Comment

Insenstivity to Close Contacts and Inability to Predict Protein Foldability

http://www.jbsdonline.com

Mittal and coworkers recently (1) analyzed the spatial organization of folded proteins backbone and advocated an alternative hypothesis on protein folding stating that specific interactions among amino acids do not drive protein folding. The authors analyzed spatial distributions of $C_\alpha$ atoms for all amino acid pairs in about four thousands proteins available in PDB database and they found that the total number of contacts of an amino acid is correlated with the average occurrence of the amino acid. Their analysis led to a conclusion that there is no preferential neighborhood for an amino acid.

The correlation that they present between the amino acids occurrence in folded proteins and the number of total contacts is not surprising, because it is trivial. Their finding that there is not any preferential neighborhood for an amino acid is, however, strange, unconventional and appears to be not entirely correct.

There are similar observations where amino-acid pairs were analyzed (2-5). From these studies one concludes that there are statistically significant preferences between amino acid pairs, in addition to specific spatial arrangements. Furthermore such preferences are extensively used to prepare knowledge-based potentials for successful prediction of protein structures (6, 7). The theoretical determination of protein structure and function by modeling and simulation, based on preferential contacts and interactions, has become routine (8, 9, and references therein).

The disagreement between these observations and the report by Mittal et al. (1) may indicate that the analysis of spatial distribution of $C_\alpha$ atoms does not provide enough information. The sigmoid curves, employed to analyze the numbers of contacts in relation to the distance, are insensitive to close contacts as the majority of the curve is defined by the long distance interactions. This can be seen in the poor interleaving of the sigmoids in close contacts where the error bars would be approximately in tens of percents. These errors would probably be even stronger for more interaction-specific amino acids like cysteine or charged residues. Therefore the analysis of Mittal et al. (1) is actually detecting the preferential interactions in close contacts even though they are saying that such preferences are not present.

Main reason for such a disagreement is probably hidden in the fact that most of the contacts are random and as such defined only by amino acid frequencies and only a small minority of contacts is spatially and energetically distinguishable as was shown previously by Berka et al. (5). This can also explain why only a small number of amino acids is “critical” for folding of a protein (10, 11); why only a minimal change of amino acid sequence with 88% sequence identity leads to a completely different fold (12).

Karel Berka*
Michal Otyepka*

Department of Physical Chemistry, Faculty of Science, Palacky University tr. 17. listopadu 12, 771 46 Olomouc, Czech Republic

*Corresponding Authors:
Karel Berka
Michal Otyepka
E-mail: karel.berka@upol.cz
michal.otyepka@upol.cz
Last but not least, the work of Mittal and coworkers (1) cannot, unfortunately, be used to predict protein foldability, nor the protein structure, as the standard deviations in their “Chargaff’s rules” are quite large and therefore not informative enough for any prediction.

References