Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Structural Biology

Protein folding: from theory to practice

D Thirumalai[1,2], Zhenxing Liu[3], Edward P O'Brien[4] and Govardhan Reddy[1]

A quantitative theory of protein folding should make testable predictions using theoretical models and simulations performed under conditions that closely mimic those used in experiments. Typically, in laboratory experiments folding or unfolding is initiated using denaturants or external mechanical force, whereas theories and simulations use temperature as the control parameter, thus making it difficult to make direct comparisons with experiments. The molecular transfer model (MTM), which incorporates environmental changes using measured quantities in molecular simulations, overcomes these difficulties. Predictions of the folding thermodynamics and kinetics of a number of proteins using MTM simulations are in remarkable agreement with experiments. The MTM and all atom simulations demonstrating the presence of dry globules represent major advances in the proteins folding field.

Addresses
[1] Biophysics Program, Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, United States
[2] Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, United States
[3] Department of Physics, Beijing Normal University, Beijing 100875, China
[4] Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK

Corresponding author: Thirumalai, D (thirum@umd.edu)

Current Opinion in Structural Biology 2013, **23**:22–29

This review comes from a themed issue on **Folding and binding**

Edited by **Jayant Udgaonkar** and **Susan Marqusee**

For a complete overview see the Issue and the Editorial

Available online 21st December 2012

0959-440X/$ – see front matter,

http://dx.doi.org/10.1016/j.sbi.2012.11.010

Introduction

A general framework based on polymer theory and statistical mechanics is in place to describe how proteins fold [1,2,3••,4], which is a central problem in molecular biology. The resulting perspective has produced a number of general predictions, several of which have found experimental support. The modern perspective of how proteins fold has played a pivotal role in establishing concepts such as the nucleation-collapse mechanism [5], the speed limit for folding [6], the thermodynamic characteristics that set apart evolved and random sequences [7], and minimum energy compact structures in directing folding [8,9], scaling of folding rates as a function of protein size [10], and the kinetic partitioning mechanism [5,11,12••] in topologically frustrated systems. Although answers to questions of generality have been largely settled, translating these ideas into practice remains a daunting task. In particular, as the protein folding field has matured it has become ever more important to develop theories and computational tools for predicting the outcomes of experiments under conditions that closely mimic those used in the laboratory. For example, majority of the experiments use the response of proteins to denaturants as a way to measure folding thermodynamics and kinetics [13] whereas most simulations are performed using temperature to control the folding reaction [14••]. The unfolding of proteins by small chemical compounds, such as urea and guanidinium chloride (GdmCl), has formed the foundation for numerous biophysical studies of *in vitro* protein folding over the past several decades, thus making it vital to provide a framework for describing their action on the folding reaction. More recently, single molecule pulling experiments have provided fundamental insights into the folding process by probing the response of proteins to the applied mechanical force either in an AFM [15] or laser optical tweezer (LOT) setup [12••]. Both well established experimental methods and advanced single molecule techniques [16••] require new theoretical and simulation tools, which can translate the theoretical ideas into practice.

Our review focuses on the substantial progress made in the inclusion of denaturants in simulations, which has for the first time, shown that theory can be used to directly predict the outcomes of experiments. The most obvious way to study the effects of denaturants on protein folding is to perform all atom molecular dynamics simulations in aqueous denaturant solutions. Although there have been impressive efforts to harness the power of computers (special purpose molecular dynamics machines and Graphic Processor Units) to generate folding trajectories of small single domain proteins at finite temperatures [17••,18–20] there has been little work on simulations of proteins in the presence of commonly used denaturants [21,22••], such as urea and GdmCl. As a consequence there is still a gap between experiments and simulations. In contrast, there have been significant developments that have combined coarse grained (CG) models for proteins [14••,23] and phenomenological ways of including the effects of denaturants and osmolytes [24] to predict folding kinetics and pathways of small and large proteins [25,26••]. Here, we illustrate that a combination of all atom and CG simulations provide not only testable predictions but also yield great insights into the nature of states that are sampled during the folding process. These developments with applications that have

included pH effects in mechanical unfolding experiments [27] show that quantitative insights into the folding process can be obtained by using a combination of modern developments in simulations and older ideas based on the transfer model.

Denaturant effects on protein folding thermodynamics

Despite significant progress in producing folding trajectories using all atom molecular dynamics simulations for peptides [28] to small proteins [17••] obtaining reliable thermodynamic properties (heat capacity curves or effect of denaturants on melting profiles) of even small proteins using detailed simulations is difficult because of inaccuracies in the force fields. To overcome some of the inherent difficulties in all atom models we [23] and others [14••] have introduced a genre of CG models, which have been used with remarkable success in the study of systems spanning a wide range of length scales [29]. In the context of folding, the Self Organized Polymer model [30] with side chains (SOP-SC) [31••] has proved to be particularly efficacious in simulating denaturant and pH effects on a variety of proteins as well as the effect of force f on proteins [32].

The MTM is an approximate theory [33] that combines the ideas based on the transfer model with CG or all atom representation of the polypeptide chain to predict the dependence of folding as a function of denaturant concentration. The MTM has been successfully used to obtain insights into the effects of denaturants and osmolytes on protein L, Cold shock protein [25], SH3 domain [31••], and most recently GFP [26••], a protein with ~230 residues. Remarkably, the MTM-based simulations capture nearly quantitatively the measured changes in FRET efficiencies of protein L and cold shock protein as well as the folding thermodynamics and the associated m-values, quantifying the stability decrease of the folded state per unit molar increase in denaturant concentration $[C]$. As a concrete example, we show in Figure 1 the results obtained using MTM for the folding thermodynamics and kinetics of src-SH_3 (Figure 1a), a well studied protein using ensemble experiments [34]. The dependence of fraction of molecules in the native basin of attraction, $f_{NBA}([C])$, on the concentration of GdmCl calculated using simulations based on the MTM is in excellent agreement with the experiments (Figure 1b). The midpoint concentration, C_m = 2.5 M, also agrees with the experimental value of 2.6 M [34]. The stability of the folded state, N, with respect to the unfolded state, U, $\Delta G_{NU}([C])(G_N([C]) - G_U([C]))$, can be calculated using an approximate two-state fit to $f_{NBA}([C])$ leading to $\Delta G_{NU}([C]) = -RT_s \ln((f_{NBA}([C]))/1 - f_{NBA}([C])))$. From the linear fit $\Delta G_{NU}([C]) = \Delta G_{NU}([0]) + m[C]$ we obtain m = 1.34–1.47 kcal/(mol M), which is in excellent agreement with the experimentally inferred m values that are in the range 1.50–1.60 kcal/(mol M). Somewhat surprisingly, the dependence of f_{NBA} on the concentration of urea (green symbols in Figure 1b) has been measured only recently. The predicted m value for urea m = 0.92 kcal/(mol M) based on MTM simulations is in semiquantitative agreement with the unpublished experimental results (S. Marqusee, private communication). The results in Figure 1b show that m values for GdmCl are higher than urea indicating that GdmCl is more efficient in unfolding proteins than urea. Another prediction of the MTM concerns the dependence of the melting temperature on the concentration of GdmCl. The melting temperature of SH3 domain, $T_m([C])$, which is identified with the peak in C_v, decreases linearly (Figure 1c) as $[C]$ increases. Thus, all of the global features obtained from the MTM simulations are in near quantitative agreement with experiments.

Perhaps, the most significant triumph of the MTM theory is that it can be used to calculate the $[C]$-dependent folding (unfolding) rates from folding (unfolding) trajectories. For apparent two-state folders lnk_{obs} (k_{obs} is the sum of the folding ($k_f([C])$) and the unfolding ($k_u([C])$) rates) has the characteristic V (or Chevron) shape (Figure 1d). For the SH_3 domain $\ln k_{obs}$ as a function of $[C]$ over the concentration range ($0M \leq [C] \leq 6.5M$) of GdmCl shows a classic Chevron shape [34]. Comparison of the simulation and experimental results (filled black circles in Figure 1d) allows us to draw three major conclusions. (i) The slopes of the folding and unfolding arms of the simulated Chevron plot are surprisingly similar to the experimental values. (ii) Within error bars in simulations and experiments, we do not find any deviation from linearity in the Chevron plot. These results represent the first simulations [31••] that capture all of the experimental features, which is remarkable given the simplicity of the MTM. (iii) However, the calculated rates differ from experiments by about a factor of sixteen at $[C] = 0$, which is due to the neglect of nonnative interactions and explicit solvent effects in the simulations.

Plastic and brittle response to mechanical force

Increasingly sophisticated single molecule pulling experiments have been particularly useful in providing insights into the folding of proteins. In LOT experiments, a mechanical force, f, is applied indirectly (through handles and beads that are trapped by light) to the ends of the proteins, and the time dependent changes in the distance between the beads are recorded. It is assumed, with some justification provided by simulations and theory [35], that the distance changes between the beads accurately reflect the changes in the end-to-end distance (projected along the f axis) of the protein. If a constant f is used then the protein hops between various states, which in the pulling experiments correspond the differing extensions of the protein. Assuming that at a fixed f the system ergodically samples the allowed conformations of the protein then an equilibrium free energy profile $F(R)$ as a function of the