Protein folding pathways from replica exchange simulations and a kinetic network model

Michael Andrec, Anthony K. Felts, Emilio Gallicchio, and Ronald M. Levy*

Department of Chemistry and Chemical Biology and BIOMAPS Institute for Quantitative Biology, Rutgers, The State University of New Jersey, Piscataway, NJ 08854

We present an approach to the study of protein folding that uses the combined power of replica exchange simulations and a network model for the kinetics. We carry out replica exchange simulations to generate a large ($\approx 10^6$) set of states with an all-atom effective potential function and construct a kinetic model for folding, using an ansatz that allows kinetic transitions between states based on structural similarity. We use this network to perform random walks in the state space and examine the overall network structure. Results are presented for the C-terminal peptide from the B1 domain of protein G. The kinetics is two-state after small temperature perturbations. However, the coil-to-hairpin folding is dominated by pathways that visit metastable helical conformations. We propose possible mechanisms for the α-helix/β-hairpin interconversion.

graph | master equation | protein G

Protein folding is a fundamental problem in modern structural biology, and recent advances in experimental techniques have helped to elucidate some of the thermodynamic and kinetic features that underlie different stages of the folding process (1–3). Computer simulations using reduced (4–11) and atomistic molecular models (12–21) have played a central role in the interpretation of experimental studies. Although reduced models have provided considerable insight into the overall mechanisms of protein folding, especially in helping to clarify the nature of folding funnels, intermediates, and kinetic bottlenecks, they are limited in their ability to explore the detailed molecular aspects of folding. However, because of the large number of degrees of freedom and the rarity of folding events even in small model systems, all-atom simulations require both considerable computational resources and ingenuity to obtain meaningful results; a variety of methods have been developed to increase simulation efficiency.

A particularly powerful technique for the calculation of free energy surfaces and other thermodynamic properties of complex systems is replica exchange molecular dynamics (REMD) (22), in which a generalized ensemble of the system is simulated in parallel over a range of temperatures. Periodically, coordinates are exchanged by using a Metropolis criterion that ensures that at any given temperature a canonical distribution is realized and allows for more efficient barrier crossing than could be obtained with conventional simulation methods. Although REMD is a powerful method for exploring free energy landscapes, it does not provide direct information about kinetics. To study folding by using all-atom effective potentials, heterogeneous distributed computing (16, 19) and transition path sampling (23, 24) techniques have been used. Although the former can enhance sampling by combining information from a large number of short MD trajectories "steered" by rare events, these techniques can introduce a bias toward fast events in the ensemble average of the reactive trajectories (25). Transition path sampling (23, 24) can yield quantitative estimates of the reaction rate, if the reactant and product macrostates have been defined properly; however, there remain practical difficulties in sampling transition paths in large molecular systems with multiple transition states (24).

Another way in which computational efficiency can be improved is by discretizing the state space and constructing rules for moving among those states. The resulting scheme can be represented as a graph or network (26). The kinetics on this graph is assumed to be stochastic, leading to a Markovian model for the time dependence of the populations of the various states (10, 11, 27–30). This approach is particularly well suited for reduced lattice models (10, 11). Similar schemes have been constructed based on the output of more conventional MD simulations (often after clustering or choosing a reaction coordinate) (26, 28–31) or based on an analysis of the minima and/or saddle points of the energy surface (27, 32, 33).

We present an approach to the study of protein folding pathways that makes use of the combined power of REMD and a network model for kinetics. REMD simulations are used to generate a "lattice" of states by using a modern all-atom effective potential function including implicit solvent that was previously found to correctly identify the native structure for well studied protein folding model systems (21). We then use these states to construct a network, with an ansatz that allows kinetic transitions between states that have sufficient structural similarity. The qualitative features of the kinetics and corresponding pathways between macrostates can be understood by analyzing the overall network structure or constructing kinetic Monte Carlo "trajectories" that consist of Markovian random walks on the lattice.

We present results for the folding of the model system corresponding to the C-terminal peptide from the B1 domain of protein G (the G peptide), which is known to form a β-hairpin structure in solution (34) and has been the subject of previous experimental (35–37) and computational (7, 8, 12–17, 20, 21, 30, 32, 33, 38) studies. Two mechanisms for G-peptide formation have been previously proposed: the zipper model (35) and the hydrophobic collapse model (20, 36, 37). These models describe the sequence of β-hairpin hydrogen bond formation and association of hydrophobic core residues. In addition, there have been recent reports that the G peptide can adopt α-helical conformations in addition to the predominant β-hairpin (15, 16, 32, 39). We investigate possible folding pathways for the G peptide by using the kinetic network model.

Methods

Molecular Simulation. A discretized state space for the uncapped G peptide (GEWTYDDATKTFTVTE, corresponding to residues 41–56 of protein G) was constructed by performing REMD sampling as implemented in the molecular simulation package IMPACT (40) following the approach proposed by Sugita and Okamoto (22). In our simulations, 20 replicas were run in parallel at temperatures between 270 and 690 K for a total of 10
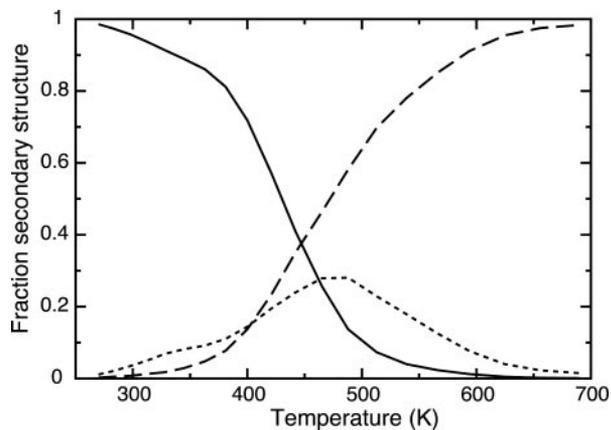
Fig. 1. Fraction of β-hairpin (solid line), α-helix (dotted line), and coil (dashed line) for the G peptide at each replica exchange temperature. A state was considered to be a β-hairpin if it contained at least one native β-hairpin hydrogen bond. A state was counted as an α-helix if at least some part of the molecule was determined to be in a helical conformation by using the program STRIDE (51). All states that did not satisfy either criterion were considered to be coil. A hydrogen bond was considered to be present when the distance between the carbonyl O and amide N was within 3.5 Å, and the angle formed by the carbonyl O and amide H and N was >130°.

ns for each replica with a time step of 1 fs. Transitions between adjacent temperatures were attempted every 250 MD steps. Immediately before attempting temperature exchanges, the configuration of each replica was stored for use in constructing the kinetic network model. MD simulations were carried out by using the OPLS-AA force field (41) and the AGBNP implicit solvent model that is based on a novel pairwise descreening implementation of the generalized Born model and a recently proposed nonpolar hydration free energy estimator (42–44). Because AGBNP is fully analytical with first derivatives it is well suited for MD sampling. For this study we have also used a modified generalized Born pair function designed to increase the dielectric screening between charged side chains as described (21). It should be noted that the current parametrization of the OPLS-AA/AGBNP effective potential function does not yield physically correct transition temperatures: the midpoint of the β-hairpin unfolding transition T_m for the G peptide is $\approx$450 K (Fig. 1), compared with the experimental value of $\approx$300 K (35, 37).

Design of the Kinetic Network Model. The REMD simulation resulted in 40,000 conformational snapshots ("states") of the G peptide at each of 20 temperatures, for a total of 800,000 states. We find that the secondary structure content of these states varies quite strongly with temperature (Fig. 1). At this point, we could group states of similar structure into clusters (26, 28–31, 33). Although such clustering would decrease the computational burden, it would also require additional steps to construct a kinetic model that preserves the correct intratemperature thermodynamics, because the states would now possess different relative free energies. We have chosen not to do clustering for this study. In our kinetic network model, the REMD snapshots can be visualized as nodes in a network. The edges connecting these nodes represent allowed conformational transitions, and the allowed conformational transitions are determined by the structural similarity of the two states involved (10, 26, 28). This network structure can be viewed as an approximate representation of effects caused by frictional interactions with the environment (10). In this work, we used a measure of similarity based on closeness in Euclidean distance space. For each state, 42 Cα–Cα distances were calculated, corresponding to all $(i, i +$

3), $(i, i + 7)$, $(i, i + 9)$, $(i, i + 10)$, $(i, i + 12)$, and $(i, i + 13)$ residue pairs. Each state was represented as a point in this 42-dimensional Cα–Cα distance space, and structural similarity was defined as the Euclidean distance in this space. The structural similarities for all sequential pairs of MD snapshots along a given REMD walker having the same temperature were tabulated. The cutoffs specifying when two structures are sufficiently similar to be connected by an edge were chosen to be proportional to the 99th percentile of the distribution of the structural similarities for each temperature. The results described below are not strongly sensitive to the choice of cutoff.

Any two states with the same REMD temperature were joined by an edge if their structural similarity was less than or equal to the cutoff value at that temperature. A state with an REMD temperature T_i and another at the neighboring lower temperature T_{i-1} were joined by an edge if their structural similarity was less than or equal to the cutoff at the higher temperature T_i. Thus, if a transition from state i to state j was allowed, then the reverse transition from j to i was also allowed. No connections were allowed between conformations not belonging to adjacent REMD temperatures. The resulting kinetic network has 800,000 nodes and 7.374×10^9 edges.

As in previous work (10, 28–31, 33), we model the kinetics on our graph as a jump Markov process with discrete states (45), where each (directed) edge is assigned a microscopic rate constant. The amount of time spent in a given state is then an exponential random variable with mean equal to the inverse of the sum of the microscopic rate constants for the edges leaving that node, and after that time the system jumps to one of the connected states with a probability proportional to the microscopic rate constant for that transition (also known as the Gillespie Algorithm). The time-dependent vector of instantaneous probabilities $\boldsymbol{P}(t)$ is given by the master equation

$$\mathrm{d}\boldsymbol{P}(t)/\mathrm{d}t = \boldsymbol{K}\boldsymbol{P}(t), \qquad [1]$$

where $\boldsymbol{K}$ is the transition matrix of microscopic rate constants (10). Although this equation can in principle be solved by the eigendecomposition of $\boldsymbol{K}$ or the direct numerical solution of the coupled differential equations, in this work we simulate actual realizations of paths that satisfy this master equation. Such simulations allow us to more directly characterize the sequence of events of folding.

The states reflect the thermodynamics of the system in that their density at each temperature is related to the energy according to the Boltzmann distribution. We choose microscopic rate constants so that the kinetic model preserves this distribution at equilibrium for each replica exchange temperature. We make the equilibrium probability of being in any given state equal to that of being in any other at the same replica exchange temperature. Such an equilibrium can be arranged by making the microscopic rate constant for each transition to be equal to the rate constant for the reverse process, for example, by setting all of the intratemperature microscopic rate constants equal to 1 in arbitrary time units.

Although the choice of equilibrium probabilities for states from the same replica exchange temperature is intuitively clear, it is less obvious how to choose the relative equilibrium populations for states from different temperatures. We would like the probability of being in states extracted from different replica exchange temperatures not to be uniform, but rather to be peaked near a "reference" or "simulation" temperature that is a parameter of the kinetic model. In the present work, we choose the equilibrium populations of states from different temperatures according to a model based on the distribution of instantaneous temperatures in a classical system. Specifically, the distribution of kinetic energy E of a molecule consisting of N atoms at a reference temperature T_0 is given by (46)

$$P(E) = \frac{1}{kT_0\Gamma\left(\frac{3N}{2}\right)} \left(\frac{E}{kT_0}\right)^{\frac{3N}{2}-1} e^{-E/kT_0}. \qquad \textbf{[2]}$$

Motivated by the fact that the average kinetic energy at a given temperature T is $(3N/2)\,k\,T$, we define the instantaneous temperature corresponding to a kinetic energy E to be $2E/(3Nk)$. If we substitute instantaneous temperature for instantaneous kinetic energy in Eq. **2**, we obtain

$$P(T) \propto \left(\frac{T}{T_0}\right)^{b-1} \exp\left(-b\,\frac{T}{T_0}\right), \qquad \textbf{[3]}$$

where b corresponds to $3N/2$ for a classical system. We choose the microscopic rates connecting conformations at different temperatures such that given a parameter b and a reference temperature T_0, the equilibrium probability of residing in a state from replica exchange temperature T will be given by Eq. **3**. This model allows a given path to sample states having instantaneous temperatures above or below the reference temperature T_0 in a physically realistic manner, and the range of accessible temperatures is controlled by the parameter b. The equilibrium distribution of Eq. **3** can be satisfied by choosing the pairs of microscopic rate constants k_{ij} and k_{ji} such that $k_{ij}/k_{ji} = P(T_j)/P(T_i)$, where k_{ij} is the rate constant for the transition from state i from replica exchange temperature T_i to state j from replica exchange temperature T_j. For the simulations described here, all of the intertemperature microscopic rate constants corresponding to a decrease in temperature were set equal to 1 (the same rate as for intratemperature transitions), whereas the rate constants for the reverse transition were determined by $k_{ij} = P(T_j)/P(T_i)$. In the subsequent discussion, the replica exchange temperature corresponding to a given state will be referred to as its instantaneous temperature. To summarize, our kinetic model consists of equal forward and reverse microscopic rate constants for all allowed intratemperature transitions, and microscopic rates $k_{ij} = P(T_j)/P(T_i)$ given by Eq. **3** for all allowed intertemperature transitions. In the calculations below we have used $b = 20$ for computational convenience. Although larger values of b more accurately reproduce the canonical potential of mean force with respect to two principal components that separate the helical and hairpin conformations of the G peptide (39), we have found by using network analysis methods that the qualitative results described below are not significantly changed.

Results and Discussion

The G Peptide Has Apparent Two-State Kinetics After a Small Temperature Jump Perturbation. Previous experimental work in the Eaton laboratory (35) has shown that the time dependence of loss of hairpin structure in the G peptide after a small temperature-jump perturbation is well fit by a single exponential. To confirm that our kinetic model is consistent with this previous experimental kinetic work, we performed a series of simulations modeling this temperature-jump experiment. We began each simulation by constructing an ensemble of starting points distributed according to an equilibrium distribution with T_0 ranging from 300 to 615 K. We then performed a Markov process simulation for 2,000–5,000 time units beginning from each starting point by using a reference temperature 60° higher than the temperature used to construct the initial starting point ensemble. For each temperature, the number of trajectories residing in a β-hairpin state were monitored as a function of time, and the resulting fractions of secondary structure were plotted as a function of time (Fig. 2). In all cases, the loss of hairpin structure is fit well by single exponential decay with the exception of a small initial "burst phase." Our results are qualitatively consistent with experimental observations (35).
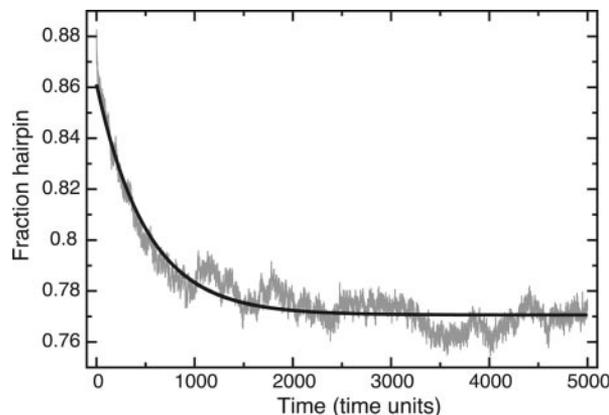
Fig. 2. Fraction of β-hairpin along an ensemble of 3,828 Markov process trajectories after a 60° temperature jump (gray line). The ensemble was constructed so that at time 0 it corresponded to the equilibrium distribution with $T_0 = 300$ K. The black curve corresponds to the least-squares fit to an exponential decay of the form $A\,e^{-kt} + B$ with parameters $A = 0.0906$, $B = 0.7706$, and $k = 0.001966$. The fraction β-hairpin at infinite time agrees with the expected fraction β-hairpin at equilibrium at 360 K. It should be noted that this fraction of native conformation is not identical to that shown in Fig. 1, because the latter does not incorporate data from multiple temperatures.

The G Peptide Has an α-Helical Intermediate During Folding from Coil Conformations. Protein folding is a process by which conformations without identifiable secondary structure adopt a native conformation. To study this process in the G peptide with our kinetic network model, we performed a temperature quench experiment similar to the temperature-jump experiment described above but for which the starting ensemble was chosen from the equilibrium distribution at $T_0 = 700$ K, and the simulation was run at a reference temperature of 300 K. The fraction of α-helix and β-hairpin states as a function of time (Fig. 3) displays a rapid rise in the amount of α-helix initially, which reaches a maximum and then decreases. Simultaneously, the amount of β-hairpin rises initially at a rapid rate, then continues to rise with a slower rate similar to the rate of decrease in the fraction of α-helix. This finding is suggestive of a mechanism in which there are a small number of fast direct paths from unfolded coil states to the β-hairpin, but that the majority quickly fold to α-helical states, which then convert into β-hairpins on a longer time scale. A similar phenomenon is not observed for the unfolding process: temperature-jump simulations from 300 to
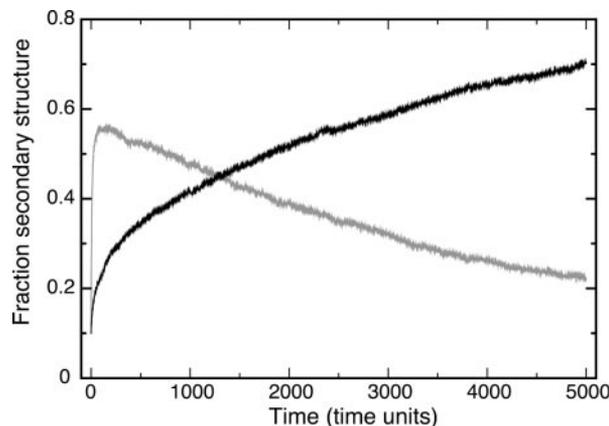
Fig. 3. Fraction of β-hairpin (black) and α-helix (gray) along an ensemble of 2,000 Markov process trajectories in which the initial ensemble corresponds to the equilibrium distribution at a reference temperature of $T_0 = 700$ K, and the trajectories were run at a reference temperature of $T_0 = 300$ K.

700 K do not show appreciable α-helix formation. That the folding and unfolding kinetic paths are different reflects the quite different nonequilibrium cooling and heating conditions that are being simulated.

We can assign approximate absolute time scales to the processes observed here by equating the "two-state" equilibration rate observed after a small temperature perturbation (Fig. 2) with that experimentally observed (6 μs) (35). Based on this finding, the appearance of β-hairpin in Fig. 3 has a time constant of ≈2,500 time units, which would correspond in physical units to ≈50 μs, whereas the rapid initial formation of α-helix occurs with a time constant of 9 time units or ≈180 ns. These rates are in qualitative agreement with experimental observations (35).

To confirm that this mechanism is indeed the basis for the "ensemble averaged" observations of Fig. 3, we performed an analogous single-molecule quenching experiment in which we chose ≈4,000 states at random from among the coil states at 690 K and used each as a starting point for a simulation at a reference temperature of 300 K. Each simulation was stopped when the random walk remained in a β-hairpin state for at least 50 consecutive steps. Each trajectory was then analyzed in terms of the total amount of time spent in α-helical conformations and the maximal extent of α-helix formation during the course of the folding trajectory. Only 9% of the trajectories reach the β-hairpin macrostate without passing through any α-helix-containing states. This finding confirms that in our kinetic network model the β-hairpin folding mechanism consists of two parallel pathways: the direct formation of the β-hairpin structure from coil states and the formation of α-helical conformations, which then interconvert into β-hairpins. On average, we found that coil starting points that formed the β-hairpin directly tended to have a somewhat lower overall radius of gyration than those that formed an α-helix first.

To qualitatively study the time scales of these processes, we performed an analysis of the first-passage times from the above quenched folding simulations. The distribution of first-passage times is multiexponential, as can be seen in the large initial spike in Fig. 4A. Furthermore, the distributions of first-passage times partitioned according to the maximal extent of helix along each folding pathway (Fig. 4B) shows that paths that have more extensive helices require on average significantly more time to reach the hairpin conformation than paths that have little or no helical content.

These observations can be rationalized by using the following schematic model. High-temperature coil structures under the influence of a low reference temperature are quickly cooled to room temperature because of their very low probability under Eq. 3. However, there are many more pathways to reach helical rather than hairpin conformations starting from those high-temperature coil states. This process quickly creates a small population of native hairpin states that folded directly from the coil and a larger number of metastable helical states. This model is also consistent with the observation of no excessive helical content along the unfolding pathway, because given a strong pull upward in temperature there exist direct routes for a hairpin structure to reach coil. Once the system has reached a helical state, though, there are no pathways accessible for a helix to convert into a hairpin at low temperature. Therefore the system must wait for a sufficiently large thermal fluctuation to occur to find a path out of the metastable helical conformation. The latter part of this scenario can be confirmed by a closer analysis of the connectivity structure of the kinetic network.

Two states (nodes) in the network are said to belong to the same connected cluster if there exists at least one path connecting the two nodes. Given our similarity criterion cutoff, such a path exists between any two nodes if we consider the complete system of all 800,000 states. However, such a path may not exist if we remove
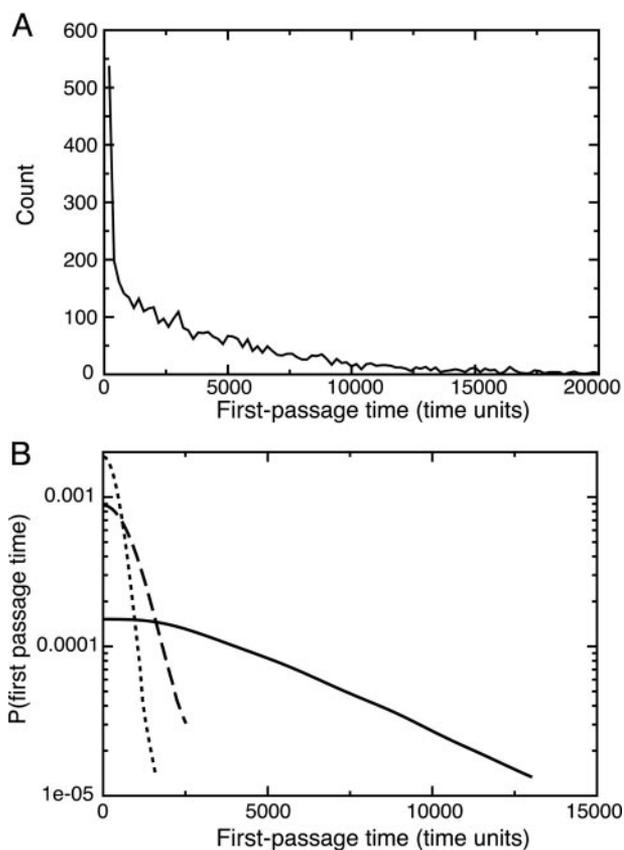

Fig. 4. First-passage time distributions corresponding to 4,214 trajectories run at 300 K beginning in randomly chosen coil states with instantaneous temperature of 690 K. Paths were terminated when they reached the hairpin macrostate. (*A*) Histogram of the overall distribution of first-passage times. The initial spike arises from direct paths to the hairpin state. (*B*) First-passage time distributions conditional on the maximal extent of α-helix formation along the path (solid line, complete helix; dashed line, partial helix; dotted line, no helix) are shown. The first-passage time distributions are approximately exponential except for the very short time limit. For first-passage times above ≈2,500 time units the distribution is almost completely dominated by paths that reach the native structure via a complete helix.

high-temperature routes. For example, if we only consider those states having instantaneous temperatures of 400 K or lower, we find that there are two major clusters. These account for 95% of the states and consist mainly of hairpin and helical states. Because none of the clusters observed for a 400-K cutoff temperature contain both a hairpin and a helical state, it follows that it is not possible to find a path connecting these two macrostates without accessing a state with an instantaneous temperature >400 K. In fact, the lowest temperature at which the hairpin and helical states coexist in the same connected cluster is 488 K, that is, slightly above T_m for the OPLS-AA/AGBNP effective potential model. Therefore, if the system is in the metastable helical macrostate at room temperature, it must experience a thermal fluctuation sufficiently large to bring its instantaneous temperature above T_m to convert into the hairpin macrostate.

A Molecular View of Kinetic Pathways. One of the advantages of the kinetic network model proposed here is that we are able to explore a large number of potential pathways that join two macrostates. The number of such paths will typically be extremely large. Furthermore, each state along the path has associated with it all of the atomic coordinates from the REMD simulation. Therefore, the molecular aspects of the
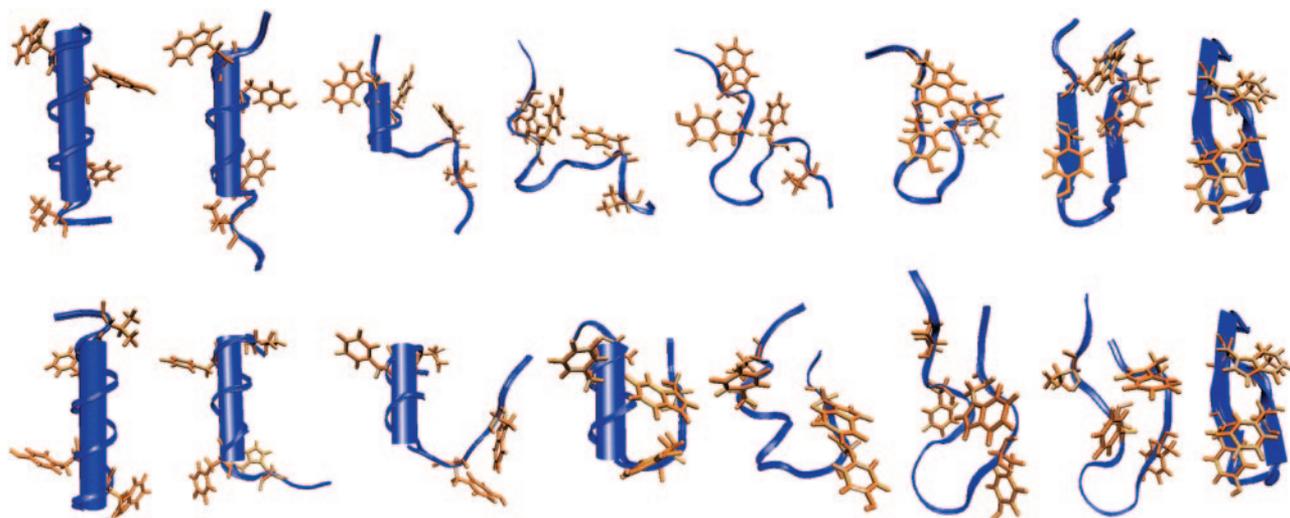
Fig. 5. Two possible pathways for the interconversion of an α-helix into a β-hairpin. Backbone trace is shown in blue, and the hydrophobic core residues (W43, Y45, F52, and V54) side chains are shown in yellow. (*Upper*) The path corresponds to an unraveling of the helix at both ends and formation of a β-turn from a residual turn of α-helix. (*Lower*) The path corresponds to an unraveling of one end of the helix, which loops back.

paths can be analyzed in detail. This ability allows us to explore the multitude of folding pathways that the system can potentially have at its disposal. One way in which this model can be used is to generate many paths by using Markovian kinetic Monte Carlo simulations. Such an approach with all-atom models has been useful for enumerating and quantifying the relative flux through parallel kinetic pathways in small systems (30, 31). Alternatively, it is possible to investigate thermodynamically favorable pathways by a detailed analysis of the structure of the kinetic network, for example, by searching for a small number of short paths connecting the two macrostates under the constraint that the instantaneous temperature remain below a predetermined maximum value. We use this approach to analyze pathways connecting the α-helix and β-hairpin macrostates in the G peptide.

Two short pathways that link the α-helical and β-hairpin macrostates without making use of microstates with an instantaneous temperature above 488 K are shown in Fig. 5. The path shown in Fig. 5 *Upper* involves the unwinding of both ends of the helix, leaving approximately one turn of helix in the middle of the molecule. This turn then serves as a nucleation point for the formation of the β-turn, which is stabilized by hydrophobic interactions between the side chains of Y45 and F52. The native hydrogen bonds nearest the turn then form, after which the remainder of the native hairpin structure forms. This pathway is similar to previously proposed mechanisms for the folding of the G-peptide β-hairpin from a coil state, which emphasize the formation of hydrophobic contacts before hydrogen bond formation (8, 12–14, 16, 20) and the persistence of the β-turn even in the unfolded state (20). The novel aspect of the path shown in Fig. 5 *Upper* is the preformation of the β-turn from a residual turn in an otherwise unfolded α-helix.

An alternative pathway (Fig. 5 *Lower*) involves the unwinding of the C-terminal half of the α-helix, which then loops back so as to be nearly parallel to the remaining helix. This proximity allows for the possibility of side-chain interactions between the helix and the C-terminal half of the molecule, including hydrophobic interactions between F52 in the helix and either W43 or Y45. This pathway is very similar to one previously identified by us on the basis of the analysis of the potential of mean force for the G peptide along two principle component degrees of freedom (39). In both pathways, it is clear that formation of native β-hairpin contacts can occur without the complete loss of helical secondary structure,

making the idea of the α-helix as an on-path intermediate in the formation of the β-hairpin physically plausible.

Conclusions

We have presented an approach to the study of protein folding that makes use of the combined power of REMD simulations with an all-atom effective potential and a kinetic network model for the kinetics. The kinetic network model is capable of following long time scale folding pathways of the G peptide under native conditions. Our observation of α-helical on-pathway intermediates in the folding of the G-peptide β-hairpin is physically plausible: high energy coil structures under folding conditions rapidly lose their excess energy. Those coil structures that are already somewhat compact and have accessible pathways will fold directly to the native state. More extended coil structures that are farther away from the native hairpin conformation are more likely to transiently fold to helical conformations, which, because they require the formation of only local contacts (47), are relatively easy to form provided that the side-chain secondary structure propensities do not strongly disfavor helix. α-helices have been experimentally observed as intermediates along the folding pathway of β-sheets in β-lactoglobulin (48, 49). Some previous computational studies of the G peptide have reported α-helix content (15, 16, 32, 39), whereas others have not (8, 13, 14, 17, 20). Although this variability may be caused by the fact that some studies focused on the unfolding reaction or on short time scale simulations, it may also be caused by differences in the helical propensities of different effective potential functions. This propensity difference could lead to changes both in the thermodynamic stability of the α-helix relative to the native hairpin and in the preferred pathways for folding. However, other simulation results show that the OPLS-AA effective potential does not seem to significantly overweight α-helical propensity (unpublished results and ref. 50).

Combining REMD simulations with kinetic network random walk models provides a powerful tool for the computational study of protein folding pathways, which is particularly effective for exploring large sets of diverse pathways. In addition, the pathways generated using this methodology could be used as starting points for further study using explicit solvent simulations with transition path sampling.

1. Eaton, W. A., Muñoz, V., Hagen, S. J., Jas, G. S., Lapidus, L. J., Henry, E. R. & Hofrichter, J. (2000) *Annu. Rev. Biophys. Biomol. Struct.* **29,** 327–359.
2. Dinner, A. R., Sali, A., Smith, L. A., Dobson, C. M. & Karplus, M. (2000) *Trends Biochem. Sci.* **25,** 331–339.
3. Fersht, A. R. & Daggett, V. (2002) *Cell* **108,** 573–582.
4. Kolinski, A. & Skolnick, J. (2004) *Polymer* **45,** 511–524.
5. Onuchic, J. N., Luthy-Schulten, Z. & Wolynes, P. G. (1997) *Annu. Rev. Phys. Chem.* **48,** 545–600.
6. Dill, K. A. (1999) *Protein Sci.* **8,** 1166–1180.
7. Kolinski, A., Ilkowski, B. & Skolnick, J. (1999) *Biophys. J.* **77,** 2942–2952.
8. Klimov, D. K. & Thirumalai, D. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 2544–2549.
9. Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E. & Shakhnovich, E. I. (2000) *J. Mol. Biol.* **296,** 1183–1188.
10. Ozkan, S. B., Dill, K. A. & Bahar, I. (2002) *Protein Sci.* **11,** 1958–1970.
11. Ozkan, S. B., Dill, K. A. & Bahar, I. (2003) *Biopolymers* **68,** 35–46.
12. Dinner, A. R., Lazaridis, T. & Karplus, M. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 9068–9073.
13. Pande, V. S. & Rokhsar, D. S. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 9062–9067.
14. Ma, B. & Nussinov, R. (2000) *J. Mol. Biol.* **296,** 1091–1104.
15. García, A. E. & Sanbonmatsu, K. Y. (2001) *Proteins* **42,** 345–354.
16. Zagrovic, B., Sorin, E. J. & Pande, V. (2001) *J. Mol. Biol.* **313,** 151–169.
17. Zhou, R., Berne, B. J. & Germain, R. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 14931–14936.
18. Brooks, C. L., III (2002) *Acc. Chem. Res.* **35,** 447–454.
19. Pande, V. S., Baker, I., Chapman, J., Elmer, S. P., Khaliq, S., Larson, S. M., Rhee, Y. M., Shirts, M. R., Snow, C. D., Sorin, E. J. & Zagrovic, B. (2003) *Biopolymers* **68,** 91–109.
20. Bolhuis, P. G. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 12129–12134.
21. Felts, A. K., Harano, Y., Gallicchio, E. & Levy, R. M. (2004) *Proteins* **56,** 310–321.
22. Sugita, Y. & Okamoto, Y. (1999) *Chem. Phys. Lett.* **314,** 141–151.
23. Elber, R. & Karplus, M. (1987) *Chem. Phys. Lett.* **139,** 375–380.
24. Bolhuis, P. G., Chandler, D., Dellago, C. & Geissler, P. L. (2002) *Annu. Rev. Phys. Chem.* **53,** 291–318.
25. Fersht, A. R. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 14122–14125.
26. Rao, F. & Caflisch, A. (2004) *J. Mol. Biol.* **342,** 299–306.
27. Ye, Y.-J., Ripoll, D. R. & Scheraga, H. A. (1999) *Comp. Theor. Polymer Sci.* **9,** 359–370.
28. Singhal, N., Snow, C. D. & Pande, V. S. (2004) *J. Chem. Phys.* **121,** 415–425.
29. Swope, W. C., Pitera, J. W. & Suits, F. (2004) *J. Phys. Chem. B* **108,** 6571–6581.
30. Swope, W. C., Pitera, J. W., Suits, F., Pitman, M., Eleftheriou, M., Fitch, B. G., Germain, R. S., Rayshubski, A., Ward, T. J. C., Zhestkov, Y. & Zhou, R. (2004) *J. Phys. Chem. B* **108,** 6582–6594.
31. Chekmarev, D. S., Ishida, T. & Levy, R. M. (2004) *J. Phys. Chem. B* **108,** 19487–19495.
32. Wei, G., Mousseau, N. & Derreumaux, P. (2004) *Proteins* **56,** 464–474.
33. Evans, D. A. & Wales, D. J. (2004) *J. Chem. Phys.* **121,** 1080–1090.
34. Blanco, F. J., Rivas, G. & Serrano, L. (1994) *Struct. Biol.* **1,** 584–590.
35. Muñoz, V., Thompson, P. A., Hofrichter, J. & Eaton, W. A. (1997) *Nature* **390,** 196–199.
36. Kobayashi, N., Honda, S., Yoshii, H. & Munekata, E. (2000) *Biochemistry* **39,** 6564–6571.
37. Honda, S., Kobayashi, N. & Munekata, E. (2000) *J. Mol. Biol.* **295,** 269–278.
38. Tsai, J. & Levitt, M. (2002) *Biophys. Chem.* **101–102,** 187–201.
39. Gallicchio, E., Andrec, M., Felts, A. K. & Levy, R. M. (2005) *J. Phys. Chem.,* in press.
40. Kitchen, D. B., Hirata, F., Kofke, D., Westbrook, J. D., Yarmush, M. & Levy, R. (1990) *J. Comput. Chem.* **11,** 1169–1180.
41. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. (1996) *J. Am. Chem. Soc.* **118,** 11225–11236.
42. Gallicchio, E. & Levy, R. M. (2004) *J. Comput. Chem.* **25,** 479–499.
43. Levy, R. M., Zhang, L. Y., Gallicchio, E. & Felts, A. K. (2003) *J. Am. Chem. Soc.* **125,** 9523–9539.
44. Su, Y. & Gallicchio, E. (2004) *Biophys. Chem.* **109,** 251–260.
45. Gillespie, D. T. (1992) *Markov Processes: An Introduction for Physical Scientists* (Academic, Boston).
46. Goodstein, D. L. (1975) *States of Matter* (Prentice–Hall, Englewood Cliffs, NJ).
47. Dill, K. A., Fiebig, K. M. & Chan, H. S. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 1942–1946.
48. Arai, M., Ikura, T., Semisotnov, G. V., Kihara, H., Amemiya, Y. & Kuwajima, K. (1998) *J. Mol. Biol.* **275,** 149–162.
49. Kuwata, K., Shastry, R., Cheng, H., Hoshino, M., Batt, C. A., Goto, Y. & Roder, H. (2001) *Nat. Struct. Biol.* **8,** 151–155.
50. Sakae, Y. & Okamoto, Y. (2004) *J. Theor. Comp. Chem.* **3,** 359–378.
51. Frishman, D. & Argos, P. (1995) *Proteins* **23,** 566–579.