Review

Protein folding: Then and now

Yiwen Chen [1], Feng Ding [1], Huifen Nie [1], Adrian W. Serohijos [1], Shantanu Sharma [1],
Kyle C. Wilcox [1], Shuangye Yin [1], Nikolay V. Dokholyan [*,1]

Department of Biochemistry and Biophysics, The University of North Carolina at Chapel Hill, School of Medicine, Chapel Hill, NC 27599, USA

Abstract

Over the past three decades the protein folding field has undergone monumental changes. Originally a purely academic question, how a protein folds has now become vital in understanding diseases and our abilities to rationally manipulate cellular life by engineering protein folding pathways. We review and contrast past and recent developments in the protein folding field. Specifically, we discuss the progress in our understanding of protein folding thermodynamics and kinetics, the properties of evasive intermediates, and unfolded states. We also discuss how some abnormalities in protein folding lead to protein aggregation and human diseases.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Protein folding; Misfolding; Aggregation; Folding kinetics; Protein thermodynamics; Discrete molecular dynamics (DMD)

Protein folding refers to the process by which a protein assumes its characteristic structure, known as the native state. The most fundamental question of how an amino-acid sequence specifies both a native structure and the pathway to attain that state has defined the protein folding field. Over more than four decades the protein folding field has evolved (Fig. 1), as have the questions pertaining to it. This evolution can be divided into two predominant phases. During the first phase, research was focused on understanding the mechanisms of protein folding and uncovering the fundamental principles that govern the folding transition. While the first phase provided general answers to the protein folding question, new and no less ambitious questions arose: what are the mechanisms of protein folding in a context, such as under the influence of other biological molecules in the cellular environment? This next set of questions defined the second phase in protein folding field evolution.

The first phase is akin to a romantic stage of research, where the final goal of studies may not be directly applicable to a broader understanding, or exploitable in a relevant science. The final goal is to determine the basic principles that relate protein sequence and structure. The second phase is a more pragmatic stage of research, where the applications drive research in the field and the rational manipulation of derived knowledge allows engineering of tools for advancement of a relevant science. For example, understanding the functional intermediates that accompany the transition of a protein *en route* to its native state may allow rational manipulation of protein structure via protein design. This example not only relates protein sequence, structure and function, but also demonstrates the engineering aspect of the modern protein folding field.

Next, we survey the questions of the modern protein folding field. We attempt to describe a number of directions where understanding protein folding offers insights into more complex questions in molecular and cellular biology as well as medicine. We also describe new approaches and tools to address complexities associated with these new areas of research. We review studies of protein stability, folding kinetics, intermediate and unfolded states, and protein self-association and aggregation.

---

[*] Corresponding author. Fax: +1 919 966 2852.
E-mail address: dokh@med.unc.edu (N.V. Dokholyan).
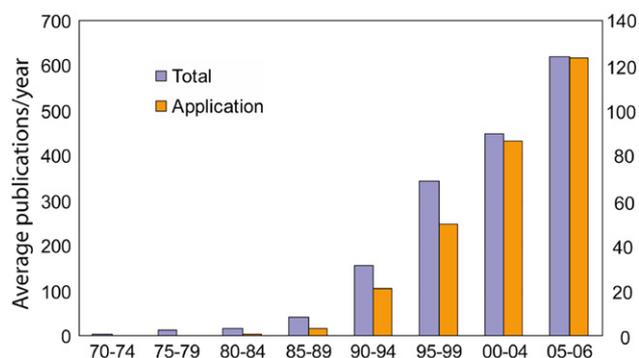[1] All authors contributed equally to this work.

Fig. 1. Growth of the Protein Folding Field. The average number of publications per year in protein folding field (left *y* axis) and the average number of publications per year that are dedicated to application (right *y* axis) were plotted every five years between 1970 and 2004, and 2005–2006. The first dataset was generated by searching articles in PubMed that contain the keyword 'protein folding' or 'protein unfolding' in either title or abstract. The second dataset was extracted from the previous dataset by searching with the following additional keywords: 'engineering', 'design', 'misfolding', 'aggregation', 'amyloid' and 'amyloid disease'.

Studying protein folding

The protein folding field has witnessed significant changes and progress since the original work of Anfinsen showing that proteins can fold spontaneously [1,2]. Early *in vitro* studies showed that the folding process typically occurs on a milliseconds-to-seconds time scale, much faster than the rate estimated assuming that folding proceeds by a random search of all possible conformations. Based upon this observation, Levinthal then proposed that a random conformation search does not occur in folding and that proteins fold by specific 'folding pathways' [3]. On these pathways, the protein molecule passes through well-defined partially-structured intermediate states. Based on this view, numerous experiments and simulations were conducted to test the existence of transient folding intermediates [4,5]. It was expected that the determination of the structures and population of folding intermediates could help elucidate protein folding mechanisms. Earlier experimental studies on protein folding kinetics monitored the structural changes through relaxation of the protein's spectroscopic properties after exposing the protein to folding or unfolding conditions. The data obtained from such experiments exhibit single- or multiple-exponential time-decay: a single-exponential decay is interpreted as a signature of two-state kinetics between the native state and the denatured state, whereas models involving more than two states are required to explain multiple-exponential decay data. These experiments generally probe only the average behavior of proteins, and they are not able to provide information about the folding/unfolding process in atomic details.

The discovery of a class of simple, single-domain proteins which fold via two-state kinetics without any detectable intermediates in the early 1990s [6,7], the development of experimental techniques with improved spatial/temporal resolution [8–13], and the application of computer simulations using simplified lattice and off-lattice models [14,15] greatly enhanced our understanding of various aspects of the protein folding problem. Based on the nucleation theory [16–18], one of the early proposed mechanisms for protein folding, the nucleation-condensation model was formulated [19–21]. In this scenario, a small number of residues (folding nucleus) need to form their native contacts in order for the folding reaction to proceed fast into the native state. The cooperativity of the protein folding process is analogous to that exhibited in first-order phase transitions, which proceed via a nucleation and growth mechanism [22]. Because of these similarities, terminology used in studies of phase transitions, such as energy landscapes and nucleation, was introduced into the discussion of protein folding. The concepts of the nucleation and the free-energy landscape have promoted much of the recent progress in understanding the process of protein folding. Proteins are generally thought to have evolved to exhibit globally funneled energy landscapes [23–25] which allow proteins to fold to their native states through a stochastic process in which the free energy decreases spontaneously. The unfolded state, transition state, native state and possible intermediates correspond to local minima or saddle points in the free-energy landscape.

Advances in experimental techniques such as protein engineering, nuclear magnetic resonance (NMR)[2], mass spectrometry, hydrogen exchange, fluorescence resonance energy transfer (FRET), and atomic force microscopy (AFM), have made it possible to obtain detailed information about the different conformations occurring in the folding process [26,27]. At the same time, computational methods have been developed to better interpret experimental data by using simulations to obtain structural information about the states which are populated during the folding process. In Table 1, we list several advances in experimental and computational methodologies used for investigating the folding of model proteins.

All-atom protein models with explicit or implicit solvents were developed to study the folding thermodynamics and the unfolding dynamics of specific proteins. Technological advances in computation allowed folding simulations of small proteins and peptides at atomic detail [28–30]. However, due to the complexity and vast dimensionality of protein conformational space, all-atom MD simulations have severe limitations on the time and length scales that can be studied. Novel simulation protocols have been proposed to improve conformational sampling

---

[2] *Abbreviations used:* DMD, discrete molecular dynamics; NMR, nuclear magnetic resonance; FRET, fluorescence resonance energy transfer; AFM, atomic force microscopy; MC, Monte Carlo; GdHCl, guanidinium HCl; DFIRE, distance-scaled finite ideal-gas reference state; CFTR, cystic fibrosis transmembrane conductance regulator; FALS, familial amyloid sclerosis; TSE, transition state ensemble; HX, hydrogen exchange; FAT, focal adhesion targeting; FAK, focal adhesion kinase; SMD, steered-molecular dynamics; ECM, extracellular matrix.

Table 1
Protein folding in select model systems

Theme	Relevant experiments and theoretical models	Contribution to protein folding research
Mechanism of folding/ unfolding	Equilibrium denaturation by guanidinium chloride [6]	CI2 unfolding and refolding follows two-state transition. Lysozyme folding intermediates obstruct formation of transition state, but does not change the folding rate [248]
	Protein engineering, Φ-value analysis [249]	CI2 folding supports nucleation-condensation model [19,20]. CI2 transition state has secondary, tertiary structure elements [121]. Native topology and hydrogen bonds mediates SH3 folding [250,251]
	Quantitative Φ-value analysis using MD simulations [122]	Postulated structure of CI2 unfolding transition state
	Free-energy landscape: protein folding funnel [23]	Statistical description of protein folding process. Role of water in facilitating protein folding [155]
Transition state structure	MD simulations of CI2 transition state [252]	Folding of CI2 is cooperative
	Multi-scalar modeling and DMD simulations	Identification of src SH3 residues critical to folding nucleus [47]
	Monte Carlo simulations on lattice models [253]	p_{fold} as a reaction coordinate for protein folding
Folding kinetics	Equilibrium and stopped-flow fluorescence [254]	Src SH3 unfolding is cooperative; its denatured state may be compact under native conditions [254]
	Relaxation dispersion NMR of Fyn SH3 [174]	Identified and characterized low-population folding intermediates
	Graph representation of CI2 and src-SH3 conformation [255]	Protein network contact topology determines proteins' ability to fold
	Φ-value analysis of SH3 [256]	Hydrophobic core composition is another determinant of protein folding rate
	Multiple MD simulations of CI2 unfolding [257]	Preferred pathway for protein folding on a funnel-like average energy surface
Unfolded proteins structure	Unfolded proteins' NMR [195]. Size of measurements of various unfolded proteins [192]	Denatured proteins have a strong local conformational bias towards native state. On the other hand, the scaling of protein sizes in the unfolded state suggests a random-coil like conformations
	Computational models of denatured proteins [192,203]	Unfolded states features local native-like structures (short-range correlations), but the correlations decays quickly. Protein behaves as a "renormalized" random coil after grouping local structures together

A sampling of experimental and theoretical approaches for probing protein folding for three model protein systems: Chymotrypsin Inhibitor 2 (CI2), Lysozyme, and src/Fyn SH3 Domain.

efficiency, including biased sampling of the free-energy surface and non-equilibrium unfolding simulations [24]. In addition, world-wide parallel computing (e.g., Folding@Home [31]) and generalized ensemble sampling techniques that involve parallel simulations of molecular systems coupled with a Monte Carlo (MC) protocol [32,33] have been successfully applied to protein folding [25,34–36].

Multi-scale modeling approaches have also been used to combine efficient conformational sampling of coarse-grained models and accuracy of all-atom models to study protein folding pathways. In this approach, iterative simulations and inter-conversion between high and low-resolution protein models are performed. Feig et al. developed a multi-scale modeling tool set, MMTSB [37], which integrates a simplified protein model with the MC simulation engine, MONSSTER [38], and the all-atom MD packages AMBER [39] or CHARMM [40]. Using a combination of CHARMM and discrete molecular dynamics (DMD) [41–46], Ding et al. reconstructed the transition state ensemble of the src-SH3 protein domain through multi-scale simulations [47]. The protein folding studies can also be facilitated by sampling protein conformations near the native state. Several native-state sampling algorithms [48,49] have been successfully utilized to study plasticity [50], cooperative interactions [51], and allostery [52] in pro-

teins. Considering native-state ensemble naturally takes into account protein flexibility, which is shown to be crucial in structure based drug designs.

During the last five years, several tools for performing web-based analyses of protein folding dynamics have been developed. The Fold-Rate server (http://psfs.cbrc.jp/fold-rate/) [53] predicts rates of protein folding using the amino-acid sequence. The Parasol folding server (http://parasol.tamu.edu/groups/amatogroup/foldingserver) [54] predicts protein folding pathways using "probabilistic roadmaps"-based motion planning techniques. The iFold server (http://ifold.dokhlab.org) [55] allows discrete molecular dynamics (DMD) simulations of protein dynamics using simplified two-bead per residue protein models. These tools facilitate the second phase of protein folding research, whereby targeted simulations may be performed for probing the dynamics of protein folding and unfolding under controlled conditions.

DMD approaches [43–46] with simplified structural models of proteins have been extensively used for investigating general principles of protein folding and unfolding [56–60]. Dokholyan et al. [61] have highlighted the differences between molecular dynamics and DMD approaches. As opposed to the traditional MD approach of iteratively solving Newtonian equations of motion for evolving protein folding trajectory, DMD simulations solve ballistic

equations of motion with square-well approximation to inter-particle interaction potentials. DMD algorithm gains efficiency over traditional MD simulations in multiple ways. First, due to ballistic modeling of particle dynamics, a larger time step can be used in DMD simulations on average, which corresponds to the time interval between fastest ballistic interactions; secondly a faster inter-particle collision detection and velocity updating algorithm is used, since only the coordinates of colliding atoms need to be updated at each collision. Additionally, faster simulation speeds are attainable with the DMD approach through simplification of protein models. Overall, an increase in simulation speed of 5–10 orders of magnitude is attainable using DMD [62]. Jang et al. [56] used DMD and simplified protein models with Gō interactions [63] to probe protein folding kinetics. Protein folding kinetics studies using DMD simulations are reviewed in [42]. Recently, DMD simulations we used in uncovering the structural mechanisms of protein aggregation [64–66]. Among the fundamental challenges in studying protein folding using computer simulations are the time scales and length scales that can be investigated. DMD simulations have been shown to be useful for investigating long-time scale folding dynamics of complex biological systems such as poly-alanine aggregation [66,67] and the nucleosome core particle [68].

In addition to the extensive *in silico* and *in vitro* studies of protein folding, significant progress has been made in understanding protein folding *in vivo*. There are two major differences between protein folding *in vivo* and *in vitro*. First, protein folding *in vivo* is usually assisted by molecular machinery, such as chaperones (in an ATP-dependent manner), and often involves small molecule cofactors. Molecular chaperones such as the heat shock protein Hsp70 and chaperonin proteins facilitate protein folding, in part, by isolating the proteins from bulk cytosol [69,70]. Hartl and Horwich pioneered the research of chaperone-mediated protein folding [71], highlighting the differences between *in vivo* and *in vitro* folding mechanisms [69]. The mechanism of chaperonin GroEL mediated folding, including *in vivo* folding intermediates, has been extensively studied by Horwich and Gierasch [72,73]. Work by Landry and Gierasch [74] showed that chaperon binding promotes α-helix formation in partially folded polypeptide chains. Horowitz et al. have investigated the role of chaperonin Cpn60-mediated hydrophobic exposure in protein folding [75,76]. Nearly one third of all proteins in living cells are coordinated to small molecule cofactors. The pioneering work of Wittung-Stafshede and coworkers on the role of cofactors in *in vivo* protein folding [77,78] demonstrated that bound metals stabilize the native fold, suggesting cofactor binding to unfolded polypeptides dramatically accelerates folding time scales [77].

A second notable difference between *in vivo* and *in vitro* protein folding is the fact that the concentrations of macromolecular solutes in cells can reach hundreds of grams per liter [79], but most *in vitro* studies are performed in buffered solution with <1% of the cellular macromolecule concentration. The crowding environment *in vivo* can have a sig-nificant impact on protein stability and the native structure by changing the energy landscape of protein folding [80,81]. Dedmon et al. [82] showed that FlgM, a 97-residue protein from *Salmonella typhimurium* is unstructured in dilute solution, but in *Escherichia coli* cells its C-terminal half is structured. McPhie et al. [83] found that a molten globular state of apomyoglobin at low pH is stabilized by high concentration of the inert polymer, dextran, compared to the unfolded state. Moreover, it was found that aggregate formation from human apolipoprotein C-II is significantly accelerated by the addition of dextran [84], suggesting a direct effect of molecular crowding on protein aggregation.

Over the past three decades, novel experimental techniques and simulations have yielded many significant insights in protein folding research. Important advances have been made, especially toward the understanding of folding and unfolding mechanisms, the structure of folding transition states, folding kinetics, the nature of folding pathways, and the structure of unfolded proteins and protein folding *in vivo*. Theoretical approaches to study protein folding have largely complemented experiments by providing experimentally testable hypotheses. In recent years, the rational manipulation of folding pathways and the association between protein folding and disease have marked a more applied phase of protein folding research.

Protein stability

The thermodynamic stability of a protein is measured by the free-energy difference between the folded state and the unfolded state ($\Delta G = G_{\text{unfold}} - G_{\text{fold}}$). It determines the fraction of folded proteins, thereby having a profound effect on protein function. Natural proteins are only marginally stable [85]. The energetic contributions from the favorable folding forces such as hydrophobic packing, hydrogen-bonding and electrostatic interactions are nearly offset by the entropic penalization of folding. As a result, the measured the ΔG values of most proteins fall in the range of 3–15 kcal/mol [86]. Due to this subtle balance between various physical interactions, a single mutation may shift the balance and significantly affect the stability of the whole protein. The accurate estimation of protein stability changes induced by mutations, measured as $\Delta\Delta G = \Delta G^{\text{WT}} - \Delta G^{\text{Mut}}$, still remains a significant challenge for computational biologists.

Experimentally, ΔG values can be obtained from denaturing experiments [6,87–90] where the protein unfolds by increasing temperature or by adding denaturing agents such as urea and guanidinium HCl (GdHCl). Theoretically, given the interactions, free energy can be obtained from statistical mechanics using the partition function, Z, as $G = -RT \ln Z$. Analytical calculations of partition functions require integration over all degrees of freedom in the protein's conformational space, which is impossible in practice, except for simple models. Advances in computational biology have made possible the direct

calculation of $\Delta\Delta G$ by MD or Monte Carlo simulations [28,91–94], a comprehensive review of which can be found in Ref. [95]. Using MD simulations, a $\Delta\Delta G$ value has been calculated for T157V mutant of T4 lysozyme which is in close agreement with experimental measurements [92].

The computational cost of direct $\Delta\Delta G$ estimation is still too high, however, preventing it from being applied to a large number of mutations for protein engineering. More heuristic approaches have been adapted which try to describe the free energy using empirical or effective functions taking advantage of the vast amount of known protein structures and stability measurements. Such simplifications significantly decrease the computational overhead and allow $\Delta\Delta G$ calculations of large numbers of mutants that can be compared with experimental results. In recent years, various methods have been proposed for large-scale $\Delta\Delta G$ predictions with reasonable prediction accuracies (which are commonly assessed by the linear correlation coefficient between the predicted and measured $\Delta\Delta G$ values). A comparison between these methods is listed in Table 2.

One approach utilizes statistical potentials that are developed using information from known protein structures, where the frequency distribution of amino acids conformations (such as pairwise distances and torsion angles) is used to extract effective potentials for free-energy evaluations [96]. Gilis and Rooman first applied database-derived backbone dihedral potentials to study the change of thermodynamic stability upon point mutations [97–99]. They found that torsion-angle potentials predict $\Delta\Delta G$ accurately for mutations of solvent-exposed residues and that distance-dependent statistical potentials are more accurate for predicting the $\Delta\Delta G$ of buried residues. They obtained correlation coefficients of 0.55–0.87 for a dataset of 238 mutations. Zhou and Zhou [100] developed a knowledge-based potential using the distance-scaled finite ideal-gas reference state (DFIRE) approach and calculated $\Delta\Delta G$ for 895 mutants, which have a correlation of 0.67 with experimental measurements. Similarly, statistical potentials utilizing side-chain rotamer libraries [101], direction- and distance-dependent distributions [102], and four-body interactions [103] were also adapted for $\Delta\Delta G$ predictions and significant agreement with experimental measurements was achieved.

Another approach for large-scale $\Delta\Delta G$ predictions uses empirical functions to describe free-energy changes induced by mutations and trains the parameters to recapitulate the experimental results. Guerois et al. [104] developed the FOLD-X energy function to study the stabilities of 1088 mutants. They used a comprehensive set of parameters to describe the van der Waals, solvation, hydrogen-bonding, electrostatic and entropic contribution to the protein stability, and obtained a correlation of 0.64 for the blind test set after training their parameters on 339 mutants [104]. Khatun et al. [105] utilized contact potentials to predict $\Delta\Delta G$ of three sets of 303, 658 and 1356 mutants and their prediction correlations varied between 0.45 and 0.78. Bordner and Abagyan [106] used a combination of physical energy terms, statistical energy terms and a structural descriptor with weight factors scaled to experimental data for $\Delta\Delta G$ predictions, and found a correlation of 0.59 on 908 test mutants. Saraboji et al. classified the available thermal denaturing data on mutations according to substitution types, secondary structures and the area of solvent accessibilities, and used the average value from each category for the prediction and obtained a correlation of 0.64 [107].

Taking advantage of the vast amount of experimental $\Delta\Delta G$ data now available, machine-learning techniques have been introduced for $\Delta\Delta G$ estimation. Capriotti et al. [108,109] trained a support vector machine using temperature, pH, mutations, nearby residues and relative solvent accessible area as input vectors. The support vector machine, when applied to a test set, gives a prediction correlation of 0.71. Cheng et al. [110] improved the support vector machine model to directly include the sequence information and obtained higher correlation prediction accuracy.

There are two significant drawbacks with training-based studies. First, improvement of the prediction accuracy relies on the available experimental stability data for parameter trainings. It is questionable whether parameters obtained from these trainings are transferable to other studies [105] since the experimentally-available mutation data may be biased (e.g., towards substitutions of large residues for small ones). Second, some mutations introduce strains in the protein backbone. To properly estimate the $\Delta\Delta G$ values, it is necessary to estimate the structural rearrangement that a protein undergoes to release the strain. To our knowledge, protein dynamics and flexibility have not been explicitly modeled in previous methods. Ignoring protein flexibility prohibits the application of current prediction methods to a wide range of mutations [100,104].

Table 2
Comparison of different $\Delta\Delta G$ calculation approaches

Methods	Speed	Parameter training	Side-chain flexibility	Backbone flexibility	Transferability
Ab initio simulation	Slow	No	Yes	Yes	Yes
Statistical potential	Fast	No	No	No	Yes
Empirical function	Fast	Yes	No	No	Yes
Machine learning	Fast	Yes	No	No	No

The approaches are evaluated base on the calculation speed, dependence on parameter training using existing stability data, side-chain and backbone flexibility modeling capability, and transferability to study other protein properties.

To address both these caveats, Yin et al. [111] developed a novel method, Eris (http://eris.dokhlab.org), for accurate and rapid evaluation of the $\Delta\Delta G$ values using the recently-developed Medusa redesign suite [112]. Eris features an all-atom force field developed from X-ray crystal structures, a fast side-chain packing algorithm, and a backbone relaxation method. The $\Delta\Delta G$ values of 595 mutants from five proteins were calculated and compared with the experimental data from the Protherm database and other sources [104,105,113,114]. Significant correlations of ≈ 0.75 were found between predicted and experimental $\Delta\Delta G$ values. Eris identifies and efficiently relaxes strains in the backbone, especially when clashes and backbone strains are introduced by a small-to-large amino-acid substitution. Interestingly, when high-resolution structures are not available, Eris allows refinement of the backbone structure, which yields better prediction accuracy. Compared with other $\Delta\Delta G$ prediction methods, the Eris method is a unique approach that combines physical energies with efficient atomic modeling, resulting in fast and unbiased $\Delta\Delta G$ predictions.

Despite remarkable progress in the last several decades, protein stability estimation methods are still imperfect. Several obstacles must be cleared in order to achieve a more reliable method for stability estimation. Due to the complexity of sampling multi-dimensional space, the entropic free energy of a protein is difficult to evaluate and is thereby often ignored or only roughly counted in current stability estimation methods. Furthermore, since protein stability is determined by the free-energy difference between the folded and unfolded states, it is crucial to model the unfolded state and its effect on protein stability (cf. section on unfolded protein states). Most importantly, the prediction of large conformational changes upon mutations remains a major challenge in protein-stability estimation.

Protein folding kinetics

To understand protein folding, important details must be taken into consideration as the protein proceeds from unfolded to native state. How fast does a protein fold? Are there multiple pathways accessed *en route* to its folded state? What are the structural characteristics that determine the path and rate of protein folding? In addition, as we realized in recent years, there is a broad class of human diseases that arises from failure of some proteins to adopt and remain in their native states, partly due to abnormal folding kinetics [115]. For example, a major cystic fibrosis-related deletion mutation of a single amino acid in the cystic fibrosis transmembrane conductance regulator (CFTR) affects its folding kinetics, but has minimal effect on its stability and structure [116]. Also, recent evidence [117] suggests that altered dynamics of superoxide dismutase (SOD1) mutants, which are non-destabilizing or even stabilizing, possibly cause the aggregation of mutants in familial amyloid sclerosis (FALS). Elucidation of protein

folding kinetics has never been more important, particularly in the context of finding molecular mechanisms of protein misfolding diseases.

Probing the structural properties of intermediate states and determining folding pathways have been major experimental challenges. Nonetheless, there have been significant advances in experimental techniques. Experimental methods with different temporal resolutions allow a more detailed dissection of the folding process (Table 3). Other methods allow for investigation of protein structure (e.g., NMR, ultraviolet/visual light CD, site-directed mutagenesis, Φ-value analysis, isotope labeling) and global properties (e.g., mass spectroscopy, quasi-elastic light scattering, and ultracentrifugation) [118].

As experimental data on folding rates of various proteins accumulated over the years, people sought the determinants of folding rates. Plaxco and coworkers [119] showed that there is a high correlation between the folding rate and the structural properties of proteins, as defined by contact order CO, $CO = \frac{1}{L \cdot N} \sum_N \Delta L_{ij}$, where L is the sequence length, N is the total number of inter-residue atomic contacts within a cutoff distance, ΔL_{ij} is the sequence separation of contacting residues i and j. Interestingly, assessment of other protein folding rate determinants such as local and long-range contacts were found to perform equally well as the contact order [53]. Also, contact order is a geometric property which does not take into account the distribution and strength of the interactions on the rate [120]. Fersht and coworkers found that specific interactions in the folding nucleus are equally important determinants of folding rates. For example, mutations in the folding nucleus of CI2 did not change its contact order, but result in a three order magnitude increase in the folding rate [90,121,122]. Sequence-based prediction of folding rates has also been proposed and was found to be of comparable performance to that of contact order [123,124]. Thus, there is still a debate as to whether structure-based or sequence-based prediction is a more reliable predictor of folding rates [123,124].

Extensive studies have likewise been made on the prediction of folding rates using molecular dynamics simulations. The primary bottleneck in this approach was sampling the time scales where the folding transitions are observable. Thus, early studies pioneered by Caflisch et al. [125–127]

Table 3
Methods for observing fast folding events (adapted from [118, p. 77])

Technique	Approximate time scale probed
LASER flash photolysis	100 fs–1 ms
Electron-transfer-induced refolding	1 μs–1 ms
Acoustic relaxation	1 ns–1 ms
Dielectric relaxation	1 ns–1 s
LASER T-jump	1 ns–100 ms
Electrical discharge T-jump	100 ns–10 s
Mixing	10 μs–∞
Pressure-jump	60 μs–1 s
NMR line broadening	100 μs–100 ms

employed continuum solvent with low viscosity to observe multiple folding transitions. However, there is a nonlinear relationship between the folding time and viscosity [128], hence, the precise effect of very low viscosity on the protein folding kinetics of various systems remains unclear. To circumvent this problem, Pande et al. used ''coupled ensemble dynamics'' to simulate the folding of a β-hairpin from protein G using continuum solvent model [129] and united atom force field [130] with water-like viscosity. In this and other subsequent simulations of other β-hairpins, the calculated folding rate was in close agreement with experimental measurements. A remarkable simulation in this class is a 1 μs folding simulation on the villin headpiece by Duan and Kollman [28]. Folding rate predictions of this type have been limited to small two-state proteins. As protein size increases, it is difficult to computationally study folding kinetics. Rate predictions have been likewise performed using molecular dynamics simulations using explicit water models such as the TIP3P [131] and SPC [132] to gain additional insight into folding kinetics. Examples of MD simulations using explicit solvent which yielded experimentally consistent rates were performed by Pande et al., who observed helix-coil transitions [133] and protein folding [34]. However, a potential drawback in the use of water models is that they are parameterized to a single temperature (~298 K), and thus may bias the dynamics in non-native temperature simulations. Overall, theoretical rate determinations and their subsequent comparisons with experiments provide a test of our understanding of protein folding kinetics.

Structural investigation of the transition state ensemble (TSE) is extremely challenging experimentally, since the TSE is an unstable state whose experimental detection is very difficult. Computationally, transition state conformations may be identified using unfolding simulations (cf. section on unfolded protein states), projection into one or two reaction coordinates, validation of putative transition states through calculation of the probability to fold (P_{fold}), and path sampling. The intuitive appeal of low-dimensional energy landscapes in explaining simple chemical reactions inspired people to develop a similar formulation for protein folding. However, unlike simple molecules, the high degrees of freedom of a protein make the analysis formidable. Thus, several groups proposed dimensional reduction by projecting the multi-dimensional energy landscape into few relevant coordinates. The proposed reaction coordinates could be the volume of the molecule [134], the fraction of amino acids in their native conformation [135], the number of contacts between amino acids [136,137], and the fraction of native contacts in a conformation [136,137]. Some others directly tackled the transition state by developing rigorous path sampling techniques [138]. They constructed a large ensemble of transition paths, and through statistical analysis, they determined conformations whose $P_{fold} = 0.5$. Although this method is computationally expensive, it is advantageous since there is no presupposed reaction coordinate.

Characteristics of the transition state ensemble have mainly been investigated by Φ-analysis [139], which involves measuring the folding kinetics and equilibrium thermodynamics of mutants containing amino-acid substitutions throughout a protein. This method provides means to identify interactions mediated by specific amino-acid side chains that stabilize the folding transition state [140]. Φ-analysis has been applied to a large number of proteins (such as BPTI, myoglobin, protein A, ubiquitin, SH3 domain, and the WW domain) and, recently, even to amyloidogenesis [141,142]. However, despite the prevalence of the methodology, there is debate regarding the validity and conventional interpretation of Φ-analysis, especially when the $\Delta\Delta G$ between wild type and mutant is less than 1.7 kcal/mol [2,42,143,144]. However, Fersht and coworkers argued that reliable Φ-values can be derived from mutations in suitable proteins with $0.6 < \Delta\Delta G < 1.7$ kcal/mol [145]. Plaxco and coworkers disproved the assumed independence of the changes in free energy of transition and folded states when calculating error estimates in Φ-values [143]. They proposed a new method of error estimation that accounts for the interdependence of changes in free energy of transition and folded states.

Using simplified protein models and rapid sampling DMD, Dokholyan et al. directly observed and characterized the transition state ensemble of Src homology 3 (SH3) (Fig. 2) [47,59,146,147]. To probe the contribution of each amino-acid residue to the transition state ensemble, they calculated the Φ-values, and found high correlation between simulation and experimental Φ-values. Moreover, they also predicted that the two most kinetically-important residues in folding are L24 and G64. Both L24 and G64 are experimentally-verified to be important kinetically [148].

Experiments suggest that proteins may be kinetically trapped en route to the native state [149–151], but how do proteins avoid kinetic traps? By using a Go-model scaled to include sequence-specific interactions, Khare et al. [152] found that the residues which contribute most
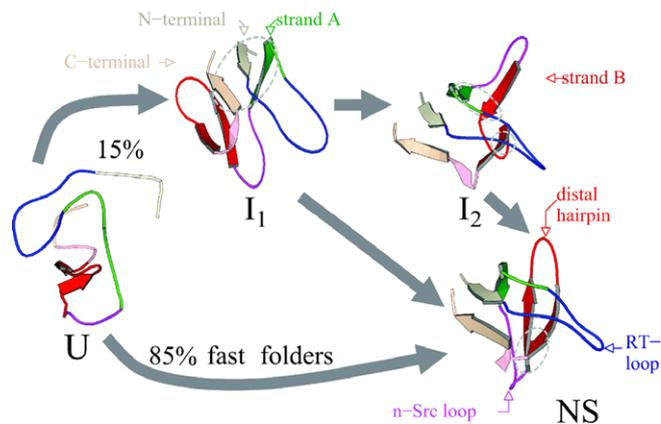

Fig. 2. Fast and slow folding pathways. Molecular dynamics simulations of the c-Crk SH3 folding show multiple folding pathways via only one or two intermediates. [Reprinted with permission from Ref. [258]. Copyright Biophysical Society.]

to Cu, Zn SOD1 stability also function as "gatekeepers" that avoid kinetic traps and protein misfolding. "Gatekeeper" residues were also identified in later computational studies of the ribosomal protein S6. Stoycheva et al. [153] and Matysiak and Clementi [154] show that the mutations of gatekeeper residues can alter the folding landscape of S6 and shift the balance between its folding and aggregation. All these computational studies are fully consistent with experimental observations and the existence of gatekeeper residues suggests a selective pressure on avoiding misfolding in natural proteins.

The close interplay of computational and experimental efforts has advanced our knowledge of protein folding kinetics, including predicting the protein folding rate, identifying the kinetically-important residues, and characterizing the multiple pathways. For example, recent studies have demonstrated an agreement between theoretical and experimental folding free-energy landscapes [154–160]. The characterization of molecular interactions responsible for different pathways opens the possibility to manipulate folding pathways. Current strategies of manipulating the folding pathway includes addition of denaturants, point mutations, and circular permutations [161]. Specifically, Kuhlman and Baker rationally engineered multiple mutations in protein L to alter its folding pathways [162]. Lowe and Itzhaki likewise recently redesigned the folding pathways of the repeat protein myotrophin [163]. Hence, one strategy to tackle kinetics-related folding abnormalities is to rationally engineer the folding pathway after the full characterization of a protein's folding kinetics.

Protein intermediate states

Proteins sample ensembles of heterogeneous conformations in solution. This emerging view of the one-to-many correspondence between protein primary sequence and its possible three-dimensional conformations also challenges the traditional paradigm that protein function is dictated by the native state. In recent years, it has been found that even for the conventionally observed small two-state proteins (∼100 amino acids or less) there exist partially unfolded intermediates on the folding pathways. These intermediates are generally undetectable in kinetic folding experiments [8,116,164–168] and, therefore, are called "hidden intermediates". In addition, mounting evidence has indicated that the intermediate states formed during protein folding and unfolding may have significant roles in protein functions (Fig. 3). The folding and unfolding intermediates impact the physiological functions of proteins by exposing cryptic post-translational modifications or ligand binding sites. The intermediates are usually weakly-populated (thermodynamic intermediates) or short-lived (kinetic intermediates), and their characterization presents a significant challenge with current experimental methods. Recent synergies between computational and experimental studies have greatly facilitated the unprecedented structural char-
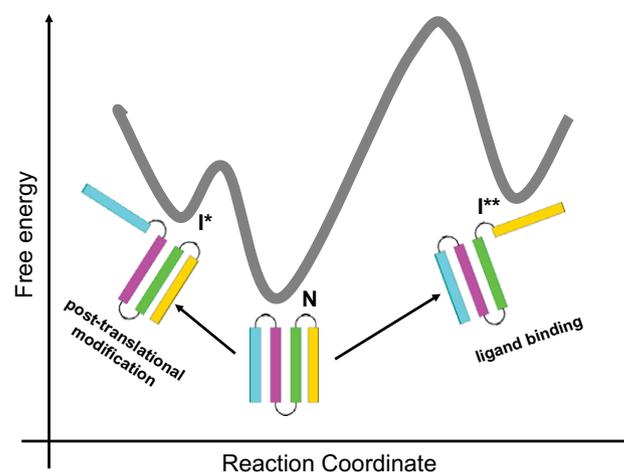

Fig. 3. Protein intermediate states. Intermediate states are meta-stable in protein free-energy landscapes (N: the native state; I*, I**: intermediates). They may play a significant role in protein function, including exposing cryptic post-translational modifications or ligand binding sites.

acterization of rare intermediates and suggested a functional role for these evasive conformations.

Among the traditional experimental methods, hydrogen exchange (HX) [169] is a unique tool that allows the detection and characterization of not only kinetic, but also weakly-populated thermodynamic intermediates. In a typical equilibrium HX experiment, the rate at which an individual main-chain amide hydrogen exchanges with solvent deuterons is measured by NMR or mass spectrometry [170]. The exchange rates of different amide hydrogens are sensitive to local and global structural changes of proteins, and thus contain useful structural information about different protein conformations. From the amide hydrogen exchange rates measured by HX it is possible to detect and characterize weakly-populated intermediates which are inaccessible by bulk methods. Over the years, HX methods have provided considerable insight into the coarse features of intermediate state conformations for a wide variety of proteins [169]. However, HX is limited in its ability to describe the detailed structures of intermediates, which severely restricts its applications. In contrast, computational methods can reveal structural information at much higher resolutions that cannot be accessed by experiments. Also, unlike bulk experimental methods like HX, computational methods are able to study protein motions at the single-molecule level, which enables the detection of heterogeneous conformations in an ensemble of molecules.

In recent studies, Gsponer et al. [171] and Dixon et al. [172] have developed new computational approaches to incorporate HX protection factors from NMR experiments as constraints into MD simulations to detect and/or characterize the conformations of intermediate states. Using their new approach, Gsponer et al. [171] were able to define the thermodynamic folding intermediates of the bacterial immunity protein Ig7. A structural comparison between this thermodynamic intermediate and kinetic intermediate determined from other methods indicates that the kinetic

and thermodynamic intermediates of Ig7 are similar. In a parallel study, Dixon et al. [172] computationally predicted and characterized a folding intermediate of the focal adhesion targeting (FAT) domain of focal adhesion kinase (FAK), which plays critical roles in cell proliferation and migration. The detected intermediate was hypothesized to expose a cryptic phosphorylation site in the FAT domain and regulate the localization of FAK. The existence of this intermediate and the predicted structural features were later experimentally confirmed [165]. Besides HX methods, the site-specific structural information obtained from other NMR-based techniques such as relaxation dispersion spectroscopy [173], has also been incorporated into MD simulations to study weakly-populated folding intermediates [174].

A major limitation of traditional experimental methods such as HX is that they can only measure the average behavior of an ensemble of molecules and often cannot distinguish individual folding and unfolding routes or intermediates in the ensemble. By contrast, single-molecule methods enable the observation of the folding and unfolding of individual molecules and play increasingly important roles in studying intermediates [175]. Two single-molecule techniques: fluorescence resonance energy transfer [176] and force spectroscopy including optical tweezers and AFM [177], have been utilized to probe intermediate states of proteins [178–180]. Direct comparisons of results from single-molecule stretching experiments by AFM and computational simulations shed light on the mechanical unfolding of proteins and how intermediates contribute to protein function.

Using AFM, Marszalek et al. [179] uncovered a force-induced unfolding intermediate of Ig domain I27 from titin, a modular protein which is responsible for muscle elasticity. This experimentally-discovered intermediate was also predicted by steered-molecular dynamics (SMD) simulations [179]. Based on the results of SMD, the authors further predicted that the hydrogen-bonding network between strand A′ and G in I27 mainly determines its mechanical stability. This prediction was verified in a mutagenesis study where Li et al. [181] showed that the point mutations on the residues participating in this hydrogen-bonding network dramatically altered the mechanical stability of I27. In a similar study of another protein, domain 10 of type III fibronectin module, which plays a pivotal role in mechanical coupling between cell surface and extracellular matrix (ECM), Gao et al. [182] predicted force-induced unfolding intermediates using SMD simulations. The discovered intermediates were considered to expose binding sites which are necessary for the assembly of ECM fibronectin fibrils. One of the computationally-predicted unfolding intermediates is in excellent agreement with the one observed in a more recent AFM stretching experiment [178].

The finding that folding and unfolding intermediates can have significant contributions to protein function is challenging the conventional understanding of protein structure and function that is centered on the native state. It is expected that a close synergy between computational and experimental approaches will continue to play essential roles in characterizing these evasive protein states.

Unfolded protein states

Unveiling the structural and dynamic properties of denatured proteins is crucial for understanding the protein folding [183,184] and misfolding [185] problems. For example, the computational determination of a protein's thermodynamic stability requires an accurate approximation of the denatured state as the reference state. NMR hydrogen exchange experiments also rely on models of protein unfolded states to tabulate the intrinsic hydrogen exchange rate [186]. Furthermore, understanding the structural properties of unfolded proteins may shed light on the early events of protein folding and protein aggregation [187].

It has long been postulated that the denatured state of proteins is composed of an ensemble of featureless random coil-like conformations. According to Flory's random coil theory [188], the size of a random coil polymer, characterized by the radius of gyration R_g, follows a power law dependence on the length of the polymer chain, n, $R_g = R_0 n^v$. Here, R_0 is the scaling constant, which is a function of persistence length, and v represents the power law scaling exponent. Flory predicted the exponent to be 0.6 and later a more accurate renormalization calculation obtained $v = 0.588$ [189]. Tanford et al. first confirmed this random coil scaling behavior for denatured proteins [190]. Using intrinsic viscosity measurements for 12 proteins denatured by 5–6 M GuHCl, the authors obtained a scaling exponent $v = 0.67 \pm 0.09$. Wilkins et al. [191] showed that the hydrodynamic radii of sets of 8 highly denatured, disulfide-free proteins follow a power law scaling with $v = 0.58 \pm 0.11$. Recently, Kohn et al. [192] reassessed the scaling behavior of denatured proteins using small angle X-ray scattering for 17 proteins of lengths varying from 8 to 549 residues. They found the scaling exponent to be $v = 0.598 \pm 0.029$. All these experimental results confirm the random coil scaling of denatured proteins. In a random coil model of the denatured state, proteins are believed to lack persistent structure both locally and globally. The distribution of end-to-end distances or radii of gyration can be fit by a Gaussian distribution. A recent computational study [193] confirmed this behavior by generating an ensemble of protein conformations whereby only steric interactions between amino acids were considered for four different proteins. A scaling exponent of 0.58 ± 0.02 was obtained.

The random coil scaling behavior [188,194] was originally derived for homopolymers. However, proteins are heteropolymers for which specific interactions between amino acids play an important role and determine a unique native structure. Hence, while the scaling of the sizes of denatured proteins follows the random coil scaling as shown in experiments, it does not necessarily exclude the

possibility that the denatured proteins can have residual native-like structures. Under denaturing conditions, the protein can still exhibit native conformational bias and retain a certain amount of residual native structures. Mounting experimental evidence [168,195–198] supports residual native-like structural elements in the denatured state for a variety of proteins. Using residual dipolar coupling from NMR measurements, Shortle and Ackerman [195] showed that native-like topology persists under strong denaturing conditions as high as 8 M Urea for a truncated staphylococcal nuclease. By applying quasi-elastic neutron scattering on α-lactalbumin, Bu et al. [196] demonstrated residual helical structure and tertiary-like interactions even in the absence of disulfide bonds and under highly denaturing conditions. Similarly, by using triple-resonance NMR, native-like topology has also been observed in protein L [197].

Several theoretical and computational studies [199–202] have addressed the role of specific interactions in conformational biasing toward the native state in the denatured states. Using a simple force field with only steric and hydrogen bond interactions, Pappu et al. [200] demonstrated that denatured protein states have a strong preference for the native structure. It was suggested [199–202] that the conformational bias of native structures in the denatured state is a possible explanation of the Levinthal's paradox [3].

To reconcile the seemingly controversial properties of denatured proteins—the random coil scaling of their sizes and the presence of residual native structures—several computational works have recently been reported [203–207]. Fitzkee and Rose [206] reproduced the random coil scaling exponent using a denatured protein model with fixed secondary structure elements for a set of proteins. Tran et al. [204,207] constructed the protein denatured state ensemble at atomic resolution in the excluded volume limit. Jha et al. [205] built the denatured state using a statistical coil library. Both the Pappu and Sosnick groups found that the putative denatured state ensemble features transient local structures such as turns, strand, and helices. In the mean time, the dimension of the denatured state follows the experimentally-observed random-coil scaling exponent. Ding et al. [203] developed a computational method to model denatured proteins using a structure-based potential [63]. This interaction model is commonly used in thermodynamic and kinetics studies of protein folding [58,208–210] to model amino-acid interactions. This study [203] suggested that denatured proteins follow the random coil scaling sizes *and* retain residual secondary structures akin to those observed in native protein states. Hence, these computational works provide a conceptual reconciliation between two seemingly mutually-exclusive views of protein unfolded states.

What is the physical origin of the random coil scaling of protein size along with the seemingly contradictory persistence of local structures in denatured proteins? Previous computational studies [203–205,207] suggest that the residual structures in the denatured state are limited to short-range elements, which only extend approximately seven to 10 residues. The correlated fluctuation of residual structures diminishes quickly along the sequence and long-range contact formation is purely governed by the random diffusion of peptide chains. Hence, a coarse-graining process, which groups locally-interacting amino acids along the polypeptide chain into renormalized structural units (Fig. 4) reduces a denatured protein to a renormalized homopolymer. The renormalization process will result in an effective homopolymer which forms contacts due to chain diffusion. Protein sizes follow the renormalized power law scaling as proposed by Flory [188]: $R_g = R_0(N/L)^v = (R_0 L^{-v})N^v$. Thus, we expect that the scaling exponent of denatured proteins v is the same as for homopolymers, whose structural units are locally-interacting amino acids.

The observation of residual native secondary structures in thermally-denatured protein states is consistent with a "guided-folding" scenario [211], where the rate-limiting process is the packing of the preformed secondary structures into the correct fold. In contrast, a random coil model of the denatured state without residual native-like structures implies that a protein has to overcome an excessive entropic barrier to form both the secondary and tertiary structures upon folding. The existence of persistent native-like secondary structures in the denatured state may also be responsible for the recent success of protein structure prediction using small secondary structure segments derived from the protein data bank [212]. The existence of residual native-like structures in the denatured
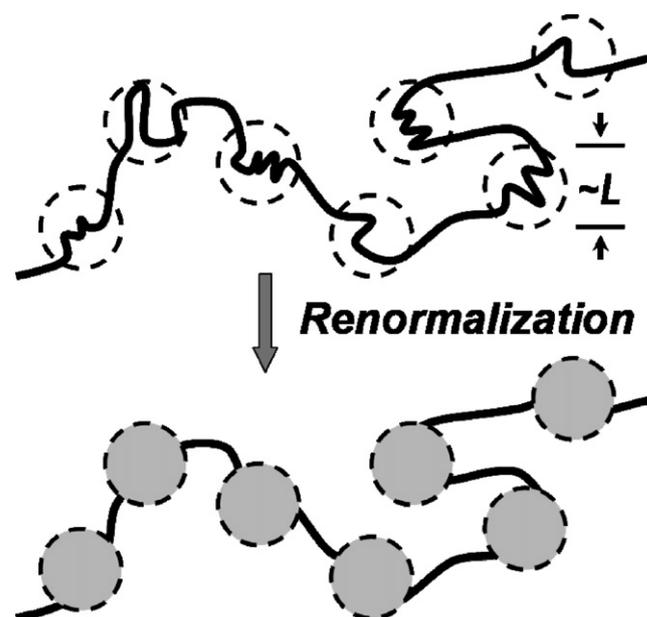

Fig. 4. Unfolded protein states. The unfolded state of a protein features residual local structures, which span a short segment of approximately 10 residues ($L \sim 10$). This structural correlation quickly decays as the segment length increases. A renormalization process, which groups the local amino-acid residues into a coarse-grained bead, reduces the unfolded protein into an effective non-interacting polymer.

state also provides a novel way to manipulate a protein's stability by stabilizing or destabilizing the residual structures [213].

Protein self-association and aggregation

While all the information needed for proteins to fold is encoded in their amino-acid sequence [1], there are many more elements that play a part *in vivo*. In a crowded cellular environment, surrounded by interacting proteins, nascent polypeptides face a formidable challenge in finding the correct interactions that result in a folded and functional protein. Many become "trapped" in meta-stable intermediate structures which are usually recognized by proteasomal machinery and degraded or refolded by chaperones. Alternatively, they can associate with similar misfolded proteins to form aggregates.

Extant protein sequences are the result of a long history of evolutionary refinement establishing a set of interactions defining the native state. However, the same inherent recognition that occurs between sequences within a protein is the basis for a type of self-association termed three-dimensional domain swapping ([214], extensively catalogued in 2002 by Liu and Eisenberg [215]). Domain swapping is an important phenomenon, taking part in both normal and disease-related processes, and is intimately tied to protein folding. Domain swapping may be viewed as a natural mechanism for dealing with instability due to evolutionary changes in the amino-acid sequence [216]. For example, a mutation that rigidifies a loop connecting two parts of a protein induces strain, which can be relieved without the loss of function by "swapping" the portion of the protein on one side of the loop with the corresponding part of a similar protein. Dimerization by this mechanism has advantages including a high local concentration of enzymatic activity, since two functional proteins are joined together.

Many diseases are now associated with protein aggregation and particularly with a form of ordered aggregate called the amyloid fibrils, which, regardless of the native sequence and structure of the precursor proteins, share distinct structural characteristics. From a protein folding standpoint, the inherent properties of the polypeptide chain that allow proteins with little or no sequence or structural similarity to misfold and assemble into similar high-order structures are of vital interest. Studies of aggregate structure reveal defined characteristics such as extensive hydrogen bond networks perpendicular to the fiber axis, called a cross-β conformation [217], and an α–β transition known to occur during the oligomerization of amyloid-forming proteins with significant helical content. Evidence for domain swapping as an early step in the aggregation process has been reported for several proteins [218–220]. In several aggregation-associated diseases including familial Amyotrophic Lateral Sclerosis and the transthyretin amyloidoses, the dissociation of a protein from its multimeric native state is known to be the rate-limiting step

for aggregation, suggesting a method for preventing aggregation by stabilizing the native interfaces of these assemblies [221–223].

Amyloid fibrils have alternately been deemed responsible for the pathologies of their various associated diseases and, more recently, credited with delaying or counteracting the observed pathologies by acting as a sink for highly cytotoxic soluble oligomers [224]. The viewpoint that soluble oligomers act as cytotoxic species has garnered widespread attention since at least 1999 when it was noted that the abundance of soluble Aβ 1–42 oligomers is inversely correlated with neuronal degeneration in Alzheimer's disease whereas amyloid levels do not correlate [225,226]. In 2003, Glabe and coworkers discovered that soluble oligomeric species from several disease-related proteins shared a common structural epitope to which an antibody was developed [227]. Later studies showed that soluble oligomers are able to disrupt the polarity of cellular membranes [144,228,229], one possible basis for disease-associated toxicity.

Various cellular protective mechanisms have evolved to ensure the proper folding of proteins. Molecular chaperones, for example, recognize misfolded proteins and provide an environment conducive to the formation of the appropriate native contacts [230]. Vast proteosomal machinery clears proteinaceous debris from the cell using ubiquitin ligases to tag misfolded proteins for degradation and removal [231,232]. One hypothesis formulated to explain the prevalence of protein aggregation in neurodegenerative diseases is that the ubiquitin proteasome loses efficiency over time causing a buildup of protein aggregates and debris in post-mitotic cells such as neurons [233]. Also, parkin, an E3 ubiquitin ligase was found to be mutated in at least half of autosomal recessive juvenile parkinsonism patients, suggesting that a deficit in the clearance of its target protein leads to an early onset of symptoms [234].

Computational studies of protein aggregation have traditionally been inadequate due to the massive complexity of the system in both time and length scale. Several approaches have been used to overcome this complexity (Fig. 5). Traditional all-atom molecular dynamics simulations have been carried out to model the aggregation of disease-related peptides such as Aβ and polyglutamine [235–237]. Other techniques seek to identify which parts within larger proteins are responsible for their aggregation behavior, resulting in the identification of sequence stretches in various proteins that are "amyloidogenic," or "hot spots" for aggregation [238–240]. Underlying this work is the idea that evolution acts to prevent aggregation by burying aggregation-prone protein sequences or otherwise prohibiting their apposition in protein structures and during folding.[3] To study the nature of subunit assembly

---

[3] It is interesting to note here that in the case of Pmel17, which aggregates to form a "functional amyloid" involved in melanin biosynthesis, a protein seems to have evolved to aggregate at an incredible rate, perhaps to minimize the population time in a soluble oligomer form [241].
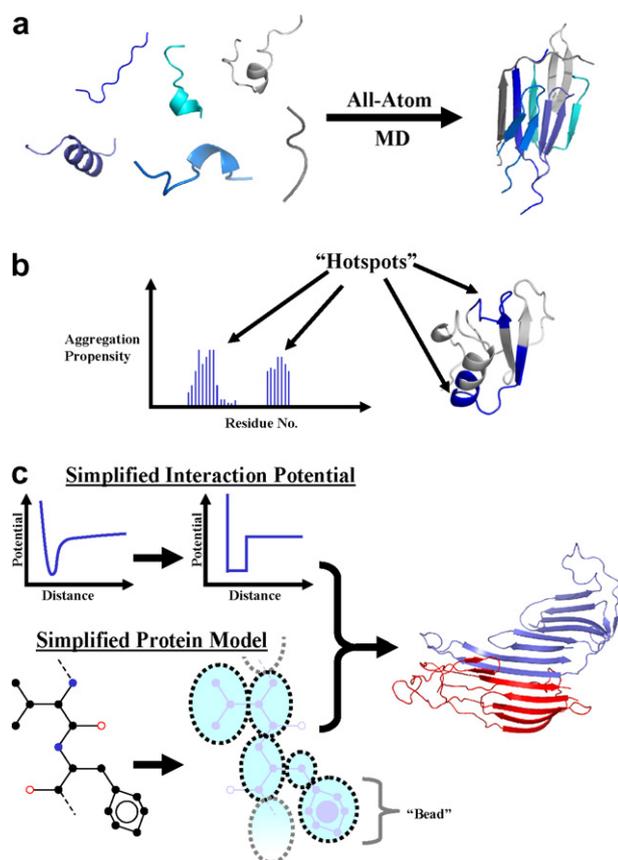
Fig. 5. Computational studies of protein self-association and aggregation. Different computational techniques have been utilized to study various aspects of protein self-association and aggregation. (a) All-atom molecular dynamics is used to model the aggregation of short peptides. (b) Simulations of peptides from within larger proteins are used to suggest aggregation "hotspots". (c) By combining simplified interaction models and protein models, aggregating systems that are inaccessible by traditional molecular dynamics due to size and time limitations can be studied. The curves show the approximation of a continuous interaction potential by a square well as used in DMD [58]. The dotted circles represent "beads" in the model which take the place of several atoms in the original protein. The structure to the right shows the self-association of two identical proteins forming an extended β-sheet.

and extension during aggregation, which are out of reach for all-atom molecular dynamics in both the size scale and time scale, simplified protein models are being utilized [65]. By accessing aggregation events that are out of the reach of experimentalists, computational studies of aggregation are an essential compliment to the experimental findings regarding aggregate structure and formation mechanisms.

Protein aggregation is now widely viewed as a fundamental property of the polypeptide chain, meaning that all of the considerations discussed in the earlier sections of this review must, to some degree, apply to the study of aggregating proteins. Therefore, the study of protein self-association and aggregation really is the study of protein folding in the context of external influences like protein concentration, localization, or evolution. As this field develops and knowledge of protein aggregation as a general phenomenon accumulates, we stand to gain not only

vital tools for treating specific diseases, but also insight into the behavior of all proteins with respect to their environment.

Conclusions

Perhaps the most significant four measures of success in the natural sciences are our abilities (i) to observe natural phenomena, (ii) to explain natural phenomena, (iii) to predict the effect of relevant variables on a specific phenomenon, and (iv) to rationally manipulate these phenomena. The protein folding field has seen significant breakthroughs according to all of these measures. Research in the protein folding field uncovered folding pathways and states that accompany the transition from unfolded to folded proteins, revealed the origin of the cooperative folding transition, allowed prediction of folding rates and changes in thermodynamic stability upon mutation, and permitted rational alteration of protein structure and folding pathways.

Given the recognition of many human maladies as "diseases of protein folding" over the past two decades, the wealth of new knowledge about the folding process is driving the study of protein folding back into its native environment, i.e., inside living organisms. The effect of the cellular environment on protein folding, e.g., how proteins fold *in vivo* and especially the behavior of "intrinsically-disordered" proteins is a highly active area of inquiry [82,242,243]. More applied research is ongoing in the design of animal models of diseases associated with protein folding such as cystic fibrosis, ALS, Alzheimer's disease, and the prion diseases [244–246], and the design of small molecules for use in clinical trials for treating diseases like the transthyretin amyloidoses [247]. The future of the protein folding field lies in its direct application to such medical problems, and for a growing number of protein systems, the future is now.

References

[1] C.B. Anfinsen, E. Harber, M. Sela Jr., F.H. White, Proc. Natl. Acad. Sci. USA 47 (1961) 1309–1314.
[2] C.B. Anfinsen, Science 181 (1973) 223–230.
[3] C. Levinthal, J. Chim. Phys. 65 (1968) 44–45.
[4] A. Ikai, C. Tanford, Nature 230 (1971) 100–102.
[5] T.Y. Tsong, R.L. Baldwin, E.L. Elson, Proc. Natl. Acad. Sci. USA 68 (1971) 2712–2715.
[6] S.E. Jackson, A.R. Fersht, Biochemistry 30 (1991) 10428–10435.

[7] S.E. Jackson, A.R. Fersht, Biochemistry 30 (1991) 10436–10443.

[8] Y. Bai, T.R. Sosnick, L. Mayne, S.W. Englander, Science 269 (1995) 192–197.

[9] Y. Bai, S.W. Englander, Proteins 24 (1996) 145–151.

[10] H. Roder, G.A. Elove, S.W. Englander, Nature 20 (335) (1988) 700–704.

[11] A. Miranker, C.V. Robinson, S.E. Radford, R.T. Aplin, C.M. Dobson, Science 262 (1993) 896–900.

[12] C.M. Jones, E.R. Henry, Y. Hu, C.K. Chan, S.D. Luck, A. Bhuyan, H. Roder, J. Hofrichter, W.A. Eaton, Proc. Natl. Acad. Sci. USA 90 (1993) 11860–11864.

[13] A. Matouschek, J.T. Kellis, L. Serrano, M. Bycroft, A.R. Fersht, Nature 346 (1990) 440–445.

[14] L. Mirny, E. Shakhnovich, Annu. Rev. Biophys. Biomol. Struct. 30 (361–396) (2001) 361–396.

[15] E. Shakhnovich, Chem. Rev. 106 (2006) 1559–1588.

[16] D.B. Wetlaufer, Proc. Natl. Acad. Sci. USA 70 (1973) 697–701.

[17] D.B. Wetlaufer, Trends Biochem. Sci. 15 (1990) 414–415.

[18] N. Go, Annu. Rev. Biophys. Bioeng. 12:183-210 (1983) 183–210.

[19] V.I. Abkevich, A.M. Gutin, E.I. Shakhnovich, Biochemistry 33 (1994) 10026–10036.

[20] A.R. Fersht, Curr. Opin. Struct. Biol. 7 (1997) 3–9.

[21] E.I. Shakhnovich, Curr. Opin. Struct. Biol. 7 (1997) 29–40.

[22] E.M. Lifshits, L.P. Pitaevskii, Physical Kinetics, Pergamon Press, Oxford, 1981.

[23] J.N. Onuchic, Z. Luthey-Schulten, P.G. Wolynes, Annu. Rev. Phys. Chem. 48 (1997) 545–600.

[24] J.E. Shea, C.L. Brooks III, Annu. Rev. Phys. Chem. 52 (499–535) (2001) 499–535.

[25] R.H. Zhou, B.J. Berne, R. Germain, Proc. Natl. Acad. Sci. USA 98 (2001) 14931–14936.

[26] A.R. Fersht, V. Daggett, Cell 108 (2002) 573–582.

[27] A.R. Dinner, A. Sali, L.J. Smith, C.M. Dobson, M. Karplus, Trends Biochem. Sci. 25 (2000) 331–339.

[28] Y. Duan, P.A. Kollman, Science 282 (1998) 740–744.

[29] A. Ghosh, R. Elber, H.A. Scheraga, Proc. Natl. Acad. Sci. USA 99 (2002) 10394–10398.

[30] A. Cavalli, P. Ferrara, A. Caflisch, Proteins 47 (2002) 305–314.

[31] V.S. Pande, I. Baker, J. Chapman, S.P. Elmer, S. Khaliq, S.M. Larson, Y.M. Rhee, M.R. Shirts, C.D. Snow, E.J. Sorin, B. Zagrovic, Biopolymers 68 (2003) 91–109.

[32] U.H.E. Hansmann, Int. J. Quantum Chem. 90 (2002) 1515–1521.

[33] U.H.E. Hansmann, Comput. Phys. Commun. 147 (2002) 604–607.

[34] C.D. Snow, N. Nguyen, V.S. Pande, M. Gruebele, Nature 420 (2002) 102–106.

[35] D. Gront, A. Kolinski, J. Skolnick, J. Chem. Phys. 115 (2001) 1569–1574.

[36] Y.M. Rhee, V.S. Pande, Biophys. J. 84 (2003) 775–786.

[37] M. Feig, J. Karanicolas, C.L. Brooks, J. Mol. Graph. Model. 22 (2004) 377–395.

[38] J. Skolnick, A. Kolinski, A.R. Ortiz, J. Mol. Biol. 265 (1997) 217–241.

[39] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, P. Kollman, Comput. Phys. Commun. 91 (1995) 1–41.

[40] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, J. Comput. Chem. 4 (1983) 187–217.

[41] Y. Zhou, M. Karplus, Proc. Natl. Acad. Sci. USA 94 (1997) 14429–14432.

[42] N.V. Dokholyan, J.M. Borreguero, S.V. Buldyrev, F. Ding, H.E. Stanley, E.I. Shakhnovich, Methods Enzymol. 374 (2003) 616–638.

[43] B.J. Alder, T.E. Wainwright, J. Chem. Phys. 31 (1959) 459–466.

[44] D.C. Rapaport, The Art of Molecular Dynamics Simulations, Cambridge University Press, Cambridge, 1997.

[45] D.C. Rapaport, J. Chem. Phys. 71 (1979) 3299–3303.

[46] A. Bellemans, J. Orban, D. Vanbelle, Mol. Phys. 39 (1980) 781–782.

[47] F. Ding, W.H. Guo, N.V. Dokholyan, E.I. Shakhnovich, J.E. Shea, J. Mol. Biol. 350 (2005) 1035–1050.

[48] V.J. Hilser, E.B. Garcia-Moreno, T.G. Oas, G. Kapp, S.T. Whitten, Chem. Rev. 106 (2006) 1545–1558.

[49] J. Ma, Structure 13 (2005) 373–380.

[50] Q. Cui, G. Li, J. Ma, M. Karplus, J. Mol. Biol. 340 (2004) 345–372.

[51] T. Liu, S.T. Whitten, V.J. Hilser, Proc. Natl. Acad. Sci. USA 104 (2007) 4347–4352.

[52] E. Freire, Proc. Natl. Acad. Sci. USA 97 (2000) 11680–11682.

[53] M.M. Gromiha, A.M. Thangakani, S. Selvaraj, Nucleic Acids Res. 34 (2006) W70–W74.

[54] S. Thomas, G. Song, N.M. Amato, Phys. Biol. 2 (2005) S148–S155.

[55] S. Sharma, F. Ding, H. Nie, D. Watson, A. Unnithan, J. Lopp, D. Pozefsky, N.V. Dokholyan, Bioinformatics (2006).

[56] H. Jang, C.K. Hall, Y. Zhou, Biophys. J. 86 (2004) 31–49.

[57] A.V. Smith, C.K. Hall, Proteins 44 (2001) 376–391.

[58] N.V. Dokholyan, S.V. Buldyrev, H.E. Stanley, E.I. Shakhnovich, Folding Des. 3 (1998) 577–587.

[59] J.M. Borreguero, N.V. Dokholyan, S.V. Buldyrev, E.I. Shakhnovich, H.E. Stanley, J. Mol. Biol. 318 (2002) 863–876.

[60] H. Jang, C.K. Hall, Y. Zhou, Biophys. J. 83 (2002) 819–835.

[61] N.V. Dokholyan, Curr. Opin. Struct. Biol. (2006).

[62] F. Ding, N.V. Dokholyan, Trends Biotechnol. 23 (2005) 450–455.

[63] N. Go, H. Abe, Biopolymers 20 (1981) 991–1011.

[64] B. Urbanc, J.M. Borreguero, L. Cruz, H.E. Stanley, Methods Enzymol. 412 (2006) 314–338.

[65] S. Peng, F. Ding, B. Urbanc, S.V. Buldyrev, L. Cruz, H.E. Stanley, N.V. Dokholyan, Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 69 (2004) 041908.

[66] H.D. Nguyen, C.K. Hall, Proc. Natl. Acad. Sci. USA 101 (2004) 16180–16185.

[67] H.D. Nguyen, C.K. Hall, J. Biol. Chem. 280 (2005) 9074–9082.

[68] S. Sharma, F. Ding, N.V. Dokholyan, Biophys. J. 92 (2007).

[69] J. Frydman, F.U. Hartl, Science 272 (1996) 1497–1502.

[70] J. Martin, F.U. Hartl, Curr. Opin. Struct. Biol. 7 (1997) 41–52.

[71] B. Bukau, A.L. Horwich, Cell 92 (1998) 351–366.

[72] J.S. Weissman, H.S. Rye, W.A. Fenton, J.M. Beechem, A.L. Horwich, Cell 84 (1996) 481–490.

[73] J.L. Feltham, L.M. Gierasch, Cell 100 (2000) 193–196.

[74] S.J. Landry, L.M. Gierasch, Biochemistry 30 (1991) 7359–7362.

[75] P.M. Horowitz, S. Hua, D.L. Gibbons, J. Biol. Chem. 270 (1995) 1535–1542.

[76] B.M. Gorovits, P.M. Horowitz, J. Biol. Chem. 270 (1995) 13057–13062.

[77] P. Wittung-Stafshede, Acc. Chem. Res. 35 (2002) 201–208.

[78] J. Leckner, N. Bonander, P. Wittung-Stafshede, B.G. Malmstrom, B.G. Karlsson, Biochim. Biophys. Acta 1342 (1997) 19–27.

[79] K. Luby-Phelps, Int. Rev. Cytol. 192 (2000) 189–221.

[80] A.P. Minton, J. Biol. Chem. 276 (2001) 10577–10580.

[81] A.P. Minton, J. Pharm. Sci. 94 (2005) 1668–1675.

[82] M.M. Dedmon, C.N. Patel, G.B. Young, G.J. Pielak, Proc. Natl. Acad. Sci. USA 99 (2002) 12681–12684.

[83] P. McPhie, Y.S. Ni, A.P. Minton, J. Mol. Biol. 361 (2006) 7–10.

[84] D.M. Hatters, A.P. Minton, G.J. Howlett, J. Biol. Chem. 277 (2002) 7824–7830.

[85] D.M. Taverna, R.A. Goldstein, Proteins 46 (2002) 105–109.

[86] A.A. Pakula, R.T. Sauer, Annu. Rev. Genet. 23 (1989) 289–310.

[87] P.L. Privalov, Adv. Protein Chem. 33 (1979) 167–241.

[88] W.J. Becktel, J.A. Schellman, Biopolymers 26 (1987) 1859–1877.

[89] G.I. Makhatadze, P.L. Privalov, J. Mol. Biol. 226 (1992) 491–505.

[90] S.E. Jackson, M. Moracci, N. elMasry, C.M. Johnson, A.R. Fersht, Biochemistry 32 (1993) 11259–11269.

[91] P.A. Bash, U.C. Singh, R. Langridge, P.A. Kollman, Science 236 (1987) 564–568.

[92] L.X. Dang, K.M. Merz, P.A. Kollman, J. Am. Chem. Soc. 111 (1989) 8505–8508.

[93] P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S.H. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case, T.E. Cheatham, Acc. Chem. Res. 33 (2000) 889–897.

[94] Y.N. Vorobjev, J. Hermans, Biophys. Chem. 78 (1999) 195–205.
[95] P. Kollman, Chem. Rev. 93 (1993) 2395–2417.
[96] T. Lazaridis, M. Karplus, Curr. Opin. Struct. Biol. 10 (2000) 139–145.
[97] D. Gilis, M. Rooman, J. Mol. Biol. 257 (1996) 1112–1126.
[98] D. Gilis, M. Rooman, J. Mol. Biol. 272 (1997) 276–290.
[99] D. Gilis, M. Rooman, Protein Eng. 13 (2000) 849–856.
[100] H.Y. Zhou, Y.Q. Zhou, Protein Sci. 12 (2003) 2121.
[101] M. Ota, Y. Isogai, K. Nishikawa, Protein Eng. 14 (2001) 557–564.
[102] C. Hoppe, D. Schomburg, Protein Sci. 14 (2005) 2682–2692.
[103] C.W. Carter, B.C. LeFebvre, S.A. Cammer, A. Tropsha, M.H. Edgell, J. Mol. Biol. 311 (2001) 625–638.
[104] R. Guerois, J.E. Nielsen, L. Serrano, J. Mol. Biol. 320 (2002) 369–387.
[105] J. Khatun, S.D. Khare, N.V. Dokholyan, J. Mol. Biol. 336 (2004) 1223–1238.
[106] A.J. Bordner, R.A. Abagyan, Proteins-Struct. Funct. Bioinform. 57 (2004) 400–413.
[107] K. Saraboji, M.M. Gromiha, M.N. Ponnuswamy, Biopolymers 82 (2006) 80–92.
[108] E. Capriotti, P. Fariselli, R. Calabrese, R. Casadio, Bioinformatics 21 (2005) 54–58.
[109] B. Kuhlman, D. Baker, Curr. Opin. Struct. Biol. 14 (2004) 89–95.
[110] J.L. Cheng, A. Randall, P. Baldi, Proteins-Struct. Funct. Bioinform. 62 (2006) 1125–1132.
[111] S.Y. Yin, F. Ding, N.V. Dokholyan, Nat. Methods 4 (2007) 466–467.
[112] F. Ding, N.V. Dokholyan, Public Libr. Sci. Comput. Biol. 2 (2006) e85.
[113] M.H. Edgell, D.A. Sims, G.J. Pielak, F. Yi, Biochemistry 42 (2003) 7587–7593.
[114] K.A. Bava, M.M. Gromiha, H. Uedaira, K. Kitajima, A. Sarai, Nucleic Acids Res. 32 (2004) D120–D121.
[115] F. Chiti, C.M. Dobson, Annu. Rev. Biochem. 75 (2006) 333–366.
[116] P.H. Thibodeau, C.A. Brautigam, M. Machius, P.J. Thomas, Nat. Struct. Mol. Biol. 12 (2005) 10–16.
[117] S.D. Khare, N.V. Dokholyan, Proc. Natl. Acad. Sci. USA 103 (2006) 3147–3152.
[118] B. Nolting, Protein Folding Kinetics, Springer Verlag, Berlin, Germany, 2006.
[119] K.W. Plaxco, K.T. Simons, D. Baker, J. Mol. Biol. 277 (1998) 985–994.
[120] V. Munoz, W.A. Eaton, PNAS 96 (1999) 11311–11316.
[121] D.E. Otzen, L.S. Itzhaki, N.F. elMasry, S.E. Jackson, A.R. Fersht, Proc. Natl. Acad. Sci. USA 91 (1994) 10422–10425.
[122] A. Li, V. Daggett, Proc. Natl. Acad. Sci. USA 91 (1994) 10430–10434.
[123] D.N. Ivankov, A.V. Finkelstein, Proc. Natl. Acad. Sci. USA 101 (2004) 8942–8944.
[124] I.B. Kuznetsov, S. Rackovsky, Proteins-Struct. Funct. Genet. 54 (2004) 333–341.
[125] P. Ferrara, J. Apostolakis, A. Caflisch, J. Phys. Chem. B 104 (2000) 5000–5010.
[126] P. Ferrara, A. Caflisch, Proc. Natl. Acad. Sci. USA 97 (2000) 10780–10785.
[127] A. Hiltpold, P. Ferrara, J. Gsponer, A. Caflisch, J. Phys. Chem. B 104 (2000) 10080–10086.
[128] B. Zagrovic, V. Pande, J. Comput. Chem. 24 (2003) 1432–1436.
[129] D. Qiu, P.S. Shenkin, F.P. Hollinger, W.C. Still, J. Phys. Chem. A 101 (1997) 3005–3014.
[130] W.L. Jorgensen, J. Tiradorives, J. Am. Chem. Soc. 110 (1988) 1657–1666.
[131] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, J. Chem. Phys. 79 (1983) 926–935.
[132] H. Berendsen, J. Potsma, W. van Gunsteren, J. Hermans, in: Intermolecular Forces, Reidel, Dordrecht, 1981, pp. 331–334.
[133] E.J. Sorin, V.S. Pande, Biophys. J. 88 (2005) 2472–2493.
[134] E.I. Shakhnovich, A.V. Finkelstein, Biopolymers 28 (1989) 1667–1680.
[135] J.D. Bryngelson, P.G. Wolynes, Proc. Natl. Acad. Sci. USA 84 (1987) 7524–7528.
[136] E. Shakhnovich, G. Farztdinov, A.M. Gutin, M. Karplus, Phys. Rev. Lett. 67 (1991) 1665.
[137] A. Sali, E. Shakhnovich, M. Karplus, Nature 369 (1994) 248–251.
[138] P.G. Bolhuis, D. Chandler, C. Dellago, P.L. Geissler, Annu. Rev. Phys. Chem. 53 (2002) 291–318.
[139] A. Matouschek, J.T. Kellis, L. Serrano, A.R. Fersht, Nature 340 (1989) 122–126.
[140] A.R. Fersht, A. Matouschek, L. Serrano, J. Mol. Biol. 224 (1992) 771–782.
[141] R. Day, V. Daggett, Protein Simulations 66 (2003) 373–403.
[142] H.A. Scheraga, FEBS J. 272 (2005) 359.
[143] I. Ruczinski, T.R. Sosnick, K.W. Plaxco, Protein Sci. 15 (2006) 2257–2264.
[144] S. Baglioni, F. Casamenti, M. Bucciantini, L.M. Luheshi, N. Taddei, F. Chiti, C.M. Dobson, M. Stefani, J. Neurosci. 26 (2006) 8160–8167.
[145] A.R. Fersht, S. Sato, PNAS 101 (2004) 7976–7981.
[146] F. Ding, N.V. Dokholyan, S.V. Buldyrev, H.E. Stanley, E.I. Shakhnovich, Biophys. J. 83 (2002) 3525–3532.
[147] F. Ding, N.V. Dokholyan, S.V. Buldyrev, H.E. Stanley, E.I. Shakhnovich, J. Mol. Biol. 324 (2002) 851–857.
[148] A.A. Di Nardo, D.M. Korzhnev, P.J. Stogios, A. Zarrine-Afsar, L.E. Kay, A.R. Davidson, Proc. Natl. Acad. Sci. USA 101 (2004) 7954–7959.
[149] Y.R. Chen, A.C. Clark, Protein Sci. 13 (2004) 2196–2206.
[150] Y.R. Chen, A.C. Clark, Biochemistry 42 (2003) 6310–6320.
[151] J.S. Reader, N.A.J. Van Nuland, G.S. Thompson, S.J. Ferguson, C.M. Dobson, S.E. Radford, Protein Sci. 10 (2001) 1216–1224.
[152] S.D. Khare, F. Ding, N.V. Dokholyan, J. Mol. Biol. 334 (2003) 515–525.
[153] A.D. Stoycheva, C.L. Brooks III, J.N. Onuchic, J. Mol. Biol. 340 (2004) 571–585.
[154] S. Matysiak, C. Clementi, J. Mol. Biol. 363 (2006) 297–308.
[155] J.E. Shea, J.N. Onuchic, C.L. Brooks III, Proc. Natl. Acad. Sci. USA 99 (2002) 16064–16068.
[156] H. Nguyen, M. Jager, A. Moretto, M. Gruebele, J.W. Kelly, Proc. Natl. Acad. Sci. USA 100 (2003) 3948–3953.
[157] J. Karanicolas, C.L. Brooks, Proc. Natl. Acad. Sci. USA 100 (2003) 3954–3959.
[158] A.M. Fernandez-Escamilla, M.S. Cheung, M.C. Vega, M. Wilmanns, J.N. Onuchic, L. Serrano, Proc. Natl. Acad. Sci. USA 101 (2004) 2834–2839.
[159] L.L. Chavez, J.N. Onuchic, C. Clementi, J. Am. Chem. Soc. 126 (2004) 8426–8432.
[160] J.N. Onuchic, P.G. Wolynes, Curr. Opin. Struct. Biol. 14 (2004) 70–75.
[161] M. Lindberg, M. Oliveberg, Curr. Opin. Struct. Biol. 17 (2007) 21–29.
[162] S. Nauli, B. Kuhlman, D. Baker, Nat. Struct. Biol. 8 (2001) 602–605.
[163] A.R. Lowe, L.S. Itzhaki, PNAS 104 (2007) 2679–2684.
[164] H.Q. Feng, N.D. Vu, Y.W. Bai, J. Mol. Biol. 346 (2005) 345–353.
[165] Z. Zhou, H. Feng, Y. Bai, Proteins 65 (2006) 259–265.
[166] A.R. Fersht, Proc. Natl. Acad. Sci. USA 97 (2000) 1525–1529.
[167] H. Kato, H.Q. Feng, Y.W. Bai, J. Mol. Biol. 365 (2007) 870–880.
[168] T.L. Religa, J.S. Markson, U. Mayor, S.M. Freund, A.R. Fersht, Nature 437 (2005) 1053–1056.
[169] S.W. Englander, Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 213–238.
[170] S.W. Englander, J. Am. Soc. Mass Spectrom. 17 (2006) 1481–1489.
[171] J. Gsponer, H. Hopearuoho, S.B. Whittaker, G.R. Spence, G.R. Moore, E. Paci, S.E. Radford, M. Vendruscolo, Proc. Natl. Acad. Sci. USA 103 (2006) 99–104.
[172] R.D. Dixon, Y. Chen, F. Ding, S.D. Khare, K.C. Prutzman, M.D. Schaller, S.L. Campbell, N.V. Dokholyan, Structure 12 (2004) 2161–2171.
[173] A. Mittermaier, L.E. Kay, Science 312 (2006) 224–228.

[174] D.M. Korzhnev, X. Salvatella, M. Vendruscolo, A.A. Di Nardo, A.R. Davidson, C.M. Dobson, L.E. Kay, Nature 430 (2004) 586–590.
[175] X. Zhuang, M. Rief, Curr. Opin. Struct. Biol. 13 (2003) 88–97.
[176] P.R. Selvin, Nat. Struct. Biol. 7 (2000) 730–734.
[177] H. Clausen-Schaumann, M. Seitz, R. Krautbauer, H.E. Gaub, Curr. Opin. Chem. Biol. 4 (2000) 524–530.
[178] L. Li, H.H. Huang, C.L. Badilla, J.M. Fernandez, J. Mol. Biol. 345 (2005) 817–826.
[179] P.E. Marszalek, H. Lu, H. Li, M. Carrion-Vazquez, A.F. Oberhauser, K. Schulten, J.M. Fernandez, Nature 402 (1999) 100–103.
[180] C. Cecconi, E.A. Shank, C. Bustamante, S. Marqusee, Science 309 (2005) 2057–2060.
[181] H. Li, M. Carrion-Vazquez, A.F. Oberhauser, P.E. Marszalek, J.M. Fernandez, Nat. Struct. Biol. 7 (2000) 1117–1120.
[182] M. Gao, D. Craig, O. Lequin, I.D. Campbell, V. Vogel, K. Schulten, Proc. Natl. Acad. Sci. USA 100 (2003) 14784–14789.
[183] J.D. Bryngelson, J.N. Onuchic, N.D. Socci, P.G. Wolynes, Proteins 21 (1995) 167–195.
[184] R.L. Baldwin, Adv. Protein Chem. 62 (2002) 361–367.
[185] C.M. Dobson, Semin. Cell Dev. Biol. 15 (2004) 3–16.
[186] Y. Bai, J.S. Milne, L. Mayne, S.W. Englander, Proteins 17 (1993) 75–86.
[187] V.N. Uversky, A.L. Fink, Biochim. Biophys. Acta 1698 (2004) 131–153.
[188] P.J. Flory, Principles of Polymer Chemistry, Cornell University Press, Ithaca, NY, 1965.
[189] J.C. Le Guillou, J. Zinn-Justin, Phys. Rev. Lett. 39 (1977) 95–98.
[190] C. Tanford, K. Kawahara, S. Lapanje, T.M. Hooker Jr., M.H. Zarlengo, A. Salahuddin, K.C. Aune, T. Takagi, J. Am. Chem. Soc. 89 (1967) 5023–5029.
[191] D.K. Wilkins, S.B. Grimshaw, V. Receveur, C.M. Dobson, J.A. Jones, L.J. Smith, Biochemistry 38 (1999) 16424–16431.
[192] J.E. Kohn, I.S. Millett, J. Jacob, B. Zagrovic, T.M. Dillon, N. Cingel, R.S. Dothager, S. Seifert, P. Thiyagarajan, T.R. Sosnick, M.Z. Hasan, V.S. Pande, I. Ruczinski, S. Doniach, K.W. Plaxco, Proc. Natl. Acad. Sci. USA 101 (2004) 12491–12496.
[193] D.P. Goldenberg, J. Mol. Biol. 326 (2003) 1615–1633.
[194] P.G. de Gennes, Scaling Concepts in Polymer Physics, Cornell University Press, Ithaca, NY, 1979.
[195] D. Shortle, M.S. Ackerman, Science 293 (2001) 487–489.
[196] Z. Bu, J. Cook, D.J. Callaway, J. Mol. Biol. 312 (2001) 865–873.
[197] Q. Yi, M.L. Scalley-Kim, E.J. Alm, D. Baker, J. Mol. Biol. 299 (2000) 1341–1351.
[198] G.W. Platt, V.J. McParland, A.P. Kalverda, S.W. Homans, S.E. Radford, J. Mol. Biol. 346 (2005) 279–294.
[199] R. Zwanzig, A. Szabo, B. Bagchi, Proc. Natl. Acad. Sci. USA 89 (1992) 20–22.
[200] R.V. Pappu, R. Srinivasan, G.D. Rose, Proc. Natl. Acad. Sci. USA 97 (2000) 12565–12570.
[201] R. Srinivasan, G.D. Rose, Biophys. Chem. 101–102 (2002) 167–171.
[202] N.C. Fitzkee, G.D. Rose, Protein Sci. 13 (2004) 633–639.
[203] F. Ding, R.K. Jha, N.V. Dokholyan, Structure 13 (2005) 1047–1054.
[204] H.T. Tran, X. Wang, R.V. Pappu, Biochemistry 44 (2005) 11369–11380.
[205] A.K. Jha, A. Colubri, K.F. Freed, T.R. Sosnick, Proc. Natl. Acad. Sci. USA 102 (2005) 13099–13104.
[206] N.C. Fitzkee, G.D. Rose, Proc. Natl. Acad. Sci. USA 101 (2004) 12497–12502.
[207] H.T. Tran, R.V. Pappu, Biophys. J. 91 (2006) 1868–1886.
[208] C. Clementi, H. Nymeyer, J.N. Onuchic, J. Mol. Biol. 298 (2000) 937–953.
[209] S. Yang, S.S. Cho, Y. Levy, M.S. Cheung, H. Levine, P.G. Wolynes, J.N. Onuchic, Proc. Natl. Acad. Sci. USA 101 (2004) 13786–13791.
[210] N.V. Dokholyan, S.V. Buldyrev, H.E. Stanley, E.I. Shakhnovich, J. Mol. Biol. 296 (2000) 1183–1188.
[211] R.L. Baldwin, G.D. Rose, Trends Biochem. Sci. 24 (1999) 77–83.
[212] P. Bradley, D. Chivian, J. Meiler, K.M. Misura, C.A. Rohl, W.R. Schief, W.J. Wedemeyer, O. Schueler-Furman, P. Murphy, Schonbrun, C.E. Strauss, D. Baker, Proteins 53 (Suppl. 6) (2003) 457–468.
[213] B.E. Bowler, Mol. Biosyst. 3 (2007) 88–99.
[214] M.J. Bennett, S. Choe, D. Eisenberg, Proc. Natl. Acad. Sci. USA 91 (1994) 3127–3131.
[215] Y. Liu, D. Eisenberg, Protein Sci. 11 (2002) 1285–1299.
[216] J.W. Schymkowitz, F. Rousseau, H.R. Wilkinson, A. Friedler, L.S. Itzhaki, Nat. Struct. Biol. 8 (2001) 888–892.
[217] C. Blake, L. Serpell, Structure 4 (1996) 989–998.
[218] C.M. Eakin, F.J. Attenello, C.J. Morgan, A.D. Miranker, Biochemistry 43 (2004) 7808–7815.
[219] K.J. Knaus, M. Morillas, W. Swietnicki, M. Malone, W.K. Surewicz, V.C. Yee, Nat. Struct. Biol. 8 (2001) 770–774.
[220] A. Sanders, C.C. Jeremy, L.D. Higgins, S. Giannini, M.J. Conroy, A.M. Hounslow, J.P. Waltho, R.A. Staniforth, J. Mol. Biol. 336 (2004) 165–178.
[221] S.M. Johnson, R.L. Wiseman, Y. Sekijima, N.S. Green, S.L. mski-Werner, J.W. Kelly, Acc. Chem. Res. 38 (2005) 911–921.
[222] S.D. Khare, M. Caplow, N.V. Dokholyan, Proc. Natl. Acad. Sci. USA 101 (2004) 15094–15099.
[223] S.S. Ray, R.J. Nowak, K. Strokovich, R.H. Brown, T. Walz, P.T. Lansbury, Biochemistry 43 (2004) 4899–4905.
[224] K.S. Rao, M.L. Hegde, S. Anitha, M. Musicco, F.A. Zucca, N.J. Turro, L. Zecca, Prog. Neurobiol. 78 (2006) 364–373.
[225] L.F. Lue, Y.M. Kuo, A.E. Roher, L. Brachova, Y. Shen, L. Sue, T. Beach, J.H. Kurth, R.E. Rydel, J. Rogers, Am. J. Pathol. 155 (1999) 853–862.
[226] C.A. McLean, R.A. Cherny, F.W. Fraser, S.J. Fuller, M.J. Smith, K. Beyreuther, A.I. Bush, C.L. Masters, Ann. Neurol. 46 (1999) 860–866.
[227] R. Kayed, E. Head, J.L. Thompson, T.M. McIntire, S.C. Milton, C.W. Cotman, C.G. Glabe, Science 300 (2003) 486–489.
[228] A. Demuro, E. Mina, R. Kayed, S.C. Milton, I. Parker, C.G. Glabe, J. Biol. Chem. 280 (2005) 17294–17300.
[229] R. Kayed, Y. Sokolov, B. Edmonds, T.M. McIntire, S.C. Milton, J.E. Hall, C.G. Glabe, J. Biol. Chem. 279 (2004) 46363–46366.
[230] B. Bukau, J. Weissman, A. Horwich, Cell 125 (2006) 443–451.
[231] S.J. Berke, H.L. Paulson, Curr. Opin. Genet. Dev. 13 (2003) 253–261.
[232] D.C. Rubinsztein, Nature 443 (2006) 780–786.
[233] J.N. Keller, J. Gee, Q. Ding, Ageing Res. Rev. 1 (2002) 279–293.
[234] C.B. Lucking, A. Durr, V. Bonifati, J. Vaughan, M.G. De, T. Gasser, B.S. Harhangi, G. Meco, P. Denefle, N.W. Wood, Y. Agid, A. Brice, N. Engl. J. Med. 342 (2000) 1560–1567.
[235] W. Han, Y.D. Wu, J. Am. Chem. Soc. 127 (2005) 15408–15416.
[236] S.D. Khare, F. Ding, K.N. Gwanmesia, N.V. Dokholyan, PLoS Comput. Biol. 1 (2005) 230–235.
[237] A.J. Marchut, C.K. Hall, Comput. Biol. Chem. 30 (2006) 215–218.
[238] A.M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, L. Serrano, Nat. Biotechnol. 22 (2004) 1302–1306.
[239] S.D. Khare, K.C. Wilcox, P. Gong, N.V. Dokholyan, Proteins 61 (2005) 617–632.
[240] M.J. Thompson, S.A. Sievers, J. Karanicolas, M.I. Ivanova, D. Baker, D. Eisenberg, Proc. Natl. Acad. Sci. USA 103 (2006) 4074–4078.
[241] D.M. Fowler, A.V. Koulov, C. ory-Jost, M.S. Marks, W.E. Balch, J.W. Kelly, PLoS Biol. 4 (2006) e6.
[242] J.E. Bryant, J.T. Lecomte, A.L. Lee, G.B. Young, G.J. Pielak, Biochemistry 44 (2005) 9275–9279.
[243] M.J. Gething, J. Sambrook, Nature 355 (1992) 33–45.
[244] K. Hensley, M. Mhatre, S. Mou, Q.N. Pye, C. Stewart, M. West, K.S. Williamson, Antioxid. Redox Signal. 8 (2006) 2075–2087.
[245] D. Games, M. Buttini, D. Kobayashi, D. Schenk, P. Seubert, J. Alzheimers Dis. 9 (2006) 133–149.
[246] D.A. Harris, R. Chiesa, B. Drisaldi, E. Quaglio, A. Migheli, P. Piccardo, B. Ghetti, Clin. Lab. Med. 23 (2003) 175–186.
[247] Y. Sekijima, M.A. Dendle, J.W. Kelly, Amyloid 13 (2006) 236–249.
[248] G. Wildegger, T. Kiefhaber, J. Mol. Biol. 270 (1997) 294–304.

[249] L.S. Itzhaki, D.E. Otzen, A.R. Fersht, J. Mol. Biol. 254 (1995) 260–288.

[250] D.S. Riddle, V.P. Grantcharova, J.V. Santiago, E. Alm, I. Ruczinski, D. Baker, Nat. Struct. Biol. 6 (1999) 1016–1024.

[251] V.P. Grantcharova, D.S. Riddle, J.V. Santiago, D. Baker, Nat. Struct. Biol. 5 (1998) 714–720.

[252] V. Daggett, A. Li, L.S. Itzhaki, D.E. Otzen, A.R. Fersht, J. Mol. Biol. 257 (1996) 430–440.

[253] R. Du, V.S. Pande, A.Y. Grosberg, T. Tanaka, E.S. Shakhnovich, J. Chem. Phys. 108 (1998) 334–350.

[254] V.P. Grantcharova, D. Baker, Biochemistry 36 (1997) 15685–15692.

[255] N.V. Dokholyan, L. Li, F. Ding, E.I. Shakhnovich, Proc. Natl. Acad. Sci. USA 99 (2002) 8637–8641.

[256] J.G.B. Northey, A.A. Di Nardo, A.R. Davidson, Nat. Struct. Biol. 9 (2002) 126–130.

[257] T. Lazaridis, M. Karplus, Science 278 (1997) 1928–1931.

[258] J.M. Borreguero, F. Ding, S.V. Buldyrev, H.E. Stanley, N.V. Dokholyan, Biophys. J. 87 (2004) 521–533.