Stoichiometry and Preferential Interaction: Two Components of the Principle for Protein Structure Organization

A paper recently published in *J Biomol Struct Dyn* by Mittal and coworkers proposed an interesting view that protein folding is driven by the stoichiometry of amino acids rather than by the “preferential interactions” between them (1). By analyzing about 3700 backbones of folded proteins, the authors argued that something like Chargaff’s rules exist in protein folding similarly as in the organization of DNA structures. The above conclusion was made by counting the numbers of contacts between two specific residues at a variable range of neighborhood distance and by showing that they all follow a similar sigmoid trend and the total number of contacts formed by one residue is excellently correlated with the percentage occurrence of that residue. The authors’ conclusion highlights the importance of amino acid composition in the determination of protein folding as well as in the organization of the folded protein structures. This is generally true and in line with some previous reports where the correlation between protein folding rates and the amino acid composition was found as high as 0.7 (2) and the prediction accuracy of protein folding type (two-state or multi-state folding) based merely on the occurrence numbers of amino acids could be more than 80% (3). Considering the limited information (only the frequencies of amino acids) used in the predictions, the achieved accuracy is surprising. All these results emphasized the important roles played by the stoichiometry of amino acids in protein folding.

However, the above results are not a reason to deny the roles played by preferential interactions of amino acids in the organization of protein structures. A wealth of evidence showed that the protein functional conformations are accomplished by particular arrangement of amino acids. Firstly, the neighborhood occurrence of amino acids in the primary sequence is not random (4). Secondly, it is well known that amino acids preferentially occur in different types of secondary structures (α-helix, β-sheets and coils, etc.,) and the successful prediction of protein secondary structures just utilized this kind of information (5, 6). Thirdly, the stability of tertiary structures depends on the preferential interactions between specific types of residues such as aromatic, charged and hydrophobic residues for stacking, salt bridge and internal packing effects (7-10). Fourthly, the formation of quaternary structures owes to preferential contacts between amino acids in the subunit interface (11). Our recent work on the thermal stability of prokaryotic protein complexes also demonstrated the strategical use of charged residues in the subunit interface to adapt to an elevated environment temperature (12). A database called AAindex summarized the amino acid properties and many of the contained indexes or matrices involve the preferential interactions between amino acid residues (13).

Corresponding Authors:
Bin-Guang Ma
Hong-Yu Zhang
Email: mbg@mail.hzau.edu.cn
zhy630@mail.hzau.edu.cn

National Key Laboratory of Crop Genetic Improvement, College of Life Science and Technology, Huazhong Agricultural University, Wuhan 430070, P. R. China
Actually, the authors’ methodology has complicated the scenario by a sigmoid fitting. To show if there are preferential contacting patterns between residues, some direct statistical tests (for example, χ^2 test) are proper to check the deviation of amino acid pairing (contacting) frequencies from a random uniformity. Seen from the Figures 2 and 3 of the paper, the numbers of contacts between one particular residue and the other 20 (including itself) residues are not uniformly distributed. If χ^2 tests were used, the results might be different. On the other hand, it has to be noticed that whatever the result is, it’s just statistical, meaning that it is an averaged consequence over different types of protein classes, folds, secondary structure contents and primary sequences. However, for the functional conformation of each particular protein, the amino acids are arranged in a strategical or even subtle manner, far from promiscuous pairing. Universality obtained from averaged diversity doesn’t mean it holds for every single molecule individually, which just shows the limitation of statistical approach.

Considering the roles of amino acid composition in the determination of protein folding rates (2), folding types (3) and protein-protein interactions (14), the importance of the stoichiometry information could never be neglected. The authors’ finding of universal backbone spatial organization just enriched the store of evidence for this. Even with this finding, the “grand challenge” of accurate ab initio prediction for protein folding trajectory and final structure from the sole primary sequence still remains and cannot be solved all of a sudden. In their paper, the authors stressed the principle of minimizing the surface-to-volume ratio in protein folding. Although this principle might be true (at least for some globular proteins), the picture given by the authors that protein folding looks like fitting “Lego Blocks” while precluding any preferential interactions between these blocks is unrealistic. The authors mentioned in the last section of their conclusion that “This (protein folding, as we understand) must be done while satisfying the structural constrains of the primary sequence composition and constitution (i.e., the order in which a given stoichiometry of amino acids appear)”.

Actually, as suggested earlier, the primary sequence can only provide the two sorts of information: composition and permutation (or in the authors’ term, constitution) (15, 16). The permutation information is in fact a manifestation of the preferential amino acid interaction (adjacency) in the primary sequence. Unfortunately, in the given picture of protein folding, the authors missed this point (constitution constraint) that they just mentioned and arrived at a biased conclusion of stoichiometry driving protein folding without the contribution from preferential amino acid interactions. Considering the roles of amino acid composition in determining protein folding behaviors (2, 3) and protein-protein interactions (14) together with the importance of preferential amino acid interactions in the organization of different levels of protein structures (4-12, 17, 18), we can conclude that stoichiometry and preferential interactions of amino acids are two indispensable components of the principle for protein folding and structure organization. Such a conclusion makes a lot of sense because the preferential interaction between amino acids is the basis for the development of knowledge-based potentials, which in turn form the underpinning of modern day protein structure prediction by modeling and simulation (19-22 and references therein).

This work was supported by the National Key Project for Basic Research (2010CB126100) and the National Natural Science Foundation of China (30870520).

References