Comment

Are there Still Surprises Buried Inside Statistical Analysis of Protein Structure?

http://www.jbsdonline.com

The Protein Folding Problem has been a long standing challenge towards achieving a complete molecular understanding of life (1). Since the days of Anfinsen, scientists over the world have been trying to understand the basic laws governing protein folding and have come a long way. We currently understand the general forces determining the protein folding such as the hydrophobic effect, secondary structure formation and the role of preferential interactions among amino-acids. However we don’t know their relative contributions at the quantitative level. Thus we are not capable of predicting the folding properties of proteins from their amino-acid sequence yet.

The article by Mittal et al. (2) investigates this issue with a fresh twist of the classic approach of extracting rules from the analysis of protein 3D structures. Particularly the researchers perform a detailed analysis of Cα- Cα contacts at various neighborhood distances, looking for correlations up to unconventionally high distance cut-offs. Various efforts have been made to study the effect of interactions among amino-acids, but for the first time the authors study the effect and importance of these interactions in protein folding over long distances. Their hypothesis is that if there are any preferential interactions between different amino-acids as it is usually assumed, then the contacts distribution at different distance cut-offs should exhibit specific biases. But if there are no preferential interactions then the contacts distribution would just follow the amino acid frequency distribution in natural sequences. From their detailed analysis, they discover that the composition of the structural environment for any amino-acid in folded proteins is directly proportional to the natural frequency of occurrence of the amino-acid rather than any preferential interactions. In other words, the distributions of amino-acid in proteins follow simple stoichiometric relationships that they call “Chargaff’s rules” for protein folding.

Over the years, there have been numerous efforts pertaining to determination of amino-acid interactions and to the development of pair-wise interaction potentials and in their application in coarse-grained computer models of proteins (3). In fact, a popular method assumes a quasi-chemical (4) approximation to determine pair-wise interaction strengths from the frequency of amino-acids pairs in contact. The Mittal et al. study departs from this general idea because it looks for correlations at very long distances (up to 60 Å). Their observation that spatial distribution of amino-acids is identical to that dictated by the amino-acid composition is a provocative result that opens up many questions which needs to be addressed in the future.

For example, the authors have performed an analysis for neighborhood distances up to 60 Å, whereas correlations are usually investigated at much shorter dis-

Ravishankar Ramanathan
Abhinav Verma*

Centro de Investigaciones Biológicas (CSIC) Calle de Ramiro de Maeztu 9, 28040 Madrid, Spain

*Corresponding Author:
Abhinav Verma
E-mail: abhinav@cib.csic.es
tances. Therefore there is an issue regarding whether when one goes up to such distances, is it still reasonable to expect specific correlations. When one calculates pairwise correlations over such long distances the number of possible pairs is so large that it is not unreasonable for them to average out resulting in a Gaussian distribution with its peak around the average radius of the proteins within the dataset (20-25 Å). The cumulative sum of a Gaussian distribution produces a sigmoidal curve that may be indistinguishable from those observed by the authors.

On the other hand, in a recent article, Jha et al. (5) have shown the presence of preferential interactions among amino-acids in protein folding by using an analysis based on a scoring matrix of Cα-Cα contacts in different structural environments. In this connection, it should be noted that that preferential interactions between amino acids are the basis for introducing knowledge-based potentials, which in turn provide the underpinning for present day three-dimensional protein structure prediction by modeling and simulation (6-9 and references therein). Another article by Berka et al. (10) investigates the intramolecular energies for pairs of amino-acid side-chains using ab initio quantum mechanical calculations. These articles observe skewed distributions that indicate the influence of amino-acid interactions, which is contrary to the observations of Mittal et al. and therefore it is important to complement the Mittal et al. study investigating the interactions at the level of specific amino-acids pairs, look for effects at shorter neighborhood distances and possibly include side-chains in the analysis.

Also, there could be astronomically high number of sequences with the natural distribution of amino-acids, yet not all of them are observed in nature and fold into proteins. In fact, experiments suggest that most random sequences are likely to be dysfunctional folding-wise. So, the amino-acid composition of natural proteins may already include long-range correlations selected by evolution that would be missed out with the Mittal et al. approach since the natural composition is used as reference state.

Summarizing, the authors here present a novel, unconventional approach to study protein folding from the statistical analysis of protein structures and make a surprising discovery that suggests the absence of specific amino-acid pairwise interactions in protein folding. If confirmed, these findings would challenge the currently accepted view, but more research is needed to clarify several outstanding issues.

Acknowledgements

RR and AV acknowledge the support of Marie Curie Actions through Marie Curie Excellence Grant MEXT-CT-2006-042334.

References