Journal of Biomolecular Structure & Dynamics ISSN 0739-1102
Volume 28, Issue Number 4, February 2011

Comment

Is Stoichiometry-Driven Protein Folding Getting Out of Thermodynamic Control?

http://www.jbsdonline.com

Xing-Lai Ji[1,3]
Shu-Qun Liu[1,2,3*]

[1]Laboratory for Conservation and Utilization of Bio-Resources & Key Laboratory for Microbial Resources of the Ministry of Education, Yunnan University, Kunming 650091, P. R. China
[2]National Laboratory of Macromolecules, Institute of Biophysics, Chinese Academy of Science, Beijing 100101, P. R. China
[3] Sino-duomedial and Information Engineering School, Northeastern University, Shenyang 110003, P. R. China

Understanding the mechanism by which monomeric proteins fold under *in vitro* conditions is fundamental to describing their functions at molecular level. Significant advances in theory, experiment and simulation have been achieved (1), making it possible to solve the three mostly focused aspects of protein folding problems (2):

(i) The thermodynamic question of how a native structure results from inter-atomic forces acting on an amino acid sequence - the folding code;
(ii) The kinetic problem of how a native structure can fold so fast - the folding rate;
(iii) The computational problem of how to predict the native structure of a protein from its amino acid sequence – the protein structure prediction.

The views on protein folding have evolved from simple force-driven folding (3), *i.e.,* the sum of many different small interactions (such as van der Waals interactions, hydrophobic interactions, hydrogen bonds, electrostatic interactions and ion pairs), to complex, free energy-driven "folding funnel" model (4) based on the energy landscape theory of protein folding (5). The latter is essentially a thermodynamically controlled process and emphasizes that folding is driven by complex balance of enthalpy and entropy leading to global free energy minimum for the protein-solvent system, rather than by simple optimization of inter-atomic forces only within the protein.

A recent article in this *Journal* (6) by Mittal *et al*. presents interesting statistical results based on 3718 folded protein structures, showing a simple principle of backbone organization of protein structure, which is interpreted as Chargaff's Rules related to protein folding, *i.e.,* a stoichiometry-driven protein folding. One of the interesting finding is that the total number of possible contacts for C_α of a given amino acid correlates excellently with its occurrence percentage in primary sequences, leading the authors to conclude that protein folding is a direct consequence of a narrow band of stoichiometric occurrences of amino-acids in primary sequences, regardless of the size and the fold of a protein (6). However, if this is true, what is the mechanism by which the percentage occurrences of amino acids determine protein folding? The authors do not answer definitely this question, although the folding manner of "exclusion by water" to minimize the surface-to-volume ratio, which is essentially equal to hydrophobic collapse hypothesis (7) for global protein folding, is proposed to relate the stoichiometric occurrences of amino acids to protein folding. It is hard to understand how the shape characteristics of individual residues can minimize the surface-to-volume ratio through constraints imposed by amino acid occurrence frequencies. One possibility is that

*Corresponding Author:
Shu-Qun Liu
Phone: +86 871 5035257
Fax: +86 871 5034838
E-mail: shuqunliu@ynu.edu.cn
shuqunliu@gmail.com

the higher occurrence frequency an amino acid has, the more probably it will occupy the core of a folded protein? Further work is needed to examine the relationship between the burial extent of amino acids and their occurrence percentages.

The "n" and "k" values for all the neighborhood sigmoids are independent of the occurrence percentages of amino acids, suggesting that there is no preferential interaction in short and medium distance ranges. Accompanying the absence of long range interactions, the authors concluded that the "preferential interactions" between amino-acids do not drive protein folding (6). It must be mentioned that preferential interactions between amino acids were the basis for introducing knowledge-based potentials, which in turn provided the underpinning for present day 3D protein structure prediction by modeling and simulation (8-10 and references therein). However, the authors' (6) conclusion of lack of preferential interaction between amino acids is drawn from analyses of 3718 already folded protein crystal structures. Therefore, it is easy to conclude the overall structural stability of already folded proteins can be maintained by the non-preferential/random inter-residue interactions, but the question that to what extent the non-preferential/random interactions contribute to the forces that drive protein folding is hard to answer based on current data.

The process of folding of a polypeptide chain, either newly synthesized from mRNA or denatured/unfolded from its native state, must be driven by certain forces such as hydrophobic side-chain interaction (11) or backbone hydrogen-bonding interaction (12). Interestingly, the hydrogen bond between protein backbone >C=O···H-N< has a potential to form between any two amino acids and can be considered as non-preferential interactions. Rose and colleagues (12) have recently proposed that the energetics of the backbone hydrogen bonds dominates the folding process, supporting the role of non-preferential interactions in driving protein folding. Nevertheless, the interactions between side-chain groups of two amino acids can be considered preferential because different amino acids have distinct side chains. At first glance, the "preferential interaction" of a given amino acid with another amino acid seems to come from specifically favorable side chain contacts such as hydrophobic stacking, hydrogen bonding (side chain-side chain hydrogen bond and side chain-backbone hydrogen bond) and electrostatic interactions. However, a further deep-thinking reveals that the so-called "preferential interaction" is to a large extent the consequence of protein desolvation effect (solute exclusion by water as mentioned in (6)) rather than specifically favorable side chain contacts. When an unfolded polypeptide chain interacts with the aqueous solvent under physiological conditions, the massive water molecules will exclude and squeeze the polypeptide to bring about the hydrophobic collapse (7, 13, 14). Upon collapse, the number of hydrophobic side chains exposed to water is minimized and the entropy of the solvent is maximized, thus lowering the total free energy of the protein-solvent system. Therefore the burial and packing of hydrophobic side chains, as the consequence of the hydrophobic collapse, increase the probability of observing hydrophobic interactions. Furthermore, it is interesting to note that not all polar or electrostatically charged side chains/groups are exposed to water, many of which can be buried inevitably in the interior of the folded structures. It has been suggested that the loss of stability by burying polar or charged groups can be gained back through forming hydrogen bonds (side chain-side chain hydrogen bonds or side chain-backbone hydrogen bonds) or salt bridges within the protein interior (15). The strength of hydrogen bonds depends on their environment; and therefore hydrogen bonds enveloped in a protein interior contribute more than those exposed to the aqueous environment to the stability of the native structure (16). This has led to the proposal that the protein folding is associated with a systematic desolvation of hydrogen bonds by surrounding hydrophobic groups (17). At stages after hydrophobic collapse but before reaching the native state, *i.e.,* the molten globular state (18) and the glass transition state (5, 19), further conformational rearrangements, which are obtained through favorable energetic contacts or preferential interactions between certain groups, are required to further lower the free energy of the protein-solvent system. The interactions between surface-exposed polar side chains and water molecules are not a negligible contributor to energetic enthalpy term of free energy change. Conclusively, the process of protein folding, which is driven by decrease in total free energy, is dictated by a delicate balance of the mechanisms of opposing effects involving entropic and/or enthalpic contribution.

For the folding process of an individual protein, the favorable/preferential interaction between any two amino acids would undoubtedly contribute to the enthalpic term of the free energy. If such a preferential interaction is "correct" (which means that the interaction is preserved in the final native structure), it contributes really to lowering free energy; if such a preferential interaction is "incorrect" (which means that the interaction is not presented in the final native structure), it contributes to the "trapped" free energy in the folding funnel of energy landscape. The entropic effect from solute and solvent and the competitive interactions (enthalpic effect) can help the protein jump out the "trap".

The statistical absence of preferential interaction between amino acids can be explained. On the one hand, the result is based on a large sample set of folded structures and therefore the simple count of number of C_α contacts within varied neighborhood distances would shield preferential interactions between side chain groups of amino acids. On the other hand, it is possible that the contacts between any two

amino acids, regardless of hydrophobic or hydrogen bonding/electrostatic interactions, can satisfy to some extent the requirement of lowering free energy during protein folding. Therefore, the observation of lack of preferential interactions can only be considered as a consequence of protein folding — a process that is driven by combined effect of enthalpy and entropy of the system — rather than the cause of protein folding.

Elucidating the folding mechanism is crucial for development of effective protein structure prediction methods, which in turn will improve our understanding of protein structure-function relationship and facilitate drug discovery and development. A very recent work by Sasisekharan and coworkers (20) reveals that the folding code is actually a network of inter-atomic interactions within the core regions of protein domains, and that the application of such a network signature to structure prediction has achieved great successes (for details, see (20)). This work also shows that each protein fold family has its own unique protein core atomic interaction network (PCAIN), implying that there must be preferential inter-atomic interactions. Such specific PCAIN is also the consequence of thermodynamically controlled folding process, and is not contradictory with the statistical result of the lack of preferential inter-residue interactions found by Mittal *et al.* (6) because most of the so-called preferential inter-atomic interactions can be observed between any two residues when the sample set is large enough.

In summary, we conclude that:

(i) The statistical method used by Mittal *et al*. is not sensitive to identify preferential/specific inter-atomic interactions.
(ii) The statistical phenomena observed in this work are the consequences of thermodynamics-driven folding rather than the driving force of protein folding.
(iii) It seems impossible to apply the "stoichiometry-driven" folding principle to protein structure prediction unless stoichiometric occurrences of residues can be translated into position constraint information.

References

1. D. Thirumalai, E. P. O'Brien, G. Morrison, and C. Hyeon. *Annu Rev Biophys 39*, 159-183 (2010).
2. K. A. Dill, S. B. Ozkan, T. R. Weikl, J. D. Chodera, and V. A. Voelz. *Curr Opin Struct Biol 17*, 342-346 (2007).
3. C. B. Anfinsen and H. A. Scheraga. *Adv Protein Chem 29*, 205-300 (1975).
4. P. E. Leopold, M. Montal, and J. N. Onuchic. *Proc Natl Acad Sci USA 89*, 8721-8725 (1992).
5. J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes. *Ann Rev Phys Chem 48*, 545-600 (1997).
6. A. Mittal, B. Jayaram, S. Shenoy, and T. S. Bawa. *J Biomol Struct Dyn 28*, 133-142 (2010).
7. P. Sklenovský and M. Otyepka. *J Biomol Struct Dyn 27*, 521-539 (2010).
8. M. J. Aman, H. Karauzum, M. G. Bowden, and T. L. Nguyen. *J Biomol Struct Dyn 28*, 1-12 (2010).
9. Y. Tao, Z. H. Rao, and S. Q. Liu. *J Biomol Struct Dyn 28*, 143-157 (2010).
10. V. R. Agashe, M. C. Shastry, and J. B. Udgaonkar. *Nature 377*, 754-757 (1995).
11. W. Kauzmann. *Adv Protein Chem 14*, 1-63 (1959).
12. G. D. Rose, P. J. Fleming, J. R. Banavar, and A. Maritan. *Proc Natl Acad Sci USA 103*, 16623-16633 (2006).
13. M. Brylinski, L. Konieczny, and I. Roterman. *Comput Biol Chem 30*, 255-267 (2006).
14. M. S. Cheung, A. E. Garcia, and J. N. Onuchic. *Proc Natl Acad Sci USA* 99, 685-690 (2002).
15. C. N. Pace, B. A. Shirley, M. Mcnutt, and K. Gajiwala. *FASEB J 10*, 75-83 (1996).
16. S. Deechongkit, H. Nguyen, E. T. Powers, P. E. Dawson, M. Gruebele, and J. W. Kelly. *Nature 430*, 101-105 (2004).
17. A. Fernandez, T. R. Sosnick, and A. Colubri. *J Mol Biol 321*, 659-675 (2002).
18. M. Ohgushi and A. Wada, *FEBS Lett 164*, 21-24 (1983).
19. J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes. *Proteins 21*, 167-195 (1995).
20. V. Soundararajan, R. Raman, S. Raguram, V. Sasisekharan, and R. Sasisekharan. *Plos One* 5, e9391 (2010).