Side-chain Dynamics and Protein Folding

Edo Kussell, Jun Shimada, and Eugene I. Shakhnovich*
Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts

ABSTRACT The processes by which protein side chains reach equilibrium during a folding reaction are investigated using both lattice and all-atom simulations. We find that rates of side-chain relaxation exhibit a distribution over the protein structure, with the fastest relaxing side chains located in positions kinetically important for folding. Traversal of the major folding transition state corresponds to the freezing of a small number of side chains, belonging to the folding nucleus, whereas the rest of the protein proceeds toward equilibrium via backbone fluctuations around the native fold. The postnucleation processes by which side chains relax are characterized by very slow dynamics and many barrier crossings, and thus resemble the behavior of a glass. Proteins 2003;52:303–321. © 2003 Wiley-Liss, Inc.

Key words: side-chain dynamics; protein folding; nucleation mechanism; glass transition; side-chain packing

INTRODUCTION

Theoretical approaches in protein folding have sought to discover how the three-dimensional backbone fold is coded in the sequence and acquired by the heteropolymer chain. Compounding the difficulties posed by backbone folding is the presence of amino-acid side-chain degrees of freedom (the $\chi$-angles). These degrees of freedom have been modeled explicitly only in molecular dynamics simulations and have generally been omitted in other treatments as a simplification. Recent progress in a Monte Carlo approach has made it possible to observe complete protein folding trajectories in an all-atom simulation that includes precise side-chain representation and dynamics.1 Those simulations demonstrated that accurate backbone topology is acquired by the protein long before thermodynamic equilibrium is reached (i.e., side chains do not reach equilibrium concomitantly with backbone folding). Full side-chain ordering cannot presently be observed computationally in high-resolution models because of the very long simulation times needed. As such, the complex dynamics which bring side chains to equilibrium have not been thoroughly investigated. The aim of this article is to elucidate those dynamics using both a simple model, which captures the basic, physical properties of the system, as well as an all-atom Monte Carlo folding simulation.

Two observations form the basis of our work: (i) it is well known that upon folding, buried side chains are usually found in a single, well-defined rotamer state,2–5 and (ii) steric interactions in the compact state severely restrict side-chain motion, preventing free rotamer interconversion.6,7 At which point in the folding process are the native side-chain rotamers found? Correct backbone topology and side-chain conformations may be acquired in a single, cooperative step. Alternatively, side-chain ordering may constitute a separate process that occurs during and after backbone folding.

We introduce a modification of the usual lattice model of protein folding that accounts for the two basic observations mentioned above. The model we obtain therefore possesses the minimal physical requirements for studying side-chain ordering. Additionally, the simulation is fast enough to observe full relaxation to equilibrium. We can study all aspects of side-chain ordering in this setting. To verify that our simple model gives meaningful physical insight, we compare some of its predictions with all-atom simulations of Protein G. We also provide a statistical-mechanical derivation of the thermodynamic behavior of the model in an appendix. Our main conclusion is that side-chain ordering proceeds by way of the fast freezing of folding nucleus side chains, organizing the protein topology, followed by slow relaxation of the rest of the side chains via backbone fluctuations.

MODEL

We modified the standard lattice model of protein folding8 by adding $n$ spin states to each monomer, mimicking the internal degrees of freedom of protein side chains. This representation is consistent with the observation that protein side chains usually populate discrete rotameric configurations.2,3 The state of each monomer is a number between 0 and $n - 1$. Of these $n$ states, only one state (the 0 state) was designated as native for each monomer. When two monomers came into contact during simulation, they interacted only if both were in their native spin state; a contact formed with one or both monomers in a non-native spin state did not contribute to the energy of the conformation (see Methods). We later modify these simple energetics and discuss the consequences.

Upon polypeptide collapse, side-chain motion is restricted because of the excluded volume effect.9,10 In order for side chains to repack in the protein interior, the backbone must perform a “breathing motion,”11 allowing side chains some extra room to move, and thus making...
certain side-chain configurations momentarily available. Any model of side-chain dynamics must incorporate this effect in some way. All-atom simulations contain this effect explicitly. In the lattice simulation, we incorporate this effect by introducing constrained dynamics of spins. In addition to the usual lattice backbone moves, we allow the spin states of a given monomer to interconvert only when there are no other monomers in contact with it (see Methods). Thus, when the chain is fully compact, the spins are frozen until a backbone fluctuation frees some monomers and allows their states to change. We use this constrained dynamics model in all lattice simulations, except when explicitly noted otherwise. Although models with side chains or spins have been previously investigated, our treatment of spin dynamics constitutes a significant departure from other studies.

A note on terminology: In our discussion of the lattice model, we have tried to use the term “spins” when discussing technical aspects of the simulation and the term “side chains” when discussing implications of the results to proteins, but in both cases we have protein side chains in mind.

METHODS
Lattice Model with Spin States

We use a standard three-dimensional lattice model in which each monomer occupies a single lattice site. For the 27-mer simulations, a sequence was designed to fold to a unique native conformation as described in Ref. 16. For the 48-mer simulation, a fast-folding sequence was obtained from a lattice protein evolution study described in Ref. 17. The standard Miyazawa-Jernigan parameter set was used to compute the energy of a conformation. Two monomers are said to be in contact if they are nearest neighbors on the lattice and are not sequence neighbors. Additionally, each monomer has n spins, where n is a parameter of the model. The spin state of a monomer is stored as a number from 0 to n − 1. The 0 state is the native state, whereas the states 1 through n − 1 are non-native. If n = 1, then all monomers remain native throughout the simulation, and the model is equivalent to the standard lattice model. In our first formulation, non-native spins do not contribute to energy: two monomers in contact will contribute to energy only if they are both in their native spin state (the 0 state). In our later formulation, non-native spin interactions are assigned a fraction of the native energy.

The standard cubic lattice move-set is used to evolve the backbone conformation, and a Metropolis criterion with temperature T is used to accept/reject moves. In addition to backbone moves, the spin states of the monomers are evolved. After each backbone move is attempted, n − 1 spin moves are attempted. At each such move, a random monomer is chosen. If the monomer is making more than c contacts with other monomers, its spin is not allowed to change. Otherwise, its spin is randomly flipped to one of the other n − 1 states, the change in energy of the conformation is computed, and the move is accepted/rejected based on the Metropolis criterion. The parameter c can take the values 0 through 4. When c = 4, spins can interchange freely and are not affected by the conformation of the backbone; we call this the unconstrained model. If c = 0, spins can interchange only if the monomer makes no contacts; we call this the constrained model. In this study we use the constrained model throughout, except for the results indicated by diamonds in Figure 2. Folding simulations are started from random backbone conformations generated by an infinite temperature simulation. The spins are initialized randomly.

All-atom Protein Folding Simulations

The folding data for protein G (pdb code: 1IGD) previously presented in Ref. 29 was used for the present analysis. In the all-atom Monte Carlo simulation that was used, all side-chain and backbone heavy atoms were represented as hard spheres. As such, the protein was simulated as a polymer with excluded volume interactions, where chain crossings are strictly prohibited. The energy of a conformation was given by the atom-atom Go energy

\[ E_G = \sum C(A, B)\Delta(A, B) \]

with \( \Delta(A, B) = 1 \) if the heavy atoms A and B were in contact and zero otherwise, C(A, B) was −1 if A and B were in contact in the native conformation, 1 if they were not, and ∞ if they were sterically clashed.

Fitting of Residue Relaxation Curves

We averaged the state of each monomer at each time step over all runs, assigning 1 if the residue was native and 0 otherwise. For all-atom simulations, 50 runs were used, and averaging over runs was performed by assigning 1 to each residue whose χ-angles were all within 60° of the native angles and 0 otherwise. A value of 1 thus corresponded to observing the native rotamer. All fits were done using the nonlinear least-squares Marquardt-Levenberg algorithm.

RESULTS

Results for a 27-mer Sequence: Thermodynamics and Kinetics

We tested the lattice model with 1, 2, 4, and 8 spin states per monomer using a 27-mer sequence designed to fold into a 3 × 3 × 3 cube (see Methods). The n = 1 model corresponds to the standard lattice model and is shown here for reference. The thermodynamics of the models is shown in Figure 1(A). All exhibit a cooperative temperature transition, with the transition temperature getting progressively lower as the number of spin states of each monomer increases. The lowered transition temperature is to be expected as the increased entropy of the model (due to more spin states) necessarily leads to some destabilization. The transition region becomes narrower as n increases, because of the increase in entropy of the unfolded state relative to the folded state.

We studied the kinetics of the various models by plotting the median folding time as a function of temperature [see Fig. 2(A–C)]. Note that throughout this article, time is always measured using the number of backbone moves attempted (i.e., the spin-flip moves are not counted). In the models with n > 1 the backbone can reach full nativity
Toptime for reaching native energy is approximately twice the median time for reaching the native backbone. Thus, the ratio of side-chain ordering to backbone folding time is about 1:1. The ratio is about 4:1 for the \( n = 4 \) model and more than 10:1 for \( n = 8 \). As \( T \) decreases, this ratio increases from zero (for \( T > T_{\text{opt}} \)) to some limiting value (for \( T \ll T_{\text{opt}} \)). The range of temperatures over which the ratio increases is approximately \( 0.8 T_{\text{opt}} < T < T_{\text{opt}} \) for all models. We find, then, that the side-chain entropy of the chain leads to a severe side-chain-ordering trap. This trap becomes particularly prominent as temperature is lowered and as the entropy of the model increases.

The mechanism for reaching equilibrium at temperatures lower than \( T_{\text{opt}} \) seems to be one in which the native backbone structure is formed, followed by side-chain ordering via backbone fluctuations around the native structure. It is possible, however, that the native backbone structure is reached during the folding trajectory but unfolds immediately because too few side chains are native. This, in fact, is the case even at low temperatures. At some point in time, however, the native backbone structure is reached with enough native side chains to remain stable long enough to allow the rest of the side chains to become ordered. It is the ordering of side chains after this stable native backbone is reached that we identify as an important kinetic step at temperatures below \( T_{\text{opt}} \). Accordingly, we plot the average time of the last pass to the native backbone conformation in all figures. The time of the last pass is defined as the first time the chain reached the native backbone without losing more than 50% of its native contacts before reaching the native energy. We found that our results did not change significantly when we varied the fraction of native contacts used in this definition.

As a control, we investigated an unconstrained model in which the spin states of each monomer could interconvert freely, regardless of how many contacts it was making. Such a model represents protein folding in molten globule conditions, in which side-chain rotamers can easily interconvert.\(^22\) This model is shown using diamonds in Figure

### Table I. Two-state Fits to Thermodynamic Data

| \( n \) | \( a_0 \) | \( a_1 \) | \( a_2 \) |
|------|-------|-------|-------|
| 27-mer |     |     |     |
| 1    | -111 | 518  | 16.9 |
| 2    | -71.7| 500  | 23.4 |
| 4    | -44.2| 433  | 26.3 |
| 8    | -18.2| 320  | 24.0 |
| 48-mer |     |     |     |
| 1    | -82  | 263  | 13.8 |
| 2    | -3.6 | 222  | 16.0 |
| 4    | 38   | 151  | 14.2 |

\(^1\)Thermodynamics shown in Figure 1 was fit using the form \( f(x) = a_2 + (a_0 - a_2)\exp(a_2 - a_1/T)/[1 + \exp(a_2 - a_1/T)] \), where \( a_2 \) is the native state energy for each model. For the 27-mer, \( a_2 = -1219 \); for the 48-mer, \( a_2 = -1361 \).
2(A–C). The overlapping open and filled diamonds at all temperatures, and for all \( n \), indicate that in the unconstrained model, the slow ordering of side chains is not observed at any temperature. Instead, once the backbone reaches nativity, any non-native side chains can immediately become native, and they do because it is energetically favorable. This control demonstrates, then, that the side-chain ordering trap is a feature of folding under conditions in which a tight native state, with low side-chain mobility, is the free-energy minimum.

Comparing the unconstrained models with \( n = 2, 4, \) and \( 8 \) to the standard lattice model \( [n = 1, \text{Fig. 2(D)}] \), we find at their respective optimal folding temperatures, the average time to reach the native energy is nearly the same. The unconstrained spin states, then, have only a small effect on the kinetics of backbone folding. On the other hand, the folding time of the constrained models at \( T_{opt} \) is significantly slower than that of the \( n = 1 \) model. The \( n = 8 \) model, for example, folds 100 times more slowly than the \( n = 1 \) model at \( T_{opt} \). This makes sense in light of the observation that the constrained models are more difficult to nucleate: if a proper nucleus forms and one or more of its spins are in a non-native state, the nucleus will break apart because the energy of each of its contacts is crucial for its ability to function as a nucleus. The spins cannot interconvert while the nucleus contacts are present, and therefore the nucleus dissolves. In the unconstrained model, the non-native spins of the nucleus can become native with the nucleus intact, so the folding time in this model is minimally affected by the presence of spins.

**Results for 48-mer Structure**

Having observed that folding to the correct backbone structure occurs significantly before spin ordering, we asked the following question: Do some residues reach their native spin state faster than others, and if so what determines the ordering rate? We decided to use a 48-mer structure whose folding in the standard lattice model has been studied exhaustively. In order to maximize the temperature range in which we could study folding of this
structure, we used a sequence that has been optimized for fast folding. In addition to having a fully characterized nucleus, using a 48-mer sequence allowed us to see whether our results were sensitive to the size of the structure.

The thermodynamics of the 48-mer sequence is shown in Figure 1(B). The kinetics for the 4-state model are shown in Figure 3. Because folding times for the 48-mer are much longer than for the 27-mer, the temperature range shown here is narrower than the one shown in Figure 2(B). For the 27-mer with \( n = 4 \), we found that side-chain ordering was 4 times slower than backbone folding at 85\% of Topt. For the 48-mer, at 85\% of Topt, side-chain ordering is approximately twice as slow as backbone folding. Because of prohibitively long simulation times at lower temperatures, we cannot make a complete comparison with 27-mer results. It appears, however, that in the vicinity of Topt, 27- and 48-mer models have qualitatively similar behavior.

**Side-chain Ordering Rates**

To obtain individual side-chain ordering rates for each residue, we performed many long folding runs. For each residue we averaged its spin state over all runs: we assigned a value of 1 to the native spin state of a given residue and a value of 0 to all other spin states and at each time-step averaged these values over runs. Two traces obtained after averaging are shown in Figure 4. We fit a single exponential to each trace and obtained time constants for each of the 48 residues.

The distribution of rate constants for two temperatures is given in Figure 5, and the fast residues are labeled by number. The first striking feature is that these distributions span two orders of magnitude. At the lower of the two temperatures \( T = 7.4 = 81\% \) Topt, most residues, but not all, exhibit slow relaxation rates, as seen by the sharp peak near zero. At the higher temperature of Topt = 9.1, the height of the peak is reduced and more residues are seen with faster rates.

A few residues have exceptionally fast rates at both temperatures. Interestingly, most of these residues belong to the folding nucleus for this structure that was identified in another study using the standard lattice model. In Figure 6 we show the 48-mer structure colored by rate of side-chain freezing at \( T = 7.4 \), and we indicate the original nucleus by large spheres. Of the 10 fastest residues that become fully ordered at \( T = 7.4 \), 7 belong to the old folding nucleus. We note that because of the presence of spins in our model, the folding nucleus in our simulations may be different from the previously characterized nucleus. Although some of the nucleus positions no longer seem to be kinetically important in our model, a strong signature of the old nucleus has remained. Importantly, with the exception of residues 9 and 25, all of the fast positions that reach full nativity are located in or near the original nucleus. We find this to be compelling evidence of strong correlation between fast positions and the present folding nucleus. A complete characterization of the folding nucleus in our model should be performed in the future.

It appears, then, that at temperatures at or below Topt, a small group of residues reaches full nativity quickly, thus organizing a critical piece of structure which remains stable, allowing the rest of the chain to gradually order its side chains. At \( T = 7.4 \), the formation of the stable piece traps many spins in non-native states, which take a very long time to reorganize via backbone fluctuations. On the other hand, at Topt more residues are found in the fast tail of the rate distribution (Fig. 5), indicating that backbone fluctuations are sufficient to allow side-chain ordering to occur more quickly once enough of the native structure has formed. Additionally, the higher temperature requires a larger amount of native structure to be formed in order to remain stable.

**Ensemble Relaxation**

Another way to see that side-chain dynamics becomes markedly different as temperature is lowered is given in Figure 7. The red line indicates the equilibrium energy at each of the two temperatures; the green and blue lines are a time trace of energy averaged over all runs. The median time to form the stable native backbone is \( 9.3 \times 10^6 \) at \( T = 9.1 \) and \( 5.3 \times 10^7 \) at \( T = 7.4 \), and is marked by an arrow in the figure. For \( T = 9.1 \) the arrow indicates that at the time of native backbone formation, the energy of the chain is already very close to its equilibrium value. In contrast, at low temperature there is a significant gap between the energy of the folded chain and the equilibrium energy.

We tried to fit the relaxation of energy by a single exponential as well as a double exponential. Both fits converged to the same curve, which is shown as a dashed line in Figure 7. The fit is poor at any time scale. At short times, the data may resemble a single-exponential process, but it relaxes very slowly at long times. We found that a single power-law (solid curve) fits all of the data for \( T = 9.1 \), which spans four decades in time. Residuals for both fits are shown in Figure 8.
A stretched-exponential of the form \( \exp(-b(t-t_0)^\alpha) \) fit the tail of the distribution \( (10^6 < t < 2 \times 10^8) \) equally well, but was not appropriate at short times. Interestingly, \( \alpha \) tended to zero during nonlinear fitting, whereas the quantity \( b \alpha \) remained constant. This indicates that the appropriate fit for the tail is a power-law, since for \( \alpha \to 0 \), \( \exp(-b(t-t_0)^\alpha) \sim (t-t_0)^{-\alpha} \). At \( T = 7.4 \), a single exponential is clearly inappropriate. Although a power-law fits well in the tail (not shown), we cannot rule out other possibilities because full relaxation could not be observed.

It is also interesting to follow the relaxation of the number of native backbone contacts (backbone \( Q \)) as shown in Figure 8. We notice that backbone relaxation is likewise nonexponential, and a power-law fits very well. The exponents for both backbone and energy relaxation, \( -1.29 \) and \( -0.94 \), respectively, are close. If we calculate the amount of time to reach within 1\% of either native backbone \( Q \) or native energy, we find times of \( 3.3 \times 10^7 \) and \( 4.6 \times 10^7 \), respectively. Backbone relaxation is thus somewhat faster than side chains, but the difference is small at this temperature, consistent with the calculation based on median folding times given in Figure 3.

**Effect of Non-native Rotamer Interaction**

In our original formulation, a contact contributed energetically only if both monomers were in their native spin state. To model protein side-chain energetics more faithfully, we added a non-native rotamer interaction in the following way: if two monomers are in contact and at least one of them is in a non-native spin state, we assign energy...
$E$ to this contact. If $E$ is the native interaction energy between these two monomers, we choose $E = E$ when $E > 0$, and choose $E = qE$ when $E < 0$, where $0 < q < 1$ is a uniformly distributed random number. This choice is based on the following picture. If two residues have an unfavorable interaction, their native rotamers would tend to minimize that repulsion and other rotamer configurations are not likely to decrease the repulsion; we therefore assign repulsive interactions their full contribution when non-native spins interact. If the pair has a normally favorable interaction, it arises either from directional polar interactions or from a hydrophobic interaction, both of which would diminish if side chains are in a non-native state; we assign a random fraction of the native energy in this case. The random fractions are fixed before simulation.

The thermodynamics of the new model (which we will term the non-native model) are shown side-by-side with the thermodynamics of the original model in Figure 9. To understand the changes resulting from addition of non-native spin interactions, we can consider the simple two-state approximation of thermodynamics. In this approximation, the protein is either folded or unfolded, and has free-energies $E_F - TS_F$ and $E_U - TS_U$ in these states, respectively. The transition temperature is found to be

$$T_\ell = \frac{E_U - E_F}{S_U - S_F}.$$
and the width of the transition is proportional to $\left(\frac{\partial E}{\partial T}\right)^{-1}$ evaluated at $T = T_f$:

$$\left(\frac{\partial E}{\partial T}\right)^{-1}
\quad\bigg|_{T = T_f} = \frac{T_o}{(E_o - E_f)^2} = \frac{4T_o^2}{(E_o - E_f)^2} = \frac{4}{(S_o - S_f)^2}.$$  \hfill (1)

Introducing non-native rotamer interactions will have a far greater effect in the folded state than in the unfolded state, because the unfolded state contains few contacts. In the folded state, non-native interactions will lead to increased entropy because more spins are now allowed to be non-native than before, without significantly destabilizing the structure energetically. The quantity $S_o - S_f$ is diminished, and therefore $T_f$ will be higher, as can be seen in Figure 9. This also results in a wider transition region, as seen via Equation 1.

We can further characterize the equilibrium ensembles in both models by plotting average backbone $Q$ vs. average spin $Q$ as temperature is varied (Fig. 10). In the original model, full backbone ordering necessitates almost complete side-chain ordering. In the non-native model, we find a striking difference: the backbone can become almost fully ordered with many disordered spins. There then follows a temperature range in which the excess spin entropy is quenched to zero, and this brings the average backbone $Q$ to 1. A dramatic manifestation of this effect is seen for the 27-mer with $n = 4$, in which a backbone $Q$ of 0.97 is achieved with only 50% of the spins ordered. Reaching a backbone $Q$ of 1 then requires ordering of the rest of the spins, corresponding to the long plateau in the graph.

The length of the plateau is determined by the energy gap between the native state and the lowest energy...
misfolds. When a disordered spin is introduced in the native structure, it results in some energetic destabilization. As long as the total energy of the spins in the native structure is below the continuous spectrum of the sequence, the given spin configuration contributes to the folded state. The bigger the gap, the more spins may be disordered in the native state. We can infer from Figure 10 that the 27-mer structure has a bigger effective gap (e.g., measured via $Z$-score) than the 48-mer structure. This is in fact correct, because the 27-mer sequence was obtained using a stability-maximizing method, whereas the 48-mer sequence was optimized for folding speed rather than stability. Additionally, the disordered spin configurations stabilize the folded state entropically. Thus when $n$ is increased, this stabilization increases in turn, and more spins can be disordered in the highly native backbone. This results in a longer plateau when comparing $n = 4$ to $n = 2$ for both 27-mer and 48-mer sequences in Figure 10.

These results indicate that the non-native rotamer interactions act to decouple backbone folding from spin ordering. We would therefore assume that this would be manifested in kinetics as well. In fact the effect is very noticeable, to a degree that we could not observe full relaxation to equilibrium in the $n = 4$ model even at optimal folding temperature. We present results for the $n = 2$ model near $T_{opt}$ in Figure 11. Although energy relaxation remains non-exponential in this model (see residuals of exponential fit), we find that backbone folding now exhibits single-exponential kinetics. This is further evidence of the partial decoupling between spins and backbone. The backbone reaches equilibrium at approximately $10^7$ steps, whereas spin relaxation is at least 2 orders of magnitude slower. In the original model near $T_{opt}$, backbone and spin relaxation times were on the same order of magnitude (see Fig. 8).

Protein folding experiments on small, single domain proteins have generally recorded single-exponential kinetics for backbone folding. Relaxation to equilibrium is a far more difficult observation to make in experiments, as discussed below. Of the two models presented here, the model with non-native rotamer interactions is clearly more protein-like in that exponential backbone relaxation is observed. This result is consistent with the idea that the exclusion of water from the protein core is the main driving force in the folding reaction, and this process is largely independent of side-chain orientations. A large energetic contribution comes simply from bringing hydrophobic groups together, without attaining perfect packing. The non-native rotamer interactions account for this effect, whereas the original model requires perfect packing and is therefore somewhat unrealistic.

**Varying the Number of Spin Moves**

In defining our lattice model, we have arbitrarily fixed at $n = 1$ the number of spin flips attempted after each backbone move. It is critical to verify that this choice does not affect our results. To this end, we varied this number and plotted, in Figure 12, the median time (measured in backbone steps, as before) for reaching both native backbone and native energy. As the number of spin moves increases, we see that the median time decreases and eventually plateaus. The plateau exists because for a large number of spin moves, the spins are essentially at equilibrium for the given backbone. Further increase in the
number of spin moves will not change the ensemble and therefore, the folding rate. The change in folding time between 1:1 ratio (spin:backbone moves) and 64:1 ratio is approximately 10% at $T_{opt}$ and 20% at a lower temperature ($<75\%$ of $T_{opt}$). Although the overall folding time is thus somewhat faster, the nature of the folding kinetics do not appear to be affected. We found that the ratio of the median times for reaching native backbone and native energy remained constant as the number of spin moves was increased. Importantly, this result did not change as we varied either temperature or number of states per monomer.

The number of spin moves at which the folding time is no longer sensitive to this parameter (i.e., the plateau is reached) is approximately $8n$, where $n$ is the number of states in the model. Ideally, we would like to run simulations at this value. This would come at a big computational cost, however, because simulation time scales almost linearly with number of spin moves, whereas the folding time decreases by only 10%. We therefore continued to run simulations using our original choice of $n - 1$ spin moves per backbone move. This seems reasonable given our demonstration that only the overall folding time is sensitive to our choice, whereas the kinetics of side-chain versus backbone ordering are not. As a final test, we repeated the collection of ensemble folding data as given in Figure 7 using $8n$ spin moves per backbone, and found a nearly identical time trace. We concluded that phenomena of interest to us in this work are not affected by this parameter.

**Side-chain Relaxation in All-atom Simulations**

Because lattice models can give only a schematic view of the folding process, we proceeded to investigate side-chain dynamics in an all-atom simulation of Protein G, an $\alpha/\beta$ protein that has featured in numerous experiments. A full characterization of the folding kinetics of this protein, as well as simulation methodology and extensive comparison to experiments, are given in Ref. 29. Our goal here is to see how the results obtained from our simplified lattice model compare with a much more realistic simulation and whether the same kind of analysis can shed light on the kinetics of a real protein.

In the lattice model, a set of microscopic dynamics were postulated for the spin states of each residue. In the all-atom simulation, all side-chain atoms and torsions are modeled explicitly. Because rotations around side-chain $\chi$ angles are continuous, inter-conversion between side-chain rotamers becomes restricted when a residue is buried. Slowing down of side-chain dynamics upon compactification emerges from the excluded volume interaction in

![Fig. 10. Average backbone Q vs. average spin Q plotted as temperature is varied for various models. Measurements of averages were done at equilibrium as described in Figure 1. Backbone Q is the fraction of the total number of native contacts present, and spin Q is the fraction of spins that are native. The solid lines correspond to the model with non-native interactions, whereas the dotted lines correspond to the original model.](image)
We analyzed the 50 folding trajectories of Protein G described in Ref. 29. These were started from random backbone and side-chain conformations and simulated at the same fixed temperature. All runs were terminated after $2 \times 10^9$ steps, by which time 88% of the runs reached the native backbone fold to 3 Å. We applied a time series analysis similar to the one we used for the lattice model: at each time step and for each residue we recorded a value of 1 if the side-chain was in its native rotameric state and a value of 0 otherwise. We averaged these values for each residue over the 50 trajectories and fit a two-state kinetic model to the resulting traces. The parameters for the fits for each residue are given in Table II, and two sample fits are shown in Figure 13.

The residues with fastest ordering rates are shown in Figure 14. Interestingly, they are located in topologically important positions: phenylalanine 30 and leucine 5 have a strong hydrophobic interaction that secures the first strand of the first hairpin against the helix; phenylalanine 52 contacts tyrosine 45, stabilizing the second hairpin internally, as well as phenylalanine 30, bringing the second hairpin into position with the helix.

The transition state for Protein G folding has been studied previously by obtaining conformations that have a 50% probability to fold during a tiny fraction of the entire folding time.29 This is currently the most rigorous characterization of folding transition states available in simulation studies.30 From this analysis, it was concluded that the rate-limiting step for reaching the native backbone involved the formation of a specific nucleus31 consisting of...
six residues: Y3, L5, F30, W43, Y45, and F52. As shown in Table II, these same residues have the fastest side-chain ordering rates.

The data obtained from the all-atom simulation are thus in excellent agreement with lattice results: there is a wide distribution of residue relaxation rates, with the fast residues located in topologically important positions. Remarkably, the nucleus residues are precisely the ones with fastest ordering rates. This observation suggests that measurement of side-chain ordering rates may be a good method for gauging participation in the folding transition state.

DISCUSSION

Extracting information about the dynamics of individual side chains is relatively easy in computational studies and veritably challenging in experiments. There are several difficulties to overcome in experiments. First, specific probes that measure properties about a single residue are scarce: tryptophan can be probed by fluorescence, and cysteine can be probed by thiol-disulfide exchange. Although dynamic NMR techniques can in principle report on many residues simultaneously, their application requires very slow folding reactions. Hydrogen exchange experiments can report on the protection of individual backbone amide groups, but backbone protection factors do not directly measure side-chain mobility. Second, it is desirable to have probes in several different parts of a structure in order to measure the distribution of side-chain rate constants over the whole fold. This, again, is in principle possible but usually requires introducing sequence mutations (adding a tryptophan or cysteine). Results must therefore be handled with care because the structure and folding pathways may be altered in subtle ways from sequence to sequence. Finally, the presence of kinetic intermediates in the folding of many proteins complicates analysis considerably.

Several recent studies have attacked the side-chain dynamics question using a variety of techniques. Staniforth and coworkers used a form of cystatin in which disulfide bonds were reduced, thus creating a molten globule whose compactness and unfolding properties were similar to folded wild type, but whose side-chain mobility was significantly increased. The size of the rate-limiting barrier for folding of the two forms was measured and found to be similar. The authors conclude that because the reduced and wild type forms differ mainly in side-chain mobility, whereas the barrier height for folding is the same, the immobilization of side chains occurs after the major folding transition in wild type cystatin. Additional
experiments on cystatin are probably needed in order to completely solidify the argument. Specifically, the connection between fluorescence quenching upon folding and full side-chain immobilization in wild type cystatin has not been made. Any conclusions about side-chain immobilization rest on the assumption that nativity of tryptophan fluorescence gives information about side-chain dynamics across the entire core.

Ha and Loh\(^3\) introduced cysteine mutations in several key places in apomyoglobin and, using pulsed thiol-disulfide exchange at different times during the folding reaction, measured the progression of side-chain ordering at each site. They found that certain locations, stabilizing the fast-forming folding intermediate, were as well packed as native protein long before folding was complete. It would be interesting to obtain similar site-specific time courses for other positions to see whether positions that become ordered in the postintermediate step exhibit a distribution of relaxation times.

Our finding that nucleus residues become well-ordered fastest is consistent with recent experimental results on the Fyn SH3 domain.\(^3\) Mutations were made in hydrophobic core positions of the protein, and folding kinetics were measured. The change in activation free-energy for folding and unfolding, \(\Delta G^\text{f-s}\) and \(\Delta G^\text{f-u}\), respectively, was computed for each mutated sequence, and these numbers were plotted against the change in side-chain volume due to the mutation. The authors assumed that in a given ensemble (transition state or folded state), disruption of well-packed residues will lead to free-energy changes that are largely independent of residue volumes. This assumption is guided by the intuition that well-packed positions participate in many specific interactions, the exact nature of which will determine the change in free energy when the residue is mutated. Poorly packed residues are nonspecifically stabilized, and therefore their free-energy contribution will be strongly correlated with changes in volume. It was found that \(\Delta G^\text{f-s}\) showed little correlation with volume changes, consistent with the fact that most core positions in a folded protein are highly ordered. Interestingly, the authors observed that \(\Delta G^\text{f-u}\) was strongly correlated to volume changes for non-nucleus positions and only weakly for nucleus positions. They concluded that nucleus positions are well packed in the transition ensemble.

In an elegant series of experiments using time-resolved fluorescence anisotropy measurements, Sridevi and coworkers\(^2\) demonstrated that barstar’s tryptophan 53 becomes fully ordered approximately eight times faster than the rate of the slow folding reaction of the protein. By observing fluorescence lifetime decay, they could watch the initially evenly populated rotamers of tryptophan reach nativity in which one rotamer is 88% populated. The authors suggest that rapid relaxation of tryptophan indicates the existence of an intermediate during the slow folding of barstar. It is not clear, however, that this must be the case. An alternate explanation is that there exists a

### Table II. Two-state Fits of Relaxation Dynamics for Protein G Side Chains\(^\ast\)

| No.  | a     | b     | c     | err | a    | b     | c     | err |
|------|-------|-------|-------|-----|------|-------|-------|-----|
| 5L   | 2.859 | 0.466 | 1.000 | ±0.01 | 53T  | 1.335 | 0.513 | 0.945 | ±0.006 |
| 30F  | 2.837 | 0.778 | 1.000 | 0.01  | 24E  | 1.169 | 0.037 | 0.102 | 0.05  |
| 45Y  | 2.686 | 0.391 | 1.000 | 0.01  | 19K  | 1.094 | 0.062 | 0.087 | 0.02  |
| 52F  | 2.630 | 0.494 | 1.000 | 0.01  | 46D  | 1.075 | 0.215 | 0.723 | 0.01  |
| 3Y   | 2.597 | 0.474 | 1.000 | 0.01  | 47D  | 1.067 | 0.118 | 0.678 | 0.02  |
| 1T   | 2.427 | 0.373 | 0.972 | 0.01  | 33Y  | 0.967 | 0.817 | 0.996 | 0.002 |
| 43W  | 2.374 | 0.737 | 1.000 | 0.006 | 31K  | 0.934 | 0.415 | 0.511 | 0.005 |
| 54V  | 2.368 | 0.574 | 1.000 | 0.007 | 44T  | 0.854 | 0.335 | 0.776 | 0.006 |
| 22D  | 2.321 | 0.582 | 0.841 | 0.01  | 42V  | 0.780 | 0.177 | 0.533 | 0.01  |
| 25T  | 2.063 | 0.502 | 0.994 | 0.007 | 39V  | 0.742 | 0.542 | 0.993 | 0.003 |
| 6V   | 2.06  | 0.623 | 0.995 | 0.006 | 56E  | 0.670 | 0.623 | 0.707 | 0.003 |
| 18T  | 1.828 | 0.572 | 0.995 | 0.006 | 15E  | 0.649 | 0.335 | 0.447 | 0.005 |
| 51T  | 1.827 | 0.259 | 0.198 | 0.01  | 32Q  | 0.551 | 0.120 | 0.199 | 0.01  |
| 40D  | 1.753 | 0.058 | 0.532 | 0.009 | 28K  | 0.448 | 0.024 | 0.064 | 0.02  |
| 21V  | 1.729 | 0.385 | 0.791 | 0.01  | 55T  | 0.512 | 0.368 | 0.769 | 0.004 |
| 17T  | 1.595 | 0.511 | 0.907 | 0.007 | 5N   | 0.479 | 0.207 | 0.350 | 0.006 |
| 2T   | 1.585 | 0.332 | 0.649 | 0.01  | 36D  | 0.464 | 0.178 | 0.318 | 0.007 |
| 50K  | 1.457 | 0.406 | 0.538 | 0.009 | 24E  | 0.424 | 0.037 | 0.190 | 0.08  |
| 49T  | 1.430 | 0.144 | 0.933 | 0.02  | 35N  | 0.436 | 0.039 | 0.190 | 0.08  |
| 27E  | 1.424 | 0.242 | 0.532 | 0.008 | 37N  | 0.336 | 0.660 | 0.844 | 0.002 |
| 16T  | 1.420 | 0.575 | 0.970 | 0.005 | 12L  | 0.222 | 0.628 | 0.876 | 0.002 |
| 4K   | 1.367 | 0.234 | 0.284 | 0.01  | 13K  | 0.183 | 0.053 | 0.084 | 0.009 |
| 0V   | 1.361 | 0.411 | 0.882 | 0.01  | 11T  | -0.534 | 0.029 | 0.810 | 0.03  |
| 7I   | 1.341 | 0.528 | 0.982 | 0.005 |       |       |       |       |       |

\(^\ast\)Individual residue relaxation curves were fit to the following two-state kinetic model: \(f(x) = c - b \exp(-ax/10^6)\). The parameter \(c\), corresponding to the fully equilibrated value of residue ordering, was obtained from long equilibrium simulation and was not varied in the fitting process. Standard nonlinear fitting was used to calculate \(a\) and \(b\). Asymptotic error on parameter \(a\) is listed. The table is sorted by the rate constant \(a\).
significant spread among side-chain relaxation rates within a single folding reaction.

Our work demonstrates that the presence of side-chain degrees of freedom leads to a wide distribution of residue relaxation rates, even within two-state cooperative folding reactions. Figure 15 gives a schematic overview of the relaxation mechanism we observed. Both in lattice and in all-atom simulations, we found a small number of residues becoming fully ordered much faster than the rest of the protein. This observation is consistent with the nucleation-condensation view of protein folding in which the major transition state of the folding reaction involves a few residues reaching their native conformation. Importantly, in our simulations, we find that these nucleating residues are not only in correct spatial geometry with respect to each other, but additionally their native rotamer has been singled out and practically frozen. Once nucleation has occurred, the native chain topology is strongly stabilized and certain experimental measures, such as radius of gyration and perhaps fluorescence, might indicate that the reaction is complete, and equilibrium has been reached (see Fig. 15 after nucleation barrier). This, however, is not the case because there exist many side chains that have become partially ordered, yet have still not reached equilibrium. Because the nucleating residues have frozen and are rigid and many other partially ordered residues are significantly stabilizing the fold, the nonequilibrated side chains are not able to convert easily to their native rotamer. They remain in a non-native state until a backbone fluctuation momentarily allows them to interconvert.

At first glance this observation seems to be at odds with a two-state transition in which all parts of the structure reach nativity at the same rate. It is crucial, however, to note that the core residues, which are observed in simulation to freeze fastest, also happen to be in key organizing positions. The ensemble of conformations consistent with their freezing is highly native and therefore extremely small compared with the ensemble of unfolded conformations. The major transition of protein folding occurs between these two ensembles and is a two-state transition in simulation as in reality. The entire molecule does not, however, necessarily reach equilibrium concomitantly with this barrier crossing. There can be many other barriers associated with backbone fluctuations that need to be crossed in order for all side chains to reach equilibrium (see Fig. 15). We note that during in vivo folding, the presence of proteases, combined with the long side-chain relaxation times we observed, may preclude the possibility of reaching equilibrium.

From our analysis of all-atom simulation data, it is clear that the dynamics of side-chain ordering occurring on the timescales of folding are intimately tied to the overall fold topology. In the protein we examined, the rate-limiting nucleation event is earmarked by the rapid ordering of the nucleating residues. This suggests that the combined degree of sidechain ordering at several key positions might be appropriate for use as a reaction coordinate for reaching the native backbone. We note that, if such a reaction coordinate exists, it would certainly not be appropriate in the postnucleation regime, where progress should be gauged using the slowest residues.

We emphasize that monitoring ordering at a single position is likely to give misleading results regarding progress toward the native backbone. One of the fastest ordering side chains we observed was W43, which has been used as the experimental probe in all stopped- and continuous-flow experiments of protein G to date. It was noted in Ref. 29 that the three-state, multipathway folding of protein G is incorrectly recast as a simple two-state process when folding is monitored by W43 alone. Likewise, our present analysis (Table II) also revealed single exponential relaxation for all residues. This observation is best explained by the fact that the ordering of any single residue corresponds to the ordering of its local environment, which does not require the global protein G fold to be complete. A good reaction coordinate for folding must
therefore follow the ordering of most of the nucleus residues, particularly when the selected protein has a complicated topology and nontrivial (i.e., not two-state and single-pathway) folding kinetics.

It is important to note that temperature plays a key role in making side-chain relaxation possible in a reasonable amount of time. At low temperatures, backbone fluctuations are small and side-chain relaxation is a very noticeable and very long process, as seen in Figure 7. At optimal folding temperature, however, the energy of the postnucleation ensemble is very close to its equilibrium value. Side-chain relaxation is still very slow, following power-law kinetics, but the product of the major transition is significantly closer energetically (and therefore structurally) to equilibrium. The relatively small energy gap between mispacked and native molecules at these temperatures suggests that relevant experiments must be sensitive enough to detect such differences. Because we employ a Monte Carlo simulation, the relationship between simulation temperature and experimental conditions must be

![Fig. 14. Protein G residues exhibiting fastest ordering. These residues are seen to form a cluster that organizes the entire topology of the protein and are located in the positions identified as the folding nucleus for this structure.](image)

![Fig. 15. Schematic diagram of barriers and their significance during the folding reaction. The first barrier corresponds to the nucleation event which organizes the backbone topology. Associated with this barrier is the freezing of a small group of residues, the nucleus, into their native side-chain states (blue dots). Other residues may still be partially disordered (red dots). The disordered residues become increasingly native-like via barriers corresponding to backbone fluctuations, which momentarily free a few residues (see small arrows), and allow their side-chain states to interchange. Barriers become higher as chain approaches equilibrium.](image)

Legend:
- **Ordered Sidechain**
- **Disordered Sidechain**
- **Free Sidechain State Can Interconvert**

| Reaction Coordinate | Free Energy |
|---------------------|-------------|
| **Nucleation**      |             |
| **Fluctuations**    |             |
| **Native Backbone** |             |
| **Equilibrium**     |             |
established by matching transition temperatures to experiment. If we take the transition temperature for a typical protein to be roughly 340K, we find that $T_{opt}$ in our models lies in the neighborhood of 300K.

There are several aspects of side-chain motion that are not captured by our lattice model. For example, it is well known that because of the steric structure of amino-acid residues, there exist barriers to side-chain rotamer interconversion even in the unfolded state. In our model we do not model internal barriers between different spin states. These could naturally be added and studied. We note, however, that in the folded state, the barrier to rotamer interconversion is dominated by the excluded volume effect of surrounding residues and not by the internal side-chain barriers. Because the constrained dynamics of spins models this dominant effect in the collapsed state, and because time scales for rotamer interconversion in the unfolded state are fast compared to backbone folding times,9 we elected not to include internal spin barriers in our model. Another potentially interesting addition to our model would be heterogeneity in $n$ across the fold. This modification would represent the diversity in the number of side-chain rotamers available to different amino acids and could lead to some useful insights regarding nucleus positions.

The lattice model we used bears some resemblance to lattice-gas models that have been recently studied.37–40 These models use constrained dynamics to simulate a gas of particles on a lattice. Moving a particle in or out of a site with more than three occupied neighbors is not allowed. Such models have been shown to exhibit a dynamic transition and other glassy characteristics, and are hence attractive for the study of structural glasses. In contrast, our model applies constrained dynamics to spins, and this is feasible precisely because we are studying a polymer. There is a critical density in the lattice-gas models above which the diffusion coefficient vanishes and dynamics become nonergodic. This would be the case in our model as well, if backbone motions were not allowed. Backbone motion is an ergodicity restoring move, similar to the one mentioned in Ref. 37. It guarantees that the system is ergodic at all densities. As discussed in Ref. 37, we do not expect a sharp ergodic to nonergodic transition in this case, but rather should expect glass-like behavior in the short and intermediate time scales. That is, at very long times the system reaches equilibrium, but the relaxation is clearly glass-like (i.e., nonexponential).

The relaxation behavior we observe is consistent with models of dynamics proposed in Refs. 41–43 in which a system with random free energies falls into progressively deeper traps, yielding stretched-exponential11 or power-law42 kinetics. Barriers to side-chain interconversion in our model correspond to backbone loop openings, with an energetic cost roughly proportional to the number of ordered side chains in the loop. As more side chains become ordered, the cost of loop opening becomes greater, and the traps thus become deeper. Detailed characterization of the postnucleation ensemble is needed to further substantiate this picture.

Using both all-atom and lattice simulations, we have demonstrated that full side-chain relaxation during protein folding can be a process whose time scale is significantly longer than that of attaining the native fold. Although backbone folding is organized by the freezing of a small number of residues, equilibrium is reached via a large set of barrier crossings that correspond to backbone fluctuations. The heterogeneities inherent in protein structures give rise to a distribution of side-chain relaxation rates that can span more than an order of magnitude. Our findings are consistent with a number of recent experiments. We hope that this work will spur further dialogue between simulations and experiments to elucidate the complex processes that bring proteins to equilibrium.

**ACKNOWLEDGMENTS**

We thank Will Chen, Gabriel Berriz, and Leonid Mirny. Research was supported by NIH grant GM52126.

**REFERENCES**

1. Shimada J, Kussell E, Shakhnovich EI. The folding thermodynamics and kinetics of crambin using an all-atom Monte Carlo simulation. J Mol Biol 2001;308:79–95.
2. Richards FM, Lim WA. An analysis of packing in the protein folding problem. Quat Rev Biophys 1994;26:423–498.
3. Herings J, Argos P. Strain in protein structures as viewed through nonrotameric side chains. I. Their position and interaction. Proteins 1999;37:30–43.
4. Levitt M, Gerstein M, Huang E, Subbiah S, Tsai J. Protein folding: the endgame. Annu Rev Biochem 1997;66:549–579.
5. Dunbrack RL, Karplus M. Conformational-analysis of the backbone-dependent rotamer preferences of protein side-chains. Nat Struct Biol 1994;1:334–340.
6. Fersht A. Structure and Mechanism in Protein Science. New York: WH Freeman and Company; 1999.
7. Shakhnovich EI, Finkelstein AV. Theory of cooperative transitions in protein molecules. I. Why denaturation of globular proteins is a first-order phase transition. Biopolymers 1989;28:1667–1680.
8. Sali A, Shakhnovich EI, Karplus M. How does a protein fold? Nature 1994;369:248–251.
9. Brooks CL, Karplus M, Pettitt BM. Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics. New York: John Wiley & Sons; 1998.
10. Kussell E, Shimada J, Shakhnovich EI. Excluded volume in protein side-chain packing. J Mol Biol 2001;311:183–193.
11. Bremberg S, Dill KA. Side-chain entropy and packing in proteins. Protein Sci 1994;3:997–1009.
12. Kimrov DK, Thirumalai D. Cooperativity in protein folding, from lattice models with sidechains to real proteins. Fold Design 1999;3:127–139.
13. Li L, Mirny LA, Shakhnovich EI. Kinetics, thermodynamics, and evolution of non-native interactions in protein folding nucleus. Nat Struct Biol 2000;7:336–342.
14. Garel T, Orland H, Orlandini E. Phase diagram of magnetic polymers. Eur Phys J B 1999;12:261–268.
15. Muñoz V, Eaton WA. A simple model for calculating the kinetics of protein folding from three-dimensional structures. Proc Natl Acad Sci USA 1999;96:11311–11316.
16. Morrissey M, Shakhnovich EI. Design of proteins with selected thermal properties. Fold Design 1996;1:391–406.
17. Mirny LA, Abkevich V, Shakhnovich EI. How evolution makes proteins fold quickly. Proc Natl Acad Sci USA 1998;95:4976–4981.
18. Miyazawa S, Jernigan R. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. Macromolecules 1985;18:534–552.
19. Sali A, Shakhnovich EI, Karplus M. Kinetics of protein folding. A lattice model study for the requirements for folding to the native state. J Mol Biol 1994;235:1614–1636.
20. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller...
SIDE-CHAIN DYNAMICS 319

APPENDIX

Statistical-mechanical Derivation of Model Thermodynamics

Figure 1 shows that simple two-state thermodynamics do not apply in the low-temperature regime of the models with \( n > 1 \). Increase of \( n \) leads to a narrower transition and the development of a kink in average energy just below the transition temperature. The following simple analytical model explains these thermodynamic observations.

We assume that monomers are independent and can exist in a bound or an unbound state. Additionally, we stay within the two-state approximation for the backbone transition, thus identifying a folded and an unfolded ensemble. In the bound state, we imagine the monomer in its native spin state. We assign energy \( E_b \) or \( E_u \) to a bound monomer in the folded or unfolded ensemble, respectively, and entropy \( S_b \) or \( S_u \). This entropy comes only from backbone motion, because the monomer is in a single spin state. In the unbound state, we assign energy of zero to a monomer, and entropy of \( rS_b \) (or \( rS_u \)), where \( r \) is the number of non-native spin states. It is clear, then, that values of energy are relative to the energy of an unbound monomer. The total backbone entropy in the folded and unfolded states is given by \( S = S_b + S_u \).

We write the partition functions for the folded and unfolded ensembles:

\[
Z_f = \gamma r e^{-\beta S_b} = \Omega_f(r + e^{-\beta S_b})
\]

\[
Z_u = (\gamma r + e^{-\beta S_u}) = \Omega_u(r + e^{-\beta S_u})
\]

and we define the total partition function \( Z = Z_f + Z_u \). If we let \( M \) denote the number of native spins, we can use the binomial expansion to write the partition function as

\[
Z = \sum_M \left( \binom{N}{M} \right)^N \Omega_f^{e^{-\beta S_b}} + \Omega_u^{e^{-\beta S_u}}.
\]

Using this expression, the average number of native spins (i.e., the average magnetization, \( \langle M \rangle \)) is seen to be given by

\[
\langle M \rangle = \left( \frac{\partial}{\partial q} + \frac{\partial}{\partial q_u} \right) A,
\]

where \( A = -T \ln(Z) \) is the free-energy. We compute the average magnetization as follows:

\[
\frac{\partial A}{\partial q} = -T \frac{\partial Z_f}{\partial q} = -T \frac{\partial}{\partial q} \Omega_f(r + e^{-\beta S_b})^{-1} e^{-\beta S_b} \langle \beta \rangle
\]

\[
= N \tilde{\Omega}_f \frac{e^{-\beta S_b}(r + e^{-\beta S_b})}{Z}.
\]

The average magnetization per monomer, \( m \), is therefore given by

\[
\langle m \rangle = \frac{\Omega_f e^{-\beta S_b}(r + e^{-\beta S_b})^{-1} \Omega_u e^{-\beta S_b}(r + e^{-\beta S_b})^{-1}}{\Omega_f(r + e^{-\beta S_b})^N + \Omega_u(r + e^{-\beta S_u})^N}.
\]

We make the definition

\[
Q = \frac{\gamma_r(r + e^{-\beta S_u})}{\gamma(r + e^{-\beta S_u})^N}
\]

and then rewrite the average \( m \) as

\[
\langle m \rangle = \frac{e^{-\beta S_b} \frac{1}{r + e^{-\beta S_u}} + e^{-\beta S_u} \frac{1}{r + e^{-\beta S_u}}}{1 + Q^N}.
\]

Taking the thermodynamic limit, \( N \to \infty \), we find
In other words, for $r > 0$ there is a first-order transition at $\beta = \beta_*$, and the discontinuity becomes more pronounced as $r$ increases. A phase diagram is shown in Figure 16.

The jump is zero exactly when $Q = 1$, as expected. When $Q > 1$, there is a jump in the magnetization across the point $Q = 1$. To see this, we compute the change in magnetization across the point $Q = 1$:

$$
\langle m \rangle = \begin{cases} 
\frac{e^{-\beta m}}{r + e^{-\beta m}} & Q < 1 \\
\frac{e^{-\beta m}}{r + e^{-\beta m}} + 1 & Q > 1.
\end{cases}
$$

where $\beta_s$ is the temperature at which $Q = 1$. From Eq. 3 we find

$$
\gamma_u(r + e^{-\beta_0}) = \gamma(r + e^{-\beta_s})
$$

$$
\gamma = \frac{\gamma u e^{-\beta_0} - \gamma e^{-\beta_s}}{\gamma u - \gamma f},
$$

and using this we obtain

$$
\langle m \rangle = \frac{\gamma u e^{-\beta_0} - \gamma e^{-\beta_s}}{(\gamma u e^{-\beta_0} - \gamma e^{-\beta_s})(\gamma u - \gamma f) + e^{-\beta_0}}.
$$

The jump is zero exactly when $r = 0$, and as $\beta_s$ increases, the jump becomes non-zero. As $\beta_s \to \infty$, using the fact that $\gamma_u > \gamma_f$, we see that the jump reaches a limiting value given by

$$
\lim_{\beta_s \to \infty} \langle m \rangle = \frac{\gamma u}{\gamma u - \gamma f} = 1 - \frac{\gamma f}{\gamma u}.
$$

In other words, for $r > 0$ there is a first-order transition at $\beta = \beta_s$, and the discontinuity becomes more pronounced as $r$ increases. A phase diagram is shown in Figure 16.

As $r$ increases, we expect the transition to become sharper. To see this, we can compute the slope of the magnetization in the folded state, when $Q = 1^-$, at the transition temperature $\beta_s$:

$$
\frac{\partial \langle m \rangle}{\partial T} \bigg|_{\beta_s} = \frac{\partial \langle m \rangle}{\partial T} \bigg|_{\beta_s} = \frac{(r + e^{-\beta_0}) e^{-\beta m} - e^{-\beta m} (e^{-\beta_0} e^{-\beta m})}{(r + e^{-\beta m})^2}.
$$

Since $r$ and $\beta_s$ are related via Eq. 5, we can recast this expression in terms of $\beta_s$ only:

$$
\frac{\partial \langle m \rangle}{\partial T} \bigg|_{\beta_s} = \frac{e^{-\beta_0} e^{-\beta m} (\gamma u e^{-\beta m} - \gamma e^{-\beta m})}{(\gamma u e^{-\beta_0} - \gamma e^{-\beta_m})^2}
$$

This expression can be shown to monotonically decrease toward $\infty$ as $\beta_s$ increases (using the fact that $\gamma_u > \gamma_f$ and $\gamma_f < \gamma_u$). Additionally, from Eq. 5 we can see that $r$ increases monotonically as $\beta_s$ increases. The slope of the magnetization in the ordered state therefore diverges to $\infty$ for increasing $r$ just before the transition discontinuity. Qualitatively, this analysis is in full agreement with the thermodynamic curves shown in Figure 1. Notice that for the 48-mer with $n = 4$, there is a very sharp transition, nearly vertical, and that the slope of the curve for $\beta > \beta_s$ is greater as $n$ increases.

In order to fit the thermodynamic curves to our analytical model, we give the expression for average energy, which is very similar to average magnetization derived in Eq. 2:

$$
\langle E \rangle = \frac{\epsilon_\Omega e^{-\beta_0 (r + e^{-\beta_0}) N - 1} + \epsilon_\Omega e^{-\beta_0 (r + e^{-\beta_0}) N - 1}}{\Omega_0 (r + e^{-\beta_0}) N + \Omega_0 (r + e^{-\beta_0}) N - 1}.
$$

The fits of the 48-mer models using Eq. 9 are shown in Figure 17, and the fit parameters are given in Table III. We see that the model presented here reproduces the correct shape of the thermodynamic data. As $n$ increases, our model improves significantly, reaching a very good fit already at $n = 4$. Although our model converges to the usual two-state model when $r = 0$ in Eq. 9, we do not see a good fit of the $n = 1$ case. This may seem surprising at first, but the explanation is simple: the usual two-state model provides a good fit to average energy only within the transition region. The baseline slopes are not properly fit by the simple two-state approximation. Because we let $r$ vary in our model, nonlinear fitting has simply found a closer fit than the
This explains why $r$ is a bit bigger than it should be for $n = 1, 2, \text{ and } 4$ (see Table III). The two-state approximation for backbone degrees of freedom does not give the correct shape away from the transition region, and therefore in order to gain a slightly higher slope, nonlinear fitting inflates $r$ somewhat. If we were to use a more sophisticated thermodynamic approximation for the backbone, we should be able to match the $n = 1$ model well and improve the accuracy of the estimates of $r$.

TABLE III. Parameters for Fits in Figure 17

| $n$ | $r$  | $\epsilon_u$ | $\Delta S$ |
|-----|------|---------------|------------|
| 1   | 1.14 | -7.54         | 35.5       |
| 2   | 3.1  | -6.18         | 38.5       |
| 4   | 6.4  | -4.83         | 41.2       |

$r = 0$ two-state solution. This explains why $r$ is a bit bigger than it should be for $n = 1, 2, \text{ and } 4$ (see Table III). The two-state approximation for backbone degrees of freedom does not give the correct shape away from the transition region, and therefore in order to gain a slightly higher slope, nonlinear fitting inflates $r$ somewhat. If we were to use a more sophisticated thermodynamic approximation for the backbone, we should be able to match the $n = 1$ model well and improve the accuracy of the estimates of $r$.