Free energy landscape analysis of protein folding dynamics

Polina Banushkina and Sergei Krivov

Introduction

Understanding how proteins fold to their native state remains a problem of fundamental interest in biology, in spite of the fact that it has been studied for many years. While, the general principles of protein folding have been established, much controversy remains on fundamental topics such as: the nature of folding steps, the height of folding barriers and the value of the pre-exponential factor, the diversity of folding pathways, and the importance of residual structure in the denatured state. Moreover, now that misfolding has been shown to be the source of a range of diseases, a detailed understanding of what determines whether a polypeptide chain will fold to its native state or aggregate has become all the more important. In principle, this questions can be answered rigorously by determining the protein folding free energy landscape - the fundamental determinant of protein folding (and any other) reaction. However, in spite of their fundamental importance the quantitatively accurate free energy landscapes of proteins are yet to be determined.

State of the art experimental techniques lack the necessary spatial and temporal resolution for determination of such landscapes. Properties of the landscapes can be probed only indirectly. Simulation, in principle, can provide high spatial and temporal resolution, necessary for the determination of the quantitatively accurate free energy landscapes. Recently, due to advances in the hardware and simulation methodology realistic simulation of folding of small fast-folding proteins became computationally affordable. Notably, D.E. Shaw and co-workers reported realistic folding-unfolding simulations for a number of fast folding proteins with up to 100 residues in size. Which brings an exciting opportunity to perform detailed rigorous analysis of protein folding dynamics and to resolve the controversial issues by determining accurate free energy landscapes.

Quantitative analysis of protein dynamics in terms of the free energy landscapes is notoriously difficult. A poorly chosen reaction coordinate may hide the complexity of the free energy landscape and associated dynamics. Many approaches, though being based on solid physical intuition, often construct sub-optimal reaction coordinates. Recently we have developed a rigorous approach to construct optimal reaction coordinates and the high resolution free energy landscapes which provide accurate description of the dynamics.

Results

After the publication of the first realistic protein folding simulation, namely that of FIP35, by D.E. Shaw and co-workers we applied our rigorous methods to reanalyse the trajectory. It was found (see Figure 1) that the coordinate used by Shaw *et al.* is sub-optimal and the associated free energy landscape does not provide accurate description of the folding dynamics. In particular, we found that FIP35 is not a "barrier-less" folder but folds via a populated on-pathway intermediate separated by high free energy barriers; the high free energy barriers rather than landscape roughness are a major determinant of the rates for conformational transitions; and that the pre-exponential factor for folding kinetics $1/k_0$ is 10 ns rather than 1μs. While the latter value is generally accepted and is supported by a large body of indirect experimental evidences, the former is a first estimate obtained by a rigorous analysis in a direct manner.

The existence of the multiple approaches for reaction coordinate construction, naturally, poses the question of how different reaction coordinates can be compared. Which reaction coordinate provides better description of the dynamical process when different coordinates

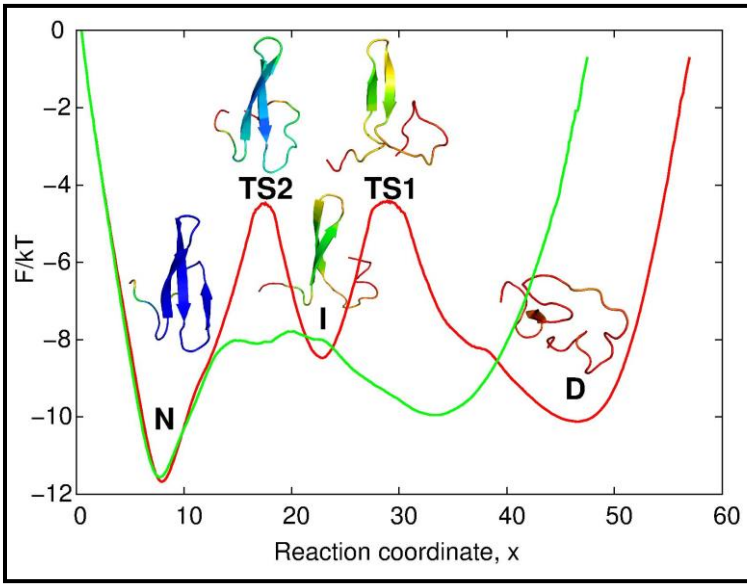

Figure 1: the free energy profiles of FIP35 protein determined by us (red) and by Shaw *et al* (green). Structures show the average conformations in configuration space region. Colors code the root mean square fluctuations of atomic positions around the average structure from 1.5 Å (blue) to 7 Å (red).

lead to different descriptions? A related question is whether one can establish that a putative optimal reaction coordinate is indeed the optimal one? We have developed a new fundamental criterion which is easy to apply. Reaction coordinate is optimal if its cut free energy profile, determined using length-weighted transitions, is constant, i.e., it is position and sampling interval independent. The observation leads to a number of interesting results. In particular, the equilibrium flux between two boundary states can be computed exactly as diffusion on a free energy profile associated with the coordinate for any equilibrium Markov process. It means that kinetics of protein folding on whatever complex free energy landscape can be computed as diffusion along the optimal reaction coordinate. The mean square displacement, for the trajectory projected onto the coordinate, grows linear with time. That for the same trajectory, projected onto a suboptimal coordinate, grows slower than linear with time, indicating sub-diffusion. The criterion showed that the coordinate used in the analysis of FIP35 is optimal, while that of Shaw is suboptimal.

Recently, the method was applied to (re)analyse simulation of all-helical protein HP35 (wild-type and mutant) to see whether the results found for all-beta FIP35 are transferable. In particular, the reaction coordinate and associate free energy landscape, determined with our method, provide quantitatively accurate description of the folding dynamics which is in good agreement with available experimental data. The pre-exponential factor is about $1/k_0 \sim 20$ns, in agreement with our previous estimate.

Funding
This work was supported in part by an RCUK fellowship and a BBSRC grant.