence of naturally occurring anagramic protein sequences, strongly supports the idea that protein structure and function is governed by the stoichiometric ratios of amino acids.