ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT

Polymer 45 (2004) 547–555

polymer

www.elsevier.com/locate/polymer

The energy landscape for protein folding and possible connections to function

Margaret S. Cheung[a], Leslie L. Chavez[b], José N. Onuchic[b,*]

[a]Institute of Physical Science and Technology, University of Maryland, College Park, MD 20742, USA

[b]Department of Physics, Center for Theoretical Biological Physics, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0319, USA

Received 10 July 2003; received in revised form 9 October 2003; accepted 9 October 2003

Abstract

In this article we review and discuss the state-of-the-art methods using minimalist models in the context of energy landscape theory to study protein folding. As good agreement between computational/theoretical studies and experimental observations in vitro continues to emerge, many research groups are asking how this structural and dynamical information can be used to understand proteins in vivo. This is a non-trivial question drawing from very limited in vivo studies. From the perspective of theory, it is a new horizon for theoreticians to test or revise their theories by making connections to experiments on this matter. We present a short discussion of several recent efforts that include factors reflecting the cellular environment in computer simulations—and that may provide some insight into the behavior of protein dynamics inside the living cell as well as inspire the development of new experimental approaches for a better understanding of the molecular mechanisms for function.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Desolvation; Minimalist models; Funnel concept

1. Introduction

Proteins perform numerous biological activities in the cell, making it vital for us to understand the structural and dynamical properties of these molecules. Physically, proteins are biopolymers composed of a sequence of amino acids that encode their three-dimensional structure. Some proteins (mostly those with less than 200 amino acids) can fold into their native structure by themselves without the aid of chaperones or other proteins. These self-assembling proteins have been studied extensively in vitro, and experiments show their folding kinetics to be relatively fast for both small two-state proteins and larger proteins with more complex folding kinetics. How proteins fold from one-dimensional chains to unique structures that regulate cellular function has been a question which first challenged researchers in biology, but has required knowledge from other fields such as chemistry and physics to answer; it has taken an interdisciplinary effort to reveal the folding mechanism of proteins.

The theoretical framework of free energy landscape theory [1–3] and the funnel concept [4] has successfully helped us to understand the mechanisms of proteins folding. Computational simulations of reduced protein models with Hamiltonians designed to provide a minimally frustrated energy landscape have shown excellent agreement with mutagenetic experiments, allowing us to sketch a big picture of the folding process; fast folding proteins have sufficiently reduced energetic frustration that the structural properties of their folding transition state ensemble (even of intermediates for more complex folders) are mostly determined by the folding geometry [5–7]. Advances in NMR techniques and hydrogen exchange experiments are providing experimental insight into a wider range of the folding mechanism, adding detail and depth to this general picture. Here we review the principles emerging from the most recent experimental results and how minimalist models are both learning from and predicting experimental data. With this clearer understanding of protein folding in hand, we may now begin to question: is it possible to relate the protein dynamics in the in vitro landscape to its functional activity? We conclude this review with a discussion of protein dynamics that may be geared for molecular function and the use of minimalist models to study these dynamics. We limit our focus to dynamics due to the more coarse-grained protein global

* Corresponding author.
E-mail address: jose@guara.ucsd.edu (J.N. Onuchic).

0032-3861/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.polymer.2003.10.082

motions, rather than investigating specific chemistry details that are needed in reactions such as electron transfer.

The review is organized as follows: The first half provides a short tour of the energy landscape theory and the funnel concept, then relates the synergistic partnership taking shape between experiment and theory in the context of energy landscape theory. In the second half of the review we discuss several topics from theory and experiment that address protein dynamics beyond typical diluted conditions and begin to explore dynamics which may be associated with molecular function: (1) single molecule stretching and (2) folding in a crowded/confined environment. We then conclude with a short discussion on the future study of folding and function.

2. The free energy landscape

The free energy landscape represents the configuration space of energy and entropy available to a protein, from its unfolded to folded structure. The protein does not have to follow a specific path through the configuration space, but instead it may travel any number of ways, taking on various partially formed structures that form the free energy landscape. Several reviews of energy landscape theory and the funnel concept have been presented elsewhere [1,3,8,9]. Here we would like to highlight how this theoretical framework underlying the general folding mechanism has been useful to the study of protein dynamics in experiments and theory.

2.1. The funnel landscape in experiment and theory

A protein shapes its energy landscape with the energetic interactions found between its residues, and through the attractive contacts forms its native structure. The funnel landscape represents a landscape with few non-native contacts, meaning that the energetic frustration is small (any energetic traps present are small enough that they do not compete with the global energetic minimum that defines the native structural ensemble). The diagram in Fig. 1 shows the funnel shape of a minimally frustrated landscape. This landscape reduces the search through configuration space and enables proteins to fold in a time reasonable for biology. In physical proteins, numerous experiments have shown the class of two-state, independently folding proteins to fold quickly and be robust folders that are able to tolerate many mutations [10–13]. Off- and on-lattice systems have helped verify the ‘funnel-like’ landscape of physical proteins [13–15]. In the simplest of these models, each amino acid is reduced to a single monomer bead (Fig. 2, models a and b). These minimalist models trace the backbone of the protein, and can be used with various energetic potentials [14,16–26], essentially tuning the amount of energetic frustration on the energy landscape. Studies comparing models with more or less energetic frustration have found

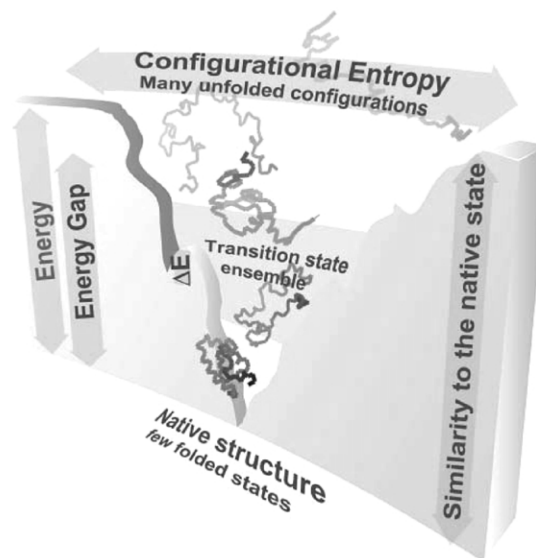

Fig. 1. The funnel landscape: the energetic bias to fold is much greater than the energetic frustration.

that the amount of energetic frustration on the free energy landscape determines whether a particular amino acid sequence is a good or a bad folder; highly frustrated sequences may never fold to a consistent minimum and therefore are not protein-like [13,14,20,27].

These studies combined with experimental data have shown that proteins capable of self-assembly have an energetic bias toward the native fold that is greater than the energetic frustration of their landscape. For this reason, protein models with a native-state biased, energetically unfrustrated Hamiltonian (Gō-like potentials [28]) can represent the folding mechanism of both two and three-state proteins. Using a Cα model with a Gō potential, the

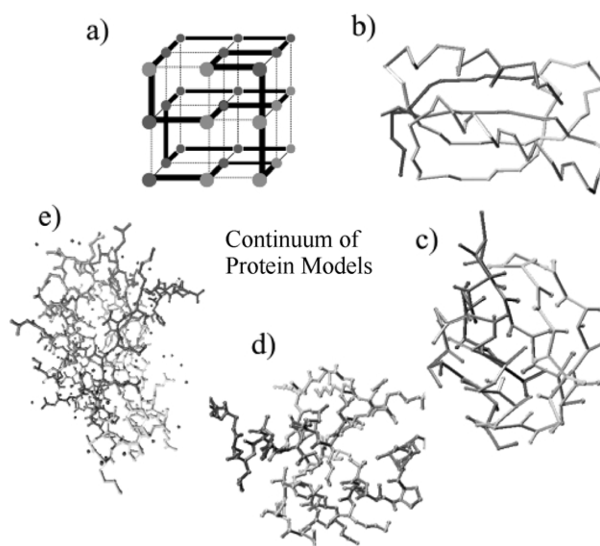

Fig. 2. Protein representation: (a) on-lattice, (b) C-α, (c) C α-β, (d) all-atom (heavy atoms only; no solvent), and (e) fully solvated all-atom model. Minimalist models continue to have a prominent role in simulations and data interpretation.

structural aspects of the transition state ensemble of CI2 and SH3 were accurately reproduced solely from backbone topology considerations [29]. Even more impressively, this simple model captured both the transition state and intermediate structures of five larger proteins: barnase, Rnase H, CheY, IL-1β, and DHFR [29,30]. A larger study of 18 proteins (representing many motifs) found this model to predict the transition state of half the studied proteins [31]. Recent evidence also shows the funnel landscape may be applied to the folding and binding of dimers [32–34]. A study of 11 dimers could predict the 2 or 3-state folding mechanism found in numerous experiments [33].

2.2. Circular permutants: weighing the importance of energetics and topology

The success of the energetically unfrustrated Cα model supports the experimental and theoretical observations that, since many proteins have sufficiently small levels of energetic frustration, topology is responsible for many features on the free energy landscape and plays a role in determining the folding rate [35–38]. The importance of topology may be seen in structurally similar proteins with low sequence identity, such as src-SH3 and α-spectrin SH3. Although they share only 30% of their sequence identity, they have a similar transition state [39–42]. A study of immunoglobulin-like β-sandwich proteins shows they have similar folding mechanisms, but does not find gross features of topology to be a simple predictor of folding rates among this family of proteins [43]. There are also notable exceptions in which topology and folding mechanism do not agree. Protein L and protein G are homologous, symmetric proteins, yet their folding mechanisms differ [44,45]. In the case when proteins have a high degree of symmetry, small variations in sequence may differentiate between possible mechanisms [22,46–49]. Small changes in energetics may be studied with mutagenetic experiments. What about an experiment that conserves the energetics but distorts the topology? One way to study differing topologies is with circular permutants [50–53]. A circular permutant is a protein in which the terminal ends have been connected with a short linker and an incision made somewhere in the protein interior. Although the linker region and the new termini may have a few non-native residues, the bulk of the amino acid sequence is identical to that of the wild-type protein.

Experimental studies of circular permutants have been performed with CI2 and SH3 [50,51,54]. Protein engineering experiments show the transition state ensemble of CI2 to be diffuse, building toward the native structure by forming many local contacts [12]. SH3 has a more polarized transition state, in which interactions between the distal β-hairpin and the diverging turn are the rate-limiting step [40, 41]. The two folding mechanisms respond very differently to circular permutation [55]. A permutation of CI2 made by cleaving the scissile bond M40-Glu41 retains the diffuse transition state of the wild-type protein. In SH3, however, the transition state ensemble of permutant N47-D48 remains polarized, but the structured region shifts from the distal loop (the cleavage occurs in this loop) to the n-Src loop and N terminus. Another SH3 permutant S19-P20 has a similar, but less dramatic shift (contacts between the distal loop and diverging term weaken, and the N terminus shows no change from that of the wild-type). These results may seem quite intuitive from a loop entropy point of view. Since the transition state ensemble of CI2 closely resembles the native structure with partially formed structure throughout the protein, the cleavage of any interior bond does not disrupt the distribution of energetic and entropic interactions. In SH3, the more structured region involves local residues, so a cleavage made there disrupts this structure by forcing the energetically favored residues to be far apart, disfavoring them entropically. The different transition state ensembles for the two proteins suggest that the distribution of energetic interactions is important for understanding the change upon circular permutation [53]. Interestingly, energetically unfrustrated Cα models are able to reproduce these experimental results [56]. This result supports the idea that the chain connectivity of the native state significantly determines the folding mechanism.

An exciting study of circular permutants of the ribosomal protein S6 has a more surprising result. The most documented permutant to date has the incision made between wild-type residues 13 and 14 [57]. Φ-value analysis[1] [58] shows the permutant N13-L14 to have a sharply polarized transition state, quite different from the diffuse transition state observed in the wild-type S6 [59]. The Cα and Cα/Cβ minimalist models (see later section for a description of the Cα/Cβ model) are able to replicate this difference, showing the folding mechanisms of the wild-type and permutant to differ (unpublished results by the authors). The diffuse transition state of S6 behaves quite uniquely, differing from the behavior previously described for CI2.

2.3. Accessing multiple folding routes

Having a more diffuse or more polarized transition state is one component of topological frustration. On-lattice and analytical theory has suggested that having a more structurally polarized transition state ensemble could speed folding [60], however most experimental data shows proteins to have a more diffuse transition state ensemble. Recent simulations using the Cα model with a Gō potential show that many fast-folding proteins have a polarized distribution of structure before the transition state.

[1] Φ-value analysis is an experimental technique which tries to relate how structured a specific residue is in the transition state ensemble. If $\Phi = 1$, the residue is structured similarly to the native state in the transition state while if $\Phi = 0$, it is not structured at all. Most amino acids show intermediate values.

Because this limits the search through configuration space as the reaction progresses, this aids the protein in folding. If the polarization occurs during the transition state barrier, however, the polarization creates a bottleneck in the energy landscape and the protein folds more slowly [61]. Being a fast-folder is not the only consideration for a protein. The protein should also be robust to mutation and resistant to unfolding by diffusive motion, and it must function biologically. These considerations may favor a transition state ensemble that is more diffuse; the permuted S6 P13-14 folds more quickly than the wild-type, yet evolution selected the more homogeneous transition state ensemble [57].

2.4. Are folding routes optimized in proteins?

S6 folds via different portions of the free energy landscape depending on specifics of its amino acid sequence and its environment [62]. Specific residues called 'gatekeepers' may prevent the wild-type protein from accessing inefficient portions of the free energy landscape and stand guard against aggregation [63,64]. Gatekeepers are charged residues that break the continuity of long hydrophobic sequences, possibly disrupting the formation of amyloid fibrils and depopulating an off-pathway intermediate with unfavorable interactions. A recent theoretical study of gatekeepers performed with a 46-mer β-barrel model has investigated the idea of gatekeepers as agents that steer the protein to more effective folding routes [65]. Charge was introduced to the Thirumalai-Honeycutt three letter potential of the extensively studied β-barrel model to determine if strategically placed salt-bridges could lead to faster folding [66,67]. Well-located salt bridges did enhance the folding rate of the energetically frustrated wild-type β-barrel by leading the protein toward more efficient folding routes early in the folding process. Theoretical results support the idea of gatekeeper residues that assist folding by discouraging routes through rugged portions of the energy landscape.

2.5. Addressing microscopic desolvation effects

Inspired by pressure-denatured protein unfolding experiments, recent theoretical studies of hydrophobic interactions show that when two non-polar solutes come together to form a contact, there is a free energy cost, the size of which depends on the hydrostatic pressure [68–70]. This free energy penalty of contact formation is a direct consequence of the granularity of water molecules in the first hydration shell. Theoretically, it may be represented by a desolvation barrier in the profile of the potential of mean force, a salient feature which has also been observed in neutron scattering experiments and all-atom simulations (refer to the top of Fig. 3 for a detailed diagram and interpretation of this potential). The implementation of this desolvation potential in the study of protein-related

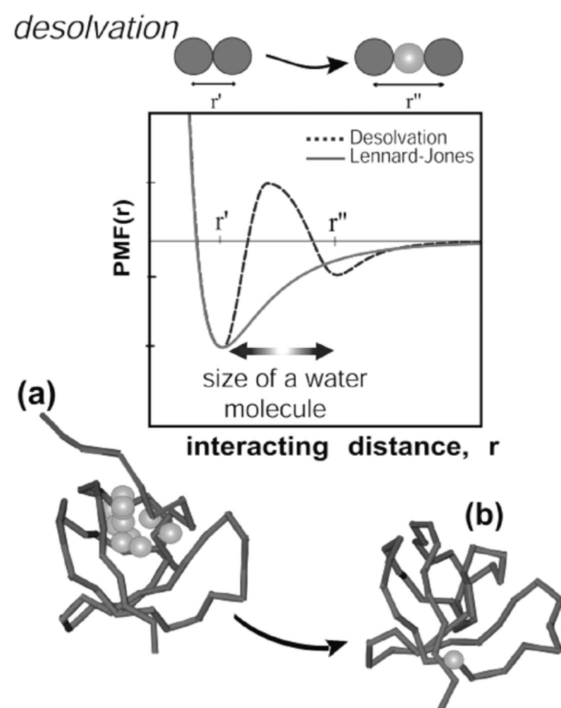

Fig. 3. Desolvation of the hydrophobic core of SH3. Top: potential of mean force for desolvation. The snapshots from (a) to (b) were taken from a typical kinetic folding trajectory of the SH3 using this potential. This transition captures the desolvation of the hydrophobic core: blue spheres in (a) marking the water separated contacts are 'expelled' upon the formation of the native hydrophobic core (b).

problems has yielded many interesting results with rich physical interpretations [24,26,71,72].

Using desolvation potentials, Cheung et al. suggest that the folding mechanism of the SH3 domain first undergoes a structural-search collapse (panel a in Fig. 3) followed by the desolvation of the hydrophobic core (panel b in Fig. 3) [24]. The latter feature has been compared to the loosely compact and structured, partially denatured ensemble identified in NMR experiments carried out under near-native conditions [73]. A study of all-atom simulations of src-SH3 also agrees with this lubricated hydrophobic core at the late stage of folding [74]. The authors would like to point out that under physiological conditions, the conformational changes associated with desolvation dynamics might be one way in which SH3 conveys molecular function. This is, however, only a possibility with no direct supporting evidence. This desolvation study has motivated experimentalists to design a novel protein engineering approach that looks for a 'solvated' or 'lubricated' hydrophobic core of SH3 in the transition states. Excellent agreement has been found on the distribution of simulated folding rates of SH3 mutants without any previous knowledge of experimental kinetic observations [75]. Mutants V44T and V53T (residues 44 and 53 participate in the folding mechanism) have slower folding rates than other core Valine to Threonine mutants. Simulations in turn provide molecular explanations for kinetic traps that cause the decrease of folding rates:

mutation at position 44 disrupts the structural search collapse, and mutation at position 53 hinders the desolvation of the hydrophobic core. This combination of experiment and theory ascertains more of the validity of the energy landscape theory as a framework to understand protein folding mechanisms.

Desolvation potentials have not only been useful in the study of SH3: a recent study by Karanicolas et al. [26] captures the folding mechanism of protein G and L using desolvation potentials. This study examined the lubricated hydrophobic core as folding events approach the native state. Furthermore, others are applying desolvation models to better understand several protein-related questions. Guo [71] et al. suggest a possible mechanism to describe protein aggregation, in which expelling 'water' molecules (i.e. desolvation) between β sheets accounts for much of the collective behavior in fibril formation. Chan et al. used desolvation potentials to examine the cooperativity of CI2 folding simulations by matching profiles of Chevron plots [72]. Both studies suggest that the desolvation description is necessary to explain several basic properties of proteins.

Recently, there has been an increase in attention from both theoreticians and experimentalists to de/solvation dynamics in protein-related problems. The water dynamics associated with de/solvation direct protein fluctuations around their native state at the bottom of the folding funnel. Assuming that most proteins remain close to their native structures under physiological conditions, studies of protein fluctuations and dynamics in the native basin of funnel-like energy landscapes may provide non-trivial insights of protein dynamics relevant to function. In addition to fluctuations in the native basin, large, partially unfolding fluctuations in the funnel may also be relevant to function.

3. Beyond Cα models

In the previous section, all analyses have been done using Gō-like Cα models. These models lack much of the detail needed to represent physical proteins, however, all atom models still cost too much computationally to be a reasonable resource for further studies and comparison with experiment. To address a compromise between the stark representation of the Cα model and the expense of all atom models, in this section we comment on intermediate level models.

How complex do intermediate level models need to be in order to provide a quantitative explanation of protein dynamics that are observed in experiments? Cα models are able to provide a qualitative description of the global folding mechanism. To be considered an improvement, an intermediate level model should give more quantitative information than a Cα model, and yet be more physically tractable and less computationally costly than all-atom models at conveying the full folding history. In this regard, several reduced protein representations that go beyond Cα models have been designed to answer several pending biological questions: What is the role of the excluded volume effects due to side-chains in folding? How does backbone hydrogen bonding contribute to protein dynamics? The former case has been extensively studied by using a so-called all-atom Gō model (see Fig. 2, model d) [46,76,77] which retains all heavy atoms in the protein representation; however, the driving force for this folding system is still designed to favor the native state. In Clementi's study, the reduction in configuration entropy due to the excluded volume effects of the all-atom model is compared with the Cα model. The side-chains introduce more topological frustration, but also serve to steer the protein into the proper configuration and thus enhance the cooperativity of folding. This guidance from the side-chains is able to break the symmetry of the folding mechanisms of protein L from that of protein G; the Cα Gō model used in this study was unable to capture any difference between the two topologically similar proteins. Interestingly, a Cα Gō model with sequence-dependent interactions and a desolvation barrier also reproduced the folding mechanisms of these proteins [22]. It would be interesting to introduce some non-nativeness in the side-chain interactions of the all-atom model to discuss the energetic competitions of native and non-native contact formation [78].

Another issue neglected by single bead models is the role of backbone hydrogen bonding, which is an interesting subject because of its unique directional behavior in main-chain interactions [79]. Hydrogen bonds play an important part in the formation of common secondary structures in proteins, such as α helices, parallel β sheets, and anti-parallel β sheets [80]. Several phenomenological models have been constructed to capture this directional feature by introducing an angle dependent term into the contact formation [71,81–88]. Using two heavy atoms on the backbone to define the backbone orientation, the angular term implicitly employs many-body effects in the backbone interactions. This many-bodied potential in backbone interactions reinstates the cooperativity of folding in real proteins into minimalist models [71,87,89–91]. This kind of microscopic cooperativity may also play a crucial role in protein aggregations where bulk structures are dominated by β-sheet formation [71,82,91,92].

Several groups have investigated the directional behavior of backbone hydrogen bond interactions for protein folding based on the energy landscape theory. Hardin et al. [86] and Cheung et al. [87] used an intermediate level representation that includes at most two beads per residue to investigate the interplay of native side-chain interactions and secondary structure formation (Fig. 2, model c). In Hardin's study, they altered the backbone hydrogen bonding to introduce non-specific interactions to the Gō-like Hamiltonian and discussed the role of chirality of atoms (improper chirality can cause long-lived kinetic traps if not properly assigned). In Cheung's study, they estimated the free energy costs to order backbone hydrogen bonds into particular orientations

based on the energy landscape theory and the funnel concept. In their model, backbone hydrogen bonding is allowed to be non-specific, while tertiary contacts and dihedral contributions remain minimally frustrated. Their results suggest that as long as the folding landscape remains funnel-like, backbone hydrogen bonds are more likely to adopt native configurations if their angular term can aid by disfavoring random backbone conformations. This study implies that the directionality of backbone hydrogen bonding in real proteins may discourage transient frustrated tertiary contacts in random secondary conformations throughout the folding event, thereby retaining the robustness of a smooth folding landscape.

4. Protein folding dynamics possibly reflecting molecular function

4.1. Unfolding by force

Structural proteins in our muscles and cellular membranes must withstand more than thermal agitation by surrounding solvent; they must have a kind of elasticity to withstand external force. Fibers in our muscles stretch and contract, and our cellular membranes yield and bend as required by bodily function. Exciting work on titin I27 and tenascin has shown that unfolding single proteins by force can still be understood with free energy landscape theory [93–96]. The applied force accesses a different free energy landscape from that of chemical or thermal denaturation, yet minimalist protein models are still able to capture the unfolding of proteins by force.

Minimalist models help interpret the data from experiments and steered molecular dynamics simulations. Backbone models reduce the degrees of freedom and examine the importance of backbone topology, pulling location, and shearing forces in forced unfolding (in Fig. 4(a)). Additionally, an unfolding simulation can be finished at a pulling speed closer (only three orders of magnitude greater than) to experimental speeds, as opposed to the all-atom steered molecular dynamics simulations that use a pulling velocity six orders of magnitude greater than the pulling velocity in experiment [97].

One lattice model study of a 27-mer, though not modeling the β-sheet structure of titin I27, raised the question of whether length, the experimentally preferred reaction coordinate in stretching experiments, is a faithful reaction coordinate for proteins, and whether force, a vector, can probe a scalar quantity, the free energy landscape [93]. These concerns are difficult to address experimentally, but are well suited for minimalist model simulations. This study found the unfolding to occur in two steps: (1) an unfolding event in which contacts in the lattice broke and (2) a lengthening to the fully unfolded length. The transition state detected in this study was similar in nature to the native state, agreeing with experimental data on titin I27, where the transition state length is much closer to the native state length than to the fully extended protein. This study found that length is not a straightforward reaction coordinate, differing for example from the fraction of native contacts because it is unable to clearly differentiate between folded and unfolded protein forms.

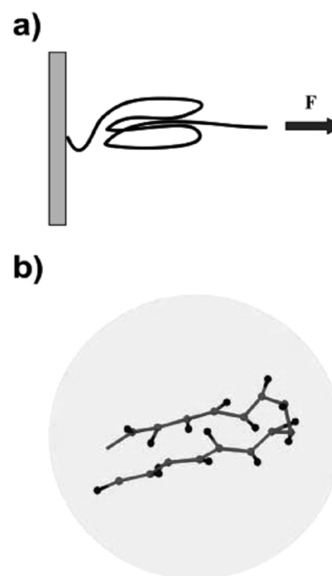

Fig. 4. (a) Single molecule stretching simulations performed with minimalist models reproduce and explore many results seen experimentally and in all-atom solvated simulations. (b) Folding simulations of a minimalist model hairpin (Cα's are green balls and Cβ's are blue balls) (see web version) are carried out in a spherical cage which mimics the confinement condition in the cellular environment.

Klimov and Thirumalai have built a theoretical model for predicting force/pulling speed relations, placement of the transition state, and the unfolding mechanism in stretching experiments. Titin I27 has its N and C-terminals on opposite sides of its structure, so as those ends are pulled apart a kind of shearing between beta-sheets occurs [97–99]. Lattice models provide an opportunity to experiment with different topologies and different placements of the end terminals. In a study of two lattice models, the model that mimicked titin 127 by having its terminal ends on opposite sides unfolded abruptly, with no intermediate structure. The other model with its terminal ends on the same side unfolded quickly to a partially structured configuration, then unfolded more slowly to the fully stretched out configuration. This shows that the geometry of the protein affects its unfolding kinetics. These studies have also explored hysteresis in forced unfolding experiments and the relationship between work and force loading rate first proposed by Evans and Ritchie [100,101]. The pulling location has also been shown to be a factor in shaping the free energy landscape [95]. Because the rotational relaxation rate of the tandem strings of titin I27 is not known, this may be relevant to unfolding experiments.

4.2. Folding in a crowding or confined environment

Because the aqueous environment in a living cell is crowded by macromolecules such as proteins, lipids, and nucleic acids, it resembles a gel more than the dilute solvents used in vitro [102]. Protein dynamics in a crowded environment have shown interesting thermodynamic and kinetic behavior different from that found in typical experimental conditions [103]. This difference has motivated the study of macromolecular properties in vivo. Studies of crowding agents can provide insight to other biological problems where dynamics occur in confined spaces, such as in chaperonin-mediated proteins. For example, chaperonins aid protein folding by encapsulating misfolded proteins into their central cavities, confining the protein until it refolds to its native state.

The restriction of crowded and confined spaces calls for theoreticians to study the role of excluded volume effects in eliminating denatured protein conformations. One of the pioneers of this problem, Minton [103], applied statistical-thermodynamic models to investigate the change of folding stabilities on the level of crowding agents. Later, Zhou et al. [104] provided analytical solutions of the partition functions for an ensemble of Gaussian chain configurations contained within a cage. In Zhou's study, the authors discuss the effect of confinement on the folding free energy as a function of the size and shape of a cage. The confined space should increase the folding stability of the proteins; this fits experimental observations. The question of crowding effects on protein folding and aggregation has also been addressed by Kinjo et al. They develop a density functional theory as a theoretical framework to describe the static and dynamic properties of proteins in a crowded environment [105,106]. The phase diagrams calculated by Kinjo et al. not only include states for protein folding, but also aggregated states. The level of crowding agents and protein densities can adjust the phase separation between these two populated states. Kinetically, crowding agents can accelerate the onset of protein aggregation and protein folding. Too high a level of crowding, however, can inhibit both folding and aggregation.

Several groups have begun to investigate the effects of confinement and crowding on protein folding, using computer simulations to better approximate folding conditions in vivo. Using minimalist models with Gō-like potentials (only native contacts are energetically favored), Klimov et al. [107] studied the confinement effect on both the kinetic and thermodynamic aspects of protein folding (in Fig. 4(b)). Thermodynamically, confinement enhanced the stability of the β-hairpin formation due to a decrease in the configuration entropy of the unfolded states. Kinetically, the encapsulation generally accelerates folding rates. However, there is a non-monotonic dependence of folding rates on the size of the confinement sphere, which reflects the changes in the mechanisms of β-hairpin formation from that of the bulk phase. Friedel et al. [108] used minimalist models with a more energetically frustrated Hamitonian. They observed that in addition to impacting the folding temperatures, the collapse temperatures also depends on the condition of confinement.

5. Energy landscape theory and function (discussion)

From a thermodynamics perspective, the evolution of protein dynamics on the free energy surface can be described by several order parameters, including the fraction of native contacts formed, radius of gyration, RMSD (root-mean-square deviation), or any parameter that can unambiguously measure the similarity of the ensemble structures to the native state. For a good folder, the 'bumps' on the funnel-like energy landscape due to frustration from kinetic or energetic traps are sufficiently small. Hence there is a general agreement that in fast-folding proteins, the folding dynamics must occur on a minimally frustrated energy landscape that probably has been selected by evolution. Proteins, however, have to perform other biologically relevant functions in addition to folding. Moving towards function, we may need to use order parameters that allow us to characterize physical interactions and sort out relevant conformations, which may reveal physical interactions that assist molecular function. For example, we were able to discriminate some protein conformations sensitive to the structure of hydration layers in the folding events by using an additional order parameter that describes the fraction of single-water-separated contacts to partition the folding energy landscape. This information gave us a better understanding of the sensitivity of protein dynamics in the native state basin to solvent perturbation. Likewise, if we can gain a better physical intuition of how proteins behave in the cellular environment, we can try to integrate these ideas into the folding energy landscape, and use it as a framework to potentially extract essential information for making connections to function. These possible mechanisms should then be tested against theoretical observations to determine if they are functionally relevant. In this combined effort, analyses of both experimental measurements and theoretical calculations/simulations have the potential to increase our understanding of the mechanism of protein folding dynamics in the cellular environment.

6. Conclusion

Studies over the past decade have focused on the establishment of frameworks of modeling and theories in tackling the protein folding problem. With greater computational power and faster algorithms, we now have a better understanding of the science of simulating particles/polymers as a means to reproduce protein dynamics. Whereas all-atom simulations are still too time-consuming to

complete folding trajectories, minimalist protein representations remain popular not only because they fold in a manageable amount of time, but also because they can bring a clear physical intuition to experimental details.

The next challenge is how to further develop minimalist models that would achieve protein dynamics that are functionally relevant and potentially applicable in the cellular environment, particularly since there is little information of structural dynamics revealed by in vivo experiments. Albeit unknown factors, we still wish to extend such studies, based on what we have already learned from in vitro observations. It will remain to be seen if the validity of theories developed in response to in vitro results holds in vivo. Provided that biomolecules are well-designed to function, however, we believe that the funnel concept of the energy landscape is still pertinent for protein dynamics in vivo, because the Levinthal-like [109] view by which all possible conformations are equally permissible simply would not satisfy biological purposes which demand specificity of interactions. In response to specificity, proteins may exploit different conformations that deviate from the native state. To elaborate this point, understanding the protein landscape should go beyond the folding of a single native state and relate protein dynamics to function, where traversals between native and near-native states should not be overlooked. For example, recent NMR studies have shown conformational changes in RIIα, a type of docking and dimerization domain (D/D) of Protein Kinase A (PKA), when its ligand (any of the A-kinase anchoring proteins: AKAP) binds [110]. To gain insight to these changes, we [111] utilize several toy models of protein-ligand interactions to better understand how the specificity of RIIα-AKAP can possibly be justified by exploring different states on the energy landscape. As an attempt to address protein dynamics to function, an extended view of the 'landscape and function' relationship should be established, instead of the 'structure and function' one in which discussions of function focus on a single native state.

In recent years, high-resolution experimental techniques have added important details to the protein dynamics: an understanding of protein folding under physiological conditions is now possible. We hope that the integrated efforts of theoretical and computational studies with future cellular experiments will connect the energy landscape with function and help identify the landscape features required for function, in addition to those pertinent for folding. Current progress in computer methods provides a link between analytical theories and experiments, and plays a pivotal role in testing the molecular mechanisms behind complex macroscopic phenomena, perhaps revealing the molecular basis for biological function.

Acknowledgements

Work in San Diego was funded by the NSF (biocomplexity program MCB-0083704 with additional support from MCB-0084797, PHY-0216576 and 0225630). M.S. Cheung additionally extends her gratitude to the Sloan Foundation/DOE for a postdoctoral fellowship (2003-7-3 CMB). L.L. Chavez thanks the NIGMS (a division of NIH) for the MARC Predoctoral Fellowship.

References

[1] Onuchic JN, Nymeyer H, Garcia AE, Chahine J, Socci ND. Adv Protein Chem 2000;53:87–152.
[2] Bryngelson JD, Wolynes PG. Proc Natl Acad Sci USA 1987;84: 7524–8.
[3] Shea JE, Brooks III CL. Annu Rev Phys Chem 2001;52:499–535.
[4] Leopold PE, Montal M, Onuchic JN. Proc Natl Acad Sci USA 1992; 89:8721–5.
[5] Eaton WA, Munoz V, Thompson P, Chan CK, Hofrichter J. Curr Opin Struct Biol 1997;10–14.
[6] Telford JR, Wittung-Stafshede P, Gray HB, Wrinkler JR. Acc Chem Res 1998;31:755–63.
[7] Gruebele M. Ann Rev Phys Chem 1999;50:485–516.
[8] Socci ND, Onuchic JN, Wolynes PG. Proteins 1998;32:136–58.
[9] Onuchic JN, Luthey-Schulten Z, Wolynes PG. Annu Rev Phys Chem 1997;48:545–600.
[10] Jackson SE, elMasry N, Fersht AR. Biochemistry 1993;32:11270–8.
[11] Jager M, Nguyen H, Crane JC, Kelly JW, Gruebele M. J Mol Biol 2001;311:373–93.
[12] Otzen DE, Itzhaki LS, elMasry NF, Jackson SE, Fersht AR. Proc Natl Acad Sci USA 1994;91:10422–5.
[13] Nymeyer H, Socci ND, Onuchic JN. Proc Natl Acad Sci USA 2000; 97:634–9.
[14] Nymeyer H, Garcia AE, Onuchic JN. Proc Natl Acad Sci USA 1998; 95:5921–8.
[15] Ozkan SB, Bahar I, Dill KA. Nat Struct Biol 2001;8:765–9.
[16] Levitt M, Warshel A. Nature 1975;253:694–8.
[17] Friedrichs MS, Goldstein RA, Wolynes PG. J Mol Biol 1991;222: 1013–34.
[18] Irback A, Potthast F. J Chem Phys 1995;103:10298–395.
[19] Guo Z, Thirumalai D, Honeycutt JD. J Chem Phys 1992;92:525–35.
[20] Shea JE, Onuchic JN, Brooks III CL. Proc Natl Acad Sci USA 1999; 96:12512–7.
[21] Sorenson JM, Head-Gordon T. Proteins 2002;46:368–79.
[22] Karanicolas J, Brooks III CL. Protein Sci 2002;11:2351–61.
[23] Miyazawa S, Jernigan RL. J Mol Biol 1996;256:623–44.
[24] Cheung MS, Garcia AE, Onuchic JN. Proc Natl Acad Sci USA 2002; 99:685–90.
[25] Portman JJ, Takada S, Wolynes PG. J Chem Phys 2001;114: 4702–16.
[26] Karanicolas J, Brooks III CL. Proc Natl Acad Sci USA 2003;100: 3954–9.
[27] Guo Z, Brooks CL. Biopolymers 1997;42:745–57.
[28] Udea Y, Taketomi H, Go N. Biopolymers 1978;17:1531–48.
[29] Clementi C, Nymeyer H, Onuchic JN. J Mol Biol 2000;298:937–53.
[30] Clementi C, Jennings PA, Onuchic JN. Proc Natl Acad Sci USA 2000;97:5871–6.
[31] Koga N, Takada S. J Mol Biol 2001;313:171–80.
[32] Papoian GA, Wolynes PG. Biopolymers 2003;68:333–49.
[33] Levy Y, Wolynes PG, Onuchic JN. Proc Natl Acad Sci USA 2003. In press.
[34] Shoemaker BA, Portman JJ, Wolynes PG. Proc Natl Acad Sci USA 2000;97:8868–73.
[35] Riddle DS, Grantcharova VP, Santiago JV, Alm E, Ruczinski I, Baker D. Nat Struct Biol 1999;6:1016–24.
[36] Plaxco KW, Simons KT, Baker D. J Mol Biol 1998;277:985–94.

[37] Fernandez A, Colubri A, Berry RS. Proc Natl Acad Sci USA 2000; 97:14062–6.
[38] Chiti F, Taddei N, White PM, Bucciantini M, Magherini F, Stefani M, Dobson CM. Nat Struct Biol 1999;6:1005–9.
[39] Grantcharova VP, Riddle DS, Santiago JV, Baker D. Nat Struct Biol 1998;5:714–20.
[40] Grantcharova VP, Riddle DS, Baker D. Proc Natl Acad Sci USA 2000;97:7084–9.
[41] Martinez JC, Pisabarro MT, Serrano L. Nat Struct Biol 1998;5: 721–9.
[42] Martinez JC, Serrano L. Nat Struct Biol 1999;6:1010–6.
[43] Clarke J, Cota E, Fowler SB, Hamill SJ. Struct Fold Des 1999;7: 1145–53.
[44] Scalley ML, Yi Q, Gu H, McCormack A, Yates III JR, Baker D. Biochemistry 1997;36:3373–82.
[45] Kim DE, Fisher C, Baker D. J Mol Biol 2000;298:971–84.
[46] Clementi C, Garcia AE, Onuchic JN. J Mol Biol 2003;326:933–54.
[47] Shimada J, Shakhnovich EI. Proc Natl Acad Sci USA 2002;99: 11175–80.
[48] Song G, Thomas S, Dill KA, Scholtz JM, Amato NM. Pac Symp Biocomput 2003;240–51.
[49] Sorenson JM, Head-Gordon T. J Comput Biol 2000;7:469–81.
[50] Viguera AR, Blanco FJ, Serrano L. J Mol Biol 1995;247:670–81.
[51] Viguera AR, Serrano L, Wilmanns M. Nat Struct Biol 1996;3: 874–80.
[52] Martinez JC, Viguera AR, Berisio R, Wilmanns M, Mateo PL, Filimonov VV, Serrano L. Biochemistry 1999;38:549–59.
[53] Li L, Shakhnovich EI. J Mol Biol 2001;306:121–32.
[54] Otzen DE, Fersht AR. Biochemistry 1998;37:8139–46.
[55] Weikl TR, Dill KA. J Mol Biol 2003;332:953–63.
[56] Clementi C, Jennings PA, Onuchic JN. J Mol Biol 2001;311: 879–90.
[57] Lindberg M, Tangrot J, Oliveberg M. Nat Struct Biol 2002;9: 818–22.
[58] Fersht AR. Structure and Mechanism in Protein Science. San Francisco: W.H. Freeman & Co.; 1999.
[59] Thirumalai D, Lorimer GH. Ann Rev Biophys Biomol Struct 2003; 30:245–69.
[60] Plotkin SS, Onuchic JN. Proc Natl Acad Sci USA 2000;97:6509–14.
[61] Chavez LL, Onuchic JN, Clementi C. 2003, in progress.
[62] Otzen DE, Oliveberg M. J Mol Biol 2002;317:613–27.
[63] Otzen DE, Oliveberg M. Proc Natl Acad Sci USA 1999;96: 11746–51.
[64] Otzen DE, Kristensen O, Oliveberg M. Proc Natl Acad Sci USA 2000;97:9907–12.
[65] Stoycheva AD, Onuchic JN, Brooks CL. J Chem Phys 2003;119: 5722–9.
[66] Honeycutt JD, Thirumalai D. Proc Natl Acad Sci USA 1990;87: 3526–9.
[67] Honeycutt JD, Thirumalai D. Biopolymers 1992;32:695–709.
[68] Hummer G, Garde S, Garcia AE, Paulaitis ME, Pratt LR. Proc Natl Acad Sci USA 1998;95:1552–5.
[69] Hillson N, Onuchic JN, Garcia AE. Proc Natl Acad Sci USA 1999; 96:14848–53.
[70] Garcia AE, Hillson N, Onuchic JN. Prog Theor Phys Suppl 2000; 138:282–91.
[71] Guo C, Cheung MS, Levine H, Kessler DA. J Chem Phys 2002;116: 4353–65.
[72] Kaya H, Chan HS. J Mol Biol 2003;326:911–31.
[73] Zhang O, Forman-Kay JD. Biochemistry 1997;36:3959–70.
[74] Shea JE, Onuchic JN, Brooks III CL. Proc Natl Acad Sci USA 2002; 99:16064–8.
[75] Fernandez-Escamilla AM, Cheung MS, Vega MC, Wilmanns M, Onuchic JN, Serrano. 2003 submitted.
[76] Shimada J, Kussell EL, Shakhnovich EI. J Mol Biol 2001;308: 79–95.
[77] Linhananta A, Zhou Y. J Chem Phys 2002;117:8983–95.
[78] Kussell E, Shimada J, Shakhnovich EI. Proc Natl Acad Sci USA 2002;99:5343–8.
[79] Leach AR. Molecular modelling: principles and applications. London: Longman; 1996.
[80] Branden C, Tooze J. Introduction to protein structure. New York: Garland; 1991.
[81] Vieth M, Kolinski A, Brooks CL, Skolnick J. J Mol Biol 1995;251: 448–67.
[82] Takada S, Luthey-Schulten Z, Wolynes PG. J Chem Phys 1999;110: 11616–29.
[83] Klimov DK, Thirumalai D. Proc Natl Acad Sci USA 2000;97: 2544–9.
[84] Irback A, Sjunnesson F, Wallin S. Proc Natl Acad Sci USA 2000;97: 13614–8.
[85] Liwo A, Czaplewski C, Pillardy J, Scheraga HA. J Chem Phys 2001; 2323–47.
[86] Hardin C, Luthey-Schulten Z, Wolynes PG. Proteins 1999;34: 281–94.
[87] Cheung MS, Finke JM, Callahan B, Onuchic JN. J Phys Chem B 2003;107:11193–200.
[88] Buchete NV, Straub JE, Thirumalai D. J Chem Phys 2003;118: 7658–71.
[89] Czaplewski C, Rodziewicz-Motowidlo S, Liwo A, Ripoll DR, Wawak RJ, Scheraga HA. Protein Sci 2000;9:1235–45.
[90] Shimizu S, Chan HS. J Chem Phys 2001;115.
[91] Eastwood MP, Wolynes PG. J Chem Phys 2001;114:4702–16.
[92] Ding F, Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI. Biophys J 2002;83:3525–32.
[93] Socci ND, Onuchic JN, Wolynes PG. Proc Natl Acad Sci USA 1999; 96:2031–5.
[94] Klimov DK, Thirumalai D. Proc Natl Acad Sci USA 1999;96: 6166–70.
[95] Li FY, Yuan JM, Mou CY. Phys Rev E Stat Nonlin Soft Matter Phys 2001;63:021905.
[96] Klimov DK, Thirumalai D. Proc Natl Acad Sci USA 2000;97: 7254–9.
[97] Gao M, Lu H, Schulten K. Biophys J 2001;81:2268–77.
[98] Rief M, Gautel M, Oesterhelt F, Fernandez JM, Gaub HE. Science 1997;276:1109–12.
[99] Carrion-Vazquez M, Oberhauser AF, Fowler SB, Marszalek PE, Broedel SE, Clarke J, Fernandez JM. Proc Natl Acad Sci USA 1999; 96:3694–9.
[100] Evans E, Ritchie K. Biophys J 1997;72:1541–55.
[101] Evans E, Ritchie K. Biophys J 1999;76:2439–47.
[102] Ellis RJ. Curr Opin Struct Biol 2001;11:114–9.
[103] Minton AP. Curr Opin Struct Biol 2000;10:34–9.
[104] Zhou HX, Dill KA. Biochemistry 2001;40:11289–93.
[105] Kinjo AR, Takada S. Phys Rev E Stat Nonlin Soft Matter Phys 2002; 66:031911.
[106] Kinjo AR, Takada S. Phys Rev E Stat Nonlin Soft Matter Phys 2002; 66:051902.
[107] Klimov DK, Newfield D, Thirumalai D. Proc Natl Acad Sci USA 2002;99:8019–24.
[108] Friedel M, Sheeler DJ, Shea J-E. J Chem Phys 2003;118:8106–13B.
[109] Dill KA, Chan HS. Nat Struct Biol 1997;4:10–19.
[110] Newlon MG, Roy M, Morikis D, Hausken ZE, Coghlan V, Scott JD, Jennings PA. Nat Struct Biol 1999;6:222–7.
[111] Cheung MS, Yang S, Onuchic JN. 2003, in progress.