Is Side-Chain Entropy an Important Factor in Protein Folding?

Joshua Ballanco

November 11, 2008

Abstract

Protein folding is a process which involves thermodynamic contributions from forces which favor the folded state and other forces which favor the unfolded state. When the balance of these forces tips in favor of the folded state, a protein will fold spontaneously into a unique conformation representing a thermodynamic minimum. The difficulty in predicting just how, and when, a protein will fold stems from a fundamental lack of knowledge about the forces involved. This is especially true in the case of the conformational entropy of amino acid side-chains (side-chain entropy).

Because of the difficulty associated with either directly measuring side-chain entropy or calculating its value from molecular simulations, there is uncertainty as to the effect it has on protein folding. The side-chains of amino acids range in size and flexibility, and some of the longer side-chains contain polar or even charged groups. It's been proposed that side-chains might counteract a loss of entropy by a gaining interactions with other amino acids. Additionally, the hydrophobic residues commonly found in the core of a folded protein are not the most flexible, and so they will not be affected as significantly by the corresponding space constraints.

On the other hand, buried residues will loose some degree of side-chain entropy. Even though the cores of proteins rarely represent a closest-packed configuration, the side chains of long, charged residues found in internal salt bridges will undoubtedly be restricted in their motion. Furthermore, it is now understood that what is commonly refered to as the "native protein structure" is actually an ensemble of closely related structures. It is possible that side-chain entropy may not be important in determining the global fold of a protein, but may be vital in understanding the dynamics associated with function.

In this paper I will look at a variety of experiments which approach the question of side-chain entropy from differing perspectives. The first series of approaches involves investigating the structure and sequence of natural proteins, specifically focusing on side-chain entropy. The alternative approach is to perform *de novo* protein design either using or neglecting side-chain entropy in the calculations. Both of these approaches are, necessarily, indirect, and each leads to a different conclusion. I will look at the results in context, and discuss whether or not a real controversy exists. Finally, I will propose a course of investigation which might shed more light on the role of side-chain entropy in protein folding.

Topics

Thermodynamics

Protein Folding

Statistical Mechanics

Molecular Dynamics

X-ray Crystallography

de novo Protein Design

Introduction

Proteins are the work-horses of biology. They are responsible for the vast majority of the biochemical processes that take place in living organisms, and consequently there is keen interest in understanding how they function. Normally, this would be a question of brute-force experimentation, approaching and investigating proteins one at a time. Two aspects of proteins, however, make this a problem that is tantalizingly close, and yet still frustratingly far, from a complete solution.

The first useful aspect of proteins is the relationship between their structure and their function. Knowing a protein's structure goes a long way to understanding how it might function. This is particularly important when developing pharmaceutical agents to treat disease. Second, a protein's structure is completely determined by its amino acid sequence, which can in turn be determined from available genetic data. This was originally shown in 1961 by Anfinsen [6] in his work with ribonuclease. This notion, of 3-dimensional structure data encoded in a 1-dimensional array of elements, has been verified for proteins in all but a few rare cases.

Thus, since knowing a protein's structure was already more than half the battle, and because it should be possible to determine that structure with readily available information, work began almost immediately on elucidating the mechanism drives proteins to form their unique structures. This work has been ongoing for the 47-years since.

Shortly after Anfinsen's discovery, Brandts proposed a two-state model for protein folding where a protein exists in an equilibrium, $D \rightleftarrows N$, between the denatured and native state [3]. This two-state model of protein folding is still the prevailing accepted model for folding, and it has some interesting implications. One of the first implications was recognized by Levinthal in 1968, and has since become known as Levinthal's paradox.

Essentially, what Levinthal realized was that Brandts' two-state model implied an absence of long-lived or marginally stable conformations. In other words, the denatured protein, which is essentially a random coil, must find the single lowest energy conformation from all of the possible conformations in one step. Levinthal's rough calculations implied that a robust search of all possible conformations would require more than the age of the universe to complete yet, paradoxically, proteins fold on a millisecond to second time scale. Obviously, proteins do not randomly sample conformation space. Instead, there must be a driving force that guides a denatured protein, no matter its conformation, toward the native state.

This understanding of Levinthal's Paradox lead Dill to propose the concept of the folding energy landscape as a funnel [4], which has become an iconic symbol of the protein folding problem. The funnel captures the essence of the relationship between energy and conformational space. At high enough energies, proteins are free to explore all conformations, and are therefore "denatured" (in fact, the "denatured" state allows for some fraction of protein molecules to explore a native or native-like conformation, but this fraction is small enough to be inconsequential). As the energy of the protein is lowered, molecules with conformations near the center of the funnel will simply decrease in energy. However, those along the sides of the funnel will be steered toward more and more native-like conformations. In this way, Levinthal's paradox is resolved since all the molecules of a protein, regardless of their starting conformation, will converge on a point in conformation space as the energy is lowered.

Dill's funnel also captures another important aspect of protein folding: entropy. At any given energy, the volume of conformation space which remains inside the funnel corresponds to the conformational entropy of the protein. As a protein works its way down the ever narrower funnel, its entropy is decreasing. In order to offset this effect, the folded state of a protein must contain a host of favorable interactions. These will be the same sorts of interactions which define the funnel in the first place, and thus the problems of determining protein folding pathways, protein unfolding pathways, and native structures are all linked.

Ultimately, the question of entropy's effect on protein folding is a rather complicated one. Entropy is, unlike enthalpy or potential energy, an ensemble property. That is, it cannot be determined by looking at a single folded molecule, but rather depends on how many different folds are possible for a given molecule at

a given energy [5]. This poses a significant problem in the calculation of entropy from molecular dynamics experiments. An exact calculation of the entropy of folding ($\Delta\mathbf{S_{fold}}$) would require not only enumerating all of the possible native-like conformations accessible to a folded protein, but also to enumerating over all of the possible denatured protein conformations. While the former is merely an extremely computationally intensive problem, the later is intractable. A number of techniques have been developed to provide good estimates for this value, but even some of these techniques are prohibitively difficult.

Entropy is also not very easy to separate into component contributions in the same manner as enthalpy or potential energy [1]. Where two hydrogen bonds in a protein would be expected to have twice the stability of a single hydrogen bond, the same cannot be said for the entropy of two side-chains. This makes it difficult to attack the problem of side-chain entropy with a piecemeal approach without making certain assumptions. This also means that directly measuring the contribution of one component of a protein to the protein's entropy via experimental methods is practically impossible.

In particular, free energy, a term which captures both the potential energy of a protein and its entropy, depends on quadratic and higher order terms of the potential energy [2]. If we can integrate the system from absolute zero to a given temperature, a technique known as Thermodynamic Integration, then these higher order terms become temperature derivatives of singular potential energy contributions. In this way, it is possible to separate free energy into contributory components, but only along certain informative paths.

To fully understand the contribution of entropy, specifically side-chain entropy, to the process of protein folding creative approaches are needed. Below, I will look at two different, yet complementary, approaches to this question. As noted above, the forces that govern the process of protein folding are the same forces that determine the native structure of a given sequence of amino acids. If we look at the amount of side-chain entropy in native protein structures, we should be able extrapolate the relative importance of side-chain entropy in the folding process. This can be accomplished either by calculation of the side-chain entropy for known structures, or by analyzing proteomic sequences with an eye toward an amino acid bias toward side-chains with more or less degrees of freedom.

Another means of assessing how much side-chain entropy participates in protein folding is by recreating, in a sense, protein evolution. Evolution is restricted by the physics of protein folding, and so it will favor certain amino acid sequences over others because they contain the right combination of physical properties to fold to the needed structure. By attempting to design proteins from scratch, we can investigate which aspects of amino acid physics are important by either including or excluding them in the selection process. If the design algorithms come to a solution which is close to the one resulting from natural evolution, we can assume that the algorithm contains the physical properties important for folding.

Side-Chain Entropy of X-ray Structures

One way of approaching the question of side-chain entropy is to look at a collection of existing protein structures.

- Review logic
- Outline methods
- Summarize conclusions

Side-Chain Entropy in Protein Design

Another approach to the question of side-chain entropy, and one that has only become feasible with recent advances, is to look at its role in the *de novo* design of proteins.

- Review logic
- Background on method
- Summarize conclusions
- Compare to new paper with contradictory results

Is Side-Chain Entropy Important?

What can we conclude about the importance of side-chain entropy in determining the folded structure of a protein?

- Design vs Structure
- Not yes or no, but how much?
- Both important and unimportant simultaneously

How Important Is Side-Chain Entropy?

With the possibility that side-chain entropy may be important for some aspects of protein folding and not for others, it is vital that we understand better the roles that side-chain entropy plays.

- Mechanical experiment for validation
- New techniques for simulating proteins

References

[1] G P Brady and K A Sharp. Entropy in protein folding and in protein-protein interactions. *Curr Opin Struct Biol*, 7(2):215–21, Apr 1997.

[2] G P Brady, A Szabo, and K A Sharp. On the decomposition of free energies. *J Mol Biol*, 263(2):123–5, Oct 1996.

[3] A Clay Clark. Protein folding: are we there yet? *Arch Biochem Biophys*, 469(1):1–3, Jan 2008.

[4] Ken A Dill, S Banu Ozkan, M Scott Shell, and Thomas R Weikl. The protein folding problem. *Annual review of biophysics*, 37:289–316, Jan 2008.

[5] Hagai Meirovitch. Recent developments in methodologies for calculating the entropy and free energy of biological systems by computer simulation. *Curr Opin Struct Biol*, 17(2):181–6, Apr 2007.

[6] Robert F Service. Problem solved* (*sort of). *Science*, 321(5890):784–6, Aug 2008.