HFSP Journal

ISSN: 1955-2068 (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tfls19

Characterization of the unfolded state of repeat proteins

Amit Mor , Gilad Haran & Yaakov Levy

To cite this article: Amit Mor , Gilad Haran & Yaakov Levy (2008) Characterization of the unfolded state of repeat proteins, HFSP Journal, 2:6, 405-415, DOI: 10.2976/1.3021145

To link to this article: https://doi.org/10.2976/1.3021145

Published online: 07 Sep 2010.

Characterization of the unfolded state of repeat proteins

Amit Mor,[1] Gilad Haran,[2] and Yaakov Levy[1]

[1]Department of Structural Biology, Weizmann Institute of Science, Rehovot, 76100, Israel
[2]Department of Chemical Physics, Weizmann Institute of Science, Rehovot, 76100, Israel

(Received 27 August 2008; accepted 20 October 2008; published online 12 November 2008; corrected 9 December 2008)

The unfolded state ensemble of proteins has been described as a structurally featureless state. While this approach is supported by the fact that many unfolded proteins follow the scaling law behavior of a random coil, there is evidence that the unfolded states of various proteins are stabilized by native or non-native interactions. Recently, the existence of extensive non-native structure was reported for a repeat protein, which resulted in a scaling law exponent that is significantly smaller than that of a random polymer [Cortajarena *et al.*, *J. Mol. Biol.* 382(1), 203–212 (2008)]. It was concluded that the high compactness of this protein stems from a significant fraction of interacting PP_{II} helical segments in the unfolded state. In this study, we aim at providing possible molecular understanding of this anomalous compactness of the unfolded state and to investigate its origin. Using a hierarchy of computational models, we ask whether in general the unfolded state of a repeat protein is likely to be intrinsically more compact than the unfolded state of globular proteins, or whether this phenomenon depends mostly on the occurrence of a specific sequence that promotes PP_{II} conformations. Our results suggest that the formation of the PP_{II} conformation is indeed essential, yet the recurring sequence of repeat proteins promotes the interactions between these PP_{II} segments and the formation of non-native interactions in the unfolded state. [DOI: 10.2976/1.3021145]

CORRESPONDENCE
Dr. Yaakov Levy:
koby.levy@weizmann.ac.il

The unfolded state of proteins has evolved lately from being viewed as an elusive entity to an acknowledged player in the folding process. For decades, researches in the fields of protein science and molecular biology focused on the characterization of the native states of proteins with the main goal of understanding protein function from structure. Later, numerous protein-folding studies have dealt with deciphering folding mechanisms and pathways by characterizing the intermediates and the transition state ensembles that govern folding kinetics (Baldwin, 2008; Dill *et al.*, 2008; Onuchic *et al.*, 1997; Onuchic and Wolynes, 2004; Shakhnovich, 2006). In recent years, the unfolded state of proteins, which was often described as a random coil, has started to be appreciated as a state with much richer physics. Accordingly, unfolded proteins are now considered to be different from random flexible polymers that have no stabilizing interactions. Therefore, they may have nonzero enthalpy and their entropy may be lower than that of a random coil.

One of the widely accepted models for the description of the unfolded state of proteins regards the polypeptide chain as a self-avoiding walk (SAW). In such a description (de Gennes, 1979; Flory, 1953), the polypeptide chain has properties that allow it to wander freely through conformational space, with the only restriction being the inability of the chain to intersect itself. This model provides a scaling law relationship for the size of such molecular species, measured by the relation between an effective radius (e.g., the radius of gyration—R_g) and the number of monomers (i.e., amino acids) comprising the chain (N). This scaling law may be written as $R_g = R_0 N^{\nu}$, where R_0 is a constant that is a function of the persistence

length of the chain and ν is a scaling exponent. The theoretical scaling exponent suggested by Flory for such SAWs is 0.6 (de Gennes, 1979; Flory, 1953). In the case of denatured proteins, the Flory scaling law was corroborated by both experimental (Kohn *et al.*, 2004; Tanford, 1968; Wilkins *et al.*, 1999) and by further theoretical investigations (Ding *et al.*, 2005; Fitzkee and Rose, 2004) and was found to be of the scales 0.58 to 0.67.

It is now understood that SAW statistics do not preclude the possibility of residual structure in the denatured states. For example, it has been shown that in the presence of rigid segments (i.e., secondary structure regions) in as much as 92% of the protein sequence, SAW statistics still prevail (Fitzkee & Rose, 2004; Wang *et al.*, 2007). This indicates that, while denatured proteins follow the polymer physics of a random coil, this does not exclude the possibility that the denatured proteins can have structural elements, either native or non-native. Several experimental studies support residual nativelike structural elements in the denatured state (Religa *et al.*, 2005; Shortle and Ackerman, 2001; Yi *et al.*, 2000). Theoretical and computational studies have addressed the role of native interactions in the denatured state and proposed a significant bias toward the native structures (Ding *et al.*, 2005; Pappu *et al.*, 2000; Wang *et al.*, 2008). It was also shown that non-native interactions (e.g., non-native salt bridges) can be formed in the denatured state (Cho, JH *et al.*, 2008). The apparently contradictory observation (McCarney *et al.*, 2005) of the random-coil scaling of proteins with formed local structures can be resolved if one treats the denatured proteins as an ensemble of conformations with locally interacting residues that lack ordered long-range contacts (Chen *et al.*, 2008). Thus, while locally the chain may have (transient) residual structure, perhaps resembling the native state, global statistics reflects thorough self-averaging over various conformations, which is manifested as random-coil-like scaling.

Any deviation of the local or global statistical properties of unfolded proteins from that of random-coil polymers may have many implications for the biophysical properties of proteins. It was shown that by modulating non-native electrostatic interactions within the unfolded state ensemble, one can significantly change protein stability (Alston *et al.*, 2008; Anil *et al.*, 2006; Cho and Raleigh, 2006; Cho, JH *et al.*, 2008; Grimsley *et al.*, 1999; Pace *et al.*, 2000). Tuning the thermodynamic features of the unfolded state can affect folding kinetics as well (Trefethen *et al.*, 2005). Variations in crowding conditions and the degree of confinement level are reported to have a direct effect on the free energy of the unfolded state (Baumketner *et al.*, 2003; Cheung *et al.*, 2005; Cheung and Thirumalai, 2007; Takagi *et al.*, 2003; Xu *et al.*, 2005). In a recent study, it was shown that attaching oligosaccharides to proteins can introduce thermodynamic and kinetic stabilization because the unfolded state becomes less structured and more extended (i.e., higher free energy due to higher enthalpy) as the degree of glycosylation increases (Shental-Bechor and Levy, 2008). These examples illustrate why understanding the properties of the unfolded state may shed light on various biophysical characteristics of proteins.

Repeat proteins, which are widespread in nature and have a broad range of functions (Blatch and Lassle, 1999; Grove *et al.*, 2008; Main *et al.*, 2003), are useful to study the properties of the unfolded state ensemble of proteins. Repeat proteins consist of tandem arrays of structurally similar elements. Commonly they possess simple topologies dominated by short-range interactions; these are known to form nonglobular, elongated structures. Due to their simplicity and periodicity, repeat proteins have been used in several studies to explore fundamental questions in protein folding such as the effect of the number of repeats on protein stability, folding rate, and cooperativity (Barrick *et al.*, 2008; Cortajarena *et al.*, 2008b; Courtemanche and Barrick, 2008; Ferreiro *et al.*, 2005; Ferreiro and Komives, 2007; Ferreiro *et al.*, 2008; Ferreiro and Wolynes, 2008; Kajander *et al.*, 2005; Kloss *et al.*, 2008; Mello and Barrick, 2004; Tripp and Barrick, 2008). Similarly, the modular nature of repeat proteins is advantageous to study the unfolded state since one can obtain the length dependence of the unfolded state while keeping the content of the secondary structure unchanged. In this study, we use the tetratricopeptide repeat (TPR) class of proteins to investigate the nature of the unfolded ensemble of proteins and the scaling law relationship between their size and the number of repeats with respect to various possible assumptions on the nature of interactions within the denatured state ensemble.

The basic repeat unit of the TPR protein family is an α helix-linker-α helix motif that spans 34 amino acids. As a result of the short linker length, the helices within each repeat are oriented in an antiparallel manner relative to each other, whereas the packing between adjacent repeating units is of a parallel, stacklike form. Main *et al.* designed a TPR domain based on the consensus sequence (called CTPR herein) of that class (Main *et al.*, 2003). The designed module formed the canonical TPR fold and the folding of proteins made of as many as 10 concatenated CTPR domains was shown to concur with an Ising model (Kajander *et al.*, 2005). In addition, CTPR proteins were shown to retain the established ligand binding functionality of the family, as was demonstrated by binding to agonist peptides (Cortajarena and Regan, 2006).

In a recent paper coauthored by one of us (Cortajarena *et al.*, 2008a), it was demonstrated that CTPR proteins, under denaturating concentrations of GuHCl and urea, populated a structured, yet non-native ensemble of conformations. The radius of the denatured conformers scaled with N with a power law typical to compact globular polymers ($\nu \sim 0.37$) rather than with the scaling expected for polypeptides under such conditions (i.e., $\nu \sim 0.6$). As evident

Table I. Simulated models and their potential energy terms.

Model	Attribute	Sizes[a]	Potential energy terms					
			Bonds	Bond angles	Dihedral angles	Contact	Excluded volume	Contact source
1	Self-avoiding walk	2,3,4,6,8,10	+	—	—	—	+	—
2	Helical	2,3,4,6,8,10	+	+	+	i, $i+4$	+	CTPR
3	Native topology	2,3,4,6,8,10	+	+	+	+	+	CTPR
4	Mirror	3,4,6,8,10	+	+	+	Mirrored[b]	+	CTPR
5	PP_{II} rigid segments[c]	3,4,6,8,10	+	PP_{II}	PP_{II}	—	+	—
6	Mirrored PP_{II}[c]	3,4,6,8,10	+	PP_{II}	PP_{II}	Mirrored PP_{II}[d]	+	Collagen

[a]Sizes are expressed in terms of number of repeats (n). The protein length (y) confirms to the expression: $y=(n\times 34)+15$, which includes the additional solvating helix length.
[b]The mirroring of contacts was accomplished by incrementing the indexes of the contact pairs and allowing each interrepeat contact (i,j) to occur between all combinations of two repeats in the protein.
[c]PP_{II} bonds, bond angles, dihedral angles, and contacts were taken from the first nine residues of each of the three helices of collagen (pdb_id: 1k6a).
[d]PP_{II} mirrored contact were created by adding a single increment to a set of native contacts of type [i:j, i:k, j:k] to get contact sets of type [i:k, j:k, k:$k+1$], etc.

from CD spectra, the non-native structure formed in the denatured CTPR chains consisted of poly-proline II (PP_{II}) segments, present in about 50% of the chain length. The existence of the postulated PP_{II} segments is supported by a bioinformatic sequence analysis and experimental investigation of the region connecting the repeating CTPR domains that corresponds to the PP_{II} segment (Cortajarena *et al.*, 2008a). The authors used a three-dimensional chain-growth lattice model to rationalize the unexpected experimental observations of the high compactness of the denatured proteins. It was demonstrated by Cortajarena *et al.* (2008a) that the introduction of a rigid segment in a fixed position within each repeat could already lead to a reduced scaling exponent (close to 0.5) when the rigid segment length was ~15 amino acids (Cortajarena *et al.*, 2008a). The addition of an attractive term between the rigid segments in the lattice model resulted in a scaling law as low as 0.35. Thus, it was implied that the PP_{II} segments need to have an additional higher order mode of arrangement than just the linear-sequential one in order to reproduce the experimental results.

In this article, we wish to broaden the theoretical discussion regarding the phenomenon described by Cortajarena *et al.* (2008a) and to provide molecular insight on the possibilities to achieve high compactness for denatured repeat proteins in general and particularly for CTPR proteins. Using potentials of increasing complexity, we test various scenarios for denatured CTPRs. We particularly focus on possibilities for non-native interactions between segments in different repeat units. Intuitively, the self-similarity of the structure of repeat proteins can be a cause of long-range interactions. Indeed, we demonstrate here that denatured-state compactness can be partially achieved by taking into account the identical interface natively found between repeating units. As a further model, building on the experimental finding that non-native PP_{II} structure appears in denatured CTPRs, we propose that the tertiary interactions formed between PP_{II} segments, similar to the ones found in the native collagen fold, may lead to a compact unfolded state (the formation of noninteracting PP_{II} segments does not explain the compactness of the denatured state). Our simulations suggest that the presence of multiple compacting contributions derived from either the inert structural repetitiveness of the CTPR domains or from the propensity to form tertiary non-native PP_{II} based folds, can reproduce to varying extents the reported experimental observations.

METHODS

To study the unfolded ensemble of the CTPR, we designed several models to address the origin of the high compactness found in the experiment conducted by Cortajarena *et al.* (2008a). Some of the models are universal and can represent the unfolded state of any protein; other models, however, are more unique to represent the unfolded repeat proteins. For the construction of the model sets we used a potential similar to the one thoroughly described by Clementi *et al.* (2000b). To construct model sets with properties different from the nativelike model, we varied the potential energy by alternately enabling or disabling different energy terms and by changing the parameter values of the active terms. Generally, all models can be visualized as "beads on strings" where the beads represent the center of mass of the C_α atoms and the strings are the virtual bonds between the atoms.

The first model (model #1; see Table I) makes use of the complete native topology-based model (Gō-like potential). The potential in this model rewards conformations that resemble the native fold. The potential of a particular conformation [$V(\Gamma,\Gamma_0)$, where Γ denotes a particular conformation and Γ_0 denotes the native conformation] along the molecular dynamics simulation trajectory consists of the following terms:

$$V(\Gamma,\Gamma_0) = \sum_{\text{bonds}} K_b(b_i - b_{0i})^2 + \sum_{\text{angles}} K_\theta(\theta_i - \theta_{0i})^2 + \sum_{\text{dihedrals}} K_\phi^{(n)}[1 - \cos(n \times (\phi_i - \phi_{i0}))] + \sum_{i<j-3}^{\text{native contacts}} \epsilon_n \left[5\left(\frac{R_{ij}^0}{R_{ij}}\right)^{12} - 6\left(\frac{R_{ij}^0}{R_{ij}}\right)^{10} \right] + \sum_{i<j-3}^{\text{non-native contacts}} \epsilon_{nn}\left(\frac{C_{ij}^0}{R_{ij}}\right)^{12},$$

where b_i is the C_α-C_α bond distance between residue i and $i+1$; θ_i represents the angle between subsequent three residues; Φ_i represents the dihedral angle formed by subsequent four residues. The dihedral potential has either a period (n) of 1 or 3, consequently allowing both the *anti* and *gauche* arrangements; R_{ij} is the distance between the Cα atoms of residues i and j, and R_{ij}^0 is the distance between residues i and j that are in contact in the native state; C_{ij} defines the repulsion distance between residues i and j (equals to 4 Å, i.e., the radius of each Cα bead is set to 2 Å). ε_n and ε_{nn} denote the energetic contributions of the native and non-native terms, respectively, and in this work are set to unity. The values of the parameters that are exclusively derived from the native structure are denoted by a "0" subscript or superscript. The details of the molecular dynamics protocol can be found elsewhere (Levy *et al.*, 2005).

Table I provides a complete list of the different terms and values constituting all model sets used in this work. Models #2, #3, and #4 all use the CTPR native fold for derivating their potential. Models #5 and #6 impose a PP_{II} structure on particular segments within each repeat, based on the PP_{II} conformation of the chains from an engineered collagen structure (pdb code: 1k6a). As stressed above, these model sets also vary in the tertiary interaction terms used in each. Figure 1 shows the interactions that are allowed in models #2, #3, #4, and #6; in models #1 and #5 no tertiary interactions are included. The residues that are allowed to interact are shown in the contact map and as green lines on the repeat structures. The contact maps and structures in Fig. 1 are shown for a CTPR protein with 8 repeating units.

Since crystal structures of CTPRs indicate great structural similarity between proteins with varying repeat numbers (Main *et al.*, 2003), we designed CTPR proteins using the coordinate files of CTPR8 (pdb code: 2avp) and CTPR20 (Kajander *et al.*, 2007) as templates for the creation of the extended-length models where their structured was not resolved experimentally. The CTPR protein composed of 2, 3, 4, 6, 8, and 10 repeats were simulated using the native topology-based model to explore their thermodynamics. In particular, the folding temperature, T_F (characterized by the peak of the heat capacity curve where $C_V = k_B T^2[\langle E\rangle^2 - \langle E^2\rangle]$), of each system was calculated using the WHAM

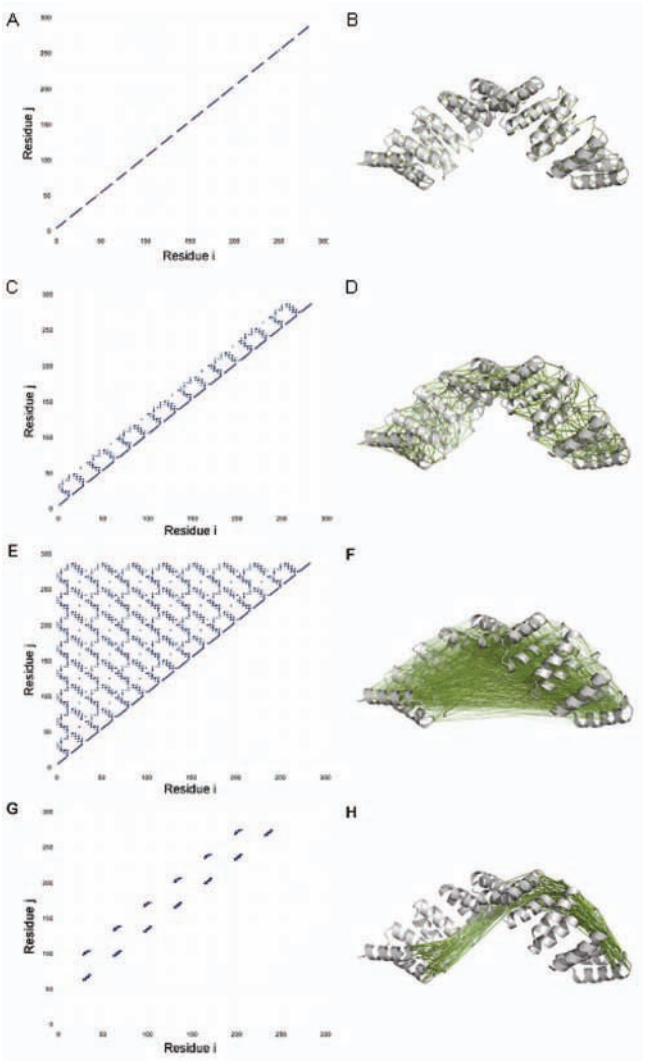

Figure 1. Contact maps and their representation on the structure of CTPR8. Contact maps of *i* and *j* residue pairs that were used for the construction of the helical model #2 (A), native model #3 (C), mirrored model #4 (E), and of the collagen model #6 (G). The corresponding *i* and *j* pairs at (B), (D), (F), (H) are connected with green line on the native structure of CTPR8.

(weighted histogram method) (Kumar *et al.*, 1992) analysis. Having the T_F and the thermal melting curve of each CTPR protein, the temperature at which the protein is mostly unfolded was more easily defined. The folding simulations were performed at various temperatures and include a few folding/unfolding transitions. To explore the denatured state, additional simulations were collected at temperatures characterized by the folding trajectories. Detailed description of the simulation protocol can be found in previous publications (Clementi *et al.*, 2000b; Levy *et al.*, 2004).

To obtain the hydrodynamic radius, R_h, from the simulations, we used an approach originally due to Kirkwood, who showed that the diffusion coefficient (D) can be calculated from an ensemble of configurations as

$D \equiv 1/n^2 \Sigma_{i \neq j}^{n} (\langle 1/R_{ij} \rangle)$, where n refers to the number of residues and $\langle 1/R_{ij} \rangle$ denotes the inverse average distance between residues i and j over the entire ensemble. Finally, R_h was calculated as the inverse of the diffusion coefficient, based on the Stokes-Einstein relation. The error of the scaling exponent υ is between 5 and 10%.

RESULTS AND DISCUSSION

To study the molecular characteristics of the unfolded state of repeat proteins, and in particular of CTPR proteins, which were reported recently to have a structured and compact unfolded state, we have designed a series of six models with increasing complexity that represent various scenarios of the unfolded state ensemble. In the first model the unfolded state is modeled as a self-avoiding random walk and in the second model the helical interactions that exist in the native proteins are allowed to be formed in the unfolded state. The third model assumes that any subset of the native interactions can be transiently formed in the unfolded state. The fourth model assumes that in repeat proteins non-native interactions can be formed due to the high similarity in the sequence and in the structure. Finally, the last two models rely on the experimental finding that non-native PP_{II} structures are populated in the denatured state. The fifth model for the unfolded state of CTPR therefore assumes formation of PP_{II} segments in each of the repeat units comprising the CTPR proteins, and the sixth model assumes long-range interactions between these PP_{II} segments. Following is a detailed description of these six models and their capability to reproduce the experimental results reported by Cortajarena *et al.* (2008a).

Set 1: Self-avoiding walk (SAW) models

We constructed SAW models for CTPRs of various lengths that serve as a control to examine if our simulations reproduce known scaling laws of SAW models. Details of the potential used in this model are presented in Table I, model #1. Briefly, the potential consists of two terms only: the first is Hookean potential for bond distance and the second term defines an excluded volume potential, keeping all non-neighboring ($|i\text{-}j| > 4$) beads with a minimal distance of 4 Å. We simulated these models at relative temperatures of 1.17 and 1.23 to allow comparison with model #3. The use of increased temperature is redundant here due to the lack of structure of this protein model. In both temperatures we calculated a scaling law of 0.58 with almost perfect correlation between the data and the linear fit. Indeed, this scaling is in good agreement with the known scaling law of a SAW. These results provide an additional support for the validity of use of molecular dynamics simulations in the investigation of unfolded state properties.

Set 2: Helical models

Native interactions, and in particular local native interactions, are prone to be formed in the unfolded state since the entropy cost of their formation is relatively low. In the helical model (model #2), we explicitly created a combination of a SAW backbone with interwoven helical contacts, based on the CTPR native-helical contacts [see Table I and Figs. 1(A) and 1(B)]. By that we created a model that only includes contributions from helical contacts that are thought to support the formation of the folded secondary structure. We simulated this set at the two high simulation temperatures used in both models #1 and #3 (relative temperatures of 1.17 and 1.23). It should be noted that, since the overall stability of the helical model is remarkably reduced relative to the native topology-based model (model #3), the ambient simulation temperature has a more significant impact on the model behavior than the impact most probably stressed onto model #3. Therefore, the probabilities of forming the helical contacts in the helical set are reduced relative to the probabilities of forming the identical contacts in the more stable model.

We nevertheless found out that at the high temperatures native helical contacts are still formed. The average Q at the simulation conditions is 13.7% at $T=1.17$ and ~13.3% at $T=1.23$. We continued our investigations by calculating the R_h of this set. We found that the scaling law for R_h of this set is 0.58 and 0.59, at relative temperatures of 1.17 and 1.23, respectively. Thus, it appears that the transient formation of helical contacts alone does not provide enough support for the formation of even a marginally compact fold. It is also evident that under the simulated conditions, the presence of transient contacts gives rise to a scaling law that is similar to those expected of SAW models.

Set 3: Native topology-based models

In native-topology-based models, the folded state of a molecule is considered to be the conformation with the smallest free energy out of all available conformations (Clementi, 2008) under folding conditions. Biomolecular models complying to such Gō-like potentials inherently have a perfectly funneled energy landscape. The potential therefore is biased to form native interactions as no competing basins are incorporated in the model. Modification of the different potential energy terms in a Gō-model simulation can lead to the formation of quite versatile model behaviors, both kinetically and thermodynamically. The native topology based models were shown in numerous studies to successfully reproduce folding kinetics and mechanisms (Chavez *et al.*, 2004; Cheung *et al.*, 2003; Cho, SS *et al.*, 2006; 2008; Clementi *et al.*, 2003; 2000a; 2001; 2000b; Finke and Onuchic, 2005; Gu *et al.*, 2007; Karanicolas and Brooks, 2003; Levy *et al.*, 2004; Levy and Onuchic, 2006; Matysiak and Clementi, 2008; Simler *et al.*, 2006)

Here, we used the native-topology-based model in an effort to rationalize the Cortajarena *et al.* recent results (Cortajarena *et al.*, 2008a) and to come up with a model that can loyally recapture some of the unique experimental observations brought in that report. The potential energy terms

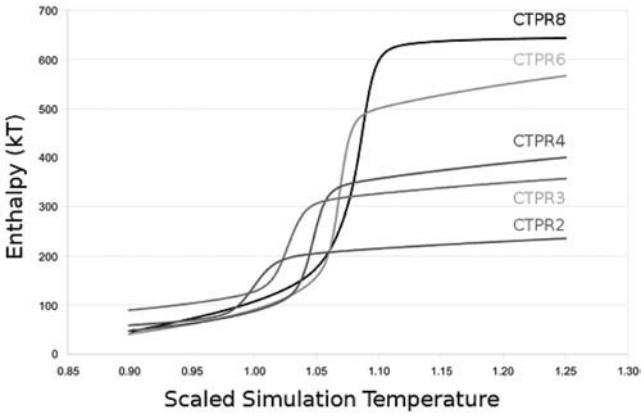

Figure 2. Denaturation of native model set. Temperatures were scaled relative to the T_f of CTPR 2. At temperatures above 1.17 the rate of transition reaches a plateau, indicating the presence of significant population of unfolded conformations.

that were actually used in the native-topology-based (Gō) set of simulations are described in detail in Table I, model #3 [see also Figs. 1(C) and 1(D)]. We initially used this model to examine whether the simple Gō-like potential can adequately describe the thermodynamic stability of CTPR proteins, but specifically we were interested in determining the temperatures at which the overall ensemble of conformations tends to be mostly unfolded. To investigate this, each of the CTPR proteins was simulated at varying temperatures which allowed monitoring of both the folding and unfolding transitions. Using this information, we were able to determine the relative stability of the different models constituting this set.

The relative stability of the CTPR proteins of varying lengths, expressed through their folding temperature (T_f, the temperature at which the folded and unfolded state are equally stable) studied using the native topology-based models is in agreement with the experimental data, the stability increases with the increase in the number of repeats (Ferreiro *et al.*, 2005; Ferreiro and Komives, 2007; Ferreiro *et al.*, 2008; Ferreiro and Wolynes, 2008; Hagai and Levy). The folding kinetics of repeat proteins have also been recently shown to be captured remarkably well by such models (Ferreiro *et al.*, 2005; Ferreiro and Komives, 2007; Ferreiro *et al.*, 2008; Ferreiro and Wolynes, 2008; Hagai and Levy, in press).

To computationally study the unfolded ensemble of proteins one has to rely on their denaturation curves to define the condition at which the proteins are unfolded. The unfolding denaturation plots of the different CTPR proteins (Fig. 2) illustrate that the unfolded state is significantly populated at temperatures above 1.17 (relative simulation temperatures, scaled to the T_F of CTPR2). Accordingly, to investigate the geometrical properties of the unfolded species in a statistically significant manner, we further simulated the CTPR series using the Gō models with increased sampling at these high temperatures. We found that the scaling law of the unfolded conformations of the CTPR models is 0.56, very close to the value expected from polymer theory. In the light of the SAW model set scaling (ν=0.58), one may also conclude that the scaling of the native-topology-based model, at similar simulation temperatures, may represent the scaling of a model set which does not reach "complete" unfolding. This provides an additional example for the difficulty in quantitatively defining "unfoldedness" and raises questions such as what is the extent of nativeness present, even transiently, if any, in chains considered as unfolded and which of the native characteristics can still be present in an unfolded chain while still providing Flory scaling laws.

Along a typical simulation trajectory, at temperatures which mostly promote unfolding, some residual presence of mostly helical contacts (between i, $i+4$ pairs) was detected. The average Q (the total number of native contacts in a given conformation) throughout a complete data set of trajectories, obtained at T=1.17, is 17.9%, and almost similar, 17%, at T=1.23. In the light of model #2, it is possible that the reduced scaling exponent we obtained for model #3 (ν=0.56) at the unfolding temperatures does not seem to originate mostly from the helical contacts which are automatically included in the native-topology-based model. If that was the case we would have expected that a model set based on helical contacts in isolation, as in model #2, would yield, at the very least, the slightest increase in the overall geometrical compactness—as measured by R_h and expressed by the scaling law.

Our results most likely demonstrate that, even in the presence of residual structure, under conditions which might not induce complete unfolding, it is possible to obtain a scaling law quite similar to the ones suggested for the self-avoiding walk models. The fact that the residual structure is composed of mostly helical contacts is in good agreement with a previous report that showed that unfolded scaling laws can be obtained even for proteins with well-formed secondary structure elements (Fitzkee and Rose, 2004). Quite expectedly, we did not witness any unusual compactness in the unfolded state of this set and therefore it could not provide ample support for the experimental results. The following models were constructed in an effort to fulfill that exact same purpose.

Set 4: Mirrored" models

We showed that modeling the unfolded state using the native interactions (models #2 and #3) under conditions that induce unfolding resulted in transient formation of native contacts. Yet, the characteristics of the unfolded state ensemble obtained from models #2 and #3 demonstrate that the unexpected scaling law reported by Cortajarena *et al.* for the CTPR proteins (Cortajarena *et al.*, 2008a) could not be accounted for by the suggested presence of helical structure or other native interactions dominating the denaturated chain. We speculate that a model that could recapture some of this

unusual unfolded-state behavior would most probably have attributes which are not readily deducible from the folded conformation of the CTPR molecules. Accordingly, nonnative interactions are expected to be formed in the denatured state.

One of the interesting features of CTPR proteins is the similarity of interfaces residing between and within the repeating units. For example, if we examine the residues in the first repeat of, say, CTPR3, which are involved in intrarepeat contacts, we would see that these are almost identical in their type and order to the residues involved in the similar intrarepeat contacts within the second repeat of that protein. This example also holds true for the interrepeat contacts. We therefore devised a model set which takes into account the possibility that upon refolding, the CTPR chain can form inter- and intrarepeat contacts between nonadjacent repeating units. The claim that an identical interface can potentially be formed between nonadjacent repeating units is based on our suggestion that both the adjacent and nonadjacent interfaces are literally identical, and thus, can alternately be formed. Model #4, which will be discussed in this section, was created to account for such interface formation possibilities. The model is designated as the "mirrored" model because the contacts appearing in the native-topology-based model of the CTPRs have been duplicated and reassigned. The origin of the mirrored inter- and intrarepeat contacts is from the contacts of adjacent repeating units and the reassigned identity of the pertaining residues is set to values of residues that belong to repeats that are nonadjacent. Thus, the interaction between residues which form a contact in the native state, say in repeats i and $i+1$, is now mirrored so that it can exist between repeat i and all other repeats. The contact map of model #4 is depicted in Figs. 1(E) and 1(F) and the different terms used in this model are summarized in Table I.

We found that the unfolded state modeled by the mirror model under the simulation unfolding conditions does possess an additional degree of compactness, with a scaling exponent of 0.53 at $T=1.17$ and 0.52 at $T=1.23$ [Fig. 3(A)]. This exponent is rather reduced relative to the native model (model #3), yet the magnitude of this reduction fails to fulfill the much anticipated decrease in the scaling law. We further explored the degree of foldedness of this model and found that the degree of contact formation at temperature of 1.17 is ~18.5%, and at 1.23 it is 17.6%. These values are quite similar to the ones calculated for model #3 ($Q=17.9\%$ and 17%), yet the scaling exponent is different (0.56 in model #3 and 0.52 in this model), especially when taking into account that the total number of contacts that define Q in model #4 is larger than in model #3. Again, these results provide an example for the terminological difficulty in defining the unfolded state of a chain, in which a reaction coordinate, such as Q, provides here imperfect predicting abilities for the geometrical properties of that state.

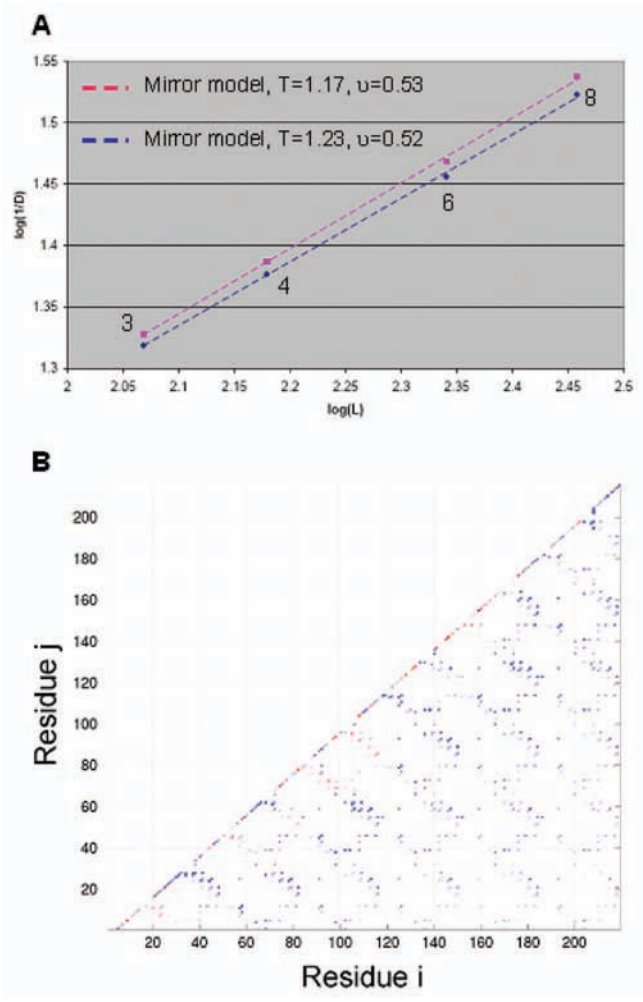

Figure 3. Modeling the unfolded state of CTPR using the mirror model. (A) The scaling law exponent of the mirror model (model #4) at temperatures of 1.17 and 1.23. (B) Difference of p^{ij} between native model and mirrored model sets at $T=1.27$. Blue contacts represent contacts present in the mirror model (model 4) of CTPR6 that are not formed in the native model (model 1). Red contacts are contacts that are more frequently present in the native model than in the mirror model.

In addition to the analysis of the average ratio of native contacts formation (Q), we wanted to quantitatively examine the type of contacts contributing to the reduced scaling law of model #4. To facilitate this, we calculated p^{ij}, the probability of contact formation between residues i and j. p^{ij} is in effect the Q of a given type of interaction (native, helical, or mirrored) for a pair of residues i and j. The p^{ij} values of most of the mirrored contacts was low relative to the CTPR-based native contacts ($P^{ij}<0.03$). To focus on the p^{ij} of the mirror model we calculated the difference-contact map, by subtracting from the p^{ij} values of the mirrored model the same values obtained from model #3. The difference-contact map reveals nicely [Fig. 3(B)] that the probability to form the CTPR-based contacts (native), and mainly the helical contacts, de-

creased in model #4 relative to model #3 (diagonal and near-diagonal contacts). It is seen that the majority of the mirrored contacts was indeed formed and that their formation resulted in destabilization of the helical contacts. It is also apparent that the prevalence of these contacts decreased as the distance between the i and j residues constituting the contacts increased, which can be rationalized by the substantial entropic cost that must be paid to form a contact between remote i, j residues.

It would be interesting to speculate on the possible role of mirroring interactions in biological modules that are composed of repeating structural interfaces. The same idea can even be applied to modules on the level of repeating multidomain proteins. Keeping a protein in this semifolded semiunfolded conformation, in situations where the mirrored contacts do form, can have major advantages such as protection from degradation, kinetic control over the folding reaction, or narrowing down the conformational search quest to a smaller region of the energy landscape by keeping the chain entropy relatively low prior to forming the native fold. On the other hand, such interface mirroring can take place between different molecules (similarly to domain swapping) and thus can promote aggregation.

Set 5: Poly-proline II models

Cortajarena *et al.* (2008a) suggested that the compact structure of denatured CTPRs is due to interaction between rigid PP_{II} segments formed upon chemical denaturation. In the next sections of this article, we wish to examine some aspects of the suggested PP_{II} model, using a more realistic, protein-related theoretical approach than the lattice simulations presented in Cortajarena *et al.* (2008a). We first simulated a model that is based on the assumption that under unfolding conditions segments with PP_{II} conformations are formed.

Model #5 was constructed in order to introduce the possibility that a CTPR chain, upon unfolding, can adopt a PP_{II} fold at local and defined regions, and test its effect on the scaling law of R_h. Particularly, we were interested to see whether the decrease in the scaling exponent observed in the lattice simulations of Cortajarena *et al.* (2008a) survives the transition to real space simulations.

We constructed nine-residues-long PP_{II} segments based on the dihedral and bond angles of actual an PP_{II} segment taken from the collagen X-ray structure (pdb code: 1k6a) (see Table I, model #5). We used the SAW skeleton as a template into which the native PP_{II} properties were introduced. To form a continuous stretch of nine residues to accommodate the basic PP_{II} unit, we assigned the nine residues at the junction of adjacent repeat sequences (the region that comprises the C-terminal of a repeat and the N-terminal of the subsequent repeat was probed experimentally to be critical for PPII formation) the PP_{II} native properties.

We found that the scaling exponent of the models at $T=0.23$ is 0.6 and the scaling at $T=0.67$ is 0.58. These results stand in conflict with the on-lattice model of Cortajarena *et al.* (2008a) that showed that the scaling exponent of a model consisting of nine rigid residues is approximately 0.53. We rationalize these disagreements on the basis of the methodological differences between the two approaches: the on-lattice model defines PP_{II} segments as fully rigid, while the current simulation uses an off-lattice model with much more detailed representation of chain residues. The PP_{II} model (model #5) is in fact not dissimilar to the helical model (model #2): both include structural elements with local interactions and no long-range interactions.

Set 6: Collagen fold model

In this section, we further elaborate on the PP_{II} model with interactions that was suggested by Cortajarena *et al.* (2008a) as the origin for the compactness of denatured CTPR. In their paper, they showed that introducing an attractive interaction term between the rigid segments (representing the PP_{II} regions) led to the formation of increasingly compact conformations with respect to the model length. The attractive potential was applied to an increasingly larger fraction of the repeat protein unit that corresponds to the defined rigid PP_{II} segments. While the simulations indicated significant compactness due to long-range attraction between rigid segments, supporting the experimental results, there are open questions regarding the molecular nature of these attractions and their relevance to protein structures.

To circumvent the need to arbitrarily determine inter-PP_{II} default distances and fractions of PP_{II} segment involved in the inter-PP_{II} network, we looked for a natural high-order arrangement of PP_{II} segments. One possibility is the superhelical collagen fold, which is assembled exclusively out of PP_{II} segments. We therefore created a potential that defines such a structure embedded in a SAW chain [Figs. 4(A)–4(C)]. Model #6 [see Table I and Figs. 1(G) and 1(H)] resembled in all of its details to model #5, except in the critical contacts-list term that was absent in model #5 and which is here crucial for the definition of the higher-order super structure. The contact list we constructed allowed the formation of several combination of three-helix PP_{II} segments, requiring the participation of three subsequent segments of PP_{II}. Three segments form a single triple-helix or at the extreme possibility between any arrangement of adjacent three-helix combinations that could be grouped together to form multiple, overlapping, three-helical elements. In addition, we set all the distances in the contacts list based on the ones from the native structure of an engineered collagen molecule (pdb code: 1k6a).

We found that the scaling exponent obtained with this model at $T=1.23$ was reduced to a value of 0.44, which is not far from the experimental results ($\nu=0.37$) which we set out

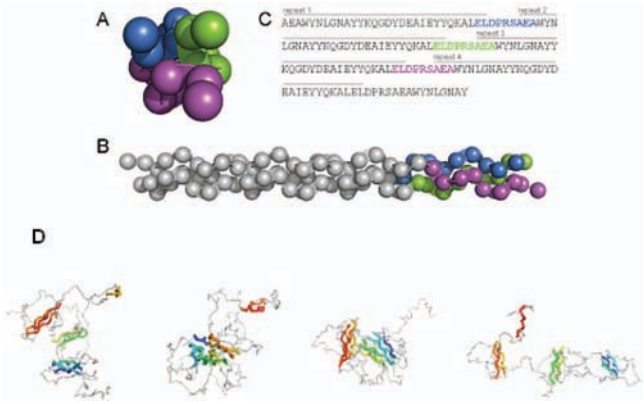

Figure 4. Modeling the unfolded state of CTPR using the collagen fold model. Collagen model construction (model #6). X-ray structure of collagen (pdb code: 1k6f) was used as the origin for the potential terms used in the simulation. The conformation of the first nine residues of the collagen sequence and their contacts were cloned to the regions in the CTPR flanking the repeat regions. Vertical (A) and horizontal (B) views of the crystal structure of synthetic collagen. (C) The individual PP_{II} segments constituting the collagen structure and their set of native contacts were cloned onto the CTPR structure in regions marked with the cognate PP_{II} colors on the CTPR sequence (here, onto CTPR4). (D) Snapshots of a simulated collagen model. Arbitrary conformers of a collagen model with ten adjacent PPII segments demonstrating the ability of the collagen models in general to form different combinations of collagen superhelices in an unbiased manner. Colored segments represent the PPII regions, with color code changing from "cold" (blue) to "hot" (red) according to the sequential order of the segments. SAW scaffold is colored in gray.

to explain. Possible reasons for the discrepancy between the simulated and experimental scaling law are that the collagen-type structures are formed from larger regions with PPII conformation and that chemical denaturation may support higher compactness than thermal denaturation which was used in the simulations to model the unfolded state. Alternatively, it is possible that the combination of models #4 and #6 would result with higher compactness and smaller scaling law. Our simulations at $T=1.67$ revealed, unsurprisingly for this system, that at such conditions, for the given potential of model #6, the chain mostly demonstrates self-avoiding behavior, with a scaling of 0.59. Examination of the conformations sampled throughout the simulations of this model revealed that indeed, different arrangements and combinations of the PP_{II} helices are in fact formed at the lower temperature. Figure 4(D) depicts some of the different arrangements of denatured CTPR10 obtained from model #6. We therefore demonstrate here that by applying an unbiased potential, in the sense that it lacks any arbitrarily decisions regarding the local concentration of the collagen segments that are to be formed, we were still able to obtain a scaling which is sensible in terms of the experimental data. Thus, we suggest this model set to be an additional valid solution to the experimental enigma Cortajarena *et al.* have raised.

CONCLUSIONS

In this study, we explored the unfolded state ensemble of CTPR repeat proteins with the goal of providing molecular insights on the high compactness of these proteins that was recently reported. We designed several coarse-grained models to address the significant deviation of CTPR proteins from the random-coil characteristics that are often found for the unfolded state of many proteins. Accordingly, our computational models aim at addressing the question of whether the high degree of structure in the unfolded state of the CTPR is unique to this protein or might be a common feature of repeat proteins due to the existence of repeating units with identical sequence.

The scaling law relationship between the average radius of the protein in the unfolded state and the number of repeats was tested for a series of CTPRs composed of 2–10 repeating units. Simulating the CTPR proteins using the native-topology-based model (at unfolding conditions) or using a model that included the formation of helical interactions only, resulted in a random-coil-like scaling. Similarly, a model that simulated the formation of PP_{II} conformation in the unfolded state yielded random-coil characteristics. These models indicate that native or local interactions are not enough to explain the anomalous features of the unfolded state of CTPR and, most likely, the high compactness originates from long-range non-native interactions.

Due to the repetitive nature of repeat proteins, we designed a model that includes interactions between any pair of repeating units and not only between neighboring repeats that form interface in the native structure. One can imagine that in the unfolded state where nonconsecutive repeats with consensus sequence are close in space, various repeats can interact with each other as their interfaces are optimized for that. Accordingly, we designed a mirror model that allows each interfacial contact to be formed between any pair of repeats. The unfolded state modeled using the mirrored model is more compact than that of a random-coil polymer, but not as compact as found for CTPR. Yet, the mirror model suggests that the unfolded state of any consensus repeat proteins may deviate from random-coil behavior due to the existence of non-native interactions stemming from the high symmetry of these proteins.

While our results suggest that the existence of PP_{II} segments in the unfolded CTPR is not sufficient to explain the experimental scaling exponent of $\nu=0.37$, we found, in agreement with previous modeling, that interactions between the PP_{II} segments can yield significantly compact conformations in the unfolded state. We modeled the interactions between the PP_{II} regions based on collagen that is composed of three chains with PP_{II} conformation. We found that two and three segments of PP_{II} can transiently associate and form collagenlike structures.

It should be noted that polymer theory predicts universal

scaling exponents, namely 1/3 for the globular compact state and 0.588 for the coiled state (de Gennes, 1979). The appearance in this work and in the experimental work of scaling exponents with values in between the two universal values should not be interpreted as contradicting polymer theory. Rather, these scaling exponents are most probably manifestations of finite-length effects. In other words, in the limit of very long CTPR proteins the universal scaling exponents should be regained (Cortajarena *et al.*, 2008a).

Our study indicates that repeat proteins and in particular those with consensus sequence can adopt relatively compact conformations in the unfolded state that are stabilized by long-range non-native interactions and their formation is facilitated by the symmetry in the structure and in the sequence.

ACKNOWLEDGMENTS

This work was supported in part by the Kimmelman Center for Macromolecular Assemblies and the Center for Complexity Science (Y.L.) Y.L. is the incumbent of the Lilian and George Lyttle Career Development Chair. The work of Gilad Haran is partially supported by NIH Grant R01GM080515 and by the Human Frontier Science Program.

REFERENCES

Alston, RW, Lasagna, M, Grimsley, GR, Scholtz, JM, Reinhart, GD, and Pace, CN (2008). "Tryptophan fluorescence reveals the presence of long-range interactions in the denatured state of ribonuclease Sa." *Biophys. J.* **94**(6), 2288–2296.

Anil, B, Li, Y, Cho, JH, and Raleigh, DP (2006). "The unfolded state of NTL9 is compact in the absence of denaturant." *Biochemistry* **45**(33), 10110–10116.

Baldwin, R (2008). "The search for folding intermediates and the mechanism of protein folding." *Annu. Rev. Biophys.* **37**, 1–21.

Barrick, D, Ferreiro, DU, and Komives, EA (2008). "Folding landscapes of ankyrin repeat proteins: experiments meet theory." *Curr. Opin. Struct. Biol.* **18**(1), 27–34.

Baumketner, A, Jewett, A, and Shea, JE (2003). "Effects of confinement in chaperonin assisted protein folding: rate enhancement by decreasing the roughness of the folding energy landscape." *J. Mol. Biol.* **332**(3), 701–713.

Blatch, GL, and Lassle, M (1999). "The tetratricopeptide repeat: a structural motif mediating protein-protein interactions." *BioEssays* **21**(11), 932–939.

Chavez, LL, Onuchic, JN, and Clementi, C (2004). "Quantifying the roughness on the Free Energy Landscape: entropic bottlenecks and protein folding rates." *J. Am. Chem. Soc.* **126**(27), 8426–8432.

Chen, Y, Ding, F, Nie, H, Serohijos, AW, Sharma, S, Wilcox, KC, Yin, S, and Dokholyan, NV (2008). "Protein folding: then and now." *Arch. Biochem. Biophys.* **469**(1), 4–19.

Cheung, MS, Finke, JM, Callahan, B, and Onuchic, JN (2003). "Exploring the interplay between topology and secondary structural formation in the protein folding problem." *J. Phys. Chem. B* **107**(40), 11193–11200.

Cheung, MS, Klimov, D, and Thirumalai, D (2005). "Molecular crowding enhances native state stability and refolding rates of globular proteins." *Proc. Natl. Acad. Sci. U.S.A.* **102**(13), 4753–4758.

Cheung, MS and Thirumalai, D (2007). "Effects of crowding and confinement on the structures of the transition state ensemble in proteins." *J. Phys. Chem. B* **111**(28), 8250–8257.

Cho, JH and Raleigh, DP (2006). "Electrostatic interactions in the denatured state and in the transition state for protein folding: effects of denatured state interactions on the analysis of transition state structure." *J. Mol. Biol.* **359**(5), 1437–1446.

Cho, JH, Sato, S, Horng, JC, Anil, B, and Raleigh, DP (2008). "Electrostatic interactions in the denatured state ensemble: their effect upon protein folding and protein stability." *Arch. Biochem. Biophys.* **469**(1), 20–28.

Cho, SS, Levy, Y, and Wolynes, PG (2006). "P versus Q: structural reaction coordinates capture protein folding on smooth landscapes." *Proc. Natl. Acad. Sci. U.S.A.* **103**(3), 586–591.

Cho, SS, Weinkam, P, and Wolynes, PG (2008). "Origins of barriers and barrierless folding in BBL." *Proc. Natl. Acad. Sci. U.S.A.* **105**(1), 118–123.

Clementi, C (2008). "Coarse-grained models of protein folding: toy models or predictive tools?" *Curr. Opin. Struct. Biol.* **18**(1), 10–15.

Clementi, C, Garcia, AE, and Onuchic, JN (2003). "Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism all-atom representation study of protein L." *J. Mol. Biol.* **326**(3), 879–890.

Clementi, C, Jennings, PA, and Onuchic, JN (2000a). "How native-state topology affects the folding of dihydrofolate reductase and interleukin-1 b." *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5871–5876.

Clementi, C, Jennings, PA, and Onuchic, JN (2001). "Prediction of folding mechanism for circular-permuted proteins." *J. Mol. Biol.* **311**(4), 879–890.

Clementi, C, Nymeyer, H, and Onuchi, JN (2000b). "Topological and energetical factors: What determines the structural details of the transition state ensemble and En-route intermediate for protein folding? An investigation of small globular proteins." *J. Mol. Biol.* **298**, 937–953.

Cortajarena, AL, and Regan, L (2006). "Ligand binding by TPR domains." *Protein Sci.* **15**(5), 1193–1198.

Cortajarena, AL, Lois, G, Sherman, E, O'Hern, CS, Regan, L, and Haran, G (2008a). "Nonrandom coil behavior as a consequence of extensive PPII structure in the denatured state." *J. Mol. Biol.* **382**(1), 203–212.

Cortajarena, AL, Mochrie, SG, and Regan, L (2008b). "Mapping the energy landscape of repeat proteins using NMR-detected hydrogen exchange." *J. Mol. Biol.* **379**(3), 617–626.

Courtemanche, N, and Barrick, D (2008). "Folding thermodynamics and kinetics of the leucine-rich repeat domain of the virulence factor Internalin B." *Protein Sci.* **17**(1), 43–53.

de Gennes, PG (1979). *Scaling Concepts in Polymer Physics*, Cornell University Press, Ithaca, NY.

Dill, K, Ozkan, S, Shell, M, and Weikl, T (2008). "The protein folding problem." *Annu. Rev. Biophys.* **37**, 289–316.

Ding, F, Jha, RK, and Dokholyan, NV (2005). "Scaling behavior and structure of denatured proteins." *Structure (London)* **13**(7), 1047–1054.

Ferreiro, DU, and Komives, EA, (2007). "The plastic landscape of repeat proteins." *Proc. Natl. Acad. Sci. U.S.A.* **104**(19), 7735–7736.

Ferreiro, DU, and Wolynes, PG (2008). "The capillarity picture and the kinetics of one-dimensional protein folding." *Proc. Natl. Acad. Sci. U.S.A.* **105**(29), 9853–9854.

Ferreiro, DU, Cho, SS, Komives, EA, and Wolynes, PG (2005). "The energy landscape of modular repeat proteins: topology determines folding mechanism in the ankyrin family." *J. Mol. Biol.* **354**(3), 679–692.

Ferreiro, DU, Walczak, AM, Komives, EA, and Wolynes, PG (2008). "The energy landscapes of repeat-containing proteins: topology, cooperativity, and the folding funnels of one-dimensional architectures." *PLOS Comput. Biol.* **4**(5), e1000070.

Finke, JM, and Onuchic, JN (2005). "Equilibrium and kinetic folding pathways of a TIM barrel with a funneled energy landscape." *Biophys. J.* **89**(1), 488–505.

Fitzkee, NC, and Rose, GD (2004). "Reassessing random-coil statistics in unfolded proteins." *Proc. Natl. Acad. Sci. U.S.A.* **101**(34), 12497–12502.

Flory, PJ (1953). *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY.

Grimsley, GR, Shaw, KL, Fee, LR, Alston, RW, Huyghues-Despointes, BM, Thurlkill, RL, Scholtz, JM, and Pace, CN (1999). "Increasing protein stability by altering long-range coulombic interactions." *Protein Sci.* **8**(9), 1843–1849.

Grove, TZ, Cortajarena, AL, and Regan, L (2008). "Ligand binding by repeat proteins: natural and designed." *Curr. Opin. Struct. Biol.*, **18**(4), 507–515.
Gu, Z, Rao, MK, Forsyth, WR, Finke, JM, and Matthews, CR (2007). "Structural analysis of kinetic folding intermediates for a TIM barrel protein, indole-3-glycerol phosphate synthase, by hydrogen exchange mass spectrometry and Go model simulation." *J. Mol. Biol.* **374**(2), 528–546.
Hagai, T, and Levy, Y (2008). "Folding of elongated proteins: Conventional or anomalous?" *J. Am. Chem. Soc.* **130**, 14253–14262.
Kajander, T, Cortajarena, AL, Main, ER, Mochrie, SG, and Regan, L (2005). "A new folding paradigm for repeat proteins." *J. Am. Chem. Soc.* **127**(29), 10188–10190.
Kajander, T, Cortajarena, AL, Mochrie, S, and Regan, L (2007). "Structure and stability of designed TPR protein superhelices: unusual crystal packing and implications for natural TPR proteins." *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **63**(Pt 7), 800–811.
Karanicolas, J, and Brooks, CL (2003). "Improved Go-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions." *J. Mol. Biol.* **334**(2), 309–325.
Kloss, E, Courtemanche, N and Barrick, D (2008). "Repeat-protein folding: new insights into origins of cooperativity, stability, and topology." *Arch. Biochem. Biophys.* **469**(1), 83–99.
Kohn, JE, *et al.* (2004). "Random-coil behavior and the dimensions of chemically unfolded proteins." *Proc. Natl. Acad. Sci. U.S.A.* **101**(34), 12491–12496.
Kumar, S, Rosenberg, JM, Bouzida, D, Swendsen, RH, and Kollman, PA (1992). "The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method." *J. Comput. Chem.* **13**(8), 1011–1021.
Levy, Y, Caflisch, A, Onuchic, JN, and Wolynes, PG (2004). "The folding and dimerization of HIV-1 protease: evidence for a stable monomer from simulations." *J. Mol. Biol.* **340**, 67–79.
Levy, Y, Cho, SS, Onuchic, JN, and Wolynes, PG (2005). "A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes." *J. Mol. Biol.* **346**, 1121–1145.
Levy, Y, and Onuchic, J (2006). "Mechanisms of protein assembly: lessons from minimalist models." *Acc. Chem. Res.* **39**, 284–290.
Main, ER, Jackson, SE, and Regan, L (2003). "The folding and design of repeat proteins: reaching a consensus." *Curr. Opin. Struct. Biol.* **13**, 482–489.
Matysiak, S and Clementi, C (2008). "Mapping folding energy landscapes with theory and experiment." *Arch. Biochem. Biophys.* **469**(1), 29–33.
McCarney, ER, Kohn, JE, and Plaxco, KW (2005). "Is there or isn't there? The case for (and against) residual structure in chemically denatured proteins." *Crit. Rev. Biochem. Mol. Biol.* **40**(4), 181–189.
Mello, CC and Barrick, D (2004). "An experimentally determined protein folding energy landscape." *Proc. Natl. Acad. Sci. U.S.A.* **101**(39), 14102–14107.
Onuchi, JN, Luthey-Schulten, Z, and Wolynes, PG (1997). "Theory of protein folding: the energy landscape perspective." *Annu. Rev. Phys. Chem.* **48**, 539–594.
Onuchi, JN and Wolynes, PG (2004). "Theory of protein folding." *Curr. Opin. Struct. Biol.* **14**, 70–75.
Pace, CN, Alston, RW, and Shaw, KL (2000). "Charge-charge interactions influence the denatured state ensemble and contribute to protein stability." *Protein Sci.* **9**(7), 1395–1398.
Pappu, RV, Srinivasan, R, and Rose, GD (2000). "The Flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding." *Proc. Natl. Acad. Sci. U.S.A.* **97**(23), 12565–12570.
Religa, TL, Markson, JS, Mayor, U, Freund, SM V, and Fersht, AR (2005). "Solution structure of a protein denatured state and folding intermediate." *Nature (London)* **437**(7061), 1053–1056.
Shakhnovich, E (2006). "Protein folding thermodynamics and dynamics: where physics, chemistry, and biology meet." *Chem. Rev. (Washington, D.C.)* **106**(5), 1559–1588.
Shental-Bechor, D and Levy, Y (2008). "Effect of glycosylation on protein folding: a close look at thermodynamic stabilization." *Proc. Natl. Acad. Sci. U.S.A.* **105**(24), 8256–8261.
Shortle, D and Ackerman, MS (2001). "Persistence of native-like topology in a denatured protein in 8 M urea." *Science* **293**(5529), 487–489.
Simler, R, Levy, Y, Onuchic, JN, and Matthews, RC (2006). "The folding energy landscape of the dimerization domain of *E. coli* Trp repressor: a joint experimental and theoretical investigation." *J. Mol. Biol.* **363**, 262–278.
Takagi, F, Koga, N, and Takada, S (2003). "How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: molecular simulations." *Proc. Natl. Acad. Sci. U.S.A.* **100**(20), 11367–11372.
Tanford, C (1968). "Protein denaturation." *Adv. Protein Chem.* **23**, 121–282.
Trefethen, JM, Pace, CN, Scholtz, JM, and Brems, DN (2005). "Charge-charge interactions in the denatured state influence the folding kinetics of ribonuclease Sa." *Protein Sci.* **14**(7), 1934–1938.
Tripp, KW and Barrick, D (2008). "Rerouting the folding pathway of the Notch ankyrin domain by reshaping the energy landscape." *J. Am. Chem. Soc.* **130**(17), 5681–5688.
Wang, S, Gu, J, Larson, SA, Whitten, ST, and Hilser, VJ (2008). "Denatured-state energy landscapes of a protein structural database reveal the energetic determinants of a framework model for folding." *J. Mol. Biol.* **381**(5), 1184–1201.
Wang, Z, Plaxco, KW, and Makarov, DE (2007). "Influence of local and residual structures on the scaling behavior and dimensions of unfolded proteins." *Biopolymers* **86**(4), 321–328.
Wilkins, DK, Grimshaw, SB, Receveur, V, Dobson, CM, Jones, JA, and Smith, LJ (1999). "Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques." *Biochemistry* **38**(50), 16424–16431.
Xu, WX, Wang, J, and Wang, W (2005). "Folding behavior of chaperonin-mediated substrate protein." *Proteins* **61**(4), 777–794.
Yi, Q, Scalley-Kim, ML, Alm, EJ, and Baker, D (2000). "NMR characterization of residual structure in the denatured state of protein L." *J. Mol. Biol.* **299**(5), 1341–1351.