PROTEINS: Structure, Function, and Genetics 46:000–000 (2002)
DOI 10.1002/prot.00000

AQ: 1

Toward Minimalist Models of Larger Proteins: a Ubiquitin-like Protein

Jon M. Sorenson[1,2] **and Teresa Head-Gordon**[3*]
[1]*Department of Chemistry, University of California, Berkeley, California*
[2]*Physical Biosciences and Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California*
[3]*Department of Bioengineering, University of California, Berkeley, California*

ABSTRACT **Our recently developed off-lattice bead model capable of simulating protein structures with mixed α/β content has been extended to model the folding of a ubiquitin-like protein and provides a means for examining the more complex kinetics involved in the folding of larger proteins. Using trajectories generated from constant-temperature Langevin dynamics simulations and sampling with the multiple multi-histogram method over five-order parameters, we are able to characterize the free energy landscape for folding and find evidence for folding through compact intermediates. Our model reproduces the observation that the C-terminus loop structure in ubiquitin is the last to fold in the folding process and most likely plays a spectator role in the folding kinetics. The possibility of a productive metastable intermediate along the folding pathway consisting of collapsed states with no secondary structure, and of intermediates or transition structures involving secondary structural elements occurring early in the sequence, is also supported by our model. The kinetics of folding remain multi-exponential below the folding temperature, with glass-like kinetics appearing at $T/T_f \sim$ 0.86. This new physicochemical model, designed to be predictive, helps validate the value of modeling protein folding at this level of detail for genomic-scale studies, and motivates further studies of other protein topologies and the impact of more complex energy functions, such as the addition of solvation forces. Proteins 2002;46:000–000.**
AQ: 2
© 2002 Wiley-Liss, Inc.

Key words: off-lattice models; ubiquitin; α/β proteins; protein folding; multiple histogram method; multi-state kinetics

INTRODUCTION

Experimental and theoretical studies on the folding of small globular proteins have established they fold by a cooperative two-state mechanism, finding evidence for the presence of an unfolded ensemble and a native-state ensemble, with little population in between these states.[1–8] This is accounted for in the theoretical framework of energy landscape theory as evidence for the principle of minimal frustration, and recent work has further advanced the idea that not only energetic, but also topological, frustration is minimized.[1,9–14]
AQ: 3

However, as we contemplate the folding of larger proteins, this picture could require modification. With larger proteins, we might expect that the need to form collapsed or partially collapsed intermediates to reduce the conformational search space becomes more important. Characterizing the nature of these intermediates, both obligatory and nonobligatory, constitutes an important part of our understanding of the folding pathways for larger and topologically more complex proteins. An investigation into the question as to whether the folding of larger proteins is also governed by topological frustration (using a Gō model that all but eliminates any energetic frustration) showed that both transition state ensembles and intermediates of larger proteins are determined primarily by topological effects,[15] in good qualitative agreement with the experiment.

Given the productive interplay between analytical theory, experiment, and protein folding simulations,[1–8] our recent work seeks to design protein models that establish stronger quantitative connections to experiments by designing sequences and structures that are faithful to the real sequences and structures, and Hamiltonians that are predictive and interpretable with respect to known physical interactions.[16–19] This approach is complemetary to off-lattice minimalist models based on Gō interactions that are not typically capable of quantitative prediction without knowledge of the known and correct interactions.

An important advantage of minimalist protein models of any sort is the ability to examine the folding of larger proteins rather than those accessible by more detailed simulations. Although more realistic detail about a specific folding or unfolding trajectory can now be achieved with atomistic simulations in explicit solvent, the ability to collect statistics over many trajectories, sampled over varying conditions and various models, is still beyond current computational limits.[20–22] The advantages of minimalist models lie in their ability to rapidly collect meaningful statistics about folding pathways and kinetics, their ease of characterization with only a few order parameters,
AQ: 4

Grant sponsor: U.S. Department of Energy (LDRD); Grant number: DEAC-03-76SFOO098; Grant sponsor: National Science Foundation.

*Correspondence to: Theresa Head-Gordon, Department of Bioengineering, University of California, Berkeley, CA 94720. E-mail: tlhead-gordon@lbl.gov

Received 23 September 2000; Accepted 10 July 2001
Published online 00 Month 2002

© 2002 WILEY-LISS, INC.

and their concentration on the essential physics of the problem, which are useful if we are to draw general conclusions applicable to more than the specific protein of interest. At the same time we would like to retain enough chemical detail in our minimalist models to make contact with experimental observables for a given target protein that we choose to design.

In recent work, using an off-lattice bead model originally pioneered by Thirumalai and coworkers,[23–25] we developed a sequence design tool based on threading for successful development of a much improved minimalist model for small proteins with all β-sheet topologies.[16] We have extended the scope of the predictive minimalist model by successfully designing for the first time a mixed α/β fold with this model, whose kinetics and thermodynamics of folding made direct comparisons with experiments on (α/β) proteins L and G.[17–19,26,27] In this article, we probe the thermodynamics and kinetics of a larger protein than has been previously studied at this level by developing a folding model of a ubiquitin[28] protein, in which our sequence design method[16] plays an instrumental role. Using the multiple multidimensional histogram method,[29,30] we are able to project the free energy landscape onto various order parameters to show that the folding of our ubiquitin-like model follows a folding pathway with a collapsed non-native intermediate, with sufficient bias in the landscape to be able to direct the chain to the proper native-state structure. We analyze these results with particular attention to experimental comparison with the observed folding of ubiquitin.[31–36] We conclude with a summary of these results and a description of future directions with this level of modeling.

METHODS AND MODELS

Energy Function

The level of description of the protein chain and choice of parameters for the energy function have been described extensively in a previous publication.[17] The essential details are described in the present study. The protein chain is modeled as a chain of beads of three flavors: hydrophobic (B), hydrophilic (L), or neutral (N). Attraction between the hydrophobic beads provides the energetic driving force for the formation of a strong core; repulsion between the hydrophilic beads and other beads is used to balance the forces and bias the correct native fold; and the neutral beads serve as soft spheres with little repulsion and typically signal the turn regions in the sequence.

AQ: 5

The potential energy for the model is

$$E = \sum_{\text{angles}} \tfrac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \Big\{ A[1 + \cos\phi] + B[1 - \cos\phi] + C[1 + \cos 3\phi] + D\Big[1 + \cos\Big(\phi + \frac{\pi}{4}\Big)\Big]\Big\} + \sum_{i,j \geq i+3} 4\varepsilon_H S_1 \Big[\Big(\frac{\sigma}{r_{ij}}\Big)^{12} - S_2 \Big(\frac{\sigma}{r_{ij}}\Big)^6\Big] \quad (1)$$

Bond lengths are held rigid, and the bond angles are maintained by a harmonic potential with force constant $k_\theta = 20\varepsilon_H/(\text{rad})^2$ and equilibrium bond angle $\theta_0 = 105°$. ε_H is the basic energy unit, corresponding to the minimum in the attraction energy between two B beads. The nonlocal interactions are given by $S_1 = S_2 = 1$ for BB interactions, $S_1 = \frac{1}{3}$ and $S_2 = -1$ for LL and LB interactions, and $S_1 = 1$ and $S_2 = 0$ for all interactions involving N residues.

Unlike the previous models, the dihedral potential is designed to simulate helical, extended, or turn regions of the protein sequence. To define a particular protein model, the sequence of beads and the sequence of dihedral propensities must be specified. The dihedral potential itself is a combination of the potentials used in the previous all-β and all-α models.[24,25] The parameter choices corresponding to the three possible dihedral states are $A = 0$, $B = C = D = 1.2\varepsilon_H$ for helical, extended ($A = 0.9\varepsilon_H$, $B = D = 0$, $C = 1.2\varepsilon_H$), or turn ($A = B = D = 0$, $C = 1.2\varepsilon_H$). Like the original dihedral potentials, these potentials possess three minima: a global minimum corresponding to the proper state, and two degenerate local minima higher in energy.

The advantages and limitations of this description have been addressed previously.[17] The dihedral potentials serve as potentials of mean force meant to reproduce the intrinsic secondary structure propensity of certain sequences of amino acids. Such potentials are required, as important determinants of secondary structure in real proteins, such as backbone hydrogen bonding and side-chain steric restrictions, are "integrated" out at our level of description.

Simulation Methods

We use a simulated annealing protocol to find the global minimum for sequences in this model. Once a global minimum is found, constant-temperature Langevin simulations are carried out for characterizing the thermodynamics and kinetics of folding to the native state. The protocol for both the simulated annealing and Langevin simulations has been described in previous publications.[16,17] The simulations are performed in reduced units, with the units of mass m, length σ, energy ε_H, and k_B all set equal to one; temperature is in units of ε_H/k_BT. The unit of reduced time is $\tau = \sqrt{m\sigma^2/\varepsilon_H}$.

The free energy landscape of the model is characterized by sampling using the multiple multidimensional histogram method.[29,30] The use of this method to extract the density of states for protein folding systems has been well documented.[16,37–39] As noted with previous uses of this method with similar off-lattice models,[16,40] one must be careful to ensure equilibrium sampling of states at low temperatures significantly below the collapse temperature. In this work, we collected six-dimensional histograms over energy and five order parameters: radius of gyration (R_g), χ, $\chi_{\beta\beta}$, $\chi_{\alpha\beta}$, and $\chi_{\beta\alpha\beta}$. χ is the order parameter for folding to the native state[24]:

$$\chi = \frac{1}{M} \sum_{i,j \geq i+4}^{K} \theta(\varepsilon - |r_{ij} - r_{ij}^{\text{nat}}|) \quad (2)$$

where the double sum is over beads on the chain, r_{ij} and r_{ij}^{nat} are the distances between beads i and j in the state for

comparison and the native state, respectively, θ is the Heaviside step function, and $\varepsilon = 0.2$ accounts for small fluctuations away from the native state structure. M is a normalizing factor to ensure that $\chi = 1$ when the chain is identical to the native state and $\chi = 0$ when the chain is in a random coil state. The other three χ values are a specialization of χ to monitor formation of specific secondary structure elements. The sum over residues for $\chi_{\beta\beta}$ only involves beads in the first β-hairpin element; the sum for $\chi_{\alpha\beta}$ involves beads in the first helix and third β-strand; and the sum for $\chi_{\beta\alpha\beta}$ involves beads in the fourth β-strand, second α-helix, and fifth β-strand. Other order parameters monitoring secondary structure formation can be proposed; those chosen in this case were found to well characterize the range of structures being sampled in the folding landscape. (Concerning computer implementation, the collection of histogram data over six dimensions requires a modification of the traditional array structure used for one-dimensional or two-dimensional histograms. Such data structures rapidly exhaust system memory when the number of dimensions increases; we avoid this problem in our work by employing sparse data structures,[41] which grow only with the number of samples taken.)

The kinetics of the folding process can be characterized by calculating first-passage times. These times were calculated by taking high-temperature unfolded structures and recording the time that they first folded to the native state at a given temperature. First-passage times for chain collapse ($R_g^2 < 8.0\sigma^2$) were also collected to investigate questions of how closely chain collapse accompanies folding to the native state.[27]

Sequence Design

A combination of positive and negative design techniques were used to converge on an appropriate ubiquitin-like sequence for study. Initially proposed sequences were modifications of the successful protein L/G-like sequence examined in our previous work.[17] This sequence formed a natural starting point, since proteins L and G are members of the ubiquitin fold class,[42] and share a very strong backbone similarity with ubiquitin. Forming a ubiquitin structure from protein L involves inserting a β-strand and α-helix into the loop region connecting β-strands 3 and 4 in the original protein L/G structure. The sequence length was extended to 68 beads from the 56 beads of the previous model; this is comparable to the length of ubiquitin, which has a sequence of 76 amino acids.

These initial sequences did not produce a final model, as characterized by simulated annealing profiles and folding studies. An ideal model has a simulated annealing profile that robustly finds the native state in multiple simulated annealing runs, indicative of a smooth underlying energy landscape.[16] It should also be able to find the native state without becoming trapped in misfolded compact states at temperatures near the folding temperature. Our original sequences failed these tests.

To improve these sequences, cycles of negative design were performed as described previously.[16] The essence of our design method is similar to many in use,[43,44] but our design method uses a library of dissimilar misfolded structures generated from running multiple trajectories at a temperature near the collapse temperature. This library is then used to suggest sequence mutations, both in the bead sequence and in the dihedral sequence, which would be expected to further separate the energy of the native state basin of attraction away from the misfolded states, an important characteristic of a robust protein folding model.[45,46] The misfolded states sampled in this way are more characteristic of the actual traps and basins of attraction of significant misfolded states than if we were a priori to choose energetically low-lying minima as our putative misfolded states to design out. The detrimental traps in the folding landscape are those that compete kinetically with formation of the native state, not necessarily those states that are thermodynamically competitive but highly unlikely to be sampled during the folding process, due to a small basin of attraction or prohibitively high barrier to formation. AQ: 6

An interesting application of this design procedure was to find regions in the dihedral sequence that would benefit by mutation from sheet- or helix-forming residues to floppy residues. Unlike most design protocols, the motivation of this mutation was not to stabilize the native state or destabilize misfolds, but rather to stabilize the transition states between misfolds and the native state to facilitate unfolding of misfolded dihedrals. A mutation in this direction for our ubiquitin model was suggested by successful results we found in proposing similar mutations on our protein L/G model.[47] The final sequence contains one floppy dihedral (14) which serves this purpose.

RESULTS

The designed sequence used in the study reported here is shown in Table I. The dihedral sequence is three shorter than the bead sequence because there are three fewer dihedral angles than beads in a given chain. Table I also includes the sequence of the naturally occurring ubiquitin protein. If we assume reasonable definitions of residues considered to be hydrophobic or hydrophilic, and define small amino acids as neutral, we see that the overall match between the reduced sequence and the aligned (with gaps) ubiquitin sequence is quite good. It is important to note that although we have designed for secondary structure and tertiary topology, we have not a priori designed a sequence to match the primary structure of naturally occurring ubiquitin. It is encouraging that the primary structure of our minimalist model is reasonably faithful to the primary structure of the real protein, an objective that is important as we try to move minimalist models toward having predictive power for new sequences whose structure is unknown. T1

The global minimum structure found from multiple simulated annealing runs is shown in comparison with the native state of ubiquitin in Figure 1. We see that the resulting structure is ubiquitin-like because of the same ordering of secondary structure elements (Table I), and because it has a very similar overall topology to the F1

TABLE I. Bead and Dihedral Sequences for the Ubiquitin-like Model*

1° Ubiquitin experiment	MQIFVKT	LTK TITLE	VPTIENVKAK	IQDKEGQRLI
1° Ubiquitin model	LBLBLBLBLB	NNLNBBBBBB	BBNNLLBBLB	BLLBNNLBBB
1° Ubiquitin experiment	FAGKQEDGRT	LSDYNIQKET	LHLVLRLR	
1° Ubiquitin model	BBNLBLBLLB	NLBBLBBLNB	BLBLBLBL	
2° Ubiquitin experiment	EEEEEET	TSC EEEEE	TTEHHHHHHH	HHHHHCEEEE
2° Ubiquitin model	EEEEEEEEH	THEHTEEEEE	HHEHHHHHHH	HHHEHTEEEE
2° Ubiquitin experiment	ETTEECTTSE	THHHTCCTTE	EEEEEC	
2° Ubiquitin model	ETHEEEETEE	THHHHHHHEH	HEEEEE	

*B = hydrophobic; L = hydrophilic; N = neutral; E = extended; H = helix; T = turn.

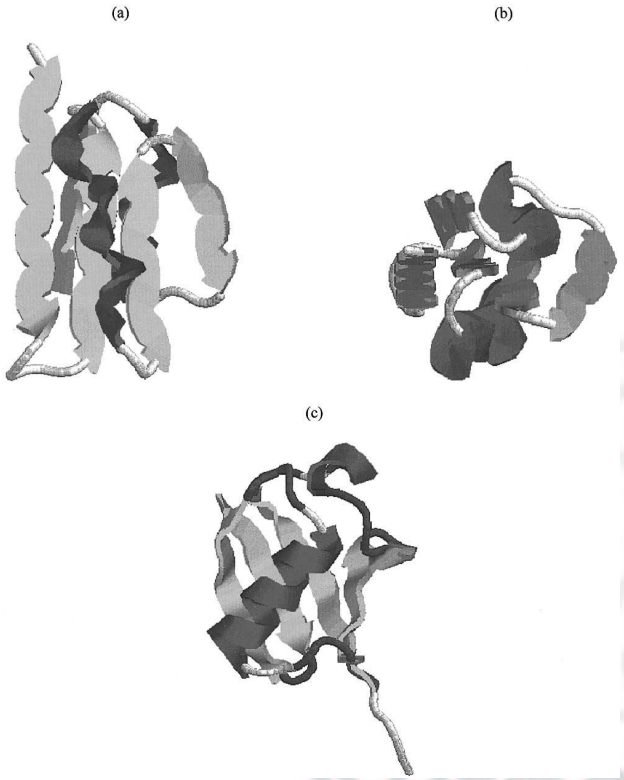

Fig. 1. Native-state structure for the ubiquitin-like model. **a:** Front view. **b:** Top view. **c:** x-ray crystal structure of ubiquitin.[26] (PDB code: lubi.) [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.] AQ: 14

ubiquitin fold. Both our model and ubiquitin consist of a mixed β-sheet with a central helix packed against it, and the ordering of elements in both structures is a β-hairpin followed by an α-helix followed by two β-strands, a short α-helix, and a final β-strand. Closer agreement with the backbone structure of ubiquitin is difficult to achieve in a model at this level of detail. For example, the characteristic twist of the β-sheet and the tilt angle of the central helix, conserved across many members of the ubiquitin-like fold class, are probably governed by hydrogen-bonding patterns and side-chain packing constraints, items not present in our model. However, our model does capture the overall topology and distribution of local and nonlocal contacts, elements recently considered to be possibly the most important in governing the overall kinetics of protein folding.[13,14,48–50]

Thermodynamics

The thermodynamics of the folding process for our model were characterized by running multiple trajectories at various temperatures and combining the resulting statistics with the multiple multidimensional histogram method.

An important indicator of the folding of the model is the relative population of the native state as a function of temperature

$$P_{\mathrm{nat}}(T) = \frac{\sum_{\chi \geq \chi_{\mathrm{NBA}}, R_g \leq R_g^{\mathrm{NBA}}} P_{\chi,R_g}}{\sum_{\chi,R_g} P_{\chi,R_g}} \tag{3}$$

where the sums are over all states, P_{χ,R_g} is the relative population of states having a similarity to the native state of χ and a radius of gyration of R_g, and χ_{NBA} and R_{NBA} define the dividing surface in (χ, R_g) space separating native states from non-native states. The actual ground-state structure itself is of such low energy and entropy that it would not be significantly populated until very very low temperatures, and therefore the definition of the native state represents an ensemble, and not only the ground-state structure. Unlike previous studies with these types of models,[16,17,51] we included the condition $R_g \leq R_g^{\mathrm{NBA}}$ to further refine the native-state basin of attraction (NBA). A small percentage of states have $\chi \geq \chi_{\mathrm{NBA}}$, but they are clearly not compact and therefore not truly native-like. χ_{NBA} can be determined by a plot of free energy versus χ[37]; alternatively, a consistent value can be found by locating the temperature of the maximum peak in the fluctuations in χ, $\Delta^2\chi(T)$, and using $\chi_{\mathrm{NBA}} = \langle\chi(T)\rangle$ corresponding to that temperature. Both methods give a value of $\chi_{\mathrm{NBA}} = 0.37$, which was used in eq. 3 and for the purpose of collecting first-passage times.

Figure 2 shows $P_{\mathrm{nat}}(T)$, as well as the fluctuations in radius of gyration as a function of temperature, $\Delta^2 R_g(T) = \langle R_g^2(T)\rangle - \langle R_g(T)\rangle^2$. From the $P_{\mathrm{nat}}(T)$ curve, we can identify $T_f = 0.51$ as the folding temperature for this model, where $P_{\mathrm{nat}}(T) = 0.5$.[37,38] The fluctuations in the R_g curve gives information on the location of collapse transitions in the folding process. We can see that a major transition occurs at $T \sim 0.56$. Looking at the plot of $\langle R_g(T)\rangle$ (J.M. Sorenson and T. Head-Gordon, unpublished observations). it is F2 AQ: 2 AQ: 7

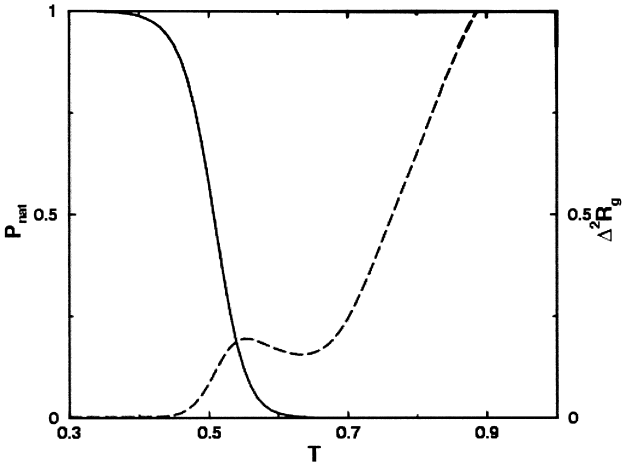

Fig. 2. Native state formation and fluctuations in R_g as a function of reduced temperature T. $P_{nat}(T)$ (—), $\Delta^2 R_g(T)$ (---).

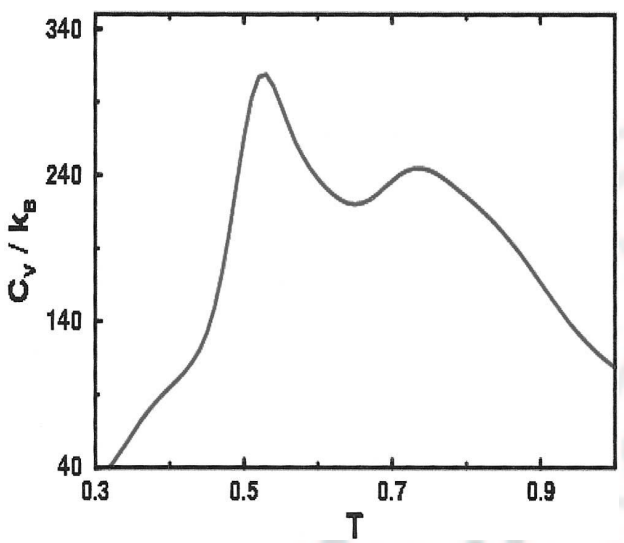

Fig. 3. Heat capacity C_p as a function of reduced temperature T. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

apparent that this corresponds to a final transition to a fully collapsed state from a partially collapsed state. The initial collapse to the partially collapsed state is a very noncooperative process, as would be expected from such a transition between two disordered states, and occurs at AQ: 2 $T \sim 0.8$. Collapse of the chain in this model can be identified as a three-state process. However, the observation that the final collapse of the chain occurs relatively close in temperature to the acquisition of the native structure is promising for demonstrating the good folding properties of the model.[52,53]

Further information on the nature of the collapse and folding process can be derived from examining the plot of F3 heat capacity C_V versus temperature, shown in Figure 3. From this plot, we can see that the early coalescence of AQ: 2 structure noted near $T \sim 0.8$ results in a broad transition signature in the heat capacity curve. As we approach the

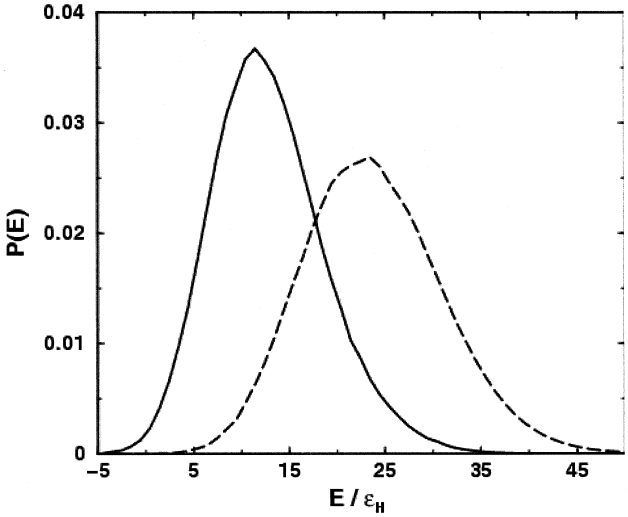

Fig. 4. Energy distributions for native (—) and non-native (---) states at $T_f = 0.51$.

folding temperature, the final collapse and formation of native contacts result in a much sharper energetic transition, as evidenced by the taller and sharper peak in $C_V(T)$ near $T = 0.55$. This finding is consistent with the temperature dependence of native-state structure formation and fluctuations in R_g shown in Figure 2.

Another common thermodynamic observable used to AQ: 8 characterize the shape of the underlying free energy landscape is the σ parameter

$$\sigma = \frac{T_\theta - T_f}{T_f} \tag{4}$$

which has been widely used previously to predict the foldability and cooperativity of folding for similar off-lattice models.[16,53,54] As noted above, the folding of the current model exhibits a less well-defined collapse transition, and it might not be appropriate to ascribe a single σ to the folding. We determine a σ value of 0.10, found by taking $T_\theta = 0.56$, but as we shall see in the kinetics section below, such a low value might be misrepresentative of the actual foldability of the system.

The Z-score,[54] defined as

$$Z = \frac{\langle E_{\text{non-nat}} \rangle - \langle E_{\text{nat}} \rangle}{\langle \Delta E_{\text{non-nat}} \rangle} \tag{5}$$

is closely related to the criterion used by Shakhnovich and coworkers[55] to predict good folders from thermodynamic information alone. A high Z-score represents good discrimination in the model between native and non-native states, a necessary criterion if the landscape is to possess a funnel-like shape directing folding chains to the native-state basin of attraction. We find a Z-score of 1.51 that is due to the large energy gap between native and non-native states found in this model (the numerator in eq. 5 with a value of 21.4 $k_B T$). Figure 4 shows the distribution of F4 energies of native and non-native states at the folding

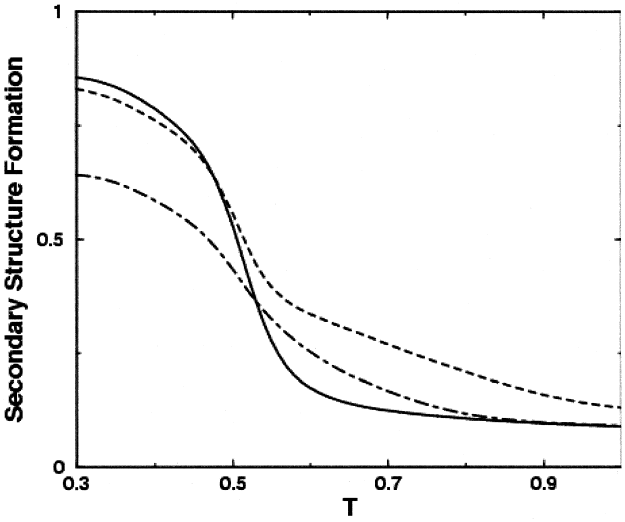

Fig. 5. Formation of secondary structure versus temperature. $\chi_{\beta\beta}$ (—), $\chi_{\alpha\beta}$ (---), and $\chi_{\beta\alpha\beta}$ (· — ·).

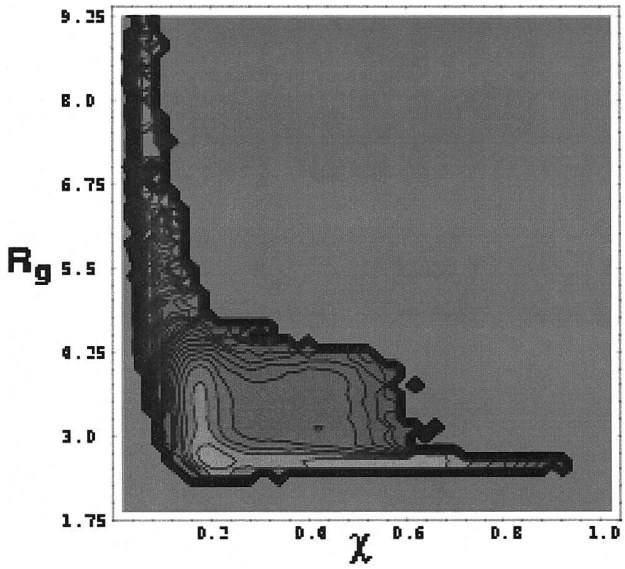

Fig. 6. Free energy as a function of R_g and χ at $T_f = 0.51$. Contour lines are spaced at intervals of k_BT. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

temperature. These distributions show that there is an energetic bias favoring native-like states, although the fact that the distributions overlap indicates the strong possibility of misfolded traps along the folding pathway.

Because our sampling is conducted over multiple order parameters, we can monitor the formation of various parts of the protein structure versus temperature (Fig. 5). We can see from the figure that the formation of the central α-helix and its adjoining β-strand (3) occurs first as the temperature is lowered and the high-temperature random coil state begins to acquire structure. It is immediately apparent that the structuring of the last part of the protein chain is last to occur as the temperature is lowered and that this process is highly noncooperative. This indicates that there are many competing states for the ordering of this portion of the chain. As we will see, in ubiquitin this portion of the chain corresponds to the most disordered segment of the protein, and it might be expected that such a disordered structure would resist full structuring until a much further decrease in temperature (or denaturant, in the case of the real protein).

As noted in the Introduction, the folding of a physicochemical minimalist protein model of this size has not been simulated to date. An interesting feature that arises from this size of protein is the division of portions of the chain into cooperative and noncooperative folding. Whereas the thermodynamics of the overall folding process is fairly two-state and cooperative, as can be seen in Figure 2, the ordering of individual portions of the protein vary from the very cooperative formation of the first hairpin, a feature also present in the earlier protein L/G model,[17] to the essentially noncooperative formation of the last β-strands and final α-helix. It has been anticipated that the folding of separate domains of multi-domain proteins might follow different kinetics, but here we see the interesting case that distinctly different kinetics can be traced down to the smaller level of supersecondary structure formation in a single domain.

The free energy landscape for folding is well captured by projecting the free energy onto order parameters monitoring collapse of the chain and formation of native-like structure. Figure 6 shows the free energy at $T_f = 0.51$ as a function of radius of gyration R_g, and native-state similarity χ. The resulting shape is L-like, with non-native unfolded states in the upper left, non-native collapsed states in the lower left, and the native-state ensemble at the lower right, is common to many of the off-lattice bead models studied so far.[4,16,37] The presence of more states toward the diagonal, representing partially collapsed, partially native states, is a favorable property of the landscape if our goal is to achieve fast-folding collapse-concomitant-with-folding scenarios.[17,56] The two favorable basins at this temperature represent collapsed non-native states, with some propensity for unraveling, and collapsed more native-like states.

The pathways for folding can be better analyzed by projecting the free energy onto the order parameters for secondary structure formation. Figure 5 indicates that the secondary structure elements corresponding $\chi_{\beta\beta}$ and $\chi_{\alpha\beta}$ show more cooperative two-state like behavior, as further confirmed in Figure 7. Here we see that the free energy is asymmetric with respect to formation of these secondary structure elements, with a marked propensity for prior formation of $\chi_{\alpha\beta}$, the packing of the central helix against strand 3. There is even evidence in the upper left corner of Figure 7 that a metastable intermediate might exist along the folding pathway, representing a structure with the α-helix formed and part of the middle of the sheet, but with the first β-hairpin not yet formed. This is very similar to the earlier observation of asymmetric β-hairpin formation in the protein L/G model.[17] The free energy is not prohibitively high for folding to proceed along the diagonal,

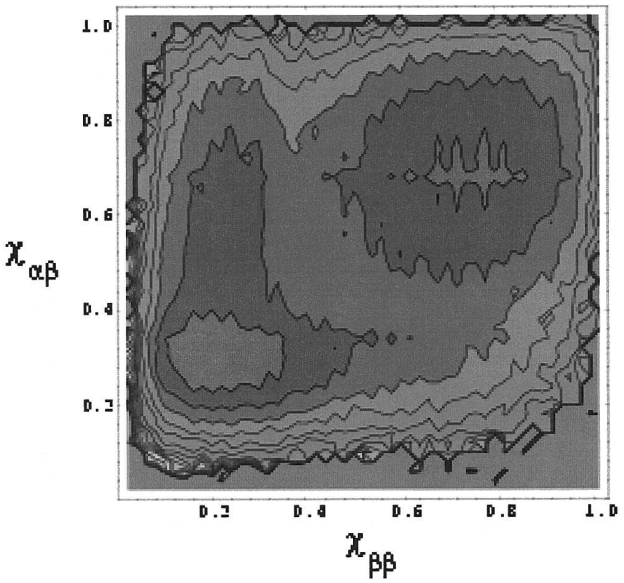

Fig. 7. Free energy as a function of $\chi_{\beta\beta}$ and $\chi_{\alpha\beta}$ at $T_f = 0.51$. Contour lines are spaced at intervals of k_BT. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

forming $\chi_{\beta\beta}$ and $\chi_{\alpha\beta}$ simultaneously, so we would identify such an intermediate as nonobligatory.

We can further characterize the nature of this intermediate state by projecting the free energy onto the radius of gyration R_g and $\chi_{\alpha\beta}$. This slice through the free energy surface is shown in Figure 8a, where we see many features similar to those in Figure 6. The free energy as a function of the $\chi_{\alpha\beta}$ coordinate clearly exhibits a non-native and native basin of attraction. The non-native basin of attraction in Figure 8a is consistent with a collapsed intermediate with no secondary structure formed.[34,57] In addition, there are significant number of states that have high values of $\chi_{\alpha\beta}$ and also relatively large values of R_g. These represent states where the α-helix and β-strand 3 are packed against each other, but the rest of the structure is non-native and fluctuating in size. This constitutes further evidence that the formation of the αβ structure is independently stable and might play a role in forming a folding intermediate.

A useful contrast is to examine the corresponding plot for $\chi_{\beta\alpha\beta}$, shown in Figure 8b. Figure 5 indicates that the secondary structure elements corresponding to the latter half of the chain, $\chi_{\beta\alpha\beta}$, does not participate actively in the folding process; rather, they seem to fold as a spectator to folding. This is verified by examining the free energy dependence on $\chi_{\beta\alpha\beta}$, which is broad and characteristic of a higher-order transition with no evidence of two-state behavior. In this case, we also see evidence for formation of native-state structure in partially collapsed states, but the formation of a native state structure is very noncooperative and is likely to flicker out of existence quickly, as evidenced by the single large basin in the free energy surface. This is consistent with the temperature profile for this order parameter, shown in Figure 5, as the slope of the

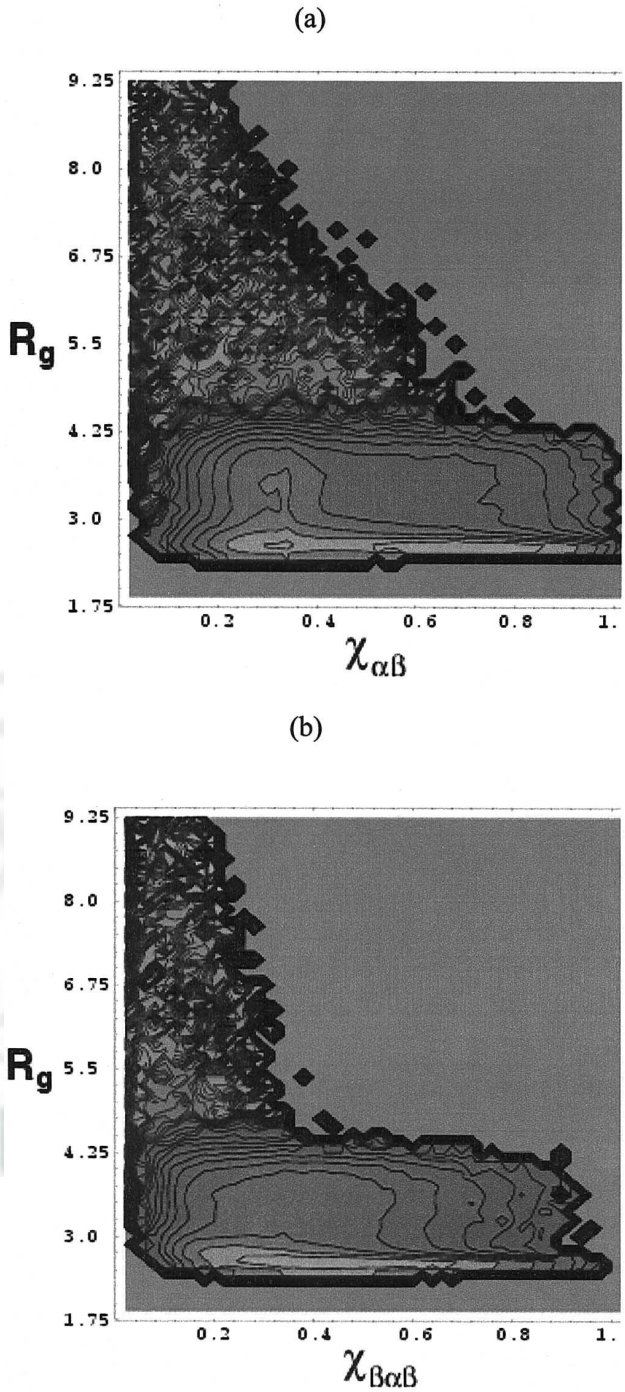

Fig. 8. Free energy as a function of R_g and native-state secondary structure formation at $T_f = 0.51$. Contour lines are spaced at intervals of k_BT. **a:** $F(\chi_{\alpha\beta}, R_g)$. **b:** $F(\chi_{\beta\alpha\beta}, R_g)$. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

$\chi_{\beta\alpha\beta}$ versus T curve is directly related to $\Delta^2\chi_{\beta\alpha\beta}(T)$, the mean fluctuations in $\chi_{\beta\alpha\beta}$.

Kinetics

The kinetics of folding to the native state were monitored at different temperatures by collecting first passage

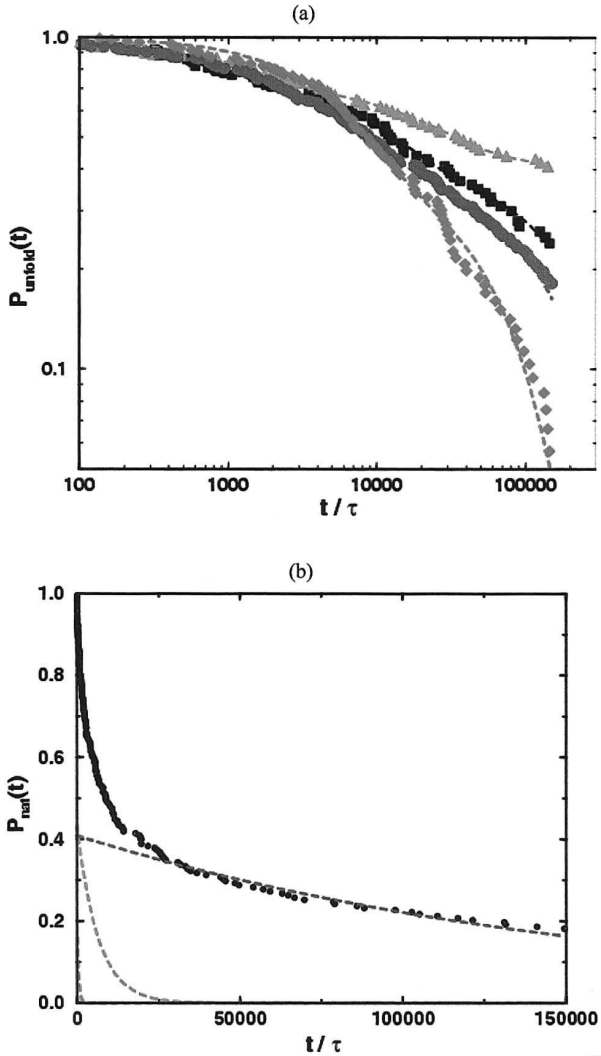

Fig. 9. **a:** Distribution of first-passage times to the native state at $T = 0.57$ (◇), $T = 0.52$ (●), $T = 0.48$ (■), and $T = 0.44$ (▲). ---, multi-exponential fits to the data points (see Table II). **b:** Distribution of first-passage times for folding at $T = 0.52$, with the kinetics broken into three exponential regimes. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

time distributions—recording the time when the folding chain first enters the native-state basin of attraction, and fitting the data to an appropriate functional form depending on the nature of the observed kinetics.

The highest temperature examined was $T = 0.57$, which is close to the final collapse transition. Folding at this relatively high temperature showed three-state behavior (Fig. 9), with the best fit of the kinetic data being a double exponential curve:

F9

$$\sum_i a_i \exp(-t/\tau_i) \tag{6}$$

T2 with parameters a_i and τ_i given in Table II. The quality of fit was measured by the mean χ^2 value:

$$\chi^2 = \frac{1}{N}\sum_i (y_i - y_i^0)^2 \tag{7}$$

where $\{y_i\}$ are the fit points and $\{y_i^0\}$ are the data points. The biexponential kinetics indicates that partially collapsed states represent long-lived intermediates on the folding pathway, with folding occurring by both fast collapse-and-folding routes as well as escape from longer-lived collapsed states.

Close to the folding temperature, the folding is best fit by three exponentials, as can be seen with the $T = 0.52$ curve in Figure 9a, with fitting parameters given in Table II. These three exponentials correspond to populations that fold on a very fast (16%), an intermediate timescale that is an order of magnitude slower than the fastest (43%), and a very slow folding population (41%) that involves a timescale close to our simulation cutoff time of 150,000τ (Fig. 9b). It should be noted that 20% of the chains have not found the native state by the cutoff time; however, the curvature of the folding kinetics leads us to conclude that more chains would fold if the allowed folding time had been extended past this cutoff.

The presence of triexponential kinetics at the folding temperature could signal poor folding kinetics for the model. We therefore considered other functional forms such as a stretched exponential

$$\exp\lfloor(-t/\tau_s)^\beta\rfloor \tag{8}$$

that is appropriate when the system moves away from activated kinetics toward behavior consistent with glassy landscapes.[40,56] However, the fit to the $T = 0.52$ data was poor.

Furthermore, the folding kinetics remain encouragingly multi-exponential to lower temperatures. If the folding were slowed down to a more glass-like regime, we would expect the kinetics to be a straight, shallow-sloped line on a log–log plot, representative of power law behavior.[16,40] Instead, Figure 9a shows that the kinetics at $T = 0.48$ do not differ significantly from the $T = 0.52$ kinetics at long times (Table II), indicating that as T drops below T_f, the protein chain is still able to find the native state by activated kinetics dominated by only a few important barriers, albeit slowly. This condition leads us to assert that the glass temperature T_g[38,58,59] is lower than the folding temperature for this model, with a $T_f/T_g \gtrsim 1.0$. The ratio T_f/T_g gives a simple characterization of the steepness of the protein folding funnel for theoretical and real proteins, and real proteins are expected to have a $T_f/T_g > 1.0$, although recent theoretical work suggests that genuine native-state stability for small two-state proteins has a signature $T_f/T_g > 6.0$. AQ: 9

The relatively straight line of the $T = 0.44$ kinetic trace is indicative of power law behavior at this temperature, representing much glassier kinetics, dominated by traversal of a rough and deep energy landscape. However, if we fit the data at $T = 0.44$ to a single stretched exponential or power law form, we find that the fits are poor. The triple exponential fit at this temperature finds an extremely large value of τ_3, indicating that the folding is

TABLE II. Fitting Parameters for the Kinetic Traces in Figure 8*

T	a_1	τ_1	a_2	τ_2	$1 - a_1 - a_2$	τ_3	$\chi^2/10^{-4}$
0.57	0.59	5,900	0.41	69,000	—	—	3.6
0.52	0.16	500	0.43	6,500	0.41	160,000	0.95
0.48	0.23	600	0.36	11,000	0.41	260,000	1.1
0.44	0.28	1,300	0.30	25,000	0.43	600,000	2.5

*Times are in units of τ.

TABLE III. Stretched Exponential Fitting Parameters for the Kinetic Traces in Figure 8*

T	a_1	τ_1	β_1	$1 - a_1$	τ_2	β_2	$\chi^2/10^{-4}$
0.44	0.42	4,400	0.57	0.58	1,100,000	0.49	0.94

*Times are in units of τ.

better described as a biexponential (Table II), characterizing the kinetics of the folded chains, plus a constant representing the population that remain unfolded beyond $t = 150{,}000\tau$. Our best fit considered two stretched exponentials, where we find values of β, the stretched exponential exponent, of 0.57 and 0.49 for the fast and slow pathways, respectively (Table III). All these determinations of exponents are consistent with the expected and observed relaxation kinetics to the native state at temperatures below the collapse and folding temperatures, a common effect seen in the folding literature.[31,60–62]

DISCUSSION

The current experimental picture for the folding of ubiquitin is in good agreement with the scenario we have described here for this ubiquitin protein model. An engineered tryptophan mutant of ubiquitin (F45W), when placed in low denaturant concentrations, gives rise to an early folding intermediate that accumulates during the first few milliseconds of refolding; the observed change in the Trp-45 fluorescence is consistent with the formation of a compact intermediate with a partially developed hydrophobic core.[34] A complementary study on ubiquitin using a variant of hydrogen-exchange labeling showed that there are no early intermediates with secondary structure formed.[57] Considering the various free energy projections (Figs. 6–8) generated from our ubiquitin model, we also see evidence for a compact intermediate with little secondary structure formed.

Several experiments have characterized the structure of an A-state of ubiquitin that is stable at low pH in 60% methanol solution.[32,33] The combined results from three-dimensional nuclear magnetic resonance (NMR)[33] and fragment secondary structure studies[32] indicate that the A-state consists of structuring of the first β-hairpin and the central α-helix, comprising the first 33–35 residues of the sequence. Results from hydrogen–deuterium exchange NMR also suggest that the first formation of secondary structure occurs cooperatively between the central α-helix and β-sheet, with the C-terminus loop region forming last in the folding process.[35] As seen in our model of ubiquitin, the secondary structure order parameters indicate that the crucial event in folding is the formation of the α-helix with the first part of the β-sheet. The cooperative formation of secondary structure occurring in the first part of the sequence is crucial in propelling the chain into the native-state basin of attraction.

Our model also reproduces the observation that the C-terminus loop structure in ubiquitin is last to fold in the folding process[35] and most likely plays a spectator role in the folding kinetics. The last part of the chain, monitored by $\chi_{\beta\alpha\beta}$, forms noncooperatively and fluctuates in and out of existence until much lower temperatures below the folding temperature. However, it is interesting that in our model this region possesses more secondary structure formation than in the corresponding loop region in ubiquitin, although the last β-strand packs itself against the β-sheet instead of inserting itself in the sheet as in the ubiquitin structure. That this region still folds noncooperatively is indicative that the lack of strong local interactions in this portion of the chain contributes to the formation of nonspecific structure. Therefore we find that the most important factors governing the overall kinetics are chain topology and the distribution of local and nonlocal contacts,[12–15,48–50] as it is specifically this aspect that is most preserved between our simplified model and ubiquitin.

CONCLUSIONS

We report the successful development of an off-lattice model for probing the folding of larger proteins based on a ubiquitin-like fold topology. Such a model permits thorough characterization of the folding thermodynamics and kinetics, while retaining enough detail of the actual protein to provide useful comparison to experiment.

The designed ubiquitin-like model possesses a funnel-like landscape, but with escape from collapsed states strongly influencing the folding kinetics at temperatures close to the folding temperature. The favorable energy gap and Z-score indicate that the native state is energetically separated from non-native folds at the folding temperature, although discrimination of the native state must compete with some overlapping low-energy non-native states (Fig. 4). This is also indicated in the $R_g - \chi$ free energy landscape shown in Figure 6, where we see the prominence of a basin of compact non-native states.

The nature of these states can be ascertained by examining the landscape as a function of radius of gyration and the secondary structure order parameters (Fig. 8). Fluctuations in R_g are large for compact non-native states, unlike the compact native state. This indicates that states formed by nonspecific collapse have many unfavorable contacts, destabilizing them and encouraging structure unraveling. This would be expected to be an important property of real proteins, enabling compact non-native states to unfold partially or completely and refold correctly to the native state. In contrast, the compact native state optimally satisfies the attractive interactions that hold the core together, and fluctuations in the radius of gyration are largely suppressed.

The folding landscape of this model shows a good energetic separation between native and non-native states, and a favorable folding funnel for directing unfolded chains to the native state. Folding proceeds through multiple pathways, fast pathways involving collapse to native-like states and rapid folding and slow pathways requiring escape from collapsed non-native states. In addition, the assembly of supersecondary structure appears to follow multiple pathways, with the ability either to proceed in stepwise fashion or to assemble concurrently.

The benefits of this level of abstraction in our model of ubiquitin, and our ability to characterize the free energy surface using projections with different order parameters, permit us to demonstrate agreement with a composite of experiments involving ubiquitin under various conditions or mutated forms. This is similar to our characterization of the free energy surface of our model of protein L/G, two proteins with identical topology but very different primary sequences, in which we find that two main pathways exist for folding of the mixed α/β structure. One pathway involves formation of an intermediate (and is the lower energy pathway in our model), while the other pathway involves a single barrier. Our model therefore encompasses a composite of experiments both for protein G, which have established the existence of a high-energy intermediate along the folding pathway,[63] and for protein L, which appears to be strictly two-state.[27] In the case of ubiquitin, our model reproduces the observation that the C-terminus loop structure in ubiquitin is last to fold in the folding process[35] and probably plays a spectator role in the folding kinetics. The possibility of a productive metastable intermediate along the folding pathway consisting of collapsed states[34] with no secondary structure[64] and intermediates or transition structures involving secondary structural elements occurring early in the sequence[32,33] is also supported by our model.

It is interesting that this longer ubiquitin chain shows biexponential kinetics above the folding temperature, while shorter chains tend to exhibit single exponential kinetics at these temperatures[16,17] (although we would expect that this model would show single exponential activated kinetics at suitably higher temperatures as well). Although the experimental heat capacity for ubiquitin does not show our pre-transition peak, the non-two-state nature of our C_V curve is not necessarily inconsistent with the observation that ubiquitin under certain conditions shows evidence for three-state behavior. Nonetheless, non-native collapsed states might play too large of a role in the folding kinetics of our model as compared with real proteins,[16,17,40] (although it should be noted that protein folding involving a highly compact non-native state has been recently reported[64]). Furthermore, there is new evidence that the temperature of the onset of glassy kinetics should be much lower than what we observe in this model.[65,66]

Ubiquitin may also be unusual in that several protein redesign experiments have indicated that interactions between hydrophobic residues in the ubiquitin core are near optimal,[48,67] which may be related to strong sequence conservation in the ubiquitin structural superfamily, suggesting that this could play an important role in the rapid folding kinetics of ubiquitin. This is further evidence that energetic frustration is likely minimized for natural ubiquitin, that our ubiquitin model successfully optimizes topological but not energetic frustration, and that further improvements in our minimalist model ubiquitin need to be pursued.

Real proteins have several structural features not modeled in this study that might provide more specific collapse. These include a larger variety of attractive interactions (polar–polar, as well as hydrophobic), more cooperative formation of secondary structure such as α-helices through backbone hydrogen bonding, and the cooperativity provided by the very specific side-chain packing of the native-state core. However, the overall faithfulness of this model to the ubiquitin fold and its favorable comparison with experiment make it viable for further improvement study in regards to the folding of longer chains and in particular to complex fold topologies. AQ: 10

In addition to the present study, we have now completed multiple studies at this simplified level of description of protein folding addressing issues of protein design,[16] the folding of small α/β proteins,[17] and the role of sequence mutations in the folding of protein L.[47] In combination, we believe they indicate the utility of this level of modeling as a productive step forward given current computational limits. Additional questions which should be amenable to further studies include the folding of other protein fold classes, the effects of parameter choices in the energy function on the formation of local and nonlocal structure, and the role of solvent effects in the thermodynamics and kinetics of protein folding[68] (J.M. Sorenson and T. Head-Gordon, unpublished observations). AQ: 7

An interesting possibility suggested by this level of modeling is the feasibility of genomic-scale modeling. Characterizing protein folding kinetics and thermodynamic stability across whole genomes is insightful because such determinations may distinguish, for example, very stable housekeeping enzymes that are required over longer time scales versus short-lived but quickly folded proteins that are expressed during times when quick response to external stress is required. While the experimental effort in structural genomics is focused on providing new fold classifications, computation and theory will and should play a complementary role of completing structural, ki-

netic, and thermodynamic information across whole genomes.[69] Minimalist models that we have developed for complex α/β folds and larger protein sizes are important validation steps in our ultimate goal of prediction across genome sequence data.

ACKNOWLEDGMENTS

J.M.S. thanks the National Science Foundation for a Graduate Research Fellowship.

AQ: 3

REFERENCES

1. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem 1997;48:545–600.
2. Shakhnovich EI. Theoretical studies of protein-folding thermodynamics and kinetics. Curr Opin Struct Biol 1997;7:29–40.
3. Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. Principles of protein folding—a perspective from simple exact models. Protein Sci 1995;4:561–602.
4. Pande VS, Grosberg AY, Tanaka T. Heteropolymer freezing and design: towards physical models of protein folding. Rev Mod Phys 2000;72:259–314.
5. Eaton WA, Munoz V, Thompson PA, Henry ER, Hofrichter J. Kinetics and dynamics of loops, α-helices, β-hairpins, and fast-folding proteins. Acct Chem Res 1998;31:745–753.
6. Dobson CM, Sali A, Karplus M. Protein folding: a perspective from theory and experiment. Angew Chem Int Ed Engl 1998;37:868–893.
7. Laurents DV, Baldwin RL. Protein folding: matching theory and experiment. Biophys J 1998;75:428–434.
8. Dill KA, Chan HS. From Levinthal to pathways and funnels. Nature Struct Biol 1997;4:10–19.
9. Fersht AR. Nucleation mechanisms in protein folding. Curr Opin Struct Biol 1997;7:3–9.
10. Pande VS, Grosberg AY, Tanaka T, Rokhsar DS. Pathways for protein folding: is a "new view" needed? Curr Opin Struct Biol 1998;8:68–79.
11. Shakhnovich EI. Modeling protein folding: the beauty and power of simplicity. Fold Des 1996;1:R50–R54.
12. Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en route" intermediates for protein fodling? An investigation for small globular proteins. J Mol Biol 2000;298: 937–953.
13. Shea JE, Onuchic JN, Brooks CL. Exploring the origins of topological frustration: design of a minimally frustrated model of fragment B of protein A. Proc Natl Acad Sci USA 1999;96:12512–12517.
14. Nymeyer H, Socci ND, Onuchic JN. Landscape approaches for determining the ensemble of folding transition states: success and failure hinge on the degree of frustration. Proc Natl Acad Sci USA 2000;97:634–639.
15. Clementi C, Jennings PA, Onuchic JN. How native-state topology affects the folding of dihydrofolate reductase and interleukin-1β. Proc Natl Acad Sci USA 2000;97:5871–5876.
16. Sorenson JM, Head-Gordon T. Redesigning the hydrophobic core of a model β-sheet protein: destabilizing traps through a threading approach. Proteins 1999;37:582–591.
17. Sorenson JM, Head-Gordon T. Matching simulation and experiment: a new simplified model for simulating protein folding. J Comp Biol 2000;7:469–481.
18. Gu H, Kim D, Baker D. Contrasting roles for symmetrically disposed β-turns in the folding of a small protein. J Mol Biol 1997;274:588–596.
19. Sorenson JM, Head-Gordon T. The importance of hydration for the kinetics and thermodynamics of protein folding: simplified lattice models. Fold Design 1998;3:523–534.
20. Brooks CL. Simulations of protein folding and unfolding. Curr Opin Struct Biol 1998;8:222–226.
21. Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science 1998;282:740–744.
22. Alonso DOV, Daggett V. Molecular dynamics simulations of hydrophobic collapse of ubiquitin. Protein Sci 1998;7:860–874.
23. Honeycutt JD, Thirumalai D. Metastability of the folded states of globular proteins. Proc Natl Acad Sci USA 1990;87:3526–3529.
24. Guo Z, Thirumalai D. Kinetics of protein folding: nucleation mechanism, time scales, and pathways. Biopolymers 1994;36:83–102.
25. Guo Z, Thirumalai D. Kinetics and thermodynamics of folding of a de novo designed four-helix bundle protein. J Mol Biol 1996;263: 323–343.
26. Kim DE, Yi Q, Gladwin ST, Goldberg JM, Baker D. The single helix in protein L is largely disrupted at the rate-limiting step in folding. J Mol Biol 1998;284:807–815.
27. Plaxco KW, Millett IS, Segel DJ, Doniach S, Baker D. Chain collapse can occur concomitantly with the rate-limiting step in protein folding. Nature Struct Biol 1999;6:554–556.
28. Vijay-Kumar S, Bugg CE, Cook WJ. Structure of ubiquitin refined at 1.8 Å resolution. J Mol Biol 1987;194:531.
29. Ferrenberg AM, Swendsen RH. Optimized Monte Carlo data analysis. Phys Rev Lett 1989;63:1195–1198.
30. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. Multidimensional free-energy calculations using the weighted histogram analysis method. J Comp Chem 1995;16:1339–1350.
31. Sabelko J, Ervin J, Gruebele M. Observation of strange kinetics in protein folding. Proc Natl Acad Sci USA 1999;96:6031–6036.
32. Cox JPL, Evans PA, Packman LC, Williams DH, Woolfson DN. Dissecting the structure of a partially folded protein: circular dichroism and nuclear magnetic resonance studies of peptides from ubiquitin. J Mol Biol 1993;234:483–492.
33. Stockman BJ, Euvrard A, Scahill TA. Heteronuclear three-dimensional NMR spectroscopy of a partially denatured protein: the A-state of human ubiquitin. J Biomol NMR 1993;3:285–296.
34. Khorasanizadeh S, Peters ID, Roder H. Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. Nature Struct Biol 1996;3:193–205.
35. Briggs MS, Roder H. Early hydrogen-bonding events in the folding reaction of ubiquitin. Proc Natl Acad Sci USA 1992;89:2017–2021.
36. Lazar GA, Desjarlais JR, Handel TM. De novo design of the hydrophobic core of ubiquitin. Protein Sci 1997;6:1167–1178.
37. Socci ND, Onuchic JN. Kinetic and thermodynamic analysis of protein-like heteropolymers: Monte Carlo histogram technique. J Chem Phys 1995;103:4732–4744.
38. Sorenson JM, Head-Gordon T. The importance of hydration for the kinetics and thermodynamics of protein folding: simplified lattice models. Fold Design 1998;3:523–534.
39. Levinthal C. In: Debrunner P, Tsibris JCM, Monck E, editors. Mossbauer spectroscopy in biological systems. Urbana: University of Illinois Press; 1969. p 22–24. AQ: 13
40. Shea JE, Nochomovitz YD, Guo Z, Brooks CL. Exploring the space of protein folding hamiltonians: the balance of forces in a minimalist β-barrel model. J Chem Phys 1998;109:2895–2903.
41. Cormen TH, Leiserson CE, Rivest RL. Introduction to algorithms. Cambridge, MA: MIT Press; 1990.
42. Lo Conte L, Ailey B, Hubbard TJP, Brenner SE, Murzin AG, Chothia C. SCOP: a structural classification of protein database. Nucleic Acids Res 2000;28:257–259.
43. Shakhnovich EI, Gutin AM. Engineering of stable and fast-folding sequences of model proteins. Proc Natl Acad Sci USA 1993;90: 7195–7199.
44. Hao MH, Scheraga HA. Optimizing potential functions for protein folding. J Phys Chem 1996;100:14540–14548.
45. Bryngelson JD, Wolynes PG. Spin glasses and the statistical mechanics of protein folding. Proc Natl Acad Sci USA 1987;84: 7524–7528.
46. Sali A, Shakhnovich EI, Karplus M. How does a protein fold? Nature 1994;369:248–251.
47. Sorenson JM, Head-Gordon T. (in press). AQ: 11
48. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol 1998;277:985–994.
49. Plaxco KW, Larson S, Ruczinski I, Riddle DS, Thayer EC, Buchwitz B, Davidson AR, Baker D. Evolutionary conservation in protein folding kinetics. J Mol Biol 2000;298:303–312.
50. Martinez JC, Pisabarro MT, Serrano L. Obligatory steps in protein folding and the conformational diversity of the transition state. Nature Struct Biol 1998;5:721–729.
51. Guo Z, Brooks CL. Thermodynamics of protein folding: a statistical mechanical study of a small all-β protein. Biopolymers 1997;42: 745–757.

52. Klimov DK, Thirumalai D. Criterion that determines the foldability of proteins. Phys Rev Lett 1996;76:4070–4073.
53. Klimov DK, Thirumalai D. Cooperativity in protein folding: from lattice models with sidechains to real proteins. Fold Design 1998;3:127–139.
54. Bowie JU, Luthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. Science 1991;253:164–170.
55. Tiana G, Broglia RA, Shakhnovich EI. Hiking in the energy landscape in sequence space: a bumpy road to good folders. Proteins 2000;39:244–251.
56. Phillips JC. Kohlrausch relaxation and glass transitions in experiment and in molecular dynamics simulations. J Noncrystallogr Solids 1995;182:155–161.
57. Gladwin ST, Evans PA. Structure of very early protein folding intermediates: new insights through a varient of hydrogen exchange labeling. Fold Design 1996;1:407–417.
58. Socci ND, Onuchic JN. Folding kinetics of protein-like heteropolymers. J Chem Phys 1994;101:1519–1528.
59. Onuchic JN, Wolynes PG, Luthey-Schulten Z, Socci ND. Toward an outline of the topography of a realistic protein-folding funnel. Proc Natl Acad Sci USA 1995;92:3626–3630.
60. Skorobogatiy M, Guo H, Zuckermann M. Non-Arrhenius modes in the relaxation of model proteins. J Chem Phys 1998;109:2528–2535.
61. Chan HS, Dill KA. Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. Proteins 1998;30:2–33.
62. Thirumalai D, Woodson SA. Kinetics of folding of proteins and RNA. Acct Chem Res 1996;29:433–439.
63. Park SH, O'Neil KT, Roder H. An early intermediate in the folding reaction of the B1 domain of protein G contains a native-like core. Biochemistry 1997;36:14277–14283.
64. Pandya MJ, Williams PB, Dempsey CE, Shewry PR, Clarke AR. Direct kinetic evidence for folding via a highly compact, misfolded state. J Biol Chem 1999;274:2682–2687.
65. Gillespie B, Plaxco KW. Nonglassy kinetics in the folding of a simple single-domain protein. Proc Natl Acad Sci USA 2000;97:12014–12019.
66. Kaya H, Chan HS. Polymer principles of protein calorimetric two-state cooperativity. Proteins 2000;40:637–661.
67. Finucane MD, Woolfson DN. Core-directed protein design. II. Rescue of a multiply mutated and destabilized variant of ubiquitin. Biochemistry 1999;38:11613–11623.
68. Sorenson JM, Hura G, Soper AK, Pertsemlidis A, Head-Gordon T. Determining the role of hydration forces in protein folding. J Phys Chem B 1999;103:5413–5426.
69. Head-Gordon T, Wooley J. Computational challenges in structural and functional genomics. IBM Res J Deep Comput Life Sci B. **AQ: 12**

Nymeyer H, Garcia AE, Onuchic JN. Folding funnels and frustration in off-lattice minimalist protein landscapes. Proc Natl Acad Sci USA 1998;95:5921–5928. **AQ: 3**

Author Proof

AQ1—OK short running title, or can you expand (within 45 characters, including spaces).

AQ2—Please check ~ vs. =. Is an equals sign intended here? Please check throughout, as queried.

AQ3—Please note: Refs. were cited out of sequence. All refs have been renumbered in order in text and on list. Former ref 47 was not cited (see end of list). Please cite it as an "a" number, or delete from ref list. Thank you.

AQ4—ok *rather* than meant?

AQ5—ok *flavors,* or other term?

AQ6—Please clarify "to design out."

AQ7—Former ref. 42 deleted from ref list and transferred to text, as shown (Journal style for unpublished observations).

AQ8—Please clarify thermodynamic "observable."

AQ9—Please check operator intended here, before 1.0.

AQ10—OK as meant? (Please check for sense.)

AQ11—Please update.

AQ12—Please complete/update reference.

AQ13—Please provide chapter title.

AQ14—Au: Online color default. No reply to color quotes sent 8/30, 11/7, and 11/29. If you wish to have figures printed in color, please return letter w/proofs.

Author Proof