Protein and Peptide Folding Explored with Molecular Simulations

CHARLES L. BROOKS III*
Department of Molecular Biology (TPC-6), The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received October 8, 2001

ABSTRACT
Molecular simulations, comprising models with atomic details of polypeptide and solvent as well as minimalist models employing only Cα atoms, are being used with specialized simulation methods from statistical mechanics to examine fundamental questions in peptide and protein folding mechanism, kinetics, and thermodynamics. Detailed calculations of free energy changes along coordinates describing the formation of hydrogen-bonding interactions in helical, turn, and β-sheet models provide insights into the time scale and mechanism of secondary structure formation. Potential roles for these processes in directing protein folding are also elucidated by such calculations. Analogous methodologies extended to more complex polypeptides with tertiary structures (proteins) are used to explore global questions about protein folding landscapes, to delineate atomic details of folding mechanism, and to elucidate putative roles for solvent in the late stages of folding.

Introduction

A cornerstone of interdisciplinary research, which envelopes chemistry, physics, and biology and includes both theoretical and experimental research programs, has been the development of a fundamental framework for understanding the processes of protein and peptide folding. In the four decades since the connection between a polypeptide's one-dimensional sequence and its functional three-dimensional structure was established by Anfinsen,[1] our views of folding have evolved from the recognition that folding cannot occur as a simple random search but must somehow be directed, as noted by Levinthal,[2] to a perspective which viewed folding as proceeding on a "chemical reaction-like" pathway,[2−5] to a global statistical view of the folding landscape.[6−11] Present day views of folding embrace the encompassing landscape perspective, which provides a "language" to analyze and categorize folding scenarios,[7] anticipate kinetic and mechanistic outcomes,[11] and formulate design strategies.[12]

Critical to the advancement of our understanding of folding has been the development and application of molecular models and simulation approaches.[13,14] These include simplified models of the polypeptide chain and its environment, such as lattice-based hydrophobic/hydrophilic (HP) representations[15] and minimalist interaction schemes.[16−21] In studies employing this class of model, often complete information, or nearly so, about the folding process (free energy, kinetic scheme, and mechanism) can be elucidated. However, this is generally at the cost of many details not present in the minimalist models, e.g., direct hydrogen bonding, competition for interactions with solvent, etc. Minimalist models thus serve to help us discover many of the gross and robust features of folding landscapes, e.g., which specific coordinates are relevant to describe the mechanism of folding and when folding may be "reduced" to characterization by a few key variables. They also provide a computational paradigm for exploring algorithms, sharpening conceptual issues, and investigating how best to "ask" questions using more complicated and detailed models.[22] This theme has influenced our own work, as we will illustrate in this Account, by demonstrating how minimalist models can be employed to examine symmetry breaking in the transition state of folding for molecules that are topologically symmetric, i.e., molecules composed of symmetrically distributed elements of secondary structure.[23]

Despite the key contributions minimalist models have made to our understanding of folding, they are generally not useful in addressing questions where the detailed atomic nature of the peptide or solvent are important in the process under study. Detailed atomic models, e.g., all-atom force fields with explicit solvent, figure significantly into the extension of conceptual and quantitative issues of folding when minimalist models fail. Detailed models have honed our understanding of folding where microscopic interpretations about specific sequences of peptide are of interest.[9,24−27] Our discussions below are motivated by the efforts in our group over the past 15 years to develop and apply methods based on detailed atomic models to problems of peptide and protein folding.

Fundamental to this effort has been the development and application of biased or umbrella sampling as a framework for calculations of thermodynamic properties for folding projected onto important (and predetermined) progress coordinates.[22,28] These free energy surfaces form the basis for exploring the time scale (kinetic aspects) and mechanism of folding for peptides and proteins as well as its thermodynamic properties. They complement methods that employ strong "denaturing" conditions to study folding, e.g., the use of extremely high temperatures to unfold peptides and proteins in solution.[24,25,27,29] While novel insights can (and have) been gained from this type of molecular simulation, this approach suffers from the potential problem of yielding only anecdotal characterizations of the process of interest, reflecting the few specific initial conditions sampled in the study.[22]

Charles L. Brooks III was born in Detroit, Michigan in 1956. He received B.S. degrees in Chemistry and Physics from Alma College in 1978 and his Ph.D. in Physical Chemistry from Purdue University in 1982. Dr Brooks studied as a postdoctoral fellow at Harvard University between 1982 and 1985 and was recipient of an NIH NRSA postdoctoral fellowship between 1983 and 1885. He joined the faculty of Carnegie Mellon University in 1985 and was appointed Professor of Chemistry in 1994. In 1994 Professor Brooks moved to the Scripps Research Institute as a Full Professor in the Department of Molecular Biology. Professor Brooks received an Alfred P. Sloan Foundation Fellowship in 1992 and was elected Fellow of the American Association for the Advancement of Science in 2001.

* E-mail: brooks@scripps.edu. Telephone: 858/784-8035. Fax: 858/784-8688.

10.1021/ar0100172 CCC: $22.00
Published on Web 04/26/2002

Table 1. Free Energy of Stability and Approximate Kinetics of Hydrogen Bond Formation for Hydrogen Bonds in Secondary Structure

secondary structure	stability per H-bond,[a] kcal/mol	folding/unfolding time scale,[b] ps	model	reference state
antiparallel β-sheet	−2.8	15/10 000	$[Ac\text{-}ala\text{-}NHMe]_2$	infinite separation
ala-gly type II turn	−0.6	N/A	Ac-ala-gly-NHMe	extended
amide H-bond	−0.3	1/25	$[formamide]_2$	infinite separation
1st helical H-bond	−0.2	100/10	$Ac\text{-}(ala)_3\text{-}NHMe$	extended
2nd helical H-bond	−0.4/−1.0	30/10	$Ac\text{-}(ala)_4\text{-}NHMe$	extended
ala-gly type I turn	2.6	N/A	Ac-ala-gly-NHMe	extended
pro-gly type I turn	2.6	N/A	Ac-pro-gly-NHMe	extended

[a] The precision of these free energy differences is on the order of ±10% (see original papers for details). [b] Time scales are from direct observation or constructed from barriers on the free energy surfaces and transition state theory using a prefactor of 1 ps.

Free energy simulations as we shall discuss them here refer to the application of simulation methods and models where the primary objective is the calculation of *changes in free energy* occurring along a small and specified set of variables describing the transformation of a physical system from one well-defined thermodynamic state into another. In the context of folding these generally refer to thermodynamic states well described by the nature of their configurational distributions. In folding of complex polypeptides such as proteins, such coordinates would distinguish the folded ensemble from the manifold of unfolded configurations. Typical variables might be the radius of gyration, R_g, or the number of native interactions, C_N. Clearly, polypeptides in their folded state will exhibit relatively small R_g values and relatively large values of C_N, whereas unfolded proteins will exist in conformational states for which C_N is small and R_g is large. Similarly in describing the formation of localized structures, such as hydrogen-bonded helical turns or cross-strand hydrogen bonds in β-sheets, the separation of the hydrogen-bonding pair, i.e., the N—H···O═C distance, is a reasonable progress coordinate to study the transitions between formed and unformed hydrogen-bonding partners.

Biased or umbrella sampling free energy simulations are relatively mature, having been a focal point in computational statistical mechanics for more than 25 years,[30] and their development and application to protein and peptide folding by our group over the past 15 years are well-documented.[26,28,31−40] What is noteworthy in the context of this Account is that these methods yield free energy curves (or surfaces in higher dimensions) that provide thermodynamic information on populations of conformational states, mechanistic interpretations of folding "pathways", and approximate kinetic time scales (through the application of transition-state theory or related kinetic models) for folding processes from detailed atomic-level descriptions of both peptide and solvent. Below we illustrate this by first describing results from work examining the competition between hydrogen-bonding interactions occurring in secondary structure in peptides and with water in aqueous solution. We progress to more complex processes in proteins.

Secondary Structure Stability and Kinetics

Experimental studies during the 1980s were motivated by the suggestion that the formation of secondary structural elements may play an important role in determining protein folding pathways and an interest in elucidating the interactions that stabilize such structures. These studies aimed to isolate the sequence of events that directed the folding of complex polypeptides comprising protein molecules and to separate the formation of nativelike secondary structure in solution from the coalescence of nascent secondary structural elements into folded proteins. Many of these studies focused on isolated peptide fragments in solution to assess their inherent "propensity" to form specific elements of localized structures: turns, helices, and β-hairpins.[41−43] Simulation studies, in turn, focused on peptide folding thermodynamics and kinetics to complement experiment, providing detailed theoretical models for secondary structure formation.

Efforts in our laboratory were directed toward addressing the question: "What is the worth of a backbone hydrogen-bond?" We developed simulation methods and models to examine the contribution of hydrogen bonding to the stability of secondary structural elements in reverse turns,[32] α-helices,[34,37,38] and model β-sheets.[31,35] These calculations represented some of the first attempts to quantify the balance of forces between solvation and direct peptide interactions and to provide an atomic level explanation of this competition. From conformational free energy pathways for formation of secondary structural elements, we were able to deduce mechanistic and time scale information for the elementary steps involved in structure formation. In Table 1 the thermodynamic and kinetic consequences of these findings are tabulated. The outcome can be summarized by noting (a) isolated hydrogen bonds, or those in exposed elements of secondary structure such as turns and helices, provide at best marginal stability ($\sim k_B T$ at 298 K) and form and dissolve in a facile manner on time scales of tens of picoseconds,[33,34,38] (b) 3_{10}-like helical interactions, often with solvent-mediated hydrogen bonding, occur mechanistically on the pathway to helix formation,[34,44] (c) the propensity for helical hydrogen bonds to form depends on the position at which they form, with the C-terminus providing less stability for nascent hydrogen bonds than the N-terminus because of the degree that solvent can penetrate and interact with the potential hydrogen bonding groups,[38] and (d) contrasting hydrogen bonding in exposed helices and turns, hydrogen bonds in modest model antiparallel β-sheets are significantly stabilized due to the ability of adjacent β-branched chemical substituents

to shield hydrogen bonds from competition with solvent.[31,35] This suggests that the mis-pairing of β-sheet interactions may be a source of kinetic traps, influencing the time scale of protein folding (or possibly misdirecting the outcome).[33,35]

Emerging from this early work were models for the formation of longer helices, built upon the calculated barriers and equilibrium constants, to examine the specific role of 3_{10}-type helical structures in helix formation[45,46] as well as the time scale for transitions between helical and coil states in alanine-based helices.[47] The former model utilized information on the relative stability of 3_{10}-versus α-helical interactions in alanine helices and a theory for the statistical mechanics of helix formation in the spirit of the Zimm-Bragg model.[48] Intermediate formation of this helical type was found to dominate in short helices under circumstances that favored 3_{10}-hydrogen-bonded interactions.[46] The latter (kinetic) theory utilized the stability for formation of sequential hydrogen bonds in helical peptides and the calculated barriers associated with this process in a sequential "zipper"-like model of helix formation kinetics.[47] The time scale for helix folding/unfolding transitions in this model was near 100 ns for moderate helices (15−20 residues), in coincidence with earliest fast spectroscopic measurements of helix formation.[49]

Folding Landscapes—Analysis of Global Features of Protein Folding Free Energy Surfaces

Results from the calculations described above illustrate how atomic insights for thermodynamic, kinetic, and mechanistic questions in secondary structure formation can be built-up from an exploration of the fundamental events that control this process. In each of these cases, simple one-dimensional progress coordinates could be identified and calculations to map the free energy change along these coordinates were carried out by using umbrella sampling techniques. While the success of such calculations was providing insights into the formation of localized structure in nascent folding chains (or in isolated peptide fragments), an emerging challenge to the computational community, growing from the theory of protein folding landscapes, was to address global aspects of folding free energy landscapes. This challenge required the development of techniques to extend biased sampling methods to large and complex systems and the use of these techniques to extract information relevant to the more statistical description of folding as suggested by the landscape theory. Of some particular early interest was the balance of energy (or more correctly solvent renormalized free energy) and conformational entropy in sculpting the folding funnel believed to dictate the nature of folding in small single-domain proteins. Critical issues which delineated the pathway description of the "old view" from the statistical nature of the "new view"[50] were the distribution of conformational states sampled by a folding protein as it approached its native conformational basin and the nature of interactions near the protein folding "transition state".

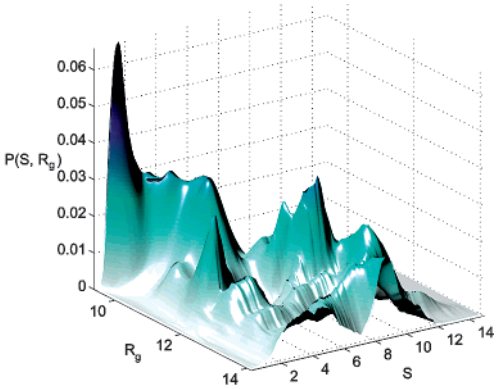

FIGURE 1. Distribution of protein conformations sampled during the folding of a three helical bundle protein, fragment B of staphylococcal protein A. The probability distribution, $P(S, R_g)$ along the folding coordinate R_g (in Å) and as a function of the conformational dissimilarity of neighboring protein conformations, S, illustrates the many-pathway nature of the folding funnel prior to the folding transition state at $R_g \sim 10.8$ Å.

Our efforts combined the approaches being used to explore protein and peptide unfolding[9,24,25,29] with the free energy simulation methods just described for simple peptide systems. Our approach utilizes the following components: sampling different regions of conformational space between the folded and unfolded states using methods of high-temperature molecular dynamics or conformational "pulling";[26,51] assessing and partitioning the space over which we sampled by clustering;[26,39,51,52] employing newly developed umbrella potentials in more collective folding coordinates, such as the radius of gyration, R_g,[26] or the number of native contacts, C_N,[39,53,54] to affect sampling in the partitioned regions of conformational space. The sampling from each of these regions is then combined en total via weighted histogram techniques to yield thermodynamic properties, including the folding free energy landscape, as a function of the progress coordinates for folding and under the thermodynamic conditions of the sampling simulations (usually 298 K or near the folding transition temperature).[22]

Folding of Small Helical Proteins Proceeds on a Smooth Funneled Landscape. In Figure 1 we illustrate one of our early calculations on the folding of a small three-helical bundle protein fragment (fragment B of staphylococcal protein A). We examine the distribution of distinct conformations ($P(S, R_g)$) sampled during the folding process. Progress in folding is measured by the value of the radius of gyration and the second dimension, S, indicates the degree of dissimilarity of conformations sampled as the progress coordinate changes.[26,51] We note that the free energy projected onto R_g for this system shows a barrier around $R_g = 10.8$ Å.[26] This three-dimensional surface illustrates that prior to the transition state for folding the protein samples a very diverse set of conformations almost uniformly; as folding progresses through the transition state, the conformational distribution narrows significantly. The suggestion following from this observation is that for protein A there are very many

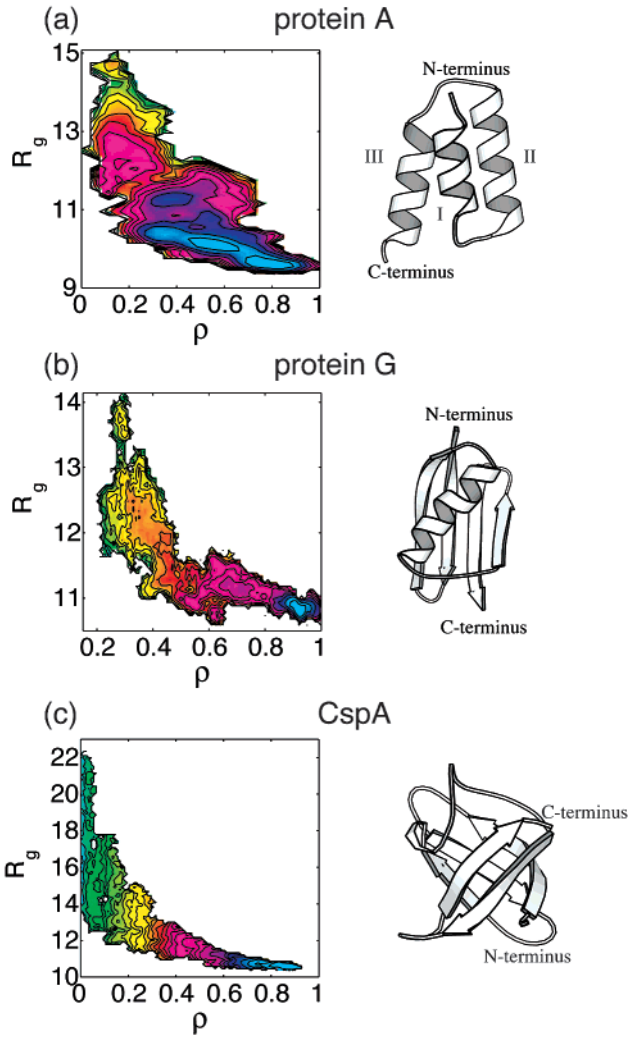

FIGURE 2. Folding free energy surfaces as a function of the fraction of native interactions (ρ) and the radius of gyration (R_g, in Å) for (a) fragment B of staphylococcal protein A (298 K), (b) segment B1 of streptococcal protein G (298 K), and (c) cold shock protein A from *E. coli* (298 K). Free energy contoured every 0.5 kcal/mol for (a) and (b) and every 2 kcal/mol for (c).

"pathways" leading to the folding transition state. After passing beyond this region, however, the number of entryways to the native basin is drastically reduced. This picture of many pathways funneling into a few is consistent with the smooth funneled landscape anticipated for helical proteins of approximately the same size.[55]

Topology Is a Key Determinant in Folding Mechanism and Kinetics. Using our approach, the folding free energy landscapes for more than five proteins (or sets of folding conditions) have been explored with atomically detailed models of protein and solvent. In Figure 2 we illustrate three representative free energy landscapes computed for proteins of overall structure that varies from all helical (fragment B of staphylococcal protein A),[26,53] to mixed α/β (fragment B1 of streptococcal protein G),[39,54] to all β-barrel (cold shock protein A from *E. coli*).[22,56] The free energy surfaces are displayed as contour maps, which indicate how the free energy varies with changes in the compactness (given by R_g) and the fraction of native interactions (denoted as ρ, the ratio of native contacts present, C_N, relative to those occupied under fully native conditions, C_N^0) changes.

Apparent in the comparison of these three representative surfaces is that their "shapes" differ from one another in the $R_g-\rho$ dimensions. Particularly, we note that for the helical protein, the folding landscape is more diagonal, providing the mechanistic interpretation that local structure (helix) formation and native tertiary packing occur commensurately. For the α/β- and all $\beta-$proteins the surfaces are "L-shaped", suggesting a nonspecific collapse precedes formation of the final native interactions. These observations suggest the possibility of kinetic differences between the folding of proteins adopting native states comprised of helical structures and those dominated by longer-range interactions, with the latter being potentially slower because of the necessity of "escaping" from compact, non-native-like traps. These insights are echoed in the findings of Baker and colleagues who illustrate that the folding kinetics and positioning of the folding transition state depend on the extent of nonlocal interactions present in the native structure's topology.[57]

Atomic Interpretation of Folding from Molecular Simulations

The results gleaned from the detailed calculations of folding free energy surfaces above reflect the global aspects of folding and grosser features of the proteins that control them. By virtue of the exquisite detail that exists in these calculations, questions may also be posed on more atomic levels. In what follows, we provide a few examples to illustrate this point, as well as to demonstrate the interplay between atomically detailed models and more minimalist representations in elucidating the underlying physical origins of folding processes not immediately apparent from fully atomic calculations.

Multiple Paths for Helix Formation and Native Packing in Fragment B of Staphylococcal Protein A. In our discussions of the global folding landscape for the small helical protein A fragment we noted that a diverse, funneled set of "pathways" to the native state exist prior to crossing the transition region for folding (see Figure 1). From the free energy simulations carried out to construct the surfaces already illustrated, we may explore the emergence of populations of localized structure throughout the folding process. Thus, we examined the free energy surface for protein A projected onto alternative coordinates, including the fraction of native hydrogen bonds and specific hydrogen bonds in each helix.[53] Our observations are consistent with early formation of structure around the helix I−helix II interface (a proline-containing turn with flanking hydrophobic residues) and later formation of the helix II−helix III interface, with concomitant rearrangement of the helix I−helix II interface.[58−60] We also find that there are distinct populations of helix (at least two) that exist prior to the folding transition state, suggesting at least two major pathways proceeding to the native basin. As illustrated in Figure 3,

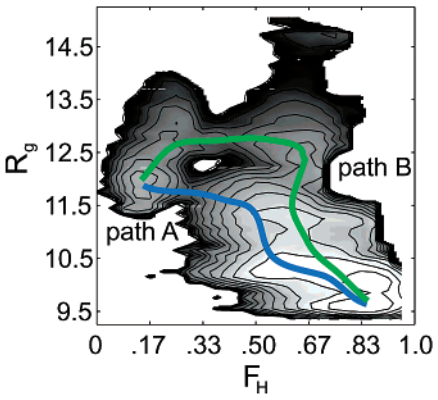

FIGURE 3. Folding free energy surface for fragment B of staphylococcal protein A (298 K) as a function of (R_g, in Å) and F_H (the fraction of native helical hydrogen bonds), illustrating alternate pathways for folding involving differing amounts of helix. Contours every 0.5 kcal/mol.

where the folding free energy surface is projected onto coordinates describing the fraction of native helical hydrogen bonds (F_H) and the radius of gyration, R_g, these pathways involve differing amounts of helix formation. In one case (labeled path A) more-or-less concomitant collapse and helical structure formation occur as folding progresses. In the other (path B), nearly all of the helix content is present before significant collapse occurs and folding progresses in a more collision−diffusion manner.[4,61]

Symmetry Breaking in Folding of Mixed α/β Proteins. In protein structures that display a topology with a high degree of symmetry, such as the mixed α/β structure of segment B1 of streptococcal protein G (GB1) discussed above (see Figure 2) or its "relative" segment B1 of peptostreptococcal protein L, one might anticipate a symmetric transition state for folding in which approximately equal amounts of structure are formed near the N- and C-terminus. This, however, is not necessarily the case because symmetry can be broken by more atomic level details of the structure and sequence, which can polarize the transition state through preferential formation of localized structure.

Kinetic folding studies of GB1 have been characterized as two-state, although experiments indicate that some structure forms early in folding. Rapid mixing measurements of fluorescence quenching suggest early formation of a relatively collapsed metastable state that buries a tryptophan residue in "rapid" equilibrium with the unfolded state preceding the rate-limiting barrier crossing to the native state[62] and support earlier work based on quenched-flow H/D exchange protection and characterization of a denaturant-induced unfolded state. These studies are interpreted to indicate that early collapse involves formation of structure around the C-terminal β-hairpin, separating β-strands 3 and 4, and in the middle and near the N-terminus of the central α-helix.[63,64] Our calculations support this finding and provide more detailed models of the nature of the interactions responsible for the polarized transition state for GB1.[39,54] For the protein L structure, kinetic analysis of interactions present in the folding transition state from experiment[65,66] suggests that structure is formed first in the N-terminal hairpin as well as in the helix.

To explore this symmetry breaking and to further focus our detailed molecular free energy simulations, we constructed minimalist Gō-type models for these two proteins and examined the similarities and differences in their folding.[23] The Gō-type models were constructed based on the Cα positions of atoms from the native folded protein structures and used sequence specific interactions between Cα atoms taken from a knowledge-based interaction scheme, "hydrogen-bonding" interactions between Cα atoms to represent secondary structure, and sequence specific pseudo-torsional potentials.[23] Because these models better capture the sequence-specific details of interresidue interactions than those used in earlier studies, we observe the differential polarization of the folding transition state as found experimentally. Some of our findings are presented in Table 2, where we provide an energetic and (chain) entropic breakdown of the relative stability of spatially localized interacting segments constructed via configurational averaging at the folding transition temperature.[23] The results show that sequence-dependent interactions between residues in the N-terminal hairpins are favored in both protein L and GB1. However, the polarization of the GB1-folding transition state to favor the C-terminal hairpin arises from a compensating (and dominate) influence in the chain entropy difference of the hairpins that does not exist in protein L. Thus, these two proteins are guided to fold via polarized, but different, transition states through details of the protein sequence.

Further support for this model derives from a recent study in which the N-terminal hairpin of protein G was redesigned to increase its stability. This redesign led to a repolarization of the transition state and folding of the N-terminal hairpin ahead of the C-terminal hairpin.[67] Within the context of the models just discussed, we modified the native Cα-based interaction strengths of GB1 to reflect the sequence differences of the two redesigned

Table 2. Thermodynamics Responsible for the Observed Symmetry Breaking in the Mechanism of Folding of Segment B1 of Streptococcal Protein G and Segment B1 of Peptostreptococcal Protein L

	protein G	protein L
ΔE_{fold} (N-terminal hairpin)	−44.8 kcal/mol	−55.1 kcal/mol
ΔE_{fold} (C-terminal hairpin)	−36.2 kcal/mol	−47.3 kcal/mol
$\Delta\Delta E_{fold}$ (N-term − C-term)	−8.5 kcal/mol	−7.8 kcal/mol
length of N-terminal hairpin	20 residues	21 residues
length of C-terminal hairpin	14 residues	21 residues
$\Delta\Delta S_{fold}$ (N-term − C-term)	−0.032 kcal/mol·K	0 kcal/mol·K
$-T_f\Delta\Delta S_{fold}$ (N-term − C-term)	11.4 kcal/mol	0 kcal/mol
$\Delta\Delta G_{fold}$ (N-term − C-term)	2.9 kcal/mol	−7.8 kcal/mol

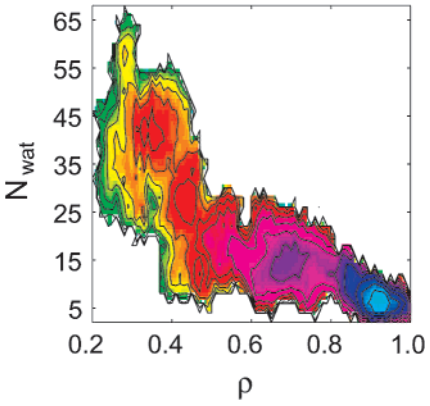

FIGURE 4. Free energy surface for core water penetration at 298 K for segment B1 of streptococcal protein G. The surface is computed as a function of the fraction of native contacts (ρ) and the number of core water molecules (N_{wat}). The free energy contours are shown every 0.5 kcal/mol.

variants. This resulted in an additional stabilization of −4.1 and −4.9 kcal/mol in the N-terminal hairpin for the two modified sequences, respectively. These values are larger than the free-energy difference between the N- and C-terminal hairpins, which gave rise to the original C-terminal polarization, suggesting (as observed in recent experiments)[67] that the N-terminal hairpin is likely to fold first in both mutants. This suggests that the redesign primarily stabilized the N-terminal hairpin microstates to overcome the entropy cost associated with the longer peptide fragment comprising this hairpin.

The Role of Solvent in Lubricating Conformational Searches during Folding. What is the bane of most molecular simulations, the inclusion of explicit solvent representations, is also their boon, by providing details not present in more simplified models and generally not easily "visible" through experimental probe. Water is a critical component in the free energy balance that yields the exquisite three-dimensional arrangement of functioning native proteins. What is also emerging from molecular simulations of protein folding free energy landscapes is that it too may play a crucial role in the kinetic and mechanistic aspects of protein folding. As was presented above, water has been identified as an important player in the formation of helical structure and possibly in the mechanism of α-helix formation/dissolution.

To probe whether water too might play a critical, explicit role in protein folding, we examined the free energy surfaces for the folding of two proteins, GB1[39,54] and the all-β protein src-SH3,[22,68] using a unique set of reaction coordinates that elucidated the role of protein interior water molecules in the late stages of folding. By projecting the free energy surface onto coordinates that describe the loss of core water molecules (N_{wat}) and the fraction of native interactions (ρ), we discovered that a significant number of water molecules existed in the protein core late in folding and that these water-impregnated cores sometimes existed as metastable states on this free energy landscape. In particular, for the GB1 surface displayed in Figure 4, the solvated core is stabilized by an excess of four to six buried core water molecules at values of ρ corresponding to a metastable minimum on the $R_g-\rho$ landscape (see the free energy surface in Figure 2b). Detailed examination of interactions between core water and the polypeptide chain reveal that they are generally bridging nascent β-sheet hydrogen bonds, often between misregistered donor/acceptor partners. From these observations and similar ones for the src-SH3 domain,[22,68] we have postulated that water is serving as a "lubricant", facilitating the "search" for correct hydrogen bonding partners or side chain pairing in the near-native configurations of the protein. Furthermore, we suggest that a free energy barrier to folding can exist from the final expulsion of these water molecules. This novel proposal points to the potential strength of complementing both experiment and theory by simulation studies. While this suggestion awaits confirmation for GB1 and src-SH3, the complex folding kinetics associated with the evolution of water-mediated hydrogen bonding interactions may be the origin of differing relaxation times for trytophan quenching and transient IR absorbance signals observed in fast laser T-jump experiments performed on the all-β protein cold shock protein A.[69]

Conclusions

In this Account we have outlined how free energy simulations are being used to illuminate the complexities of peptide and protein folding. Models utilizing complete representations of the polypeptide chain and solvent can be combined with simulation methods employing the techniques of biased sampling to yield the free energy landscapes fundamental to understanding the energetics, mechanism, and kinetics of folding. Also extremely useful are minimalist models that capture the essence of molecular topology and sequential interactions and allow complete descriptions of folding within the context of their simplified format. These models augment and complement the fully atomic models, and both provide a critical bridge between analytical theories of protein and peptide folding and laboratory experiments.

Molecular simulations using these models and methods have been employed to quantify the balance between peptide−peptide and peptide−solvent interactions in hydrogen bond formation in peptides, providing theoretical insight into the mechanism and time scales of the coil to helix transition. The general framework of protein folding landscape theory makes qualitative predictions regarding the nature of protein conformations approaching the folding transition state; free energy simulations of specific proteins provide direct confirmatory evidence for these ideas. In bridging the gap between experimental observation and molecular interpretation of these data, computational analyses of folding free energy landscapes (projected onto a variety of coordinates) have yielded detailed (and often quantitative) pictures of the order of appearance of folding "microstates" populated between the unfolded and folded manifold of states and the specific and potentially ubiquitous role of water in the late stages of folding.

Molecular models and free energy simulation methods are poised to assist in questions of design and direction in protein engineering and should prove essential in addressing key questions of atomic interactions in the formation of misfolded proteins and their attendant manifestations.[70,71] However, challenges remain in the use of such methods for more direct exploration of protein folding kinetics and folding of multidomain or multisubunit proteins.

I am indebted to a talented group of co-workers who over the years have pioneered the methods and created the results described in this work, much of the work discussed in this Account has appeared in referenced publications they coauthored. Financial support from the National Institutes of Health (GM48807, GM57513, and RR12255) is appreciated.

References

(1) Principles that govern the folding of protein chains. Anfinsen, C. B. *Science* **1973**, *181*, 223−230.
(2) Are there pathways for protein folding.Levinthal, C. *J. Chem. Phys.* **1968**, *65*, 44−45.
(3) Protein folding: evaluation of some simple rules for the assembly of helices into tertiary structures with myoglobin as an example. Cohen, F. E.; Richmond, T. J.; Richard, F. M. *J. Mol. Biol.* **1979**, *132*, 275−288.
(4) Protein folding dynamics: the diffusion-collision model and experimental data. Karplus, M.; Weaver, D. L. *Protein Sci.* **1994**, *3*, 650−668.
(5) Intermediates in the folding reactions of small proteins. Kim, P. S.; Baldwin, R. L. *Annu. Rev. Biochem.* **1990**, *59*, 631−660.
(6) Funnels, pathways and the energy landscape of protein folding: a synthesis. *Proteins: Struct. Funct. Genet.* Bryngelson, J. D.; Onuchic, J. N.; Wolynes, P. G. **1995**, *21*, 167−195.
(7) Theory of protein folding: the energy landscape perspective. Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545−600.
(8) Theoretical studies of protein folding thermodynamics and kinetics. Shakhnovich, E. I. *Curr. Opin. Struct. Biol.* **1997**, *7*, 29−40.
(9) Protein folding: a perspective from theory and experiment. Dobson, C. M.; Sali, A.; Karplus, M. *Angew. Chem., Int. Ed. Engl.* **1998**, *37*, 868−893.
(10) Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. Chan, H. S.; Dill, K. A. *Proteins: Struct. Funct. Genet.* **1998**, *30*, 2−33.
(11) Deciphering the time scales and mechanisms of protein folding using minimal off-lattice models. Thirumalai, D.; Klimov, D. *Curr. Opin. Struct. Biol.* **1999**, *9*, 197−207.
(12) Associative memory Hamiltonians for structure prediction without homology: alpha-helical proteins. Hardin, C.; Eastwood, M. P.; Luthey-Schulten, Z.; Wolynes, P. G. *Proc. Natl. Acad. Sci., USA.* **2000**, *97*, 14235−14240.
(13) McCammon, J. A.; Harvey, S. *Dynamics of proteins and nucleic acids*; Cambridge University Press: New York, 1987.
(14) Brooks, C. L., III.; Karplus, M.; Pettitt, B. M. *Proteins: a theoretical perspective of dynamics, structure and thermodynamics*; John Wiley and Sons: New York, 1988; Vol. LXXI.
(15) Protein folding kinetics from the perspective of simple models. Chan, H. S.; Dill, K. A. *Proteins: Struct. Funct. Genet.* **1997**, *8*, 2−33.
(16) Comparison of lattice Monte Carlo dynamics and Brownian dynamics folding pathways of α-helical hairpins. Rey, A.; Skolnick, J. *Chem. Phys.* **1991**, *158*, 199−219.
(17) The nature of folded states of globular proteins. Honeycutt, J. N.; Thirumalai, D. *Biopolymers* **1992**, *32*, 695−709.
(18) Thermodynamics of protein folding: a statistical mechanical study of a small beta protein. Guo, Z.; Brooks, C. L., III *Biopolymers* **1997**, *42*, 745−757.
(19) Exploring the space of protein folding Hamiltonians: The balance of forces in a minimalist beta-barrel model. Shea, J.-E.; Nochomovitz, Y. D.; Guo, Z.; Brooks, C. L., III *J. Chem. Phys.* **1998**, *109*, 2895−2903.
(20) Exploring the origins of topological frustration: Design of a minimally frustrated model of fragment B of protein A. Shea, J.-E.; Onuchic, J. N.; Brooks, C. L., III *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 12512−12517.
(21) Energetic frustration and the nature of the transition state in protein folding. Shea, J.-E.; Onuchic, J. N.; Brooks, C. L., III *J. Chem. Phys.* **2000**, *113*, 7663−7671.
(22) From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. Shea, J. E.; Brooks, C. L., III *Annu. Rev. Phys. Chem.* **2001**, *52*, 499−535.
(23) The origins of symmetry breaking in the folding of proteins with analogous topologies. Karanicolas, J.; Brooks, C. L., III *Protein Sci.*, submitted for publication.
(24) . Molecular simulations of peptide and protein unfolding: In quest of a molten globule Brooks, C. L., III *Curr. Opin. Struct. Biol.* **1993**, *3*, 92−98.
(25) Protein folding−unfolding dynamics. Daggett, V.; Levitt, M. *Curr. Opin. Struct. Biol.* **1994**, *4*, 291−295.
(26) . First principles calculation of the folding free energy of a three-helix bundle protein. Boczko, E. M.; Brooks, C. L., III *Science* **1995**, *269*, 393−396.
(27) Long time scale simulations. Daggett, V. *Curr. Opin. Struct. Biol.* **2000**, *10*, 160−164.
(28) Brooks, C. L., III. In *Int. J. Quantum Chem., Quantum Biol. Symp.* **1988,** *15,* 221−234.
(29) Simulations of protein folding and unfolding. Brooks, C. L., III *Curr. Opin. Struct. Biol.* **1998**, *8*, 222−226.
(30) Valleau, J. P.; Torrie, G. M. In *Mod. Theor. Chem.* Berne, B. J., Ed., 1977; Vol. 5; pp 169−194.
(31) Sneddon, S. F.; Tobias, D. J.; Brooks, C. L., III Thermodynamics of amide hydrogen bond formation in polar and apolar solvents. *J. Mol. Biol.* **1989**, *209*, 817−820.
(32) Reverse turns in blocked dipeptides are intrinsically unstable in water. Tobias, D. J.; Sneddon, S. F.; Brooks, C. L., III *J. Mol. Biol.* **1990**, *216*, 783−96.
(33) Tobias, D. J.; Sneddon, S. F.; Brooks, C. L., III. In *Adv. Biomed. Simul.*; AIP: Obernai, France, 1991; Vol. 239, pp 174−199.
(34) Thermodynamics and mechanism of a-helix initiation in alanine and valine peptides. Tobias, D. J.; Brooks, C. L., III *Biochemistry* **1991**, *30*, 6059−6070.
(35) Stability of a model β-sheet in water. Tobias, D. J.; Sneddon, S. F.; Brooks, C. L., III *J. Mol. Biol.* **1992**, *227*, 1244−52.
(36) Constant-temperature free energy surfaces for physical and chemical processes. Boczko, E. M.; Brooks, C. L., III *J. Phys. Chem.* **1993**, *97*, 7, 4509−13.
(37) Promotion of helix formation in peptides dissolved in alcohol and water-alcohol mixtures. Brooks, C. L., III; Nilsson, L. *J. Am. Chem. Soc.* **1993**, *115*, 11034−11035.
(38) A microscopic view of helix propagation: N and C-terminal helix growth in alanine helixes. Young, W. S.; Brooks, C. L., III *J. Mol. Biol.* **1996**, *259*, 560−572.
(39) Calculations on folding of segment B1 of streptococcal protein G. Sheinerman, F. B.; Brooks, C. L., III *J. Mol. Biol.* **1998**, *278*, 439−456.
(40) Folding free energy surface of a three-stranded β-sheet protein. Bursulaya, B. D.; Brooks, C. L., III *J. Am. Chem. Soc.* **1999**, *121*, 9947−9951.
(41) Conformation of Peptide Fragments in Aqueous Solution: Implications for Initiation of Protein Folding. Wright, P. W.; Dyson, H. J.; Lerner, R. A. *Biochemistry* **1988**, *27*, 7167−7175.
(42) Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding. I. Myohemerythrin. Dyson, H. J.; Merutka, G.; Waltho, J. P.; Lerner, R. A.; Wright, P. E. *J. Mol. Biol.* **1992**, *226*, 795−817.
(43) Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding. II. Plastocyanin. Dyson, H. J.; Sayre, J. R.; Merutka, G.; Shin, H. C.; Lerner, R. A.; Wright, P. E. *J. Mol. Biol.* **1992**, *226*, 819−835.
(44) Water-inserted a-helical segments implicate reverse turns as folding intermediates. Sundaralingam, M.; Sekharudu, Y. C. *Science* **1989**, *244*, 1333−1337.
(45) Views of helical peptides: a proposal for the position of 3(10)-helix along the thermodynamic folding pathway. Millhauser, G. L. *Biochemistry* **1995**, *34*, 3873−3877.
(46) 3_{10} Helices in peptides and proteins as studied by modified Zimm-Bragg theory. Sheinerman, F. B.; Brooks, C. L., III *J. Am. Chem. Soc.* **1995**, *117*, 10 098−10 103.
(47) Helix-coil kinetics: folding time scales for helical peptides from a sequential kinetic model. Brooks, C. L., III *J. Phys. Chem.* **1996**, *100*, 2546−2549.
(48) Theory of the phase transition between helix and random coil in polypeptide chains. Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526−535.
(49) Fast events in protein folding: helix melting and formation in a small peptide. Williams, S.; Causgrove, T. P.; Gilmanshin, R.; Fang, K. S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B. *Biochemistry* **1996**, *35*, 691−697.

(50) The nature of protein folding pathways: the classical versus the new view. Baldwin, R. L. *J. Biomol. NMR* **1995**, *5*, 103−109.
(51) Boczko, E. M. Ph.D. Carnegie Mellon University, 1996.
(52) Statistical Clustering Techniques for Long Molecular Dynamics Trajectories: Analysis of 2.2 ns Trajectories of YPGDV. Karpen, M. E.; Tobias, D. J.; Brooks, C. L., III *Biochemistry* **1993**, *32*, 412−420.
(53) Exploring the folding free energy surface of a three-helix bundle protein. Guo, Z.; Brooks, C. L., III.; Boczko, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10 161−10 166.
(54) Molecular picture of folding of a small a/b protein. Sheinerman, F. B.; Brooks, C. L., III *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1562−1567.
(55) Protein folding funnels: the nature of the transition state ensemble. Onuchic, J. N.; Socci, N. D.; Luthey-Schulten, Z.; Wolynes, P. G. *Fold., Des.* **1996**, *1*, 441−450.
(56) Brooks, C. L., III, unpublished results.
(57) Contact order, transition state placement and the refolding rates of single domain proteins. Plaxco, K. W.; Simons, K. T.; Baker, D. *J. Mol. Biol.* **1998**, *277*, 985−994.
(58) Folding thermodynamics of a model three-helix-bundle protein. Zhou, Y.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 14429−14432.
(59) Staphylococcal protein A: unfolding pathways, unfolded states, and differences between the B and E domains. Alonso, D. O.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 133−138.
(60) Monte Carlo simulations of protein folding. II. Application to protein A, ROP, and crambin. Kolinski, A.; Skolnick, J. *Proteins* **1994**, *18*, 353−366.
(61) Preorganized secondary structure as an important determinant of fast protein folding. Myers, J. K.; Oas, T. G. *Nat. Struct. Biol.* **2001**, *8*, 552−558.
(62) An early intermediate in the folding reaction of the B1 domain of protein G contains a nativelike core. Park, S. H.; O'Neil, K. T.; Roder, H. *Biochemistry* **1997**, *36*, 14277−14283.
(63) Fast folding of a prototypic polypeptide: the immunoglobulin binding domain of streptococcal protein G. Kuszewski, J.; Clore, G. M.; Gronenborn, A. M. *Protein Sci.* **1994**, *3*, 1945−1952.
(64) Structural and dynamic characterization of the urea denatured state of the immunoglobulin binding domain of streptococcal protein G by multidimensional heteronuclear NMR spectroscopy. Frank, M. K.; Clore, G. M.; Gronenborn, A. M. *Protein Sci.* **1995**, *4*, 2605−2615.
(65) Contrasting roles for symmetrically disposed beta-turns in the folding of a small protein. Gu, H.; Kim, D.; Baker, D. *J. Mol. Biol.* **1997**, *274*, 588−596.
(66) A breakdown of symmetry in the folding transition state of protein L. Kim, D. E.; Fisher, C.; Baker, D. *J. Mol. Biol.* **2000**, *298*, 971−984.
(67) Computer-based redesign of a protein folding pathway. Nauli, S.; Kuhlman, B.; Baker, D. *Nat. Struct. Biol.* **2001**, *8*, 602−605.
(68) . The folding free energy landscape of thesrc-SH3 domain studied by molecular dynamics in explicit solvent. Shea, J.-E.; Onuchic, J. N.; Brooks, C. L., III *J. Mol. Biol.*, submitted for publication.
(69) Protein folding and unfolding on a complex energy landscape. Leeson, D. T.; Gai, F.; Rodriguez, H. M.; Gregoret, L. M.; Dyer, R. B. *Proc. Natl. Acad. Sci., U.S.A.* **2000**, *97*, 2527−2532.
(70) Mutational analysis of the propensity for amyloid formation by a globular protein. Chiti, F.; Taddei, N.; Bucciantini, M.; White, P.; Ramponi, G.; Dobson, C. M. *EMBO J.* **2000**, *19*, 1441−1449.
(71) Mechanisms of amyloidogenesis. Kelly, J. W. *Nat. Struct. Biol.* **2000**, *7*, 824−826.

AR0100172