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].

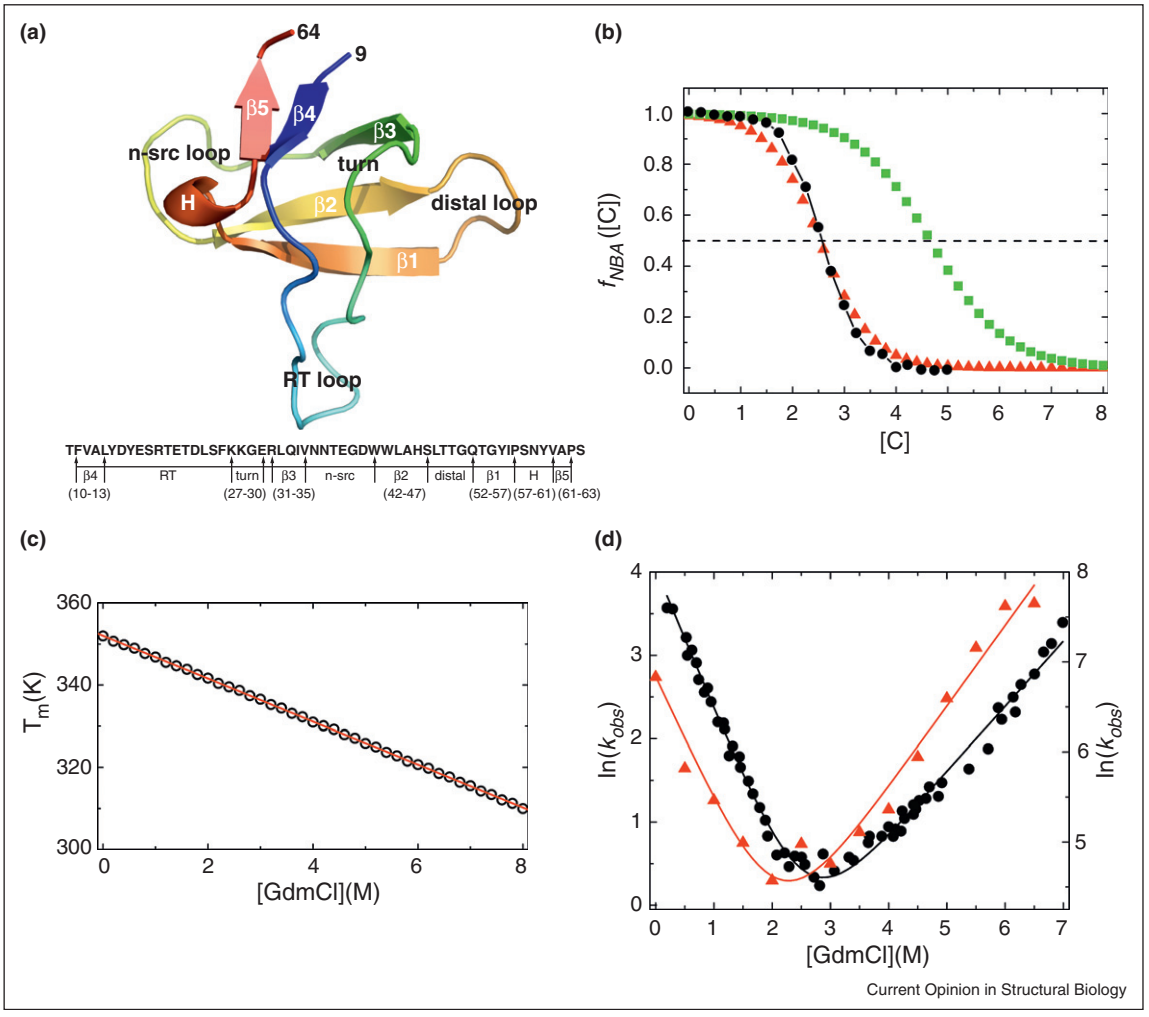
The MTM is an approximate theory [33] that combines the ideas based on the transfer model with CG or all atom representation of the polypeptide chain to predict the dependence of folding as a function of denaturant concentration. The MTM has been successfully used to obtain insights into the effects 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

Figure 1

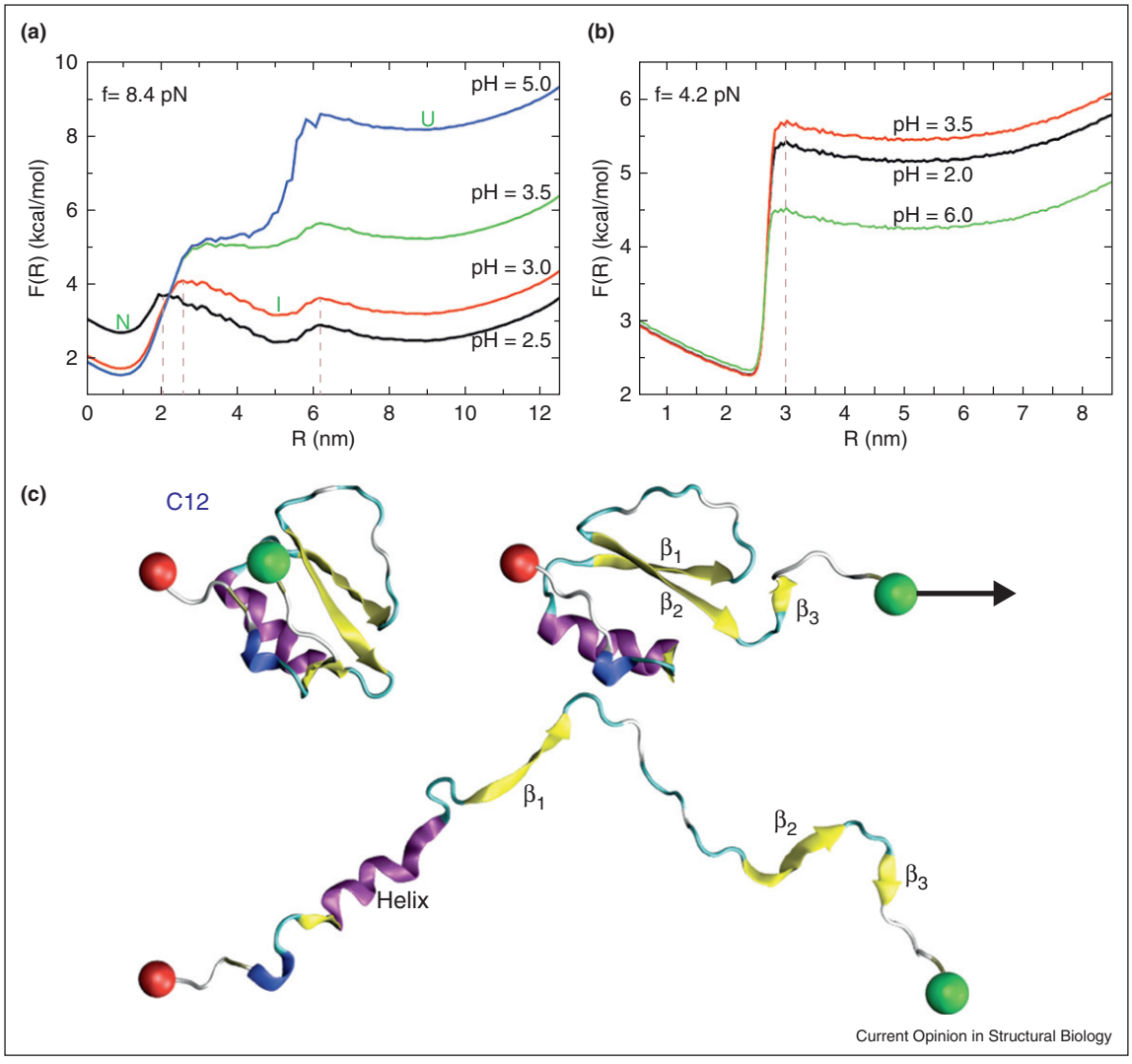


Denaturant-dependent folding thermodynamic and kinetics of src-SH_3 as a function of denaturant concentration. **(a)** Structure of src-SH_3 (PDB ID: 1SRL). The numbering of the protein sequence starts with 9 and the locations of the secondary structures along the sequence are given below the structure. **(b)** Fraction of src-SH_3 in the NBA, computed using simulations, as a function of GdmCl (red) and urea (green). The experimental data in black are for GdmCl [34]. **(c)** Melting temperature, $T_m([C])$, identified with the peak in the simulated specific heat as a function of Gdmcl concentration. The line is a fit ($T_m([C]) = 352 - 5.2[C]$) to the simulated data. **(d)** Comparison of simulated and measured Chevron plots in GdmCl. The simulation results are in red and the scale is on the right, and the experimental data are in black and the scale is on the left. The slopes of the folding and unfolding arms computed from the simulations are 0.95 kcal/(mol M) and 0.60 kcal/(mol M), respectively, which are in good agreement with the experimental values of 0.99 kcal/(mol M) and 0.45 kcal/(mol M), respectively.

extension R of the protein can be constructed from the mechanical folding trajectories. The utility of $F(R)$ in the folding problem depends very much on whether the experimentally accessible one dimensional variable R that is conjugate to the force is a reasonable reaction coordinate. To date, experiments have been performed at a fixed temperature at neutral pH and as such the response of proteins to f to varying environmental changes have not been probed. The versatility of the MTM is that the effects of pH can also be studied as illustrated in recent applications [27], which serve as predictions for future experiments. The simulations show that two small proteins, protein G and Chymotrypsin Inhibitor 2 (CI2), exhibit vastly different responses to f as pH is varied (Figure 2).

For CI2, $F(R)$ shows three basins of attraction over a range of pH values (Figure 2a), which implies that CI2 undergoes a two-stage force induced unfolding transition in which a partially folded state is populated between the fully folded and fully unfolded basins. The transition from the native to the intermediate basin (located between 2 and 6 nm corresponds to the unfolding of β-strand 3 resulting in the loss of tertiary interactions with β-strand 2 and the α-helix (Figure 2c). The transition to the unfolded basin (located at $R > 7$ nm) corresponds to

Figure 2



Effects of pH on the free energy profiles as a function of extension of proteins. **(a)** For CI2 at a constant force of 8.4 pN. **(b)** For protein G at 4.2 pN. The pH values are indicated in each figure, where R is the end-to-end distance extension of the protein. The $F(R)$ profiles show that protein G has two free energy basins corresponding to the NBA and the unfolded state whereas CI2 exhibits an additional intermediate at $R \approx 5$ nm as indicated in (a). Interestingly, the transition state does not move (dashed lines) for protein G but moves towards the least stable state for CI2 in accordance with the Hammond–Leffler postulate. **(c)** Representative structures of N, I, and U for CI2 obtained using MTM simulations are shown.

the rupture of the rest of the structural elements in the protein (i.e., β-strands 1–2, and interaction of these strands with the α-helix). In contrast, the pH-dependent free energy profiles for protein G has only two basins of attraction (Figure 2b). Force induced unfolding of protein G occurs in a single step. Remarkably, unlike CI2 the location of the transition state (TS) is invariant as pH or f is altered, which is the hallmark of a brittle material [36].

The variations in the TS locations as f or pH is changed are dramatically different for these proteins. According to the Hammond–Leffler postulate the transition state (TS) should resemble the least stable species in the reaction. Although originally proposed for reactions of small organic molecules, Hyeon and Thirumalai [37,38] showed that the Hammond–Leffler postulate is also applicable to force unfolding of biomolecules regardless of the nature of the reaction coordinate. For proteins under tension this implies that the location, R_{TS}, of the TS, should either be independent of f or move towards the native state when f increases. There are two transition states for CI2, one at $\Delta R_{N\text{-}TS}$ between the NBA and the intermediate state. The other between the intermediate and the UBA is at $\Delta R_{I\text{-}TS}$. Both $\Delta R_{N\text{-}TS}$ and $\Delta R_{I\text{-}TS}$ are independent of pH when pH exceeds 3.5. As pH increases, resulting in enhanced stability of both N with respect to I and I with respect to U, $\Delta R_{N\text{-}TS}$ and $\Delta R_{I\text{-}TS}$ increase with a dramatic jump at a pH = 3.0. The locations of the two transition

states move closer to the less stable species in accordance with the Hammond–Leffler postulate. The larger change in $\Delta R_{\text{I-TS}}$ compared to $\Delta R_{\text{N-TS}}$ as f changes is reminiscent of the plastic behavior of the I state in apo myoglobin at pH = 5.0 [39••]. As a corollary, we expect and find that upon an increase in temperature $\Delta R_{\text{N-TS}}$ and $\Delta R_{\text{I-TS}}$ should decrease as both the folded and intermediate states are destabilized relative to the unfolded state. In sharp contrast to CI2, the TS changes in the unfolding of protein G are dramatically different. The TS location $\Delta R_{\text{N-TS}}$ is independent of pH (Figure 2b), which implies that protein G behaves as a brittle material when subject to f at all pH values.

In single molecule pulling experiments, with R as the only experimentally accessible coordinate, it is unclear if the structures at the TS location correspond to the ensemble of TS structures in the multidimensional folding landscape. It is possible that when f exceeds a critical value the pulling coordinate is a good reaction coordinate because at large forces the molecule is likely to be aligned along the f direction, thus forcing it to unfold along the coordinate conjugate to f. From physical arguments the suitability of R as a reaction coordinate is determined by an interplay of compaction (determined by protein stability) and tensile (dependent on R_{TS} and the barrier to unfolding) forces [40], which is captured by the experimentally measurable molecular tensegrity parameter, $s = f_c/f_m$ [40] where the unfolding critical force, $f_c = \Delta G^{\ddagger}/R_{\text{TS}}$, and $\Delta G^{\ddagger}$ is the height of the free energy barrier. For CI2, with two transition states, the values of s_1 ($\text{N} \rightarrow \text{TS}$) and s_2 ($\text{I} \rightarrow \text{TS}_2$) are 0.019 and 0.005 at pH = 3.0. The theory of Morrison *et al.* [40] predicts that, at this pH, extension is a good reaction coordinate for both the transitions because it is likely that the ensemble of conformations starting from $\Delta R_{\text{N-TS}}$ ($\Delta R_{\text{I-TS}}$) would reach I and N (I and U) with equal probability ($p_{\text{fold}} \approx 0.5$). For protein G at pH = 6.0, $s = 0.058$, which also lies in the range for which R is likely to be a good reaction coordinate as assessed by the theory outlined in [40]. These predictions await future experiments and simulations.

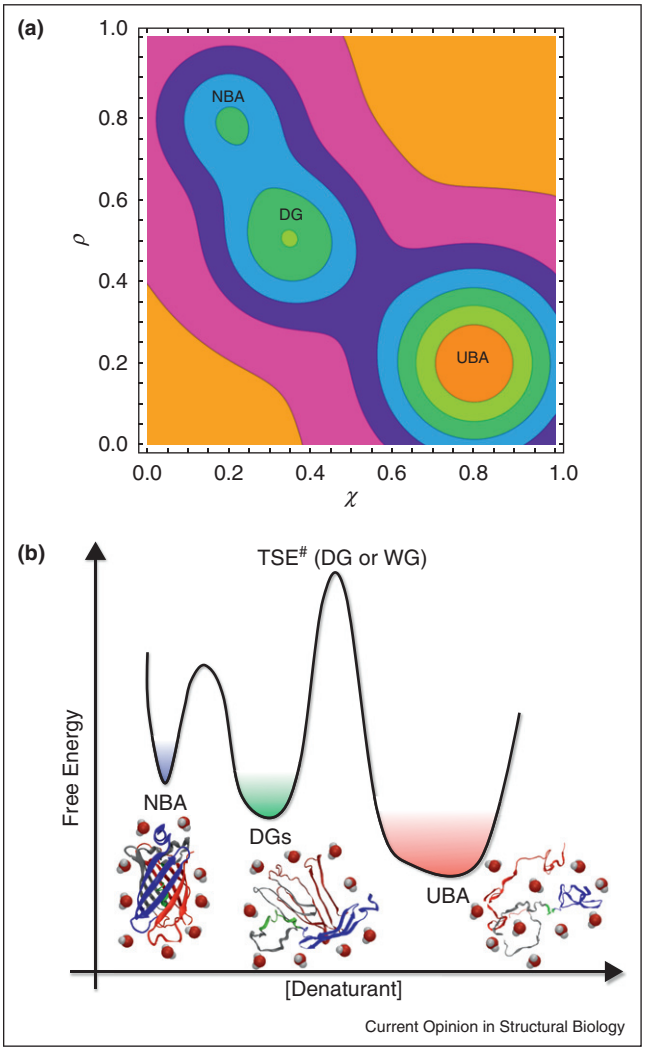
Are dry globules universal intermediates in protein folding?

The folded states are compact with the radius of gyration, $R_g \approx a_N N^{1/3}$ where N is the number of amino acid residues and $a_N \approx 3.3$ Å[41]. Using R_g as the sole order parameter it is difficult to distinguish folded states from collapsed globules, whose dimensions are only modestly larger than the folded states. Specific energetic interactions between side chains and backbone make the folded state stable. Thus, from a thermodynamic perspective there are at least two 'phase' changes as the folding condition is altered. At high denaturant concentrations, a polypeptide chain is expanded whereas at low [C]s the native state is occupied with substantial probability. Two order parameters are needed to describe the phases of a polypeptide chain. One is monomer density, $\rho = N/R_g^3$ [42], which takes on a small value at high [C] and becomes on the order of unity as the polypeptide chain becomes compact (folded state or globules). To distinguish between the globular and native states we require an overlap function χ, which is zero or small in the NBA and is larger when the protein is compact but not folded (Figure 3a). We proposed sometime ago that efficient folding occurs when solvent conditions favoring collapse (T_θ or [C_θ]) and folding (T_F or C_m) are close to each other [43,44]. In the intervening years, a number of experiments have provided evidence for protein collapse, and more recently it has been shown that $T_\theta \approx T_F$ [45••] in accord with theoretical predictions [43].

What are the structural characteristics of the compact but not natively folded conformations explored during the folding reaction by apparent two-state folders, initiated by diluting out the denaturants? The first clear cut answer to this question was provided in an experimental study on ribonuclease A [46] using one-dimensional proton magnetic resonance experiments in the presence of GdmCl. It was argued that that at least one low free energy compact structure is populated before the rate determining step in GdmCl-induced unfolding. On the basis of additional experiments it was proposed that the compact structure has enhanced side chain entropy but likely has low water content, and hence is a 'dry' globule (DG). A similar interpretation was given for the unfolding of the 159 residue *Escherichia coli* DHFR [47]. The plausible existence of DG was predicted in important theoretical papers [42,48] earlier although these authors envisioned the denaturant-induced DG as a high free energy structure, perhaps like a major transition state in folding or unfolding [49••]. The mean-field theory, using ρ as the only order parameter to describe the folding transition, also illustrated that other globular forms (wet globule and swollen globules) could be populated during the folding process under appropriate conditions [48]. More recently, DGs have been experimentally identified in other proteins [50••] that are unrelated by size, structure, or sequence, thus raising the possibility that they are universally accessed during the folding reaction [51] even for simple systems. From a statistical mechanical perspective, the DG states are unlikely to be unique and must be viewed as an ensemble of mobile conformations with low water content.

Despite the potential existence of DGs it is difficult to characterize their structures although some distance constraints have been obtained for the single chain monellin [50••]. Dry globules were predicted to be plausible structures in urea-induced solubility of linear hydrocarbon chains [51], which have long served as model systems for understanding the hydrophobic effect in proteins. In the context of proteins, it is only recently, through all

Figure 3



Folding landscapes illustrating the importance of dry globules. **(a)** 'Phases of a protein that ostensibly folds in an apparent two-state manner. Two order parameters monomer density, ρ and overlap function, χ, characterize the three phases, the Native Basin of Attraction (NBA), DG, and the unfolded basin of attraction (UBA). **(b)** Schematic of the protein folding free energy profile emphasizing dry globule (DG) formation. As the denaturant concentration increases, the compact folded structure of is destabilized results in an expanded structure in which the close packing of the side chains is lost and the core of the structure is devoid of water to form an ensemble of dry globular structures (middle basin). The loss in the packing interactions involving both side chains and backbone is compensated by enhanced entropy of the backbone and side chain motions. These structures, which are expanded form of the folded state, should be viewed as an ensemble of structures. On further increasing the denaturant concentration, the chain completely unfolds resulting in a complete loss in the packing interactions associated with the backbone and side chains. The dimensions of the unfolded state depend on the denaturant concentration. The transition state ensemble could be either dry or wet (soaked with discrete water molecules). The universal presence of DGs as intermediates, before the major unfolding (or refolding) barrier of even small protein implies that a two-state description is at best an approximate description of thermodynamics of protein folding.

atom computer simulations, exceeding μs, of a mutant of lysozyme in 8M urea there has been clear demonstration of DGs [21]. As schematically depicted in Figure 3b all atom MD simulations showed that urea-induced unfolding occurs in two distinct stages. In the first stage, urea, and not water, molecules penetrate the core creating DGs, thus disrupting the interactions involving the side chains and the backbone. The structures sampled in this stage are DGs. In the second stage water penetration leads to global unfolding. From a molecular perspective it was also established that in the first stage urea solvated the surface of the protein, interacting with both side chains and exposed backbone. Subsequently, urea penetrated the hydrophobic core of the protein, making contacts with hydrophobic side chains. Thus, it appears that urea-induced DGs have higher urea density in the core and is essentially devoid of water. Because such structures can be experimentally characterized we regard them as 'equilibrium' DGs although such a criterion may not be satisfactory to the purist.

It appears that when pressure is used to unfold proteins distinct globular structures could be populated depending on the protein. Early studies on pressure denaturation of staphylococcal nuclease showed that the transition state, not an equilibrium intermediate, has the characteristics of a DG [52]. A recent study showed that under pressure ubiquitin [53] unfolds by population an intermediate that contains discrete number of water molecules. Hence, this corresponds to a wet globule. The number of water molecules in the protein interior depends on the pressure. At values of pressure less than that needed to unfold the protein the water content increases [53], presumably continuously. Thus, we speculate that for some proteins an ensemble of wet globules is populated upon increasing pressure. From this perspective it appears that the phases of proteins under pressure and in the presence of denaturants are different. Consequently, differing folding mechanisms depend on the protocol used to initiate folding and unfolding. The existence of DGs has implications for protein folding, which are more fully explored in [49••]. (1) The usual assumption that proteins, especially moderately sized polypeptide chains, fold in a strict all-or-none transition is at best an approximation [49••]. (2) The extent of folding cooperativity is compromised as demonstrated recently for monellin [50••]. The unfolding, especially DG to UBA, occurs gradually as the denaturant concentration is increased. Because the expansion (decrease in ρ and increase in χ (see Figure 3a)) is unlikely to result in uniform loosening of the core there has to be a dispersion in the denaturant midpoints of transition depending on the residue [54], a prediction that was made based on finite size effects [55,56]. (3) We expect that the response to mechanical force of the NBA and DGs are likely to be different, with the latter being more plastic. It would be of considerable interest to quantify the mechanical response of DGs just as was

done recently for apo myoglobin at acidic pH [39••]. Such an experiment will require a phase diagram in terms of f and $[C]$.

Conclusions

Progress in physical sciences occurs when abstract ideas are translated into practice, thus enabling a direct comparison between theory, simulations, and experiments. The ideas and applications sketched here show that the conceptual advances and practical simulation design, made possible by combining old and new ideas, have ushered in a new era in protein folding in which experiments demand that the theoretical studies provide quantitative and testable predictions. With the development of new models, we are now poised to foster closer collaboration between experiments, simulations, and theory to elucidate the folding mechanisms of small and large proteins.

Acknowledgements

We are grateful to Buzz Baldwin, Cathy Royer, and Eugene Shakhnovich for useful discussions. This work was supported in part by a grant from the National Science Foundation (CHE 09-14033) and the National Institutes of Health (GM089685). ZL acknowledges financial support from the National Natural Science Foundation of China under the grant no. 11104015.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Dill KA, Ozkan BS, Shell M, Weikl TR: **The protein folding problem**. *Ann Rev Biophys* 2008, **37**:289-316.

2. Onuchic JN, Wolynes PG: **Theory of protein folding**. *Curr Opin Struct Biol* 2004, **14**:70-75.

3. •• Kubelka J, Henry ER, Cellmer T, Hofrichter J, Eaton WA: **Chemical, physical, and theoretical kinetics of an ultrafast folding protein**. *Proc Natl Acad Sci U S A* 2008, **105**:18655-18662.
A theoretical Ising-like model is used to quantitatively predict the folding thermodynamics and kinetics of an ultrafast folding protein.

4. Thirumalai D, O'Brien EP, Morrison G, Hyeon C: **Theoretical perspectives on protein folding**. *Ann Rev Biophys* 2010, **39**:159-183.

5. Guo ZY, Thirumalai D: **Kinetics of protein-folding – nucleation mechanism, time scales, and pathways**. *Biopolymers* 1995, **36(1)**:83-102.

6. Hagen S, Hofrichter J, Szabo A, Eaton W: **Diffusion-limited contact formation in unfolded cytochrome c: estimating the maximum rate of protein folding**. *Proc Natl Acad Sci U S A* 1996, **93(21)**:11615-11617.

7. Guo Z, Thirumalai D, Honeycutt J: **Folding kinetics of proteins – a model study**. *J Chem Phys* 1992, **97(1)**:525-535.

8. Camacho CJ, Thirumalai D: **Minimum energy compact structures of random sequences of heteropolymers**. *Phys Rev Lett* 1993, **71**:2505-2508.

9. Lin MM, Zewail AH: **Hydrophobic forces and the length limit of foldable protein domains**. *Proc Natl Acad Sci U S A* 2012, **109**:9851-9856.

10. Thirumalai D: **From minimal models to real proteins – time scales for protein-folding kinetics**. *J Phys I* 1995, **5(11)**:1457-1467.

11. Peng Q, Fang J, Wang M, Li H: **Kinetic partitioning mechanism governs the folding of the third FnIII domain of tenascin-c: evidence at the single-molecule level**. *J Mol Biol* 2011, **412**:698-709.

12. •• Stigler J, Ziegler F, Gieseke A, Gebhardt JCM, Rief M: **The complex folding network of single calmodulin molecules**. *Science* 2011, **334(6055)**:512-516.
A clear demonstration that the flux to the native state of Calmodulin flows through many interconnected states.

13. Schuler B, Eaton WA: **Protein folding studied by single-molecule FRET**. *Curr Opin Struct Biol* 2008, **18(1)**:16-26.

14. •• Whitford PC, Sanbonmatsu KY, Onuchic JN: **Biomolecular dynamics: order-disorder transitions and energy landscapes**. *Rep Prog Phys* 2012, **75(7)**:076601.
An excellent survey of the uses of structure-based model in a variety of problems ranging from protein folding to functions of biological machines.

15. Garcia-Manyes S, Dougan L, Badilla CL, Brujic J, Fernandez JM: **Direct observation of an ensemble of stable collapsed states in the mechanical folding of ubiquitin**. *Proc Natl Acad Sci U S A* 2009, **106(26)**:10534-10539.

16. •• Chung HS, McHale K, Louis JM, Eaton WA: **Single-molecule fluorescence experiments determine protein folding transition path times**. *Science* 2012, **335**:981-984.
The most direct measurement of the time a protein takes to cross the free energy separating the folded and unfolded states. By implication this study also provides an estimate of the speed limit in protein folding.

17. •• Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP *et al.*: **Atomic-level characterization of the structural dynamics of proteins**. *Science* 2010, **330**:341-346.
Demonstration that advances in computer technology can be harnessed to obtain a detailed movie of how proteins fold. The study could also serve as a benchmark for testing the efficacy of minimal models for protein folding.

18. Jensen MO, Jogini V, Borhani DW, Leffler AE, Dror RO, Shaw DE: **Mechanism of voltage gating in potassium channels**. *Science* 2012, **336(6078)**:229-233.

19. Freddolino PL, Harrison CB, Liu Y, Schulten K: **Challenges in protein-folding simulations**. *Nat Phys* 2010, **6**:751-758.

20. Voelz VA, Bowman GR, Beauchamp K, Pande VS: **Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1–39)**. *J Am Chem Soc* 2010, **132**:1526-1528.

21. Hua L, Zhou R, Thirumalai D, Berne BJ: **Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding**. *Proc Natl Acad Sci U S A* 2008, **105**:16928-16933.

22. •• Canchi D, Paschek D, Garcia AE: **Equilibrium study of protein denaturation by urea**. *J Am Chem Soc* 2010, **132**:2338-2344.
The first study that uses atomically detailed simulations to generate equilibrium titration curve of urea-induced unfolding.

23. Hyeon C, Thirumalai D: **Capturing the essence of folding and functions of biomolecules using coarse-grained models**. *Nat Commun* 2011, **2**:487.

24. Auton M, Bolen D: **Predicting the energetics of osmolyte-induced protein folding/unfolding**. *Proc Natl Acad Sci U S A* 2005, **102(42)**:15065-15068.

25. O'Brien EP, Ziv G, Haran G, Brooks BR, Thirumalai D: **Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model**. *Proc Natl Acad Sci U S A* 2008, **105**:13403-13408.

26. •• Reddy G, Liu Z, Thirumalai D: **Denaturant-dependent folding of GFP**. *Proc Natl Acad Sci U S A* 2012, **109**:17832-17838.
Folding GFP in the presence of denaturants is simulated using Self Organized Polymer model [30]. The study is the first to report the complex folding landscape of a large protein as a function of temperature and denaturant concentration.

27. O'Brien EP, Brooks BR, Thirumalai D: **Effects of pH on proteins: predictions for ensemble and single-molecule pulling experiments**. *J Am Chem Soc* 2012, **134(2)**:979-987.

28. Mohanty D, Elber R, Thirumalai D, Beglov D, Roux B: **Kinetics of peptide folding: computer simulations of SYPFDV and peptide variants in water**. *J Mol Biol* 1997, **272**:423-442.

29. Zhang Z, Thirumalai D: **Dissecting the kinematics of the kinesin step**. *Structure* 2012, **20**:628-640.

30. Hyeon C, Dima RI, Thirumalai D: **Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins**. *Structure* 2006, **14**:1633-1645.

31. •• Liu Z, Reddy G, O'Brien EP, Thirumalai D: **Collapse kinetics and chevron plots from simulations of denaturant-dependent folding of globular proteins**. *Proc Natl Acad Sci U S A* 2011, **108(19)**:7787-7792.
This represents to date the first and only computational study, which semi-quantitatively reproduces the experimentally measured Chevron plot in the folding of proteins. The effects of denaturants are taken into account using the Molecular Transfer Model [25].

32. Mickler M, Dima RI, Dietz H, Hyeon C, Thirumalai D, Rief M: **Revealing the bifurcation in the unfolding pathways of GFP by using single-molecule experiments and simulations**. *Proc Natl Acad Sci U S A* 2007, **104(51)**:20268-20273.

33. Liu Z, Reddy G, Thirumalai D: **Theory of the molecular transfer model for proteins with applications to the folding of the src-SH3 domain**. *J Phys Chem B* 2012, **116(23)**:6707-6716.

34. Grantcharova VP, Baker D: **Folding dynamics of the src SH3 domain**. *Biochemistry* 1997, **36(50)**:15685-15692.

35. Hyeon C, Morrison G, Thirumalai D: **Force dependent hopping rates of RNA hairpins can be estimated from accurate measurement of the folding landscapes**. *Proc Natl Acad Sci U S A* 2008, **105**:9604-9606.

36. Hyeon C, Thirumalai D: **Measuring the energy landscape roughness and the transition state location of biomolecules using single molecule mechanical unfolding experiments**. *J Phys: Condens Matter* 2007, **19**:113101.

37. Hyeon C, Thirumalai D: **Forced-unfolding and force-quench refolding of RNA hairpins**. *Biophys J* 2006, **90(10)**:3410-3427.

38. Klimov D, Thirumalai D: **Stretching single-domain proteins: phase diagram and kinetics of force-induced unfolding**. *Proc Natl Acad Sci U S A* 1999, **96(11)**:6166-6170.

39. •• Elms PJ, Chodera JD, Bustamante C, Marqusee S: **The molten globule state is unusually deformable under mechanical force**. *Proc Natl Acad Sci* 2012, **109**:3796-3801.
A nice study that shows the response to force can vary depending on the extent to which protein is folded.

40. Morrison G, Hyeon C, Hinczewski M, Thirumalai D: **Compaction and tensile forces determine the accuracy of folding landscape parameters from single molecule pulling experiments**. *Phys Rev Lett* 2011, **106(13)**.

41. Dima R, Thirumalai D: **Asymmetry in the shapes of folded and denatured states of proteins**. *J Phys Chem B* 2004, **108**:6564-6570.

42. Shakhnovich E, Finkelstein A: **Theory of cooperative transitions in protein molecules. 1. Why denaturation of globular protein is a 1st-order phase-transition**. *Biopolymers* 1989, **28(10)**:1667-1680.

43. Camacho CJ, Thirumalai D: **Kinetics and thermodynamics of folding in model proteins**. *Proc Natl Acad Sci U S A* 1993, **90(13)**:6369-6372.

44. Klimov D, Thirumalai D: **Linking rates of folding in lattice models of proteins with underlying thermodynamic characteristics**. *J Chem Phys* 1998, **109(10)**:4119-4125.

45. •• Hofmann H, Soranno A, Borgia A, Gast K, Nettels D, Schuler B: **Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single molecule spectroscopy**. *Proc Natl Acad Sci* 2012, **109**:16155-16160.
A clear study demonstrating that natural proteins may have evolved in such a way that folding and collapse occur nearly simultaneously as was first predicted in [43]. It implies that water is close a θ-solvent for many proteins.

46. Kiefhaber T, Labhardt A, Baldwin R: **Direct NMR evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease-a**. *Nature* 1995, **375(6531)**:513-515.

47. Hoeltzli S, Frieden C: **Stopped-flow NMR-spectroscopy – real-time unfolding studies of 6-F-19-tryptophan-labeled *E. coli* dihydrofolate-reductase**. *Proc Natl Acad Sci U S A* 1995, **92**:9318-9322.

48. Finkelstein A, Shakhnovich E: **Theory of cooperative transitions in protein molecules. 2. Phase-diagram for a protein molecule in solution**. *Biopolymers* 1989, **28(10)**:1681-1694.

49. •• Baldwin RL, Frieden C, Rose GD: **Dry molten globule intermediates and the mechanism of protein unfolding**. *Proteins* 2010, **78(13)**:2725-2737.
An excellent survey of the stages in denaturant-induced unfolding. It is proposed that the formation of dry globules may be a universal feature of protein folding, implying that the often used two-state description of folding of small proteins is at best an approximation.

50. •• Jha SK, Udgaonkar JB: **Direct evidence for a dry molten globule intermediate during the unfolding of a small protein**. *Proc Natl Acad Sci U S A* 2009, **106(30)**:12289-12294.
An outstanding recent study in which the presence of dry globule predicted on theoretical grounds [48] and using simulations [21,51] is clearly demonstrated.

51. Mountain R, Thirumalai D: **Molecular dynamics simulations of end-to-end contact formation in hydrocarbon chains in water and aqueous urea solution**. *J Am Chem Soc* 2003, **125(7)**:1950-1957.

52. Vidugiris G, Markley J, Royer C: **Evidence for a molten globule-like transition-state in protein-folding from determination of activation volumes**. *Biochemistry* 1995, **34**:4909-4912.

53. Day R, Garcia AE: **Water penetration in the low and high pressure native states of ubiquitin**. *Proteins: Struct. Funct. Bioinform.* 2008, **70**:1175-1184.

54. O'Brien EP, Brooks BR, Thirumalai D: **Molecular origin of constant m-values, denatured state collapse, and residue-dependent transition midpoints in globular proteins**. *Biochemistry* 2009, **48**:3743-3754.

55. Li M, Klimov D, Thirumalai D: **Finite size effects on thermal denaturation of globular proteins**. *Phys Rev Lett* 2004, **93(26)**.

56. Klimov D, Thirumalai D: **Factors governing the foldability of proteins**. *Proteins: Struct. Funct. Bioinform.* 1996, **26**: 411-441.