Deconstructing the Native State: Energy Landscapes, Function, and Dynamics of Globular Proteins

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · JUNE 2009

Impact Factor: 3.3 · DOI: 10.1021/jp810659u · Source: PubMed

3 AUTHORS, INCLUDING:

Pavel I Zhuravlev
University of Maryland, College Park
16 PUBLICATIONS **255** CITATIONS
SEE PROFILE

Garegin A Papoian
University of Maryland, College Park
83 PUBLICATIONS **1,979** CITATIONS
SEE PROFILE

FEATURE ARTICLE

Deconstructing the Native State: Energy Landscapes, Function, and Dynamics of Globular Proteins

Pavel I. Zhuravlev, Christopher Kroboth Materese, and Garegin A. Papoian*

Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3290

Received: December 3, 2008; Revised Manuscript Received: March 2, 2009

Proteins are highly complex molecules with features exquisitely selected by nature to carry out essential biological functions. Physical chemistry and polymer physics provide us with the tools needed to make sense of this complexity. Upon translation, many proteins fold to a thermodynamically stable form known as the native state. The native state is not static, but consists of a hierarchy of conformations, that are continuously explored through dynamics. In this review we provide a brief introduction to some of the core concepts required in the discussion of the protein native dynamics using energy landscapes ideas. We first discuss recent works which have challenged the structure−function paradigm by demonstrating function in disordered proteins. Next we examine the hierarchical organization in the energy landscapes using atomistic molecular dynamics simulations and principal component analysis. In particular, the role of direct and water-mediated contacts in sculpting the landscape is elaborated. Another approach to studying the native state ensemble is based on choosing high-resolution order parameters for computing one- or two-dimensional free energy surfaces. We demonstrate that 2D free energy surfaces provide rich thermodynamic and kinetic information about the native state ensemble. Brownian dynamics simulations on such a surface indicate that protein conformational dynamics is weakly activated. Finally, we briefly discuss implicit and coarse-grained protein models and emphasize the solvent role in determining native state structure and dynamics.

1. Introduction

Molecular biology has shown time and again that protein function is intrinsically connected to structure, dynamics, and thermodynamics. Rationalizing protein dynamics and its connection to the protein energy landscape is not only an important biological problem, but also presents extremely interesting physicochemical challenges. To provide context and a broader view for our discussion, we will briefly overview the celebrated protein folding problem[1–12] and then shift to computational methods used in the description of the native state.

Proteins are polymer chains consisting of amino acid residues connected by covalent peptide bonds. Many of the principles of polymer physics are directly applicable to the study of proteins. One of the most interesting thermodynamic properties of polymer chains is the "coil−globule" transition,[13] which, as with all phase transitions, is directed by the competition of energy and entropy: if the energy of attractive intramolecular interactions overcomes the conformational entropy (as occurs in a poor solvent/low temperature), the chain precipitates into a compact globular state, with the volume fraction of polymer on the order of unity. In the opposite case (good solvent/high temperature), entropy dominates and a coil state is preferred. This state tends to possess extreme fluctuations and the volume fraction of the polymer tends to zero in the thermodynamic limit ($\rho_p \sim N^{-\alpha}$ where N is the degree of polymerization and $\alpha > 0$).

* To whom correspondence should be addressed. E-mail: gpapoian@unc.edu.

The coil−globule transition in proteins[14,15] is more interesting than in homopolymers and random heteropolymers and has special features, absent in both. This transition is accompanied or followed by protein folding, during which the protein chain adopts a specific tertiary structure. The tertiary structure differs from the secondary structure in that it dictates the features on the scale of the whole chain, whereas secondary structures like α-helices and β-strands are locally determined by the sequence. A typical protein sequence is exquisitely built in order to provide a unique native state that is both lowest in energy and biologically functional.[16,17] While this is not a single static state in reality,[18,19] for the moment we shall speak in approximate terms because the native ensemble is significantly smaller than that of a homopolymer globule. The sequence also ensures that the native state is kinetically accessible from the unfolded state. These features are essential to meet the demands of the biological role of proteins. Upon examining the phase space available to a protein polymer chain (depicted in Figure 1as a schematic representation of the funneled landscape), several distinct and fundamentally important regions may be identified as a function of solvent quality. The "coil" region is the most disordered, though still not as random as a typical homopolymer coil.[20] Since there is no clearly defined structure, function should not be possible according to the traditional "structure−function" paradigm. Despite this, some disordered proteins have been shown to be functional,[21,22] though discussion of the noncompact states observed in these proteins is beyond the scope of this review. In the coil state, flickering secondary structure elements

10.1021/jp810659u CCC: $40.75
Published on Web 05/19/2009

Pavel Zhuravlev was born in Dubna, Russia, in 1982. He received his Masters of Science degree in physics from Moscow State University in 2005. During his undergradute research in Division of Polymers and Crystals Physics under Igor Potemkin, he investigated the intramolecular phase separation of a polyamphiphilic macromolecule using random phase approximation theory. He is currently working on his Ph.D. at the University of North Carolina at Chapel Hill under Garegin Papoian. His research is focused on the dynamics of filopodia and native state organization.

Christopher K. Materese was born in Binghamton, New York, in 1982. He received his Bachelor of Science in chemistry from the Pennsylvania State University in 2005 where he performed undergraduate research in the field of neuroscience under Anne M. Andrews. He is currently working on his Ph.D. at the University of North Carolina under Garegin Paopian. His research focuses on analysis of protein dynamics and nucleosomal electrostatics.

reconfigure on the time scale of about 1 ns,[23] but to what extent the residual structure exists is still ambiguous.[24] In any case, the coil state is dominated by entropy. Even without considering the solvent, the conformational entropy of the chain is much more significant than the volume interactions, analogous to gases. In fact, Rouse−Zimm theory gives an appropriate description of the dynamics,[25] taking into account viscous friction between the solvent and monomer units and hydrodynamic interactions.

As the solvent quality decreases (possibly as a result of a temperature decrease) the protein undergoes a transition to *the molten globule region*[26−31] (Figure 1). In a classical polymer science, this is already a compact state, but in protein folding terms it is an ensemble of rather general intermediate states[32−36] which are still considered unfolded and highly dynamic. Some of them contain a substantial amount of secondary helical structure. Varying degrees of helicity were observed,[37,38] as well as some tertiary structure.[27,33,39−41] An extension of the analogy above would be to parallel the dynamics of a molten globule to that of liquids.[42] Again the structure−function paradigm maintains that there should be no functionality in this state, as there

Garegin A. Papoian was born in Yerevan, Armenia in 1973. He completed 4 years of undergraduate studies in the Higher College of the Russian Accadamy of Sciences, followed by graduate work with Professor Roald Hoffmann at Cornell University. He received his Ph.D. in 1999, working on quantum chemistry of intermetallic alloys and heterogenous catalysis. He then joined the research group of Professor Michael Klein at the University of Pennsylvania as a postdoctoral associate in 2000, to study metallo-enzymes with Car−Parrinello simulations. This was followed by a National Institute of Health sponsored postdoctoral fellowship with Professor Peter Wolynes at University of California, San Diego, from 2001−2004, with research focused on protein folding, binding, and hydration. He is currently an Assistant Professor of Chemistry at the University of North Carolina at Chapel Hill. He is interested in bringing tools of physical chemistry to shed light on complex biological processes at the molecular and cell levels. Students and postdocs in his group currently work on elaborating on the energy landscapes view of protein functional dynamics, modeling the way DNA packs in cells of higher organisms, and developing physico-chemical models of cell motility.

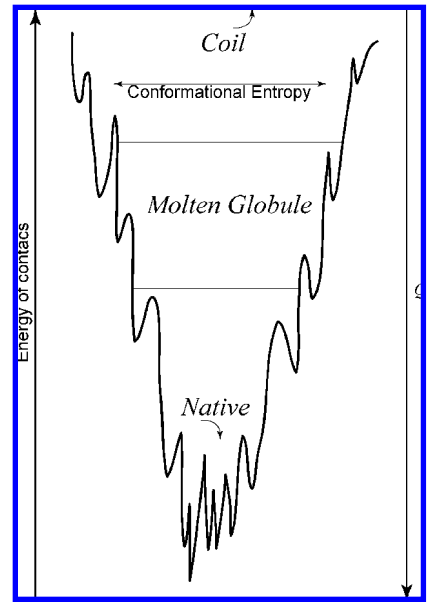

Figure 1. A funneled energy landscape of a typical protein depicting the major structural and dynamical regimes. The order parameter Q, characterizing similarity of the given conformation to the native structure, increases from 0 to 1 as the protein descends into the native state of the funnel.

is no well-defined structure. However, exceptions have been found recently, which we will discuss in this paper.

When the solvent quality decreases even further, the protein starts to feel the ruggedness of its own energy landscape. The protein spends more time in the minima of the landscape, rather than in saddle points, and the barriers between the minima become harder to overcome. Such dynamics is similar to that of a supercooled liquid.[43] Thermodynamically, it is still a molten globule, but in a real experiment, some conformational states

may remain unexplored despite being thermodynamically favorable, because of the finite run time: the mean first passage time for crossing a barrier of height E is proportional to $\exp(E/kT)$ according to Arrhenius relation.

At some temperature T_g, the barriers between the minima become insurmountable on any feasible experimental time scale. The dynamics of a protein whose temperature has reached T_g, mark the transition to *the region of glassy states*.[44,45] In this regime, a protein gets trapped in a single, random conformational basin and if these states are not functional, the protein will not be able to perform its required role. Therefore, for many naturally evolved proteins it is unsurprising that T_g is well below room temperature.[5,9,45,46]

Typically, proteins undergo folding from a molten globule to what is called *the native state* at folding temperature (T_f) which is well above T_g.[3,5,47–49] To continue our analogy, the native state can be compared to crystalline states,[42,50] as the conformational entropy is much lower and the effective energy of contacts plays the dominant role. However, here the parallel ends, since the native state possesses far more than just vibrational entropy.[4–6,9,18,51] As an aside for readers new to the topic, since the first protein models did not possess an explicit solvent, hydrophobic interactions, though substantially entropic in origin, are conventionally sorted into the energetics category. We follow the convention here, thus, when referring to the "energy" of contacts or "energy" landscape, we actually mean free energy (where fast solvent degrees of freedom are integrated out).

Though the focus of our work is on the energy landscape of the native state, it would be prudent to briefly describe how a protein arrives at this state. In 1969 Levinthal postulated a thought experiment which proved that proteins could not possibly find the thermodynamically stable functional state via *unbiased* random exploration of their phase spaces.[1,9] Consider a 100 residue protein, where each residue may adopt one of three alternative states, then the total number of possible conformations available to that protein may be roughly estimated as $3^{100} \approx 10^{47}$. If we restrict conformational transitions to occur no faster than once per picosecond (10^{-12} sec), it would take roughly 10^{28} years to complete an exhaustive search, significantly longer than the age of the universe. A protein possessing such an unbiased energy landscape could not reach a functional conformation within a biologically relevant time scale, therefore, there must be a defining feature of their energy landscape that circumvents this problem. This realization lead to the insight that the energy landscape must be biased in favor of the native state such that folding through an exhaustive search of the protein's phase space does not occur.[52] Unlike random heteropolymers, proteins possess a funneled energy landscape which drives folding to the native state[4–6,9,53–55](Figure 1). Evolution has carefully selected protein sequences which are able to quickly fold to their native state upon translation. This means in general, organization of a protein's energy landscape is dominated by the native state.[56] However, a free energy bias alone is not enough to guarantee consistent successful folding. If the landscape is funneled but very rugged, trapping still may prevent folding on reasonable time scales. This brings us to another very important, feature of the protein energy landscape. However, in order to discuss this we must first introduce the concept of frustration.

Frustration, simply put, is the inability to simultaneously achieve favorable interaction energies between all contacts. Frustration leads to an energy landscape with many local minima separated by barriers. Such a situation is also encountered in spin glasses, an elegant example given by Bryngelson et al.[6] Consider a system in which magnetic spins are arranged randomly. Interactions between spins may be either ferromagnetic or antiferromagnetic and occur at an equal frequency. Frustration arises out of a competition between these two mutually exclusive tendencies. The contacts between monomer units of a heteropolymer can also be energetically either favorable or unfavorable, and this analogy allows researchers to use methods and models developed for glasses to study proteins.

Proteins have evolved to minimize the frustration in their native states, in order to keep the funneled landscape as smooth as possible to minimize trapping in nonfunctional conformations.[52] Despite the kinetic need for a minimally frustrated landscape, some residual frustration remains at the bottom of the funnel.[57,58] The energy landscape of the native basin is degenerate with multiple ways to achieve low effective energy structures. Conceptualizing the native state as an ensemble of states rather than one single state was a significant advancement in protein science. This breakthrough in the description of the native state occurred over 3 decades ago when Austin et al. conducted low temperature flash photolysis measurements, providing experimental evidence for the existence of substates within the native landscape.[59] In this study, an anomalous myoglobin-ligand binding energy spectrum led the researchers to conclude that myoglobin must possess a multitude of conformational states with differing activation energies. Subsequent experiments by Ansari et al. suggested that the protein energy landscape is organized hierarchically with various states divided into substates which themselves may be subsequently divided further into substates *ad nauseam*.[4,18,60]

Understanding the dynamics of a protein in the native state is a separate very interesting problem with profound biological applications. Examples include fluorescence intermittency[61] and allosteric regulation.[62] The first molecular dynamics simulation of a protein suggested that protein molecules are more fluid than originally expected.[63] This fluidity provides a protein with the means to transition between conformational states through thermal and solvent fluctuations. There exist diverse viewpoints on the nature of dynamical transitions between the native substates. Some techniques, such as normal-mode analysis (NMA), treat proteins as solids possessing vibrations and phonons.[64–66] This approach works well in some cases and has provided significant insight into protein dynamics. However, it has the significant drawback that it is difficult to reconcile with the multitude of conformational minima and single minimum anharmonicities. This technique should be contrasted with principal component analysis (PCA), in which anharmonicities are retained.[67] Work has also been done on the related field of instantaneous normal-mode analysis to account for the existence of multiple minima in the energy landscape. This technique uses the differing time scales of thermal solvent motion and solute vibrations to assume the separability of the Hamiltonian into individual solvent and solute Hamiltonians in addition to a third Hamiltonian which describes their interaction. The basics of this technique have been reviewed by Schmitz et al.:[68] briefly, MD simulations are used to generate a collection of solvent cages which are then frozen for subsequent NMA, and anharmonicites can be introduced through quantum corrections. While this technique can be quite useful in some systems, and has even been applied to peptides,[69–71] translation of the work to proteins is complicated by the very large number of local minima in the energy landscape. Another viewpoint is to imagine activated hopping between different conformations

separated by energy barriers.[72,73] This can be paralleled to the dynamics of a supercooled liquid. Yet another possibility is that protein's motion is similar to flow of a normal liquid such that the system spends most of the time in saddle points rather than in minima of the energy landscape. There have been indications that activated, or even glassy dynamics, are more relevant to proteins, at least at low temperatures.[3,4,74]

Knowledge of the existence of a rugged native energy landscape, that is organized in hierarchical tiers, has inspired extensive research both experimentally and theoretically to further our understanding of its nature. Complete characterization of a protein's native landscape is often not feasible with modern computer technology but it is still possible to gain important insight with various techniques. Significant work has been done to characterize the energy landscape for peptides (mainly via models with implicit solvent), using disconnectivity graphs (DG)[73,75] which provide important insight into finer scale splittings of similar structural clusters and provide information about the transition barrier heights. A different approach to modeling the energy landscape of a protein emerges from graph theory and the analysis of complex systems such as those applied to World-Wide Web.[76] In this technique, molecular dynamics simulations are conducted and conformations are characterized by secondary structural elements. These conformations then become nodes on the network graph while transitions between different conformations become the links. A detailed, hierarchical network emerges which provides significant insight into the nature of the energy landscape. Another approach is to compute a low-dimensional map of the energy landscape. A typical result is a one- or two-dimensional free energy map of the system which can be used to study structural clustering and kinetics of conformational transitions.

The structure of the remaining sections of this article are outlined as follows. In section 2 we discuss a recent challenge to the protein structure−function paradigm in which activity was observed in a disordered protein. In section 3 we describe our recent work in elucidating a hierarchical landscape of a small globular protein derived from a combination of all atom molecular dynamics (MD) and PCA techniques. We demonstrate that the emergent hierarchy displayed arises from residual frustration and results in a differentiation between inter-residue contacts. In section 4 we discuss recent work in our laboratory related to the development of a technique for constructing high-resolution free energy surfaces (FES) for the native basin. We constructed a FES for a small protein, Trp-cage, and used it to study the kinetics of transition between two native-like structures, finding borderline behavior between activated and fluidlike dynamics. In section 5 we briefly discuss the application of FES techniques to the optimization of coarse grained force fields. In section 6 we emphasize the important role of water in sculpting the native ensemble.

2. Activity in the Disordered State

The classical view of enzymatic catalysis, pioneered by Haldane and Pauling, is based on the selective stabilization of reaction's transition state, which lowers the reaction barrier.[77,78] Design of transition state analogues has validated this approach for a large number of enzymes.[79–81] A well-defined protein structure is needed for this selective recognition, to achieve steric and electrostatic complementarity with the transition state. In this light, the recent discovery of catalytic activity by a protein in the molten globule state has been unexpected,[82] challenging the static picture of an enzyme. Although structural dynamics of the active site[83] and the protein as a whole[18,62,78,84,85] were well appreciated in the context of enzymatic catalysis, structural motions in the molten globule are on much larger scales, corresponding to orders of magnitude higher conformational entropy. Thus, the traditional structure−function paradigm needs to be extended, to account for large-scale conformational dynamics.[86] Energy landscape ideas, initially developed to describe protein folding,[3,5,6,9] provide a convenient starting point for such extension.[22]

Hilvert and co-workers mutated a homodimeric enzyme into a highly active monomer.[82] This mutant behaves like a molten globule, with an ensemble of rapidly interconverting compact structures, and yet provides almost the same catalytic power as the wild-type enzyme in the native state. This raises the question of how enzymatic activity is possible for a disordered protein. In this case it seems that an active three-dimensional structure, while present, is disguised among a multitude of other structures that the molten globule ensemble explores.[86] Upon binding of the transition state analogue (TSA), however, the enzyme becomes stabilized in its active conformation, adopting a more well-defined structure.[82] This is similar to the way binding and folding funnels are coupled for many disordered signaling and transcriptional proteins, that fold only upon finding their binding partner.[22,87–89] The folding landscape must still retain an overall smooth funneled shape, as it would otherwise result in the problems similar to the Levinthal paradox—the catalytically competent conformations need to be visited on the experimental time scale. However, while protein's energy landscape is funneled, with the active structure presumably residing at the bottom, the funnel is not deep enough to provide the native state with the overwhelming energetic stability typical of many other globular proteins.[22,86] In this case, the energy landscape being weakly funneled in the absence of the partner (Figure 2; middle panel), is deepened upon binding to stabilize the native conformation.[22] Apart from the binding induced folding, a transient excitation to a native-like functional conformation capable of catalysis may be another reason for observed enzymatic activity.[86] Even if the enzymatically active states are visited infrequently, the enormous acceleration of a chemical reaction by these states would still result in significant enzymatic activity by a molten globule.

Roca et al. explored both the mutant monomer and wild-type dimer computationally, through the construction of a 2D free energy surface, using the radius of gyration and contact order as coordinates.[90] Upon binding of the transition state analogue (TSA) the minimum region (a basin) in the free energy surface becomes more confined for both dimer and monomer, though this effect was more pronounced in the latter case. Thus, the weakly funneled landscape of the monomeric molten globule is "pulled down" by the TSA binding. By calculating empirical valence bond surfaces from X-ray and NMR structures of the wild-type dimer and the mutant monomer, respectively, bound to TSA, Warshel and co-workers computed the heights of the activation barriers, which are directly connected to the catalytic power, and these were shown to be nearly identical between forms.[90] Picking the structures from the minima on the calculated FES, they repeated the procedure and found similar barrier heights for the monomer, but for the dimer they were higher. Therefore, unlike the dimer, the monomer also has catalytic configurations in the minima of the FES.[90] The Roca et al. work demonstrates the effectiveness and usefulness of a free energy surface approach for studying functional dynamics of a disordered protein.

There may be reasons for a protein to evolve into an active molten globule: rapid turnover, faster binding kinetics, and

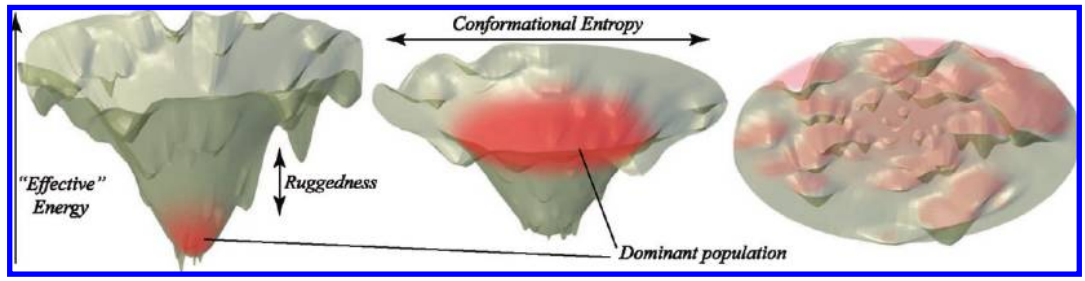

Figure 2. The interplay between protein's folding and functional landscapes. (Left) Energy landscapes of many globular proteins are thought to be funneled, such that the native state is both a thermodynamic global minimum and is also kinetically accessible. (Center) Energy landscapes of many disordered proteins are likely organized around a special state, characterized by a weakly funneled landscape. However, the driving force for folding is reduced, allowing the protein to remain disordered. Interactions with specific targets create additional favorable contacts, deepening the funnel and driving subsequent folding. Transient population of catalytically competent states, near the funnel's bottom, may allow for efficient catalysis. (Right) A random energy landscape is shown, where the search for a specific functionally competent conformation is extremely inefficient.

functional promiscuity.[89] Molten globule occupies larger effective volume, which leads to an increase in interaction cross-section with the surrounding molecules. This may be used to increase the association rate via a "fly-casting" mechanism,[91] when the binding and folding funnels are coupled.[22] Furthermore, molten globular proteins are quickly degraded by the proteasome, allowing finer temporal control over the catalytic process. This would be useful if only a short burst of catalytic activity is desired, for example, in signal transduction cascades.[86] From a practical viewpoint, it is extremely difficult to either predict or experimentally determine what specific functional role is played by a particular disordered protein, because the molten globule dynamics efficiently conceal which particular conformation is the active one.[86] These difficulties are clearly exemplified by a current lack of understanding of the function of α-synuclein, a disordered protein that is implicated in Parkinson's disease.[92]

Thus, recent works have posed a question of whether or not the structure−function paradigm is too limited and have suggested to expand it to an energy landscape−function paradigm. Although the functionally competent structures may be transient, the protein can perform catalysis as long as these states are visited frequently enough during conformational dynamics on protein's energy landscape. Stabilization of the functional state upon substrate or binding partner binding ("pulling down" of the funnel)[22,86] seems to be another ubiquitous strategy to confer function on disordered proteins.

3. Kinetic Hierarchy Elucidated by PCA

As temperature decreases, the *molten globule* folds to the *native state*. The native state, however, is still very dynamic and possesses measurable entropy due to residual frustration. This residual frustration splits the native state into an ensemble of substates, which are organized in hierarchical tiers both kinetically and energetically. Often, the investigation of the *molten globule* is focused more on the thermodynamics, with low-resolution order parameters such as radius of gyration. The *native state*, however, must be studied from a more precise structural perspective since specific structures are responsible for determining functional activity, such as enzymatic catalysis.

Much work has been accomplished toward the application of principal component analysis to protein systems. Early work by Amadei et al. laid the foundations[67] and subsequently, significant additional work followed toward characterizing molecular dynamics trajectories of peptides. García et al. used PCA to create free-energy surface maps of a small peptide as a function of temperature using the first two PCs.[93] Becker evaluated the effectiveness of PCA at capturing peptide dynamics and also used it to create energy surfaces.[94] In a later work, Levy and Becker used a PCA variant in conjunction with disconnectivity graph analysis to illustrate the effects of conformational constraints on the peptide energy landscape.[95] A paper by Altis et al. suggested that dihedral angles provide significant advantages over Cartesian coordinates in the construction of energy landscapes.[96] Recently, Hegger examined the complexity of peptide folding and compared the dimensionality of an energy landscape obtained by PCA to the dimensionality of the dynamics obtained through Lyapunov analysis.[97] Our work can be seen as an extension of these earlier efforts in that we used PCA to characterize the latent kinetic hierarchy of a globular protein (*eglin c*) from MD simulations and elucidate the way direct and water-mediated interactions sculpt this hierarchy. To achieve these goals, we conducted a long 336 ns MD simulation of the protein in explicit solvent and employed dihedral angle principal component analysis to deconvolute essential degrees of freedom and reduce the dimensionality of the system.

PCA is a powerful linear orthogonal technique used to aid in the comprehension of complex multidimensional systems, such as a protein, by reducing the phase space while retaining essential degrees of freedom[67] and accounting for anharmonicities. PCA operates through the diagonalization of a covariance matrix of the coordinate fluctuations of the system. The eigenvectors obtained by this diagonalization are the principal components which are sorted by decreasing eigenvalues. The eigenvalues represent the variance of the data along each eigenvector, meaning that the first principal component retains the greatest variance of the data followed by the second and so on. It has been shown that the majority of the degrees of freedom in a full multidimensional hyperspace of a protein's dynamics are uninteresting and contain no essential information.[67] Therefore, to simplify our problem we identified the essential degrees of freedom by projecting our MD trajectory into PC space and histogramming that data. Data projected into a PC which only characterizes fluctuations within a single state act like a harmonic oscillator and will appear as a single Gaussian distribution.[67] On the other hand, data projected onto one of the PCs belonging to the essential subspace would appear as a multipeaked distribution with different peaks representing the states accessible within that PC. Thus, the essential degrees of freedom were identified by isolating the non-Gaussian forming PCs and eliminating Gaussian forming PCs. Our results showed that over the course of the 336 ns simulation, the first four PCs revealed highly non-Gaussian and the next seven revealed

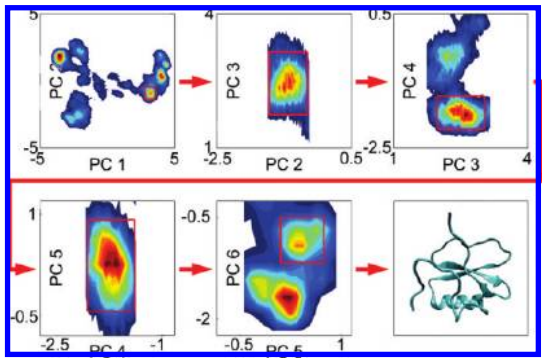

Figure 3. We isolated each basin in our 2-D PC landscapes and projected them into the next set of PC's. This had the effect of further separating the data into more basins which themselves were each isolated and projected in subsequent PCs. At the end, we arrived at a basin whose structure can be well characterized.

slightly non-Gaussian distributions of trajectory data, suggesting *eglin c* is sufficiently described in, at most, an eleven dimensional manifold or roughly 3% of the characterized degrees of freedom. This result is important because it reassures us that it is indeed possible to describe protein's native basin with a manageable number of collective degrees of freedom.

After establishing limits on the essential phase space, we proceeded with a detailed characterization of the energy landscape. Using a two-dimensional histogram of the trajectory projected in the first two principal components we observed the appearance of several densely populated regions. These regions were considered basins in the energy landscape as there is a spontaneous tendency for the system to adopt the structural conformations associated with those basins. While the totality of the protein dynamics was not restricted to reside solely within these basins, they did represent a very significant portion of the native landscape. With our knowledge of the energy landscape from the first two degrees of freedom, it was possible to extract further detail by examining each basin individually. Trajectory data residing within a single basin was then plotted as a two-dimensional histogram in the second and third PCs (thus, we zoomed into the basin using the next set of PCA dimensions). By doing this we further identified basins in the energy landscape which were eclipsed when viewed from the first two PCs as shown in Figure 3. This procedure may be repeated in an iterative fashion for all subsequent essential degrees of freedom. To the best of our knowledge this technique has not been used prior to this work. The result of this is the identification of a hierarchy of states (Figure 4) in the energy landscape whose structural significances may be determined by sampling the individual conformations residing within each basin.

Structural sampling from the observed basins provided valuable information about the protein's native dynamics. At a coarse visual level, we used the collection of conformations in each sampling to create an average structure, displaying the characteristic features of that basin. The most significant large scale structural rearrangements occurred between basins obtained from the histogram of the first two principal components. Differentiation observed from the splitting of subsequent PCs became increasingly subtle. For a more substantive description of the significance of the landscape hierarchy we examined the inter-residue contacts present in each basin. We found that basin splitting is accompanied by a change in intramolecular contacts formed within the protein as depicted in Figure 5. The observed landscape hierarchy was thus characterized by a divergence of intermolecular contacts between basins. Subordinate basins retained the defining features of their superiors while further differentiating among themselves, showing ultrametric features characteristic of spin glasses.[99,100] Interestingly, we extended our search for divergent contacts to include water-mediated interactions and found that these interactions also played an important role in basin definition. This observation is consistent with prior works on the way water-mediated interactions guide protein folding and stabilize native state architectures.[101–104]

The general techniques used in this work are useful for investigating other protein systems. Currently we are applying this technique to investigate the dynamics of a yeast nucleosome in order to determine if it undergoes any significant conformational changes. In summary, our results showed that competition between direct and water-mediated inter-residue contacts sculpts the landscape and dynamics of the native state and exemplifies the residual frustration of proteins' native ensembles.

4. Conformational Kinetics on Low-Dimensional Free Energy Surfaces

In Wu et al. we approached the investigation of protein native dynamics by studying the underlying energy landscape.[105] Even after integrating the solvent degrees of freedom, this (free) energy landscape is still a function of a multitude of conformational variables.[106] A drastic reduction of this remaining phase space is necessary to achieve a physically meaningful description of protein's conformational dynamics. We used two collective coordinates to project the high-dimensional landscape into a two-dimensional (2D) free energy surface (FES). Constructing 2D FESs to describe folding is a commonly used technique;[107–109] however, the native ensemble at the bottom of the folding funnel consists of conformations that are very similar to each other from both classical polymer and protein folding perspectives. Therefore, it was imperative that we carefully choose coordinates with sufficiently high structural resolution. A suitable coordinate commonly applied to folding problems is defined with respect to the native conformation. Specifically, it is the fraction of native contacts between atoms, or residues. This coordinate has been denoted Q and possesses a range between 1 (all the contacts are the same as in native state) and 0 (no native contacts).[110] Q has been shown to correlate with the strata of the folding funnel.[5] For a molten globule, a typical Q value is about 0.3.[5] In our work, we chose two slightly different conformations from the bottom of the funnel and defined two Qs with respect to each structure. The use of a 1D collective coordinate such as Q is highly degenerate in that a single Q value represents a large collection of states residing on a hypersphere. Thus the use of two collective coordinates significantly reduces this degeneracy by limiting the space to the ring which accounts for the overlap of the two hyperspheres. In other words, this choice of coordinates unfurls the usual 1D free energy curve to a 2D surface (Figure 6a), where the latter may show saddle points and other interesting features. In terms of studying kinetics of a transition between two reference structures, one important question is whether a dominant pathway connects these structures or many dissimilar pathways contribute with significant statistical weights.

We applied this technique to the native state of Trp-cage, a 20 residue peptide, using two most dissimilar conformation among structures reported by an NMR-study.[111,112] Since both of these conformations are native-like, with a Q between them being about 0.9, the resulting 2D free energy surface shows high resolution details of the native state ensemble. Though not done in this work, another possible choice of the reference

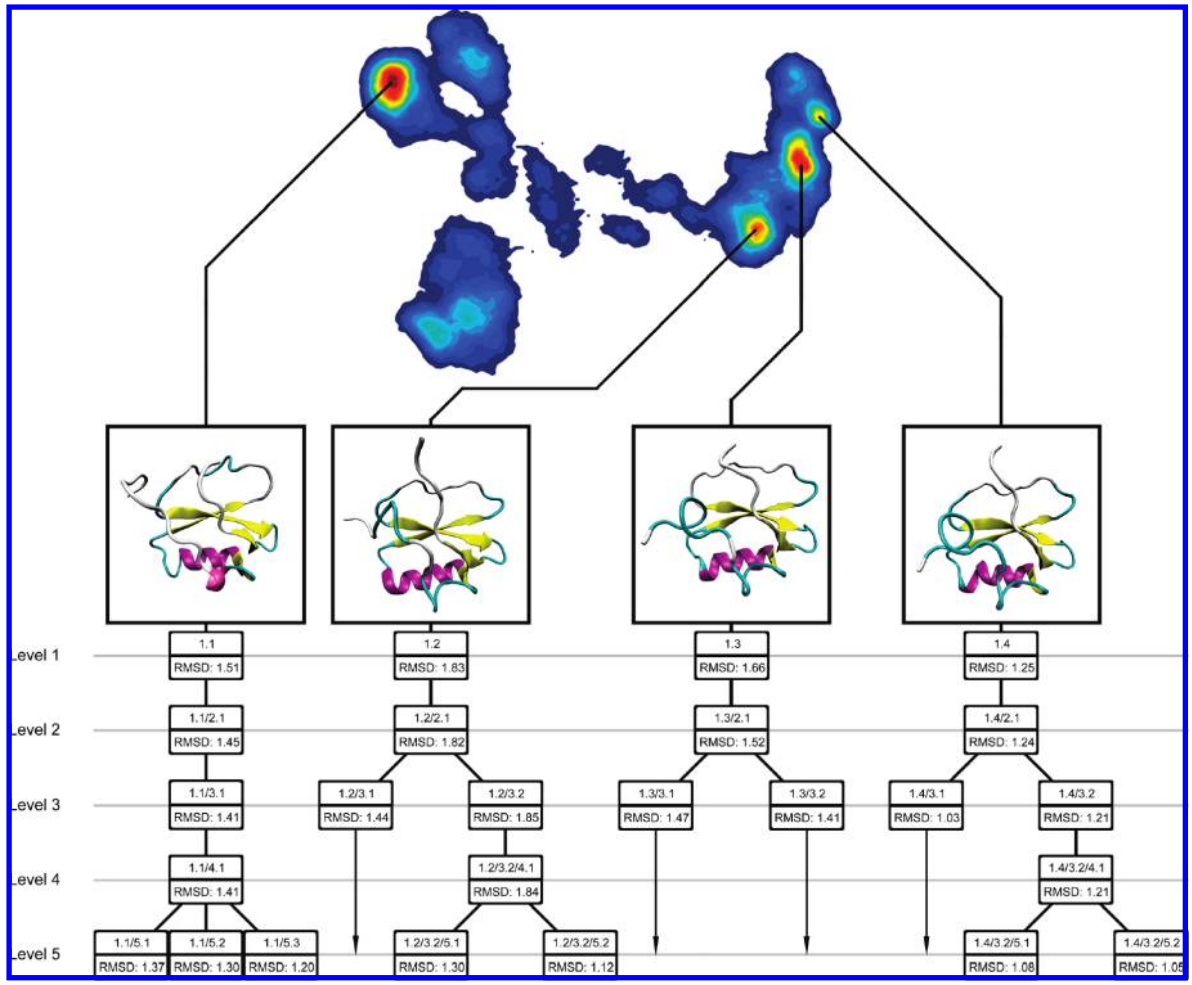

Figure 4. Tree of basin hierarchy is shown. Branches terminated before the fifth level could be continued, however, no further separation of those branches was found within the range of PCs investigated.

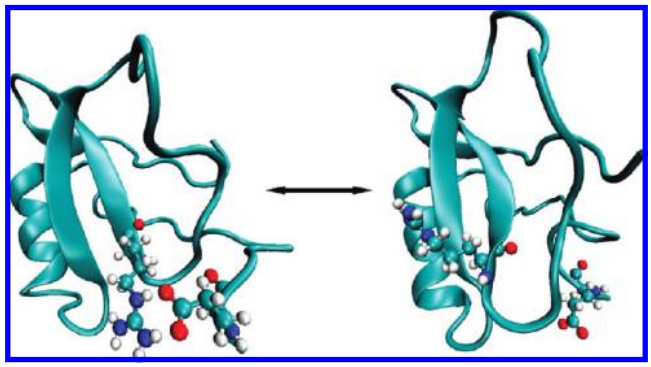

Figure 5. Upon basin splitting, significant changes in inter-residue contacts would typically occur. This figure shows a representative divergence of contacts between two basins. Structures from basins 1.1 and 1.4, shown here, display an arginine−glutamic acid salt bridge broken.

conformations is to take structures from the different basins obtained with a PCA technique developed by Materese et al.[98] Using the weighted histogram analysis method (WHAM),[113] a modification of umbrella sampling, which is a popular technique for calculating free energy differences in computer simulations, we have built a 2D FES in terms of two Q-values defined with respect to each NMR reference states using all-atom CHARMM force-field with explicit solvent[114] (see Figure 6a).

This FES allowed us to investigate kinetic questions, primarily, whether the dynamics of Trp-cage at room temperature is activated (the system spends most time in the minima of the energy landscape waiting for a fluctuation to throw it over the barrier to another minimum), or whether it is similar to a regular diffusional flow of a liquid (when the time is mostly spent in the saddle points of the energy landscape). Seeking a first passage time distribution of the transition from one of the reference conformation to another, we ran damped Brownian dynamics on the 2D FES[115] at several different temperatures. These studies suggested that ambient temperature is in the region of the crossover between the two dynamical regimes. Figure 7 shows the dependence of the mean first passage (MFPT) time on the Brownian dynamics temperature. At higher temperatures, the MFPT dependence is characteristic of diffusion in confined space with $\bar{\tau} \sim D^{-1} \sim T^{-1}$, while at lower temperatures it would be Arrhenius-like with $\bar{\tau} \sim \exp(E_A/(kT))$. The former dependence follows from the discreteness of the spectrum of diffusion equation inside a confined space, where at long times the term with the highest eigenvalue dominates the solution. The exponent of this term is proportional to the diffusion coefficient, and the latter is proportional to temperature from the Einstein−Smoluchowski relation.

The width of the FPT-distribution, which is another way to identify the dynamical regime, also shows borderline behavior: the distributions for three temperatures are given in Figure 6c with their coefficients of variation (standard-deviation-to-mean ratios) denoted as C_v. When $C_v < 1$ the distribution may be considered narrow (signature of the dominant pathway and Arrhenius behavior), while with $C_v > 1$ it may be considered

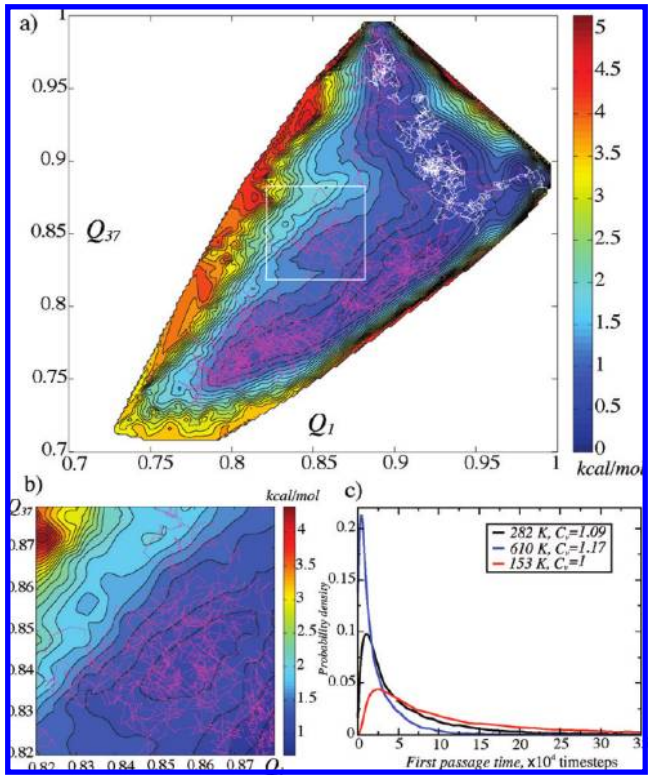

Figure 6. Brownian dynamics of the transition between states corresponding to NMR structures numbered 1 and 37. (a) Two sample trajectories: one (white) is from the peak of FPT distribution in panel c, the other one (magenta) is from the shoulder; (b) Part of the surface with trajectory shown in higher resolution revealing finer features of the surface and their influence on the trajectory; (c) First passage time (FPT) distributions for conformational transition between 1 and 37. FPTs were computed at two additional temperatures to obtain the corresponding coefficients of variation, to provide reference for categorizing the nature of the dynamical regime at room temperature.

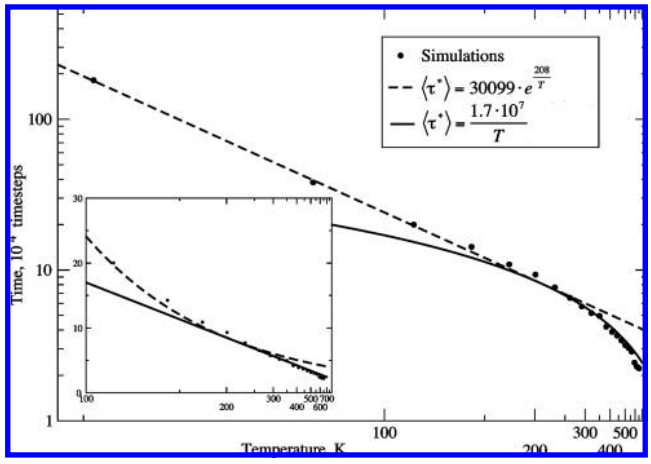

Figure 7. Arrhenius plot shows the dependence of the mean first passage time with temperature for the conformational transition from "37" to "1" on fixed FES (computed at 282 K), where the FES is temperature-independent. The inset zooms into higher temperature region using a semireciprocal plot. Two regimes are clearly seen: exponential at low temperatures and linear at high temperatures. The ambient temperature (282 K) result is near the crossover.

broad (signature of many pathways). At room temperature $C_v = 1.09$ which hints at borderline behavior. Furthermore, one may look at the BD trajectories on the FES and notice the lack of any dominant pathway, which is a characteristic of Arrhenius-type processes. Figure 6a shows two trajectories at 298 K with one taken from the peak of FPT-distribution (white) and the other taken from the shoulder (magenta). About half of the trajectories fall into the shoulder, so there is no dominant pathway, hence, no Arrhenius behavior. However, visual analysis of the trajectories reveals some weak trapping in local shallow minima (Figure 6b). Another method for analyzing kinetics and its time scales is to fit survival curves (obtained by integrating the FPT-distribution, in other words, the probability that the transition has not yet occurred at a given time) to exponential, biexponential or stretched exponential functions. This problem of single- and multiexponential behavior has been discussed in the context of both regular (with a barrier) and downhill protein folding.[116,117] In our case, the curve can be fit by a single exponential, especially at long time scales. At long time scales, such behavior might be expected for diffusion in a confined phase space. However, a somewhat better fit is achieved through fitting a double exponential, representative of dynamical behaviors discussed in the context of barrierless, or downhill protein folding where the shorter time scale is called "speed limit".[116,118] Thus, constructing 2D FESs provides many useful insights about the native dynamics, revealing it at very high resolution.

As useful as two- (or higher) dimensional FESs may be, building them requires a massive computational effort. For some purposes, just knowing the free energy difference between two native-like conformations will suffice. For instance, certain NMR structural experiments provide distance constraints between the atoms as an output.[119] Afterward, these constraints are used within simple force-fields to obtain structures that are deposited as NMR-structures in PDB. Typically, a given set of distance constraints produces many structures, thus, ranking them by free energies would yield a much clearer picture of the optimized structures. In yet another example, measuring the free energy differences between PCA-basins[98] should corroborate the constructed hierarchy, showing the assumed correspondence between kinetic and energetic hierarchies. Lastly, knowing a set of free energy differences between many conformations, one can construct a master equation to describe the system or perhaps tune a force-field. A problem with calculating free energies in computer simulations is the need to sample the whole phase space including the rarely visited regions. In case of an explicit solvent force field, the problem is exacerbated by a large number of the solvent microstates. Additionally, techniques like umbrella sampling that treat this problem by introducing and then correcting for a biasing potential require a meaningful collective coordinate, or order parameter, that will discriminate the two states whose free energy difference is measured (lets call them A and B). Identifying such a coordinate is a nontrivial task. For example, the obvious $\Delta Q = Q_A - Q_B$ coordinate tends to group together very dissimilar conformations in the vicinity of transition end points, that, in turn, may result in serious artifacts. We have recently found an appropriate 1D coordinate which is compact near the end points A and B, solving the degeneracy problem:[105,120]

$$\xi(Q_A, Q_B) = \exp\left(-\frac{(Q_A - Q_{AB})^2 + (Q_B - 1)^2}{(1 - Q_{AB})\rho^2}\right) - \exp\left(-\frac{(Q_A - Q_{AB})^2 + (Q_B - 1)^2}{(1 - Q_{AB})\rho^2}\right) \quad (4.1)$$

In addition, the ΔQ coordinate is nearly impossible to equilibrate in windows where $\Delta Q \approx 0$, because all unfolded states may need to be sampled. To circumvent this problem,

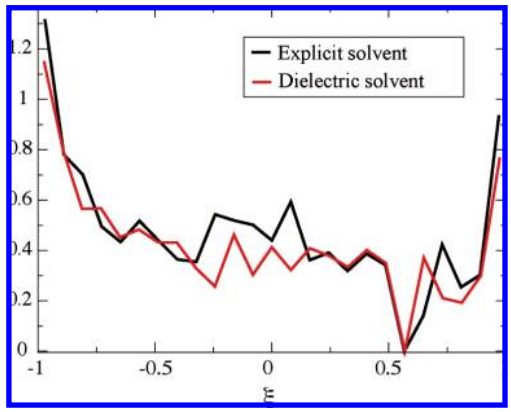

Figure 8. Free-energy profiles for dielectric solvent model and explicit solvent model obtained by 1D-coordinate technique for measuring the free energy difference between the end points ($\xi = 1$ and $\xi = -1$).

we introduced a confining tube in phase space, that allows for fast equilibration.[120] We found that this 1D technique allows free energy differences to be measured in explicit solvent simulations (see Figure 8): the difference in the results from 2D FES and the 1D technique is less than five percent with a 10-fold reduction in computation time.[120]

5. Coarse-Grained Models That Would Describe the Native State Organization

To explore the native state at a resolution as high as in the methods described above, computer simulation models must include explicit solvent. However, all-atom explicit solvent simulations are unable to capture slow (microsecond to second) processes in proteins, or handle large systems such as chromatin. Coarse-grained and implicit solvent models, which allow researchers to overcome some of the sampling and size problems, have been used to simulate dynamics of proteins, and also other hetero- and homopolymers. However, it is not clear if these models are fully adequate for describing the native state dynamics of globular proteins. One of the current goals in implicit solvent and coarse-grained force field development is to obtain a more consistent formulation that reproduces either all-atom explicit solvent simulations or experiments. In this regard, measuring free energy differences and computing FESs seems to be a promising approach for constructing or validating force fields without explicit solvent. If a CG model reproduces the hierarchical organization of the protein energy landscape, in terms of free energies, and kinetics, then it has captured perhaps the most significant essential properties of protein dynamics.

To develop a framework for systematically comparing energy landscapes obtained from explicit and implicit solvent models, in Wu et al.[105] we used a very simplistic model, that we called "dielectric solvent model" (DSM): CHARMM force-field[114] with water molecules removed and all the electrostatic terms in the Hamiltonian divided by dielectric constant of $\varepsilon = 80$. With DSM we have constructed the same 2D FES with respect to two Q-coordinates described in the previous section. Surprisingly, the surfaces from DSM and explicit solvent model were very similar in the region of very high Q values (Figure 9). The difference between the explicit solvent FES and the DSM FES given in Figure 10 demonstrates this more clearly. As Q values decrease more significant differences appear between explicit and DSM FESs, highlighting the role of the hydrophobic effect, which is completely absent in the DSM simulations. The 2D FES differences are asymmetric with respect to the diagonal line, suggesting that the hydrophobic effect favors structures on one part of the surface. To explore this suggestion, we picked several structures from different sides of the diagonal line, ($Q_1 = 0.90$, $Q_{37} = 0.81$) and ($Q_1 = 0.77$, $Q_{37} = 0.76$), and calculated the solvent accessible surface areas (SASA) for each residue (see Table 1). The SASA of the hydrophobic core for the first region turned out to be about 10% smaller which provides further support that the asymmetric shape of the difference FES is hydrophobic in origin.

We also used a 1D coordinate technique mentioned in previous section which shows similar results (see Figure 8). This helps us to better understand the role of water in the native dynamics. We will further discuss the role of water in the next section and discuss how to capture much of it with a simple modification of the DSM, for example using the generalized Born model.[109] It is quite possible that a free energy surface calculated using GBSA Hamiltonian[109] will be even more similar to that computed using explicit solvent. However, the effect of structured water within proteins[104] is difficult to capture with simple implicit solvent models, and an approach with explicit, many-body, water-mediated interaction potentials may be needed.[103]

6. Role of Water

Water plays an extremely important role in both protein folding and native dynamics.[101,121–124] The hydrophobic effect has been long known as the main driving force of protein folding,[125] and early heteropolymer models introduced it as a pairwise attraction between the hydrophobic monomer units.[126–129] As we mentioned in the introduction this tradition still has an impact on contemporary terminology, like "energy" landscape, which includes phenomena of entropic origin. As described below, our results indicate that the true thermodynamic energy landscape is much more rugged, than the landscapes studied in protein folding theories. Long-range hydrophilic contacts are also mediated by water[103]—introducing a second well into the pairwise polar−polar potential improves protein structure prediction. This second well corresponds to a water-mediated contact, a common occurrence in the native state.[103,130] Overall, water plays a significant role in smoothing the folding funnel and guiding the folding of a protein to its native state.[103] The dynamics of the protein and its hydration shell are also interconnected. In some cases protein motion is thought to be slaved to solvent motion,[59,131] as solvent effectively cages the protein. The collective modes of these coupled protein and solvent dynamics can be probed by terahertz (THz) spectroscopy. A recent study suggests that the so-called protein dynamical transition—the rapid increase of the dynamics around 200 K—originates in the motions of water and interaction between water and side-chains.[132] THz spectroscopy has also been used to study the effect of mutations on hydration shells of the proteins, showing a case where pseudo-wild-type, λ^*_{6-85}, has a much more pronounced effect on long-distance solvation water than a point mutant.[133]

In Materese et al.[98] we have shown that the presence of water-mediated contacts is correlated with hierarchical organization of the PCA derived basins, which in turn is relevant to hierarchical organization of the energy landscape. Water bridging was found to involve both side chain and backbone chemical groups. These bridges were found both between residues close enough to be considered in direct contact, and interactions at longer range. The mediation of direct contacts often included residues of like charge. We suggested that hydrogen bonding, created by favorable water interactions diminishes repulsion

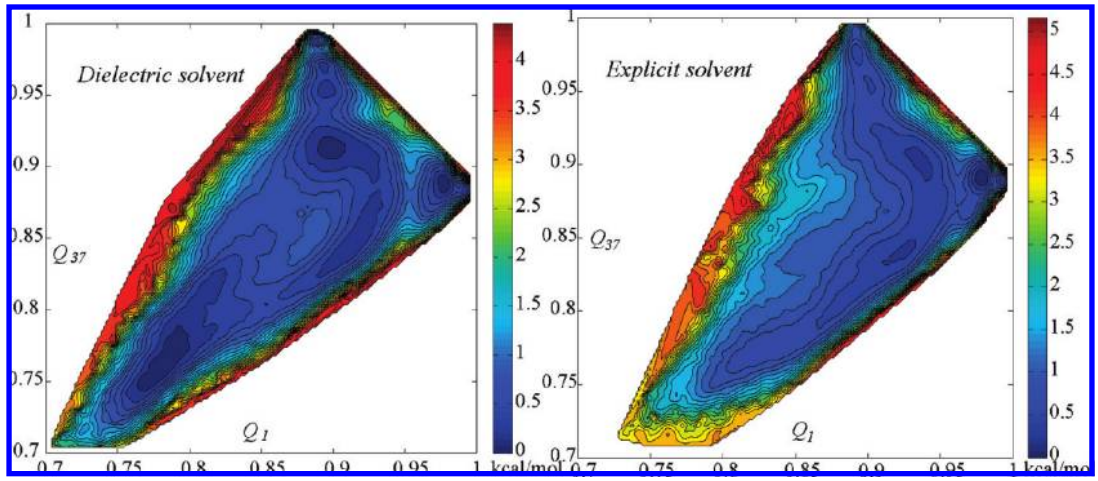

Figure 9. (left) FES computed for the dielectric solvent model on the left. (Right) Explicit solvent FES. The lower basin is absent in the explicit solvent FES. On the other hand, a number of barriers have disappeared in the DSM FES. The vicinities of the reference points 1 and 37 are very similar between explicit and DSM FESs.

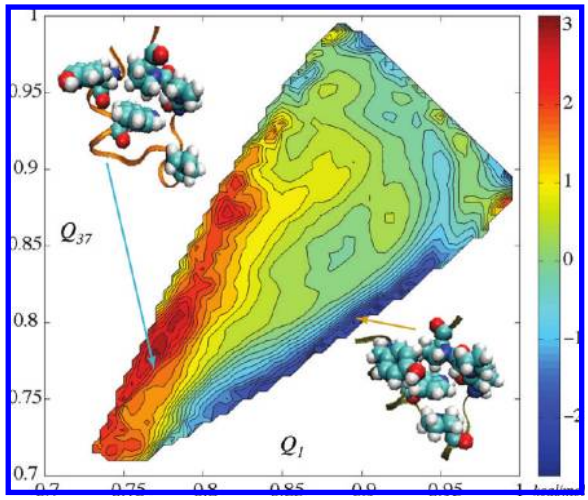

Figure 10. The difference between explicit solvent FES and DSM FES (the former minus the latter). In the Qs $\lesssim$ 0.95 region it mainly represents the hydrophobic effect. The structures in the corners are taken from the hydrophobically favorable and unfavorable regions. The hydrophobic core is shown in van der Waals spheres, so that its opening is visually noticeable.

TABLE 1: Solvent Accesible Surface Areas for Several Structures ($Å^2$) Indicated on Figure 10

(Q_1, Q_{37})	SASA of HP core	SASA of the whole protein
model 1	955	1888
Hydrophobically Favorable Region		
(0.907, 0.807)	975	1943
(0.905, 0.809)	975	1952
(0.901, 0.802)	969	1931
(0.900, 0.808)	964	1901
Hydrophobically Unfavorable Region		
(0.774, 0.762)	1089	1936
(0.772, 0.763)	1120	1932
(0.773, 0.764)	1092	1920
(0.773, 0.761)	1038	2027

between the like charged residues, allowing them to develop such a contact. We observed changes in water mediated contacts between different basins in our hierarchy suggesting that water plays a role in shaping said hierarchy. Proteins have evolved in aqueous environments and water can play important, specific, microscopic roles as shown in this work. The importance of water at specific sites in the protein structure underscores the importance of explicit solvent in fine scale structure determination. For instance, recent experimental work points to the importance of water in sculpting specific protein structure and function. In Szep et al., the authors performed crystallographic studies on wild type and mutant FKBP12 binding protein and provided evidence that a specific crystallographic water, distal to the binding pocket, plays a role in sculpting the shape of that pocket.[104] The authors suggested that despite the fact that the binding pocket is greater than 8 Å away, this specific water plays a key role on a network of interactions which shape it.

In Wu et al.[105] we have built true thermodynamic energy and entropy surfaces and explicitly shown that their ruggedness is about 10 times greater than the free energy surface plotted in terms of the parameters Q (see Figure 11). One way to rationalize this is to focus on the solvent enthalpy−entropy compensation, which is characteristic of the hydrophobic effect. When a protein chain slowly fluctuates, the resulting fluctuations of protein's solvent accessible area lead to large changes in entropy and enthalpy separately, but small changes in their combination, free energy. Because of the time scale separation between fast motions of the solvent and slower motions of the protein chain, the latter does not feel all the ruggedness of solvent's energy landscape—the protein moves in the thermally averaged field of solvent motions.[105]

This confirms the smoothing role of water. Since this effect is so important, its role must be reflected in computational models in a more rigorous way than simple pairwise potentials. Introduction of nonpairwise water has been shown to improve structure prediction.[103,130] Solvent accessible surface area seems to be good starting approach to improving pairwise hydrophobic potentials.[105] As discussed above, when we compared a Hamiltonian that does not include water molecules explicitly, but where water's dielectric permittivity was taken into account, the difference between the FES obtained from this simple model and from explicit water Hamiltonian was attributed mostly to the hydrophobic effect, and was correlated to the solvent accessible surface area calculated for several randomly picked states. In summary, water smoothes the landscape, guides folding and takes part in hierarchical organization, particularly, through both hydrophobic effect and water-mediated hydrophilic contacts.

7. Summary

The dynamics of the native state is varied and interesting ranging from fluid diffusive behavior through an Arrhenius-type activated behavior to glassy behavior. It is intrinsically connected to protein's energy landscape, the specific features of which are manifest in hierarchical organization of kinetics and free energy basins. Ruggedness of the energy landscape and to what extent it is frustrated directly influences both folding and native dynamics. Detailed knowledge of these properties is important for studying the enzymatic catalysis, allosteric switching, fluorescence intermittency, protein quakes, and other phenomena.

In recent years, some NMR experimental techniques, including nuclear Overhauser enhancement spectroscopy (NOESY), ZZ-exchange spectroscopy, and some relaxations dispersion experiments, have emerged which permit the investigation of native dynamics on time scales that are also accessible to all-atom computer simulations.[134,135] The X-ray-diffraction crystallography also yields the spatial distribution of atoms around the average structure in addition to the average structure itself.[18,136] More recently, sub-Angstrom resolutions were achieved providing some data on directionality of these atomic fluctuations through anisotropic Debye−Waller factors.[137] In addition, Laue X-ray diffraction, although not universally applicable to proteins, can measure the dynamic component of the Debye−Waller factor separately from the component caused by lattice disorder.[138] Paramagnetic relaxation enhancement measurements can also capture transient events of protein's dynamical ensemble and access low-populated regions of the landscape.[139,140]

The energy landscape framework sets a common basis between these experiments and simulations of the kind we discuss here, and experimentalists are beginning to increasingly use the energy landscape language.[139,141] This is nicely illustrated by a recent work by Romesberg et al.,[141] where the dynamics of molecular recognition of flourescein by antibodies was studied with three-pulse photon echo peak shift (3PEPS) spectroscopy. Observing distinct time scales in the dynamics of antibody−fluorescein complexes, the authors interpreted it as a manifestation of hierarchical nature of the underlying energy landscape. They have organized the observed motions into three classes corresponding to three tiers: inertial (femtosecond, local motions in microbasin), diffusive (picosecond, crossing small barriers between neighboring substates), and large-scale conformational changes. Different antibodies that have evolved to bind fluorescein do so in a divergent fashion: some have a larger enthalpic component to the binding, others smaller; and the entropic component has been shown to vary in sign. Dynamically, the proportion of motions different antibodies exhibit from different tiers varies from an antibody to antibody. For example, contribution of conformational dynamics (tier 3) may change from ∼15% to ∼45%. Thus one can envision different energy landscapes for such antibodies: one with steep barriers, so that "elastic" behavior from inertial motions of tier 1 dominates the dynamics, rendering the antibody rigid, or one with low barriers and shallow traps on the landscape, leading to a more flexible antibody.[141] From simulations, one could compute the landscape and compare the barriers with those predicted from experimental interpretation. Additionally, simulations can be used to see how the conformational entropy is correlated with the tiered composition of the dynamics, or to observe the various binding mechanisms. In this light, using energy landscape ideas as the language for discussing native state dynamics seems to be very promising.

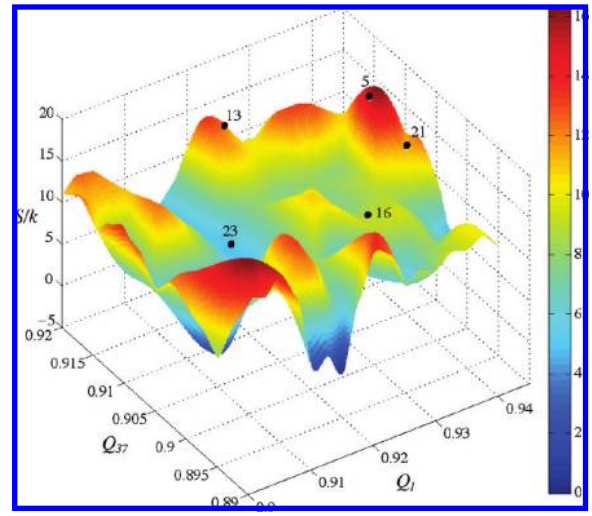

Figure 11. Rugged entropy landscape. The average energy landscape is very similar. Corresponding contributions to free energy is 10 times larger than the change in free energy itself. Various structures from the NMR-study[111] are marked.

In this review we have highlighted several important and powerful techniques for studying the organization and kinetics of the native state. These include dihedral angles principal component analysis with subsequent mapping of the dynamics onto a tree; building free energy surfaces with subsequent Brownian dynamics investigation of the surface properties; and other free energy techniques. The PCA-based technique opens perspectives for better understanding of the role of particular inter-residue contacts, including water-mediated contacts and finding the large scale topography of dynamical basins. Free energy surfaces allow much less expensive study of the type of dynamics and the time scale spectrum; both techniques may contribute to the building of the coarse-grained models in a more rigorous way. Much has been accomplished in the field of protein folding over the past 2 decades and much progress has been made in the study of the native state itself. Gaining further deeper insights into the thermodynamics and kinetics of the native state ensemble, including for proteins with weakly funneled landscapes, is interesting from the viewpoint of physical chemistry and is essential for understanding biological function of many proteins.

Acknowledgment. We thank Dr. Sangwook Wu for helpful discussions. We gratefully acknowledge support from the Camille and Henry Dreyfus New Faculty Award, Camille Dreyfus Teacher-Scholar Award, and Arnold and Mabel Beckman Foundation Beckman Young Investigator Award.

References and Notes

(1) Levinthal, C. *J. Chim. Phys.* **1968**, *65*, 44–45.
(2) Scheraga, H. A. *Biopolymers* **1983**, *22*, 1–14.
(3) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991**, *254*, 1598–1603.
(4) Frauenfelder, H.; Alberding, N. A.; Ansari, A.; Braunstein, D.; Cowen, B. R.; Hong, M. K.; Iben, I. E. T.; Johnson, J. B.; Luck, S.; et al. *J. Phys. Chem.* **1990**, *94*, 1024–1037.
(5) Onuchic, J. N.; Wolynes, P. G.; Luthey-Schulten, Z.; Socci, N. D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3626–3630.
(6) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins* **1995**, *21*, 167–195.
(7) Pande, V. S.; Grosberg, A. Y.; Tanaka, T. *Biophys. J.* **1997**, *73*, 3192–3210.
(8) Pande, V. S.; Grosberg, A. Y.; Tanaka, T. *Rev. Mod. Phys.* **2000**, *72*, 259–314.

(9) Plotkin, S. S.; Onuchic, J. N. *Q. Rev. Biophys.* **2002**, *35*, 205–286.
(10) Scheraga, H. A.; Khalili, M.; Liwo, A. *Annu. Rev. Phys. Chem.* **2007**, *58*, 57–83.
(11) Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. *Annu. Rev. Biophys.* **2008**, *37*, 289–316.
(12) Chen, Y.; Ding, F.; Nie, H.; Serohijos, A. W.; Sharma, S.; Wilcox, K. C.; Yin, S.; Dokholyan, N. V. *Arch. Biochem. Biophys.* **2008**, *469*, 4–19.
(13) Grosberg, A.; Khokhlov, A. R. *Statistical Physics of Macromolecules*; American Institute of Physics, Melville, NY: 1994.
(14) Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. *Fold Des.* **1998**, *3*, 577–587.
(15) Sherman, E.; Haran, G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11539–11543.
(16) Pande, V. S.; Grosberg, A. Y.; Tanaka, T. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12972–12975.
(17) Shakhnovich, E. I.; Gutin, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 7195–7199.
(18) Henzler-Wildman, K.; Kern, D. *Nature (London)* **2007**, *450*, 964–972.
(19) Fenimore, P. W.; Frauenfelder, H.; McMahon, B. H.; Young, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14408–14413.
(20) Plaxco, K. W.; Gross, M. *Nat. Struct. Biol.* **2001**, *8*, 659–660.
(21) Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K. *Proteins* **2001**, *42*, 38–48.
(22) Papoian, G. A.; Wolynes, P. G. *Biopolymers* **2003**, *68*, 333–349.
(23) McCammon, J.; Harvey, S. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: New York, 1987.
(24) McCarney, E. R.; Kohn, J. E.; Plaxco, K. W. *Crit. Rev. Biochem. Mol. Biol.* **2005**, *40*, 181–189.
(25) Doi, M.; Edwards, S. *Theory of Polymer Dynamics*; Clarendon Press: Oxford, U.K., 1986.
(26) Guo, Z.; Thirumalai, D. *Biopolymers* **1995**, *36*, 83–102.
(27) Gast, K.; Damaschun, H.; Misselwitz, R.; Mller-Frohne, M.; Zirwer, D.; Damaschun, G. *Eur. Biophys. J.* **1994**, *23*, 297–305.
(28) Ptitsyn, O. B.; Pain, R. H.; Semisotnov, G. V.; Zerovnik, E.; Razgulyaev, O. I. *FEBS Lett.* **1990**, *262*, 20–24.
(29) Ptitsyn, O. B. *Nat. Struct. Biol.* **1996**, *3*, 488–490.
(30) Ptitsyn, O. B. *Trends Biochem. Sci.* **1995**, *20*, 376–379.
(31) Ptitsyn, O. B.; Uversky, V. N. *FEBS Lett.* **1994**, *341*, 15–18.
(32) Ptitsyn, O. B. *Protein Folding*; Creighton, T. E., Ed.; Freeman: New York, 1992.
(33) Ptitsyn, O. B. *Curr. Opin. Struct. Biol.* **1995**, *5*, 74–78.
(34) Ptitsyn, O. B.; Bychkova, V. E.; Uversky, V. N. *Philos. Trans. R. Soc., London, B* **1995**, *348*, 35–41.
(35) Ding, F.; Jha, R. K.; Dokholyan, N. V. *Structure* **2005**, *13*, 1047–1054.
(36) Tran, H. T.; Wang, X.; Pappu, R. V. *Biochemistry* **2005**, *44*, 11369–11380.
(37) Lin, L.; Pinker, R. J.; Forde, K.; Rose, G. D.; Kallenbach, N. R. *Nat. Struct. Biol.* **1994**, *1*, 447–452.
(38) Flanagan, J. M.; Kataoka, M.; Fujisawa, T.; Engelman, D. M. *Biochemistry* **1993**, *32*, 10359–10370.
(39) Plaxco, K. W.; Dobson, C. M. *Curr. Opin. Struct. Biol.* **1996**, *6*, 630–636.
(40) Klimov, D. K.; Thirumalai, D. *Fold Des.* **1998**, *3*, 127–139.
(41) Bu, Z.; Neumann, D. A.; Lee, S. H.; Brown, C. M.; Engelman, D. M.; Han, C. C. *J. Mol. Biol.* **2000**, *301*, 525–536.
(42) Lubchenko, V. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 10635–10636.
(43) Lubchenko, V. *J. Non-Cryst. Solids* **2006**, *352*, 4400–4409.
(44) Ringe, D.; Petsko, G. A. *Biophys. Chem.* **2003**, *105*, 667–680.
(45) Ding, X.; Rasmussen, B. F.; Petsko, G. A.; Ringe, D. *Bioorg. Chem.* **2006**, *34*, 410–423.
(46) Vitkup, D.; Ringe, D.; Petsko, G. A.; Karplus, M. *Nat. Struct. Biol.* **2000**, *7*, 34–38.
(47) Plotkin, S. S.; Onuchic, J. N. *Q. Rev. Biophys.* **2002**, *35*, 111–167.
(48) Camacho, C. J.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6369–6372.
(49) Klimov, D. K.; Thirumalai, D. *Proteins* **1996**, *26*, 411–441.
(50) Schrodinger, E. *What Is Life?: The Physical Aspect of the Living Cells*; Cambridge University Press: Cambridge, New York, 1944.
(51) Shakhnovich, E. I.; Finkelstein, A. V. *Biopolymers* **1989**, *28*, 1667–1680.
(52) Bryngelson, J. D.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7524–7528.
(53) Dill, K. A.; Chan, H. S. *Nat. Struct. Biol.* **1997**, *4*, 10–19.
(54) Yue, K.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 146–150.
(55) Dobson, C. M.; Sali, A.; Karplus, M. *Angew. Chem., Int. Ed.* **1998**, *37*, 868–893.
(56) Wolynes, P. *Proc. Am. Philos. Soc.* **2001**, *145*, 555–563.
(57) Ferreiro, D. U.; Hegler, J. A.; Komives, E. A.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 19819–19824.
(58) Sutto, L.; Latzer, J.; Hegler, J. A.; Ferreiro, D. U.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 19825–19830.
(59) Austin, R. H.; Beeson, K. W.; Eisenstein, L.; Frauenfelder, H.; Gunsalus, I. C. *Biochemistry* **1975**, *14*, 5355–5373.
(60) Ansari, A. A.; Hattikudur, N. S.; Joshi, S. R.; Medeira, M. A. *J. Immunol. Methods* **1985**, *84*, 117–124.
(61) Schwille, P.; Kummer, S.; Heikal, A. A.; Moerner, W. E.; Webb, W. W. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 151–156.
(62) Kern, D.; Zuiderweg, E. R. P. *Curr. Opin. Struct. Biol.* **2003**, *13*, 748–757.
(63) McCammon, J. A.; Karplus, M. *Nature (London)* **1977**, *268*, 765–766.
(64) Bahar, I.; Rader, A. J. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586–592.
(65) Hayward, S.; Go, N. *Annu. Rev. Phys. Chem.* **1995**, *46*, 223–250.
(66) Ma, J. *Structure* **2005**, *13*, 373–380.
(67) Amadei, A.; Linssen, A. B.; Berendsen, H. J. *Proteins* **1993**, *17*, 412–425.
(68) Schmitz, M.; Tavan, P. *J. Chem. Phys.* **2004**, *121*, 12233–12246.
(69) Ham, S.; Hahn, S.; Lee, C.; Kim, T.-K.; Kwak, K.; Cho, M. *J. Phys. Chem. B* **2004**, *108*, 9333–9345.
(70) Hahn, S.; Ham, S.; Cho, M. *J. Phys. Chem. B* **2005**, *109*, 11789–11801.
(71) Schrader, T. E.; Schreier, W. J.; Cordes, T.; Koller, F. O.; Babitzki, G.; Denschlag, R.; Renner, C.; Löweneck, M.; Dong, S.-L.; Moroder, L.; Tavan, P.; Zinth, W. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15729–15734.
(72) Caflisch, A. *Curr. Opin. Struct. Biol.* **2006**, *16*, 71–78.
(73) Becker, O. M.; Karplus, M. *J. Chem. Phys.* **1997**, *106*, 1495–1517.
(74) Frauenfelder, H.; McMahon, B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 4795–4797.
(75) Wales, D. J.; Bogdan, T. V. *J. Phys. Chem. B* **2006**, *110*, 20765–20776.
(76) Rao, F.; Caflisch, A. *J. Mol. Biol.* **2004**, *342*, 299–306.
(77) Fersht, A. *Structure and Mechanism in Protein Science*; Freeman: New York, 1999.
(78) Min, W.; Xie, X. S.; Bagchi, B. *J. Phys. Chem. B* **2008**, *112*, 454–466.
(79) Borders, C. L.; Snider, M. J.; Wolfenden, R.; Edmiston, P. L. *Biochemistry* **2002**, *41*, 6995–7000.
(80) Wolfenden, R.; Snider, M. J. *Acc. Chem. Res.* **2001**, *34*, 938–945.
(81) Borchers, C. H.; Marquez, V. E.; Schroeder, G. K.; Short, S. A.; Snider, M. J.; Speir, J. P.; Wolfenden, R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15341–15345.
(82) Pervushin, K.; Vamvaca, K.; Vgeli, B.; Hilvert, D. *Nat. Struct. Mol. Biol.* **2007**, *14*, 1202–1206.
(83) Papoian, G. A.; DeGrado, W. F.; Klein, M. L. *J. Am. Chem. Soc.* **2003**, *125*, 560–569.
(84) Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. *Nature (London)* **2007**, *450*, 913–916.
(85) Eisenmesser, E. Z.; Millet, O.; Labeikovsky, W.; Korzhnev, D. M.; Wolf-Watz, M.; Bosco, D. A.; Skalicky, J. J.; Kay, L. E.; Kern, D. *Nature (London)* **2005**, *438*, 117–121.
(86) Papoian, G. A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 14237–14238.
(87) Levy, Y.; Cho, S. S.; Onuchic, J. N.; Wolynes, P. G. *J. Mol. Biol.* **2005**, *346*, 1121–1145.
(88) Levy, Y.; Papoian, G. A.; Onuchic, J. N.; Wolynes, P. G. *Israel. J. Chem.* **2004**, *44*, 281–297.
(89) Dyson, H. J.; Wright, P. E. *Curr. Opin. Struct. Biol.* **2002**, *12*, 54–60.
(90) Roca, M.; Messer, B.; Hilvert, D.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13877–13882.
(91) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8868–8873.
(92) McNulty, B. C.; Young, G. B.; Pielak, G. J. *J. Mol. Biol.* **2006**, *355*, 893–897.
(93) Garcia, A. E.; Sanbonmatsu, K. Y. *Proteins* **2001**, *42*, 345–354.
(94) Becker, O. M. *J. Comput. Chem.* **1998**, *19*, 1255–1267.
(95) Levy, Y.; Becker, O. M. *J. Chem. Phys.* **2000**, *114*, 993–1009.
(96) Altis, A.; Nguyen, P. H.; Hegger, R.; Stock, G. *J. Chem. Phys.* **2007**, *126*, 244111.
(97) Hegger, R.; Altis, A.; Nguyen, P. H.; Stock, G. *Phys. Rev. Lett.* **2007**, *98*, 028102.
(98) Materese, C. K.; Goldmon, C. C.; Papoian, G. A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 10659–10664.
(99) Rammal, R.; Toulouse, G.; Virasoro, M. *Rev. Mod. Phys.* **1986**, *58*, 765–788.

(100) Franz, S.; Mezard, M.; Parisi, G.; Peliti, L. *J. Stat. Phys.* **1999**, *97*, 459–488.
(101) Cheung, M. S.; Garcia, A. E.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 685–690.
(102) Papoian, G. A.; Ulander, J.; Wolynes, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 9170–9178.
(103) Papoian, G. A.; Ulander, J.; Eastwood, M. P.; Luthey-Schulten, Z.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 3352–3357.
(104) Szep, S.; Park, S.; Boder, E. T.; Duyne, G. D. V.; Saven, J. G. *Proteins* **2009**, *74*, 603–611.
(105) Wu, S.; Zhuravlev, P. I.; Papoian, G. A. *Biophys. J.* **2008**, *95*, 5524–5532.
(106) Gruebele, M. *Curr. Opin. Struct. Biol.* **2002**, *12*, 161–168.
(107) Krivov, S. V.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14766–14770.
(108) Bursulaya, B.; Brooks, C. *J. Am. Chem. Soc.* **1999**, *121*, 9947–9951.
(109) Zhou, R. *Proteins* **2003**, *53*, 148–161.
(110) Plotkin, S. S.; Wang, J.; Wolynes, P. G. *J. Chem. Phys.* **1997**, *106*, 2932–2948.
(111) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425–430.
(112) Ding, F.; Buldyrev, S. V.; Dokholyan, N. V. *Biophys. J.* **2005**, *88*, 147–155.
(113) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13* (8), 1011–1021.
(114) MacKerell, A. D.; Banavali, N.; Foloppe, N. *Biopolymers* **2000**, *56*, 257–265.
(115) Socci, N. D.; Onuchic, J. N.; Wolynes, P. G. *J. Chem. Phys.* **1996**, *104*, 5860–5868.
(116) Yang, W. Y.; Gruebele, M. *Nature (London)* **2003**, *423*, 193–197.
(117) Chekmarev, S. F.; Krivov, S. V.; Karplus, M. *J. Phys. Chem. B* **2006**, *110*, 8865–8869.
(118) Liu, F.; Du, D.; Fuller, A. A.; Davoren, J. E.; Wipf, P.; Kelly, J. W.; Gruebele, M. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 2369–2374.
(119) Sattler, M.; Schleucher, J.; Griesinger, C. *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 93–158.
(120) Zhuravlev, P. I.; Wu., S.; Rubinstein, M.; Papoian, G. A. *Phys. Rev. Lett*., submitted for publication.
(121) Head-Gordon, T.; Brown, S. *Curr. Opin. Struct. Biol.* **2003**, *13*, 160–167.
(122) van der Vaart, A.; Bursulaya, B. D.; Brooks, C. L.; Merz, K. M. *J. Phys. Chem. B* **2000**, *104*, 9554–9563.
(123) Kaya, H.; Chan, H. S. *J. Mol. Biol.* **2003**, *326*, 911–931.
(124) Kauzmann, W. *Adv. Protein Chem.* **1959**, *14*, 1–63.
(125) Dyson, H. J.; Wright, P. E.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13057–13061.
(126) Lau, K. F.; Dill, K. A. *Macromolecules* **1989**, *22*, 3986–3997.
(127) Yue, K.; Fiebig, K. M.; Thomas, P. D.; Chan, H. S.; Shakhnovich, E. I.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 325–329.
(128) Honeycutt, J. D.; Thirumalai, D. *Biopolymers* **1992**, *32*, 695–709.
(129) Vasilevskaya, V.; Khalatur, P.; Khokhlov, A. *Macromolecules* **2003**, *36*, 10103–10111.
(130) Zong, C.; Papoian, G. A.; Ulander, J.; Wolynes, P. G. *J. Am. Chem. Soc.* **2006**, *128*, 5168–5176.
(131) Frauenfelder, H.; Fenimore, P. W.; Young, R. D. *IUBMB Life* **2007**, *59*, 506–512.
(132) He, Y.; Ku, P. I.; Knab, J. R.; Chen, J. Y.; Markelz, A. G. *Phys. Rev. Lett.* **2008**, *101*, 178103.
(133) Ebbinghaus, S.; Kim, S. J.; Heyden, M.; Yu, X.; Gruebele, M.; Leitner, D. M.; Havenith, M. *J. Am. Chem. Soc.* **2008**, *130*, 2374–2375.
(134) Kay, L. E. *J. Magn. Reson.* **2005**, *173*, 193–207.
(135) Palmer, A. G. *Chem. Rev.* **2004**, *104*, 3623–3640.
(136) Frauenfelder, H.; Petsko, G. A.; Tsernoglou, D. *Nature (London)* **1979**, *280*, 558–563.
(137) Merritt, E. A. *Acta Crystallogr., D* **1999**, *55*, 1109–1117.
(138) Bourgeois, D.; Royant, A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 538–547.
(139) Clore, G. M. *Mol. Biosyst.* **2008**, *4*, 1058–1069.
(140) Tang, C.; Louis, J. M.; Aniana, A.; Suh, J.-Y.; Clore, G. M. *Nature (London)* **2008**, *455*, 693–696.
(141) Thielges, M. C.; Zimmermann, J.; Yu, W.; Oda, M.; Romesberg, F. E. *Biochemistry* **2008**, *47*, 7237–7247.

JP810659U