PROTEINS: Structure, Function, and Genetics 44:418–427 (2001)

SHORT REVIEW

Structured Disorder and Conformational Selection

Chung-Jung Tsai,[1] Buyong Ma,[2] Yuk Yin Sham,[2] Sandeep Kumar,[2] and Ruth Nussinov[1,3]*

[1]*Intramural Research Support Program—Science Application International Corporation (SAIC), Laboratory of Experimental and Computational Biology, NCI—Frederick, Frederick, Maryland*

[2]*Laboratory of Experimental and Computational Biology, NCI—Frederick, Frederick, Maryland*

[3]*Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel*

ABSTRACT **Traditionally, molecular disorder has been viewed as local or global instability. Molecules or regions displaying disorder have been considered inherently unstructured. The term has been routinely applied to cases for which no atomic coordinates can be derived from crystallized molecules. Yet, even when it appears that the molecules are disordered, prevailing conformations exist, with population times higher than those of all alternate conformations. Disordered molecules are the outcome of rugged energy landscapes away from the native state around the bottom of the funnel. Ruggedness has a biological function, creating a distribution of structured conformers that bind via conformational selection, driving association and multimolecular complex formation, whether chain-linked in folding or unlinked in binding. We classify disordered molecules into two types. The first type possesses a hydrophobic core. Here, even if the native conformation is unstable, it still has a large enough population time, enabling its experimental detection. In the second type, no such hydrophobic core exists. Hence, the native conformations of molecules belonging to this category have shorter population times, hindering their experimental detection. Although there is a continuum of distribution of hydrophobic cores in proteins, an empirical, statistically based hydrophobicity function may be used as a guideline for distinguishing the two disordered molecule types. Furthermore, the two types relate to steps in the protein folding reaction. With respect to protein design, this leads us to propose that engineering-optimized specific electrostatic interactions to avoid electrostatic repulsion would reduce the type I disordered state, driving the molten globule (MG) → native (N) state. In contrast, for overcoming the type II disordered state, in addition to specific interactions, a stronger hydrophobic core is also indicated, leading to the denatured → MG → N state. Proteins 2001;44:418–427.**

Key words: disorder; molten globule; stability; populations; folding binding; conformational selection

INTRODUCTION

Previously, we have used the protein energy landscape concept to rationalize binding mechanisms and protein function.[1–4] We have focused on the energy landscape at and near the bottom of the funnels around the native conformation and on enzyme catalysis. Here we illustrate that the protein energy landscape theory is capable of accounting for protein function around the native state and in native disordered unfolded states. In either case, structured conformers bind via selection, with population shifts, rationalizing the rugged, rough folding funnels observed for disordered proteins. Furthermore, whether in ordered or disordered states, given structured conformations prevail, albeit with (very) different population times, driving protein binding and complex formation through conformational selection. Here we argue that this is the function that dictates the extent of ruggedness. Hence, the ruggedness in regions away from the native folded states yields information on the type of function, such as in nucleic acid-binding domains or in proteins requiring cations.

The protein energy landscape theory, or the folding funnel concept,[5–19] suggests that the most realistic model of a protein is a minimally frustrated heteropolymer with a rugged funnel-like landscape biased toward the native structure.[20] Here we focus on the rugged nature of the protein energy landscape away from native conformations. A rugged protein energy landscape has certain immediate

Grant sponsor: BSF; Grant number 95-00208; Grant sponsor: Israel Science Foundation (Israel Academy of Sciences); Grant sponsor: Magnet; Grant sponsor: Ministry of Science (Israel); Grant sponsor: Tel Aviv University; Grant sponsor: Center of Excellence (Israel Academy of Sciences); Grant sponsor: National Cancer Institute (National Institutes of Health); Grant number: NO1-CO-56000.

*Correspondence to: Ruth Nussinov, NCI—Frederick Building 469, Room 151, Frederick, MD 21702. E-mail: ruthn@ncifcrf.gov.

Received 8 December 2000; Accepted 1 May 2001

biological consequences. A rugged surface slows protein folding and increases its susceptibility to misfolding. Yet, it has been well documented that nature does not optimize proteins for rapid folding. By studying the thermodynamics and kinetics of 12 heavily mutated variants of the small IgG-binding domain of protein L, Kim et al.[21] demonstrated that the sequences of small proteins were not extensively optimized to be fast folders. Clearly, there is selective pressure against folding rates that are too slow, as can be clearly seen from the absence of extensive conformational space search and the presence of major folding path(s). Had folding rates been too slow, they would have become the rate-limiting step in protein synthesis, jeopardizing cellular development and existence. To make proteins fold reliably to their native state and avoid deadly misfolding, smooth funnels are certainly desired. However, other operative mechanisms have evolved to achieve reliable folds, such as the availability of chaperones or the intracellular presence of small organic solutes (osmolytes). In the latter mechanism, osmolytes act by raising the chemical potential of the denatured state relative to the native state.[22,23] There are two conceivable reasons evolution has not optimized the folding funnels to be smoother. First, evolution has not been able to achieve smooth folding funnels simply because of the polymeric nature of proteins. Second, the ruggedness of the potential energy surface has a biological function.

We suggest that the ruggedness of the energy landscape is biologically essential, controlling the distribution of protein conformations along the biologically relevant landscape, not necessarily around the funnel bottom. A smooth energy landscape with unique or evenly distributed random structures cannot effectively modulate protein conformational distribution. Instead, the ruggedness of the energy landscape provides specific traps for the preferred conformers. The depth of the wells and the barrier heights of the rugged traps on the energy landscape modulate the lifetimes of different conformers, depending on the biological function. We illustrate this point by reviewing recent studies of denatured proteins and native disordered proteins. Rugged funnels reflect structured, disordered proteins binding through conformational selection.

DISORDERED STATE

Recently, the issue of disorder has become very prominent. Disordered proteins include denatured proteins and native disordered proteins (or intrinsically unstructured proteins[24]). Denatured proteins have native structures under physiological conditions; however, they are denatured under harsh physical and chemical conditions. Native disordered proteins are those that lack a well-defined ordered structure under neutral pH in vitro. The sequences of native unfolded proteins have been shown to be specifically localized within a unique region of the charge-hydrophobicity phase space and indicate that a combination of low overall hydrophobicity and a large net charge represents a unique structural feature of such proteins.[25] Here we do not distinguish between denatured and native disordered proteins. We consider them to be similar on the protein energy landscape. Both are able to fold under proper conditions. For native disordered proteins, the change in conditions is frequently the presence of a ligand.

In terms of biological function, an understanding of native disordered proteins is highly desirable given the limited structural information. A particularly useful article presents a compilation of different types of proteins, all exhibiting this phenomenon.[24] This compilation and its biological analysis provide rich data for an examination of the concept and its role in regulation. Disordered proteins have been described as intrinsically unstructured proteins. A disordered region has been taken to imply a region with a high flexibility, which may become ordered upon binding to a ligand or to another molecule in general. Hence, the implication is that binding induces a conformational change. Such molecules are regarded as being inherently unstable, locally or globally; consequently, there are no preferred structures.

Completely random coiled structures can only be achieved under harsh conditions where proteins have sufficient energies to overcome the ruggedness of the energy landscape. Under such conditions, the effective energy landscape is not as rugged. This might be the case at extremely high temperatures, low pHs, or high denaturant concentrations. Under mild denaturant concentrations or normal physiological conditions, rugged energy landscapes imply uneven conformation distributions. Thus, disorder is not a completely unstructured, flexible region. Although the molecules, globally or locally, might be unstable, for each of these there is still a preferred conformation. This conformation might have a low population time and, therefore, not be observed experimentally. Nevertheless, the preferred conformation, as implied by this description, has a higher population time than all alternate conformations.

The notion that an apparently disordered state actually consists of structured conformers is not new. Shortle and his colleagues have long argued and recently presented direct evidence that there is a nativelike structure in the denatured state. Depending on the conditions, it may become more populated in some mutants and less populated in others.[26] Shortle and his colleagues have studied the thermodynamic changes due to solvent denaturation, the so-called m values. They have defined this value as the rate of change of ΔG, which will be increased or decreased in a mutant relative to the wild type as a function of the denaturant concentration. Previously, Shortle and Meeker[27] argued that the most plausible mechanism for an increase in ΔA, the new surface area exposed upon denaturation in m^+ mutants, is due to a reduced population with a nativelike structure compared with the wild type under these given conditions. However, m^- mutants illustrate a reduced ΔA compared with that of the wild-type protein under these solvent denaturing conditions. The attractive postulate that they put forward relates the changes in the m values to the populations of structured, nativelike molecules in the denatured state. In a more recent article, Wrabl and Shortle[26] provided hydrogen-exchange data validating their proposed mechanism. Shor-

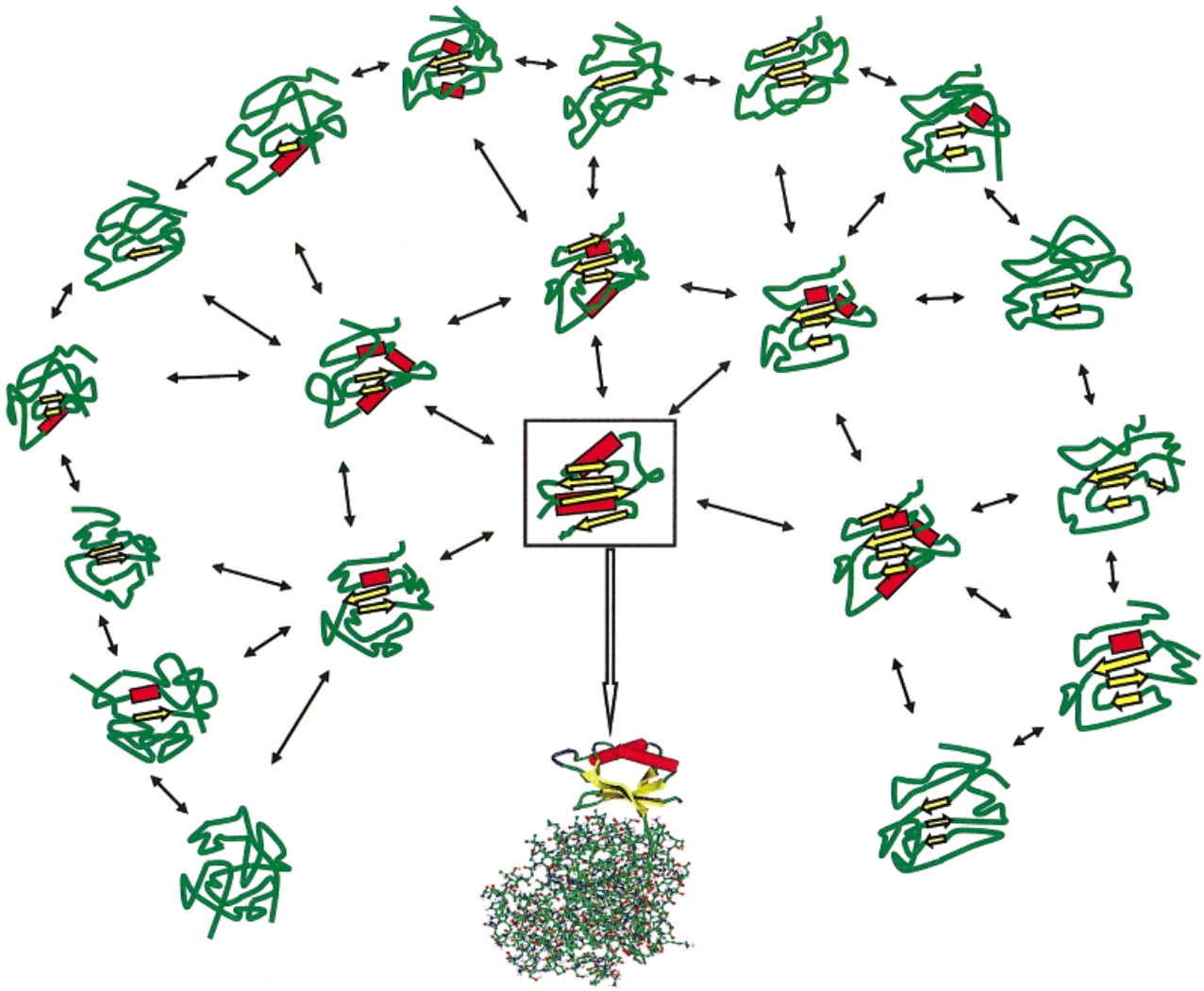

Fig. 1. Concept of the model. The figure illustrates potential conformers of the proregion of the subtilisin. Rather than there being a flexible, disordered conformation that becomes ordered upon binding, the native conformer is already present, albeit with a low population. This conformer is selected in the binding process, with the equilibrium shifting in its direction. The subtilisin is depicted on the right-hand side of the figure, with the native conformer of the proregion bound to it. Potential conformers are depicted in a half-moon form, with the native conformer boxed. Other conformers are also compact and structured, with different extents of secondary structures, and are in equilibrium with the native conformer. Subtilisin belongs to the first case type (see the text).

tle's work provides an excellent definition of the disordered state.

Consistently, Villegas et al.[28] have lent support to the notion that local interactions may persist and stabilize nativelike structured conformers present in the denatured state. Through the engineering of mutations that stabilize α-helices in the human procarboxypeptidase A2 (ADA2h), the population of the native α-helices in the denatured state increases, avoiding amyloid fibril formation. Furthermore, Demarest et al.[29] observed local interactions in structured helical regions in the molten globule state of α-lactalbumin that are not eliminated in urea. They showed that the existence of long-range stabilizing interactions between the *B*-helix and the C-terminal region[29,30] were sufficient to drive forward the formation of the secondary structure.

Analogous to the denatured states, a very well-studied case is the proregion of the subtilisin. The proregion provides strong support for the presence of a population with a native structure for native disordered proteins. Bryan and his colleagues[31,32] elegantly showed that a native, wild-type subtilisin proregion was 97% unstable. However, through the introduction of three mutations, the equilibrium for the folding of the proregion was shifted to 65% stability. Hence, if the bound conformation of the wild-type proregion were to be separated from its complexed form with the enzyme, it would be largely disordered. Nevertheless, among all conformations, the native one would have the largest population time. Hence, there is no induced fit that stabilizes the so-called disordered region; rather, there is a conformational selection of a conformer, which experimentally has not been observed

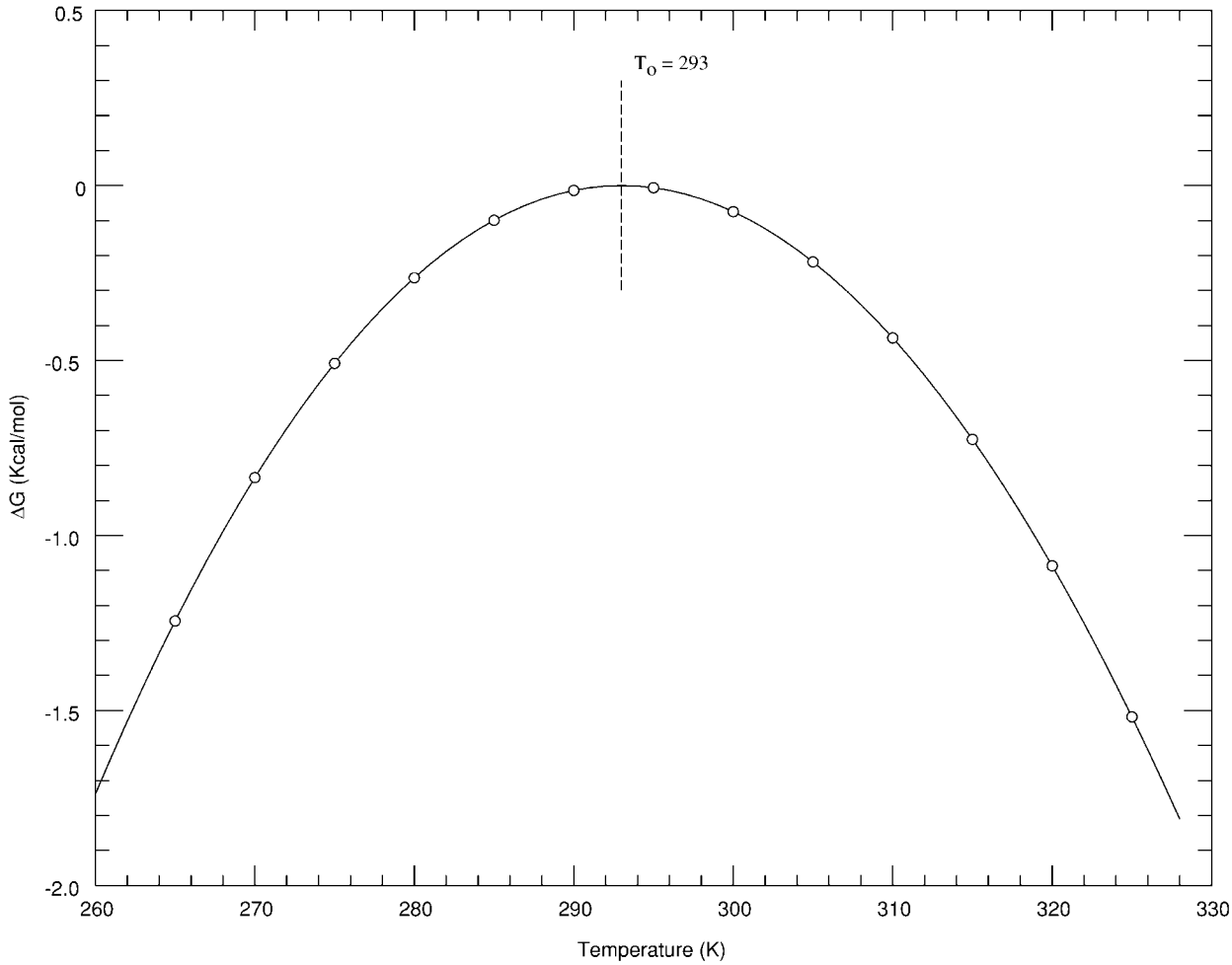

Fig. 2. Stability curve for the proregion of the subtilisin.[33] The equation used in the calculation is $\Delta G_{\text{unfolding}} = \Delta H_0 - T\Delta S_0 + \Delta C_p[T - T_0 - T \ln(T/T_0)]$, where $T_0 = 293$ K, $\Delta H_0 = 0$ kcal/mol, $\Delta S_0 = 0$ cal deg^{-1}mol^{-1}, and $\Delta C_p = 0.9$ kcal deg^{-1}mol^{-1}.

directly.[1–3] Figure 1 is a schematic illustration of the concept we are proposing. The figure depicts the proregion of the subtilisin. Figure 2 illustrates the maximum stability curve of the proregion, showing why, when it is in the unbound state, it is unstable. The figure presents ΔG as a function of the temperature for the unfolding of a double mutant (40L-57E) of the subtilisin proregion.[33] Experimentally, the curve was the outcome of measuring the population times, that is, the concentration, by far-ultraviolet circular dichroism. Here we present a theoretically fitted plot, calculated with the Gibbs–Helmholz equation. The equation and parameters are given in the legend to Figure 2.

TWO TYPES OF CASES: WITH AND WITHOUT AN APPRECIABLE HYDROPHOBIC CORE

Type I: With a Hydrophobic Core

Disordered proteins can be classified into two types of cases. In the first case, if we pulled the complexed, bound molecules apart, the conformation equivalent to that of the bound species would still be compact. This case applies to the proteins studied by Shortle and his colleagues.[26] Such conformers have hydrophobic cores and some stability. Furthermore, this type of case is applicable to the molten globule situation. A molten globule is not unstructured. Rather, the molten globule is a compact protein, with a predominant nativelike conformation. However, the population time of this conformation is low and, therefore, undetected. α-lactalbumin provides an excellent example (reviewed by Kuwamura[34]). In his review, Kuwamura highlighted several points about the α-lactalbumin molten globule, noting in particular that it is not a nonspecific, collapsed polypeptide; conversely, it has a nativelike tertiary fold. Furthermore, the fact that molten globules have also been described as heterogeneous implies that the native-state conformation is relatively unstable; therefore, although the most prevalent, it is still too low to detect. This view is shared and promoted by Fontana and his colleagues.[35,36]

An additional attractive example was provided by Fedorov and Baldwin,[37] who studied the folding of the β-subunit of bacterial luciferase. They showed that the folding of the β-subunit in the presence of the β-subunit was substantially faster than the refolding of the β-sub-

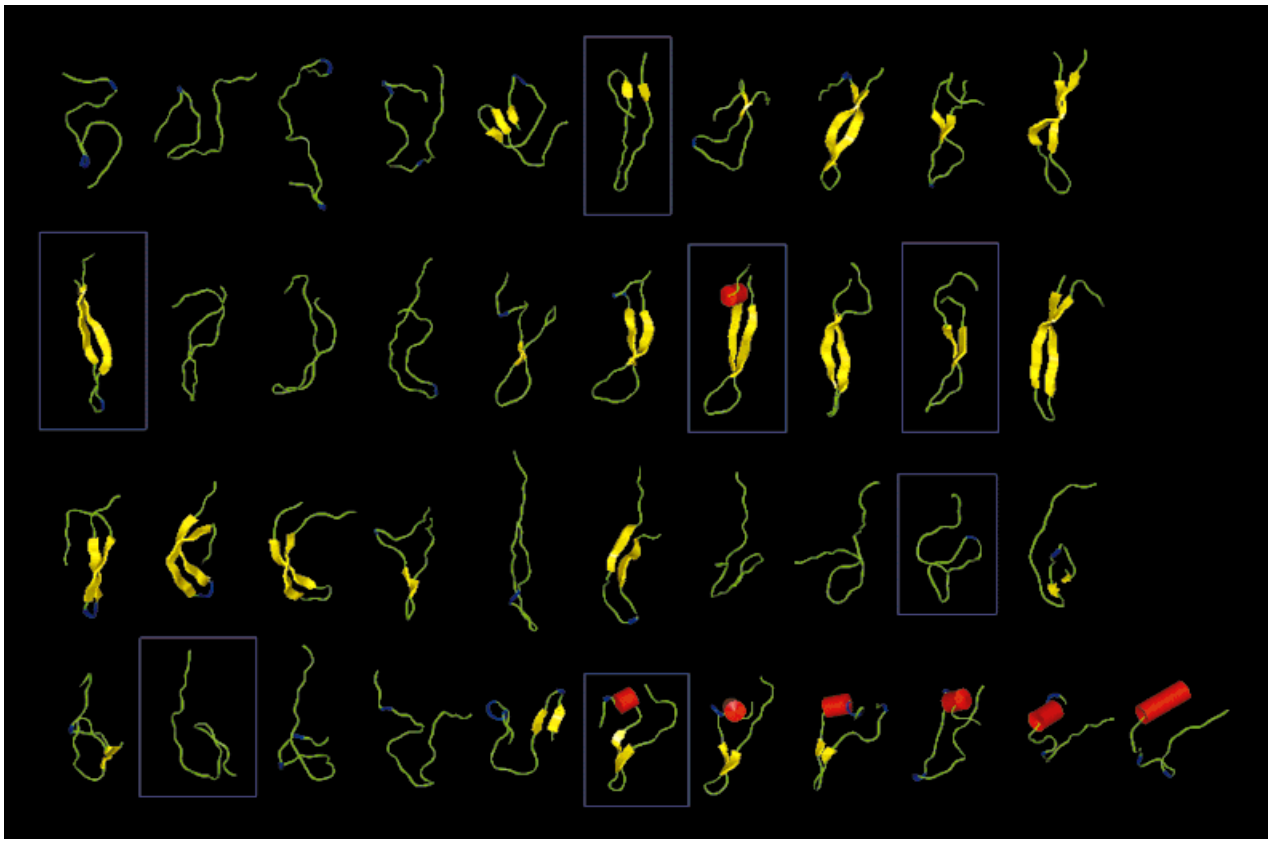

Fig. 3. The NH_2-terminal building block fragment of the *E. coli* dihydrofolate reductase belongs to the second case type, where the stability is low. Nevertheless, the seemingly disordered molecule still has a predominant conformation, even if its population is lower than in the first type and is, therefore, not observed experimentally. To identify other highly populated conformers of this fragment, we carried out a high-temperature molecular dynamic simulation. The structure of the N-terminus (residues 1–36) was taken from the native structure of *E. coli* dihydrofolate reductase from the Protein Data Bank (7DFR[50]). The simulation was carried out with the CHARMM (version c27b1) software package. A CHARMM19 polar hydrogen force field with an empirical energy function (EEF1) implicit solvent model[51] was used. All missing hydrogen atoms were added with the Hbuild algorithm. The initialization involved a 1000 steps of Adapted Basis Newton Raphson (ABNR) minimization, followed by 10 ps of heating to 300 K and a 20-ps equilibration at the same temperature. The final structure at the end of the initialization was used as the starting point for the high-temperature simulation at 500 K for 4 ns. The coordinates of the N-terminus were saved at 0.1-ns time intervals, resulting in the 40 conformers shown. The stability of each of these conformers was subsequently tested by low-temperature simulation. Each structure was subjected to a 500 ABNR minimization followed by 100 ps at 300 K. Conformers with a C_αRMSD less than 3.2 Å relative to the starting conformation of the low-temperature simulation were considered stable and are boxed in the figure.

unit after its denaturation by urea. In the absence of the α-subunit, the β-subunit was trapped in a well, inhabited by an ensemble of disordered conformations, all with population times still lower than that of the native conformation.

Type II: Without an Appreciable Hydrophobic Core

In the second case of disordered molecules, we pull apart two intertwined monomers, such as in the case of a functional dimer. In such a case, the consequences of such an action will be to produce a noncompact conformation. This situation is reminiscent of that of the fluctuating building blocks.[2,38] In such a case, there is a population of unstable, fluctuating fragments, with population times considerably lower than those observed in molecules belonging to the first case. Such cases can still be detected experimentally, as in the case of short fragments forming helices, or via computations. A good example of such a computational approach is the program Agadir, which predicts the helical behavior of monomeric peptides on the basis of the helix–coil transition theory.[39,40] The leucine zipper, monellin, and troponic C provide excellent examples of such disordered molecules. Figure 3 presents an example from our simulations. The figure presents the native conformation and 40 snapshots from high-temperature (500 K) simulations of the unstable, N-terminal building block fragment of *Escherichia coli* dihydrofolate reductase. Despite its instability, the features of the native conformation are largely preserved. We subsequently subjected all conformers to long simulations at 300 K to examine their stability. Seven conformations were preserved, and they are boxed in the figure. Hence, although the population time of this N-terminal building block fragment is low, it may still be higher than all alternate conformations. The variability in the conformations, the outcome of instability, results in a seemingly disordered, unstructured state.

The inhibitor of the aspartic proteinase from *Saccaromyces cerevisiae*[41] provides another particularly adequate example for this second case type. Although the inhibitor

is 68 residues long, the inhibitory activity resides in a short, 32-residue fragment, at the amino terminus of the molecule. Interestingly, on its own the inhibitor appears to be disordered, not possessing any defined structure. Yet, upon binding to the proteinase, the fragment assumes a near perfect α-helix conformation. Hence, rather than the proteinase inducing a conformational change in the inhibitor and folding it, the helical conformation exists in solution a priori, albeit in a low but still higher concentration than those of all alternates.

The examples presented in the review by Wright and Dyson[24] belong to both types of cases. Although varied, there is still a common theme. In all cases, one conformation predominates, having the highest population time. It might not be observed experimentally. Furthermore, it is not necessarily the conformation with the highest population time that is selected and observed when in the bound form. However, once bound, the equilibrium would change in its favor. Hence, rather than being unstructured in solution and becoming structured upon binding, the native or nativelike conformation is the most populated one, although its concentration may be too small to be observed. A particularly nice example is that of the GCN4. The conformation of the bound state of GCN4 is unstable, with a low population time. However, once bound to the DNA, the equilibrium shifts toward this conformer. Binding to the DNA lends stability to the GCN4. However, binding is via selection, instead of DNA inducing the conformational change.

Relating the Steps in the Folding Reaction to the Two Disorder Types

In a query of why native unfolded proteins are unstructured under physiological conditions, Uversky et al.[25] studied the relationship of protein folding properties, the mean hydrophobicity of the amino acid sequence, and the normalized protein net charges. In the charge-hydrophobicity phase space, native folded proteins occupy the space with a lower mean net charge and a higher mean hydrophobicity, whereas native unfolded proteins reside in the region of a higher mean net charge and a lower hydrophobicity. These two regions merge with a mixture of the two.

For the native folded protein, our recently developed building block model[38] may provide insight as well. In analyzing the anatomy of a protein structure, we probe possible folding routes by dividing the protein into building blocks of various sizes. The hydrophobicity of the building block is calculated as the fraction of the buried nonpolar area out of the total nonpolar area.[38] Figure 4 shows the average hydrophobicity and standard deviation for all fragments in the 930 representative chains as a function of fragment size. If a protein can be divided into building blocks and each of the building blocks has a hydrophobicity higher than the average values in Figure 4, it is likely that this protein has a type I unfolded state. If each of the building blocks has a hydrophobicity score lower than the average values, it may have a type II unfolded state, (i.e., without a hydrophobic core). In the type I case, the first step leading to hydrophobic collapse achieves such a hydrophobicity. In contrast, in type II the hydrophobic core is not strong enough and falls below the average. Hence, the folding reaction of type I may differ from that of type II.

As the temperature decreases from the heat-transition temperature (T_G) to the cold-denaturation-transition temperature (T_G'), the protein folds. There are two steps in the denatured (D) to native (N) folding reaction. The first involves hydrophobic collapse, leading to the noncompact molten globule (MG) state. In the second, the MG state settles down to a single native conformation. To understand these folding steps, we need to consider the Gibbs energy change, $\Delta G = \Delta H - T\Delta S$, in this $T_G' < T < T_G$ temperature range. According to Griko,[42] the first step may be driven by entropy. In this step, both the entropy and enthalpy of the system increase. Hence, the D → MG formation is under entropy control. In the next MG → N settling-down step, both the entropy and enthalpy decrease. This combinatorial rearrangement step of the already largely folded local building block elements is driven by its favorable enthalpy. The native state has lower entropy than the intermediate MG. However, this step also leads to the formation of favorable specific interactions and, therefore, lower enthalpy, putting it under enthalpy control.

The two-step folding process may be used to describe the first disordered case type, where the protein has an appreciable hydrophobic core, for example, in α-lactalbumin. However, in the second case type, in the absence of a hydrophobic core, what we observe is a single D → N step. Considering the contributions to the Gibbs energy change in the temperature zone of the D-to-N change, the water structure changes, resulting in an unfavorable reduction in the conformational entropy. However, in the native state specific favorable interactions are formed, lowering the enthalpy and putting the D → N step under favorable enthalpy control. Here the native state can be that of a building block element of the protein that by itself is in a disordered state but, upon binding to other building block elements, reaches the native conformation. Alternatively, it may be a peptide or a monomer belonging to a functional dimer.

Why then over $T_G' < T < T_G$ do we observe disordered states for types I and II? For type I, because the temperature is reduced and the molecule has a relatively appreciable hydrophobic core, the favorable hydration entropy will drive the D → MG step. However, the MG → N step is unfavorable, leading to a relatively low population of the native state. Because this step is under enthalpy control, this implies that in our case the optimization of specific interactions is not favorable enough to compensate for the unfavorable decrease in conformational entropy. We may assume that such an effect can be the outcome of repulsive electrostatic interactions, preventing compact native state formation. That this might be the case is observed both in the effect of lowering of the pH to produce an MG state and in the absence of Ca^{2+} in α-lactalbumin. This suggests that altering the protein sequence via engineered muta-

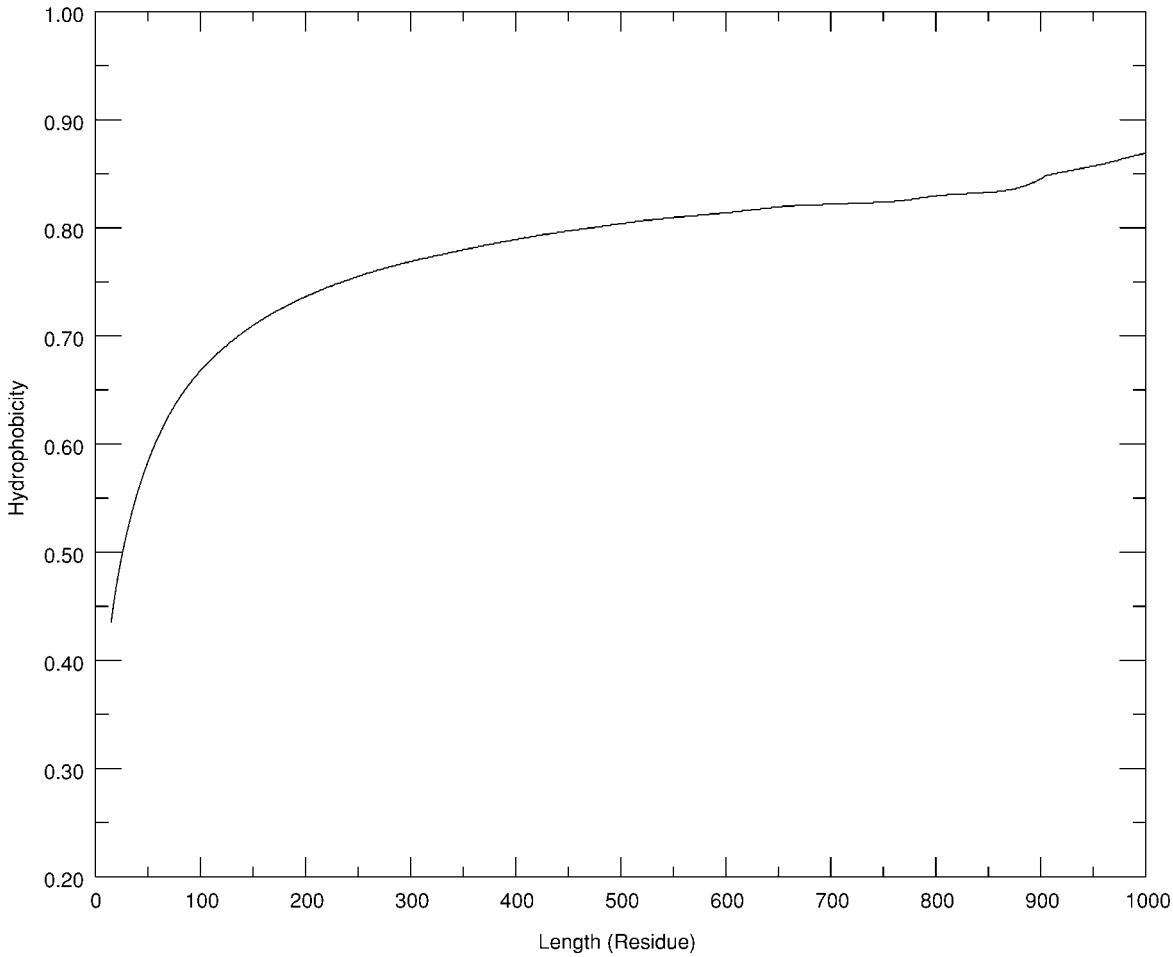

Fig. 4. Plots of averaged values of hydrophobicity based on fragments generated from a representative data set of 930 single-chain proteins. To calculate the hydrophobicity, we divided the solvent accessible surface area (ASA) of a fragment (or of a molecule) into nonpolar ASA (^{Non}ASA) and polar ASA (^{Pol}ASA) according to the atom type to which the ASA belonged. The hydrophobicity was defined as the fraction of the buried nonpolar area out of the total nonpolar area, $H = {}^{Non}ASA_{Buried}/({}^{Non}ASA_{Buried} + {}^{Non}ASA_{Surf})$, where $^{Non}ASA_{Buried}$ and $^{Non}ASA_{Surf}$ are the buried and exposed nonpolar ASAs, respectively. For a chain with size N_e residues and with a size limit of a fragment set to N_s residues, the total number of sampled fragments was $N_{total} = \Sigma(N_e - N_i + 1)$, where N_i runs the summation from N_s to N_e. In the plot, the total number of sampled fragments for 15, 400, and 900 residues is 194,127, 19,621, and 132, respectively.

tions to overcome the electrostatic repulsion should lead to driving the reaction to the native state.

Here it is particularly interesting to compare α-lactalbumin and hen egg white lysozyme, two proteins with virtually identical structures in the native state. Yet, α-lactalbumin forms an MG state at a lower pH or in the absence of Ca^{2+}, whereas no MG state has been observed in lysozyme. The difference between the two might reside in the electrostatic repulsion in the former, which is absent in the latter. In this regard, it is particularly interesting to note the recent results of Raleigh and his colleagues.[52] Previously, Demarest et al.[43] showed that peptide constructs from the α-domain of hen lysozyme are largely unstructured. However, the human α-LA peptide construct is largely helical and displays all the characteristics of an MG state. Furthermore, Demarest et al. showed that the integrity of the α-LA α-domain is critical for MG formation by the protein. A more recent analysis by Raleigh and his colleagues (D. Raleigh, personal communication, 2001) illustrates that the α-domain of lysozyme is more highly packed than the corresponding domain from α-LA. Furthermore, the lysozyme domain buries a larger extent of polar surface area than the α-domain of α-lactalbumin. However, the extent of nonpolar buried surface areas is similar in both corresponding protein domains. These suggest that the α-domain of lysozyme has more optimized specific (vdW, electrostatic) interactions, driving the MG → N step and, therefore, the experimentally unobserved MG state in lysozyme, making it a two-state D ↔ N reaction. Consistently, Freire[44] has shown that there is a larger extent of disrupted polar interactions upon MG formation in α-LA.

Type II is reminiscent of the fluctuating building blocks with relatively low population times. For type II, most of the hydrophobic core is interbuilding block or at the intermolecular interface. In the absence of an intrabuilding block or intramolecular hydrophobic core, there is no appreciable hydrophobic collapse. Thus, the population of the native state is further reduced. Here a D-to-N folding step is unfavorable. Because such a step entails an unfavor-

able reduction in hydration entropy and the favorable reduction in enthalpy due to the formation of specific interactions is not large enough to compensate for the unfavorable entropic reduction, the native state is unobservable.

Types I and II: A Continuum of Distributions of Hydrophobic Cores

The utility of any scheme is a function of its ability to provide predictive value. Here we have classified the disordered state into types I and II. Thus, it would have been particularly beneficial if we could provide guidelines for defining the two states, which would enable their prediction.

Globular proteins yield a continuous hydrophobicity function. The distinction between types I and II rests on the population times. Over the last few years, interest has largely focused on distinguishing between ordered and disordered states, rather than characterizing apparent classes within the disordered state. Although different spectroscopic techniques can probe ordered conformations with different lifetimes, the ability to detect an ordered conformation is related to the sensitivity of the type of experiment. Currently, such sensitivity is unfortunately limited, and there is no experimental evidence for this two-type classification.

Although clear definition and consequent prediction are infeasible, nevertheless, we may resort to some general guidelines. On the experimental side, large radii of gyration, or an observed lack of protection in hydrogen-exchange experiments, might provide some clue. On the computational side, if only sequence information is available, one may get some idea from plots of the mean hydrophobicity as a function of the mean net charge (e.g., Fig. 5 in Uversky et al.[25]). In such a plot, the smaller the hydrophobicity is and the larger the mean net charge is, the further away the protein is from the native state. Proteins falling in the upper left corner of such a plot are likely to belong to the type II category. If, however, their mean hydrophobicity is larger and the mean net charge is smaller, they are likely to fall into the type I category. These are the two extremes. Between them, there is a continuum of hydrophobic cores and of mean net charge with a range of corresponding population times. If the three-dimensional structure is also available, we may compute the hydrophobicity. Nevertheless, prediction based on the extent of the hydrophobic core to which the protein belongs provides mere general guidelines. If we are to use a plot such as Figure 4, we implicitly assume that the electrostatic interactions have been optimized. However, in practice this is not necessarily the case.

Here we have focused on globular proteins. For helices, population times have been both measured and computed. Unlike the situation in globular proteins, in helices the hydrophobic effect is important, but it is not the driving force. Hydrogen-bond formation has been optimized. Population times of α-helices are a function of their hydrophobic cores and of their exposed nonpolar surface areas. Polyalanine, with a relatively small hydrophobic core, is stable. However, helices with large hydrophobic side-chains and, consequently, an appreciable extent of exposed nonpolar surface area, are likely to have low population times.

Hence, the critical points to consider are (1) the continuity of the hydrophobicity function and (2) the realization that types I and II represent the two extremes of this continuum. Different disordered proteins would map into this continuum, depending on the extent of the optimization of their electrostatic interactions.

Evolutionary Advantage of Disorder for Biological Function

It is tempting to speculate on the evolutionary advantage of disorder for biological function. First, an examination of the cases presented in the review by Wright and Dyson[24] illustrates that in all examples, the proteins that are shown to be in such a disordered state are biologically active in the form of large, supramolecular assemblies. The authors presented examples of domains that appear disordered in solution but manifest a well-defined conformation when bound. Intriguing examples illustrate cases from the areas of cell cycle control, signal transduction, and transcriptional and translational regulation, suggesting that low population times of a range of conformations confer functional advantage. It is attractive to propose that this allows progressive enrichments of some conformations as the molecule goes through its binding cascade.[3,45] At successive binding stages, the conditions continuously change, illustrating a dynamic landscape, with a changed equilibrium. A nice example is that of the $p21^{\text{H-ras}}$ protein.[46] *p21* mediates a large number of binding events[47] (see also the discussion of Wright and Dyson[24]). In signal transduction and gene expression pathways, a given molecule frequently binds different signaling proteins to switch on or off different processes. Different molecules may select and bind alternate conformations.

Second, many RNA-binding proteins whose structures have been studied fall into this category. Regions in RNA are conformationally fluctuating. Under such circumstances, it is advantageous that their cognate binding proteins exist in a range of conformations as well.

Third, Fontana and his colleagues[35,36] recently put forward an intriguing proposition supported by experiment, that proteases recognize and cleave regions that in the native state are characterized by high segmental mobility. Such regions typically manifest high crystallographic B-factors, devoid of regular secondary structures. They further proposed that a mechanism of local unfolding is required for selective peptide bond fission of a protein substrate. Hence, the substrate loop appears disordered, as it is a mixture of conformational states. Nevertheless, there is a preferred binding conformation that, although having a low population time, still has a higher concentration than alternate conformers.

Fourth, complexes above all are largely stabilized by the hydrophobic effect. In general, the stronger the hydrophobic core is, the more stable the complex is. Yet, if the hydrophobic effect at the interface between two bound molecules is too extensive, such as in the case of functional

dimers, pulling them apart will inevitably result in destabilizing the separated conformation, resulting in unfolding, that is, molecular disorder.

Disordered functional proteins provide evidence that the function of a protein and its properties are not only decided by its static folded three-dimensional structure; they are determined by the distribution and redistribution of the conformational substates. A recent enumeration[48] of all sterically allowed conformations for short polyalanine chains consistently shows that in the denatured state not all conformational states are accessible. Even for alanines, local steric effects beyond nearest neighbors already restrict significantly the conformational space. For variable-sequence chains, with bulkier side-chains this effect is likely to be enhanced, likely biasing the local conformations.[48,49] It will be interesting to see how this bias relates to the local conformations existing in the native state.

CONCLUSIONS

Traditionally, disorder has been looked upon as local or global instability. The term has been applied to situations where no atomic coordinates can be derived from crystallized molecules. These regions have been viewed as inherently unstructured. However, even in seemingly disordered molecules, there are prevailing conformations, with population times higher than those of all alternate conformations. We have presented two types of cases. The first type includes molecules possessing a hydrophobic core, and so even if their native conformation is unstable, its population time is high enough to enable its experimental detection (as in the case of the proregion of the subtilisin). The second type includes molecules that do not manifest a hydrophobic core and are, therefore, less stable, with still lower population times. These types of cases cannot be detected or differentiated experimentally, and so our proposition for such a situation cannot be validated directly. To date, the only way to illustrate the existence of a predominant conformation with a population time higher than all others is through simulations.

Above all, such a view is logical. If the inhibitor of the aspartic proteinase did not have a preferred conformation (i.e., that of the helix), the enzyme would have had to search for a favorable inhibitor conformation over broad space in timescales not biologically relevant. Hence, the fact that binding is fast implies selection: The helix conformation is there, and its population is higher than all others. With its binding, the equilibrium shifts in its favor, further driving the reaction. As binding and folding are similar processes with similar underlying principles, with the only exception being chain connectivity, this principle applies to disordered molecules in binding and to unstable, conformationally fluctuating building block fragments in folding. Folding and binding imply selection, whether of building blocks, hydrophobic folding units, domains, or subunits or of any ligand, small molecule, or nucleic acid. Such an approach rationalizes rugged energy landscapes away from the native conformations.

On the practical side from the viewpoint of protein design, for engineering higher population times for an intermediate MG state in a protein that illustrates a two-step folding process (i.e., a type I case), it is not the hydrophobic effect that should be enhanced. Rather, introducing repulsive electrostatic interactions that would be overcome by binding to a ligand or a cofactor is the key. Conversely, for reducing the population of the MG state, engineering favorable electrostatic interactions would drive the MG $\rightarrow$ N step. With regard to type II, for reducing the population of the disordered state and driving the D $\rightarrow$ N reaction, engineering a stronger hydrophobic core, in addition to specific interactions, is the key.

ACKNOWLEDGMENTS

The authors thank Dr. Daniel Raleigh for communicating with them and allowing them to cite his data before publication. They also thank Dr Jacob V. Maizel for numerous helpful discussions. The research of R. Nussinov in Israel was supported in part by a grant from BSF (Israel), a grant from the Israel Science Foundation (administered by the Israel Academy of Sciences), a Magnet grant, a Ministry of Science grant, Tel Aviv University Basic Research grants, and the Center of Excellence (administered by the Israel Academy of Sciences). This project was funded in whole or in part with Federal funds from the National Cancer Institute (National Institutes of Health). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

REFERENCES

1. Ma B, Kumar S, Tsai CJ, Nussinov R. Folding funnels and binding mechanisms. Protein Eng 1999;12:713–720.
2. Tsai CJ, Kumar S, Ma B, Nussinov R. Folding funnels, binding funnels and protein function. Protein Sci 1999;8:1181–1190.
3. Kumar S, Ma B, Tsai CJ, Sinha N, Nussinov R. Folding and binding cascades: dynamic landscapes and population shifts. Protein Sci 2000;9:10–19.
4. Ma B, Tsai CJ, Nussinov R. In search of an intra-molecular chaperone-like building block fragments. Protein Eng 2000;13:617–627.
5. Bryngelson JD, Wolynes PG. Intermediates and barrier crossing in a random energy model (with applications to protein folding). J Phys Chem 1989;93:6902–6915.
6. Frauenfelder H, Sligar SG, Wolynes PG. The energy landscapes and motions of proteins. Science 1991;254:1598–1603.
7. Karplus M, Shakhnovitch EI. Protein folding: theoretical studies of thermodynamics and dynamics. In: Creighton T, editor. Protein folding. New York: Freeman; 1992. p 127–195.
8. Karplus M, Sali A, Shakhnovitch E. Comment: kinetics of protein folding. Nature 1995;373:664–665.
9. Wolynes PG, Onuchic JN, Thirumalai D. Navigating the folding routes. Science 1995;267:1619–1620.
10. Onuchic JN, Wolynes PG, Luthey-Schulten Z, Socci ND. Towards an outline of the topography of a realistic protein folding funnel. Proc Natl Acad Sci U S A 1995;92:3626–3630.
11. Onuchic JN, Socci ND, Luthey-Schulten Z, Wolynes PG. Protein folding funnels: the nature of the transition state ensemble. Folding Des 1996;1:441–450.
12. Baldwin RL. Matching speed and stability. Nature 1994;369:183–184.
13. Baldwin RL. The nature of protein folding pathways: the classical versus the new view. J Biomol NMR 1995;5:103–109.
14. Bozko EM, Brooks CL III. First principles calculation of the folding free energy for a three helix bundle protein. Science 1995;269:393–396.

15. Dill KA, Chan HS. From Levinthal to pathways to funnels. Nat Struct Biol 1997;4:10–19.
16. Karplus M. The Levinthal paradox: yesterday and today. Folding Des 1997;2:S69–S75.
17. Lazaridis T, Karplus M. "New view" of protein folding reconciled with the old through multiple unfolding simulations. Science 1997;278:1928–1931.
18. Gruebele M, Wolynes P. Satisfying turns in folding transitions. Nat Struct Biol 1998;5:662–665.
19. Dill KA. Polymer principles and protein folding. Protein Sci 1999;8:1166–1180.
20. Shea JE, Onuchic JN, Brooks CL III. Exploring the origins of topological frustration: design of a minimally frustrated model of fragment B of protein A. Proc Natl Acad Sci U S A 1999;96:12512–12517.
21. Kim DE, Gu H, Baker D. The sequences of small proteins are not extensively optimized for rapid folding by natural selection. Proc Natl Acad Sci U S A 1998;95:4982–4986.
22. Arakawa T, Bhat R, Timasheff S. Preferential interactions determine protein solubility in three-component solutions: the $MgCl_2$ system. Biochemistry 1990;29:1914–1923.
23. Qu Y, Bolen CL, Bolen DW. Osmolyte-driven contraction of a random coil protein. Proc Natl Acad Sci U S A 1998;95:9268–9273.
24. Wright PE, Dyson HJ. Intrinsically unstructured proteins: reassessing the protein structure–function paradigm. J Mol Biol 1999;293:321–331.
25. Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 2000;41:415–427.
26. Wrabl J, Shortle D. A model of the changes in denatured state structure underlying *m* value effects in staphylococcal nuclease. Nat Struct Biol 1999;6:876–882.
27. Shortle D, Meeker AK. Mutant forms of staphylococcal nuclease with altered patterns of guanidine hydrochloride and urea denaturation. Proteins 1986;1:81–89.
28. Villegas V, Zurdo J, Filimonov VV, Aviles FX, Dobson CM, Serrano L. Protein engineering as a strategy to avoid formation of amyloid fibrils. Protein Sci 2000;9:1700–1708.
29. Demarest SJ, Fairman R, Raleigh DP. Peptide models of local and long-range interactions in the molten globule state of human α-lactalbumin. J Mol Biol 1998;283:279–291.
30. Demarest SJ, Horng JC, Raleigh DP. A protein dissection study demonstrates that two specific hydrophobic clusters play a key role in stabilizing the core structure of the molten globule state of human α-lactalbumin. Proteins 2000;42:237–242.
31. Wang L, Ruan B, Ruvinov S, Bryan PN. Engineering the independent folding of the subtilisin BPN′ pro-domain: correlation of the pro-domain stability with the rate of subtilisin folding. Biochemistry 1998;37:3165–3171.
32. Ruan B, Hoskins J, Bryan PN. Rapid folding of calcium-free subtilisin by a stabilized pro-domain mutant. Biochemistry 1999; 38:8562–8571.
33. Ruvinov S, Wang L, Ruan B, Almog O, Gilliland GL, Eisenstein E, Bryan PN. Engineering the independent folding of the subtilisin BPN′ prodomain: analysis of two-state folding versus protein stability. Biochemistry 1997;36:10414–10421.
34. Kuwamura K. The molten globule state of α-lactalbumin. FASEB J 1996;10:102–109.
35. Polverino de Laureto P, Scaramella E, Frigo M, Wondrich FG, De Filippis V, Zambonin M, Fontana A. Limited proteolysis of bovine α-lactalbumin: isolation and characterization of protein domains. Protein Sci 1999;8:2290–2303.
36. Fontana A, Zambonin M, Polverino de Laureto P, De Filippis V, Clementi A, Scaramella E. Probing the conformational state of apomyoglobin by limited proteolysis. J Mol Biol 1997;266:223–230.
37. Fedorov AN, Baldwin TO. Process of biosynthetic protein folding determines the rapid formation of native structure. J Mol Biol 1999;294:579–586.
38. Tsai CJ, Maizel JV, Nussinov R. Anatomy of protein structures: visualizing how a 1D protein chain folds into a 3D shape. Proc Natl Acad Sci U S A 2000;97:12038–12043.
39. Munioz V, Serrano L. Elucidating the folding problem of helical peptides using empirical parameters. Nat Struct Biol 1994;1:399–409.
40. Lacrroix E, Viguera AR, Serrano L. Elucidating the folding problem of α-helices: local motifs, long-range electrostatics, ionic strength dependence and prediction of NMR parameters. J Mol Biol 1998;284:173–191.
41. Li M, Phylip LH, Lees WE, Winther JR, Dunn BM, Wlodawer A, Kay J, Gutchina A. The aspartic proteinase from *Saccaromyces cerevisiae* folds its own inhibitor into a helix. Nat Struct Biol 2000;7:113–117.
42. Griko YV. Energetic basis of structural stability in the molten globule state: α-lactalbumin. J Mol Biol 2000;297:1259–1268.
43. Demarest SJ, Boice JA, Fairman R, Raleigh DP. Defining the critical core of α-lactalbumin molten globule state. J Mol Biol 1999;294:213–221.
44. Freire F. Thermodynamics of partly folded intermediates in proteins. Annu Rev Biophys Biomol Struct 1995;24:141–165.
45. Tsai CJ, Ma B, Nussinov R. Folding and binding cascades: shifts in energy landscapes. Proc Natl Acad Sci U S A 1999;96:9970–9972.
46. Zhang J, Matthews CR. Ligand binding is the principal determinant of stability for the $p21^{\text{H-ras}}$ protein. Biochemistry 1998;37: 14881–14890.
47. Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE. Structural studies of $p21^{Waf1/Cip1/Sdi1}$ in the free and Cdk2-bound state: conformational disorder mediates binding diversity. Proc Natl Acad Sci U S A 1996;93:11504–11509.
48. Pappu RV, Srinivasan R, Rose GD. The Flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding. Proc Natl Acad Sci U S A 2000;97:12565–12570.
49. Baldwin RL, Zimm BH. Are denatured proteins ever random coils? Proc Natl Acad Sci U S A 2000;97:12391–12392.
50. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The protein databank: A computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542
51. Lazaridis T, Karplus M. Effective energy function for proteins in solution. Proteins 1999;35;133–152.
52. Demarest SJ, Horng JC, Raleigh DP. A protein dissection study demonstrates that two specific hydrophobic clusters play a key role in stabilizing the core structure of the molten globule state of human α-lactalbumin. Proteins 2001;42:237–242.