

One-Dimensional Barrier-Preserving Free-Energy Projections of a β -sheet Miniprotein: New Insights into the Folding Process

Sergei V. Krivov,^{†,‡} Stefanie Muff,^{‡,§} Amedeo Caflisch,^{*,§} and Martin Karplus^{*,†,||}

Laboratoire de Chimie Biophysique, ISIS F-67000, Strasbourg, France, Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland, and Department of Chemistry & Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

Received: December 18, 2007; Revised Manuscript Received: March 12, 2008

The conformational space of a 20-residue three-stranded antiparallel β -sheet peptide (double hairpin) was sampled by equilibrium folding/unfolding molecular dynamics simulations for a total of 20 μ s. The resulting one-dimensional free-energy profiles (FEPs) provide a detailed description of the free-energy basins and barriers for the folding reaction. The similarity of the FEPs obtained using the probability of folding before unfolding (p_{fold}) or the mean first passage time supports the robustness of the procedure. The folded state and the most populated free-energy basins in the denatured state are described by the one-dimensional FEPs, which avoid the overlap of states present in the usual one- or two-dimensional projections. Within the denatured state, a basin with fluctuating helical conformations and a heterogeneous entropic state are populated near the melting temperature at about 11% and 33%, respectively. Folding pathways from the helical basin or enthalpic traps (with only one of the two hairpins formed) reach the native state through the entropic state, which is on-pathway and is separated by a low barrier from the folded state. A simplified equilibrium kinetic network based on the FEPs shows the complexity of the folding reaction and indicates, as augmented by additional analyses, that the basins in the denatured state are connected primarily by the native state. The overall folding kinetics shows single-exponential behavior because barriers between the non-native basins and the folded state have similar heights.

Introduction

Protein and peptide folding from the very broad ensemble of denatured conformations to the well-defined native state is a very complex unimolecular reaction because of the many degrees of freedom of the system.¹ The conformational transitions involved, like those of other chemical reactions, are governed by the free-energy surface.² During the folding process, the loss of configurational entropy of the protein chain is approximately counterbalanced by the more favorable interactions among the protein atoms, modulated by the effect of the solvent. Thus, although the enthalpy and entropy of folding can be large at physiological temperatures, the free energy stabilizing the native state is normally only 10 kcal/mol or less, independent of the size of the protein.³ The loss of configurational entropy during folding is thought to be primarily responsible for the experimental activation barrier, observed in many so-called two-state proteins.^{4,5} Consequently, the major role played by the entropic contributions in protein folding, in contrast to most simple reactions,⁶ requires an analysis of the free-energy surface; that is, knowledge of the potential energy surface is not sufficient.^{6–10}

Although a protein has many degrees of freedom (with the φ and ψ backbone dihedral angles of the amino acids being particularly important), the common way to investigate the free-energy surface is to display it as a function of a small (usually only one to two) number of order parameters. A commonly used

coordinate is the fraction of native contacts, Q .¹¹ Q appears to be a satisfactory approximate reaction coordinate for Go-model proteins^{12,13} because favorable interactions occur only between residues in contact in the folded state.¹⁴ On the other hand, for transferable potentials (e.g., those based on physicochemical principles, such as AMBER, CHARMM, and OPLS) or statistical potentials,^{15,16} Q is adequate only for the fully folded ($Q = 1$) state.¹⁷ For example, for a structured peptide simulated by a transferable force field, some conformations with $Q \approx 0.7$ belong to the denatured state ensemble, and conformations with $Q \approx 0.3$ belong to the folded state.^{18,19}

The multidimensionality of the system makes the choice of order parameters for presenting the free-energy surface very important and often leads investigators to compare the results from several different sets. Moreover, the likelihood of hiding essential information concerning the free-energy surface by the commonly used projections has led to a search for alternative methods that are useful for studying protein folding, as well as other complex reactions. One approach that was introduced just 10 years ago is based on disconnectivity graphs²⁰ (also see ref 21). The unprojected free-energy surface is represented by a disconnectivity graph calculated from an equilibrium folding trajectory with the minimum cut (mincut) or balanced minimum cut (bmincut) procedure.²² The idea of the method is to group the coordinate sets into free-energy minima, according not to the standard geometric characteristics, but rather to the equilibrium dynamics; that is, the trajectory is used to determine the populations of the states, which provide the relative free energies and the rates of the transition between the states, which yield the free-energy barriers. Application of the method to the β -hairpin of protein G demonstrated that the free-energy surface has multiple low free-energy basins in the denatured state, in

* Corresponding authors. Tel.: +33 390 24 5123. Fax: +33 390 24 5124. E-mail: marci@tammy.harvard.edu (M.K.), caflisch@bioc.uzh.ch (A.C.).

[†] Laboratoire de Chimie Biophysique.

[‡] S.K. and S.M. made equal contributions to this study.

[§] University of Zurich.

^{||} Harvard University.

addition to the native basin, results that complement the analysis of the experimental observation of two-state folding.²³ The same simulation data as were used to reveal the complexity of the denatured state show a relatively smooth free-energy landscape when projected onto a few geometrical coordinates.²² This result demonstrates that, to obtain a projection that gives an accurate description of the essential aspects of the free-energy surface, different progress coordinates are required.

Projected free-energy surfaces are most useful if they preserve the barriers and minima in the order in which they are encountered during folding/unfolding events. Recently, a new progress coordinate that has some of the desired properties was introduced.²⁴ It uses the (normalized) partition function of a given region as the progress coordinate and determines the free-energy barriers as a function of the coordinate by a method based on p_{fold} , defined as the probability of reaching the folded state before an unfolded conformation.²⁵ The result is a one-dimensional projected free-energy profile (FEP) that preserves the barriers between the free-energy basins; given the barriers, the minima can be determined.²⁴ The method was applied to the β -hairpin of protein G, using root-mean-square deviation (rmsd) clustering, and the conclusions concerning the multimimum character of the free-energy surface obtained from the disconnectivity graph analysis²² were confirmed.

It is of interest to apply the methodology just described to a system that is more complex than the β -hairpin. An excellent candidate is the β -sheet miniprotein, called Beta3s.²⁶ Its structure corresponds to a three-stranded antiparallel β -sheet consisting of two β -hairpins.²⁷ It has been shown to fold to the native structure determined by NMR spectroscopy²⁷ in molecular dynamics simulations with a polar hydrogen molecular mechanics potential function modulated by a simple implicit solvent model.²⁶ Because folding simulations of this system are very fast (for example, close to the melting temperature, the folding time is about 100 ns and requires about 24 h on an Athlon 1.7 GHz computer), many studies have been conducted to elucidate the folding mechanism. Two main folding pathways were observed: one begins with formation of the C-terminal β -hairpin, followed by association of the N-terminal strand on the preformed β -hairpin, and the other follows the symmetry-related pathway (first formation of N-terminal β -hairpin).²⁶ Conformations in the denatured state of Beta3s were shown to contain a significant amount of non-native contacts,²⁸ and the folding mechanism of Beta3s had a weak temperature dependence.²⁹ Moreover, because of the very efficient implicit solvent model, multiple simulations of Beta3s and 32 single-point mutants (totaling 0.65 ms) were performed to directly calculate φ values³⁰ from folding and unfolding rates extracted from equilibrium folding/unfolding trajectories.³¹ More recently, an earlier network analysis of Beta3s³² was extended to determine free-energy basins.¹⁹ Secondary structure was used to coarse-grain and label the conformations visited in the simulations. The folded state and the most populated free-energy basins in the denatured state were isolated by grouping conformations according to fast relaxation in equilibrium trajectories, a procedure called kinetic grouping analysis (KGA).¹⁹ The comparative application of KGA to Beta3s and a central-strand mutant thereof (W10V) revealed how a single-point mutation can alter the population of native and non-native basins, as well as the relative accessibility of parallel folding pathways.¹⁹ It was shown that only one parameter is required for grouping, namely, the commitment time τ_{commit} , which is chosen as a typical relaxation time within the basins of the investigated system. Any two conformations are grouped into the same basin

if they interconvert within the time τ_{commit} with a probability, p_{commit} , of ≥ 0.5 . In other words, two conformations are said to be separated by a short kinetic distance if the interconversion between them is fast, which implies that the free-energy barrier between them is low. Kinetic distance is used here, in analogy to the previously introduced term of kinetic closeness,¹¹ to distinguish what is being described from structural distance.

In the present work, several different, but related, approaches for determining one-dimensional FEPs were applied to Beta3s and the W10V mutant using 20 μs of sampling for each peptide at 330 K, which is slightly above the melting temperature, to obtain adequate sampling of the native and denatured regions of the free-energy surface. The method *bpfold*, described in ref 24, is based on p_{fold} . As in the balanced mincut method,²² which finds exact barriers separating individual basins, an extra node is introduced to represent the unfolded state. The extra node is connected to all nodes in the network with a capacity proportional to a Lagrange multiplier λ . For different values of λ , different partitions into two basins with $p_i < 0.5$ and $p_i > 0.5$ are obtained in the *bpfold* procedure. One approximation to *bpfold*, referred to as *pfoldf* (which stands for *pfold fast*), requires only one value of λ .²⁴ Two new related approaches were used in this work for comparison. One is called *pfoldt* and is based on $p_{\text{fold}}(\tau_{\text{commit}})$, which is defined as the probability of reaching the folded state within the time τ_{commit} .^{33–35} In the second procedure, the mean first passage time (mfpt) to the native state is used; the procedure is called *mfpt*. The main difference from *pfoldf* is that the calculations of the progress variables, $p_{\text{fold}}(\tau_{\text{commit}})$ and *mfpt*, depend only on the native node; that is, no extra node needs to be added to represent the unfolded state in these procedures. Conversely, only *pfoldf* is suitable for calculating the barrier between two existing nodes, because only one node can be specified in the *mfpt* and *pfoldt* procedures. Further, evaluation of the exact $p_{\text{fold}}(\tau_{\text{commit}})$ values is computationally more expensive than p_{fold} or *mfpt* calculations, as detailed in the Methods section and the Supporting Information. All three approaches applied here have in common that they encode the kinetic distance to the folded (or any other representative) state and are therefore expected to give similar results. Indeed, the one-dimensional FEPs were found to be very similar and to approximate the exact mincut barrier equally well, underlining the robustness of the methods. The significance of this result is discussed. Further, we compare the results using secondary structure clustering with those obtained with *rmsd* clustering. It is shown that the barriers from the former tend to be lower than those obtained with the latter; with a proper choice of the *rmsd* value used for clustering (here found to be 2.5 Å for all-atoms), the resulting Monte Carlo (MC) kinetics agrees with that calculated directly from the trajectories.

Interestingly, a helical state with a statistical weight of about 11% is identified by the three procedures as a free-energy basin separated by a barrier from the rest of the denatured ensemble. The implications of non-native secondary structure content in the denatured state of a β -sheet peptide are briefly discussed. Further, both KGA and one-dimensional FEPs reveal a large and heterogeneous entropic region (weight of 33%) that is separated by a barrier of less than $k_{\text{B}}T$ from the native state (weight of 35%). The single-exponential behavior of Beta3s folding is shown to be due to the similar free-energy barriers to exit from the non-native enthalpic traps (total population of about 20%) or from the helical basin (weight of 11%), which is primarily stabilized by its entropy.

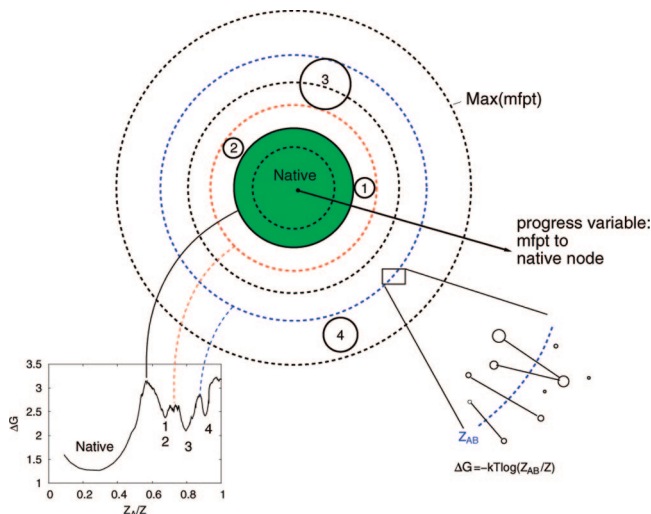


Figure 1. Schematic illustration of the one-dimensional FEP procedure using mfpt as the progress variable. Each of the four solid circles represents a free-energy basin, and the concentric dashed circles represent values of mfpt. For each value of mfpt, between 0 (native node) and $\text{max}(\text{mfpt})$, a point in the profile is obtained. ΔG of the fraction of links crossing the cutting surface at $\text{mfpt} = \text{mfpt}_c$ (bottom right) is plotted as a function of the relative partition function Z_A/Z (bottom left), where the set A contains all nodes with $\text{mfpt} < \text{mfpt}_c$. Basins 1 and 2 overlap because they have the same mfpt distance from the native state and are therefore not separated in the unfolded part of the profile. Note that the same illustration is valid if mfpt is replaced by P_{fold} or $P_{\text{fold}}(x_{\text{commit}})$, with the only difference being that the distance from native would decrease from 1 (native) to 0.

Methods

Molecular Dynamics Simulations. All simulations and most of the analysis of the trajectories were performed with the program CHARMM;³⁶ the rest of the analysis was done with the program WORDOM,³⁷ which is particularly efficient in handling large sets of trajectories. The designed 20-residue peptide Beta3s²⁷ (Thr₁-Trp₂-Ile₃-Gln₄-Asn₅-Gly₆-Ser₇-Thr₈-Lys₉-Trp₁₀-Tyr₁₁-Gln₁₂-Asn₁₃-Gly₁₄-Ser₁₅-Thr₁₆-Lys₁₇-Ile₁₈-Tyr₁₉-Thr₂₀) and its W10V mutant were modeled by explicitly considering all heavy atoms and the hydrogen atoms bound to nitrogen or oxygen atoms (PARAM19 force field³⁸ with a default cutoff of 7.5 Å for the nonbonding interactions). A mean-field approximation based on the solvent-accessible surface (SAS) was used to describe the main effects of the aqueous solvent.³⁹ It has been shown previously that this model for the solvated Beta3s peptide yields reversible folding at 330 K to the NMR conformation, irrespective of the starting structure; 23 of the 26 nuclear Overhauser effect (NOE) constraints are satisfied.²⁶ Moreover, despite the neglect of collisions with water molecules (frictional effects) in the simulations with the implicit solvent model, the *relative* rates of folding for different secondary structural elements are comparable to the values observed experimentally; i.e., helices fold in about 1 ns,⁴⁰ β -hairpins in about 10 ns,⁴⁰ and triple-stranded β -sheets in about 100 ns,³¹ compared to experimental values of ~ 0.1 ,⁴¹ ~ 1 ,⁴¹ and ~ 10 μs ,²⁷ respectively. For Beta3s and the W10V mutant, 10 molecular dynamics runs of 2 μs each with different initial distributions of velocities were performed with the Berendsen thermostat (coupling constant of 5 ps) at 330 K, which is slightly above the melting temperature of Beta3s.²⁹ A time step of 2 fs was used, and the coordinates were saved every 20 ps, for a total of 10^6 snapshots for each system. This required three weeks on a 20-CPU cluster. Using explicit water simulations, it would have been much more time-consuming to obtain the 40 μs of

simulation time required to sample a statistically significant number of equilibrium folding/unfolding transitions.

Coarse-Graining and Equilibrium Kinetic Network (EKN).

For the purpose of using the finite-time simulation data to obtain free-energy surfaces, it is necessary to coarse-grain the snapshots in some way because each conformation is visited only once; the trajectory, per se, is nothing but a long string of configurations. There are several meaningful methods for clustering individual coordinate sets from the trajectory to obtain coarse-grained conformations, and different approaches are likely to be most useful for different types of analysis. For a system such as Beta3s or a β -hairpin, rmsd and secondary structural coarse-graining are obvious possibilities.^{22,32,42} The coarse-graining used in this work is based on secondary structure strings.⁴³ A “coarse-grained conformation” or node is a single string of secondary structure; for example, the most populated conformation of Beta3s, which corresponds to the native state, is -EEEEES-SEEEEESEEEEE-.³² There are eight possible “letters” in the secondary structure “alphabet”: H, G, I, E, B, T, S, and -, standing for α -helix, 3_{10} -helix, π -helix, extended, isolated β -bridge, hydrogen-bonded turn, bend, and unstructured, respectively.⁴³ Because the N- and C-terminal residues are always assigned as unstructured,⁴³ a 20-residue peptide can, in principle, assume $8^{18} \approx 10^{16}$ conformations. We note that there is no relation between the Hamming distance (number of different entries, i.e., letters, in two strings of equal length) and the kinetic distance (see also the Results section). The secondary-structure-based coarse-graining is used to permit comparison with earlier work.¹⁹ It has the advantage over approaches based on the rmsd of the atomic coordinates of being more efficient because it scales with the number of snapshots whereas rmsd is a pairwise measure. Also, each node is uniquely defined by its secondary structure string, which serves as a useful conformational “label”. However, as we show here, secondary-structure-based coarse-graining can, in some cases, lead to overlapping of regions that are distant in terms of rmsd (i.e., far from each other in configuration space), which can result in barriers for the basins that are too low; this is the case for Beta3s and is referred to as “pseudotunneling” hereafter.

The number of snapshots with a given secondary structure string i is called the weight of the node and is denoted as \tilde{w} . The statistical weight w of a node is given by $w = \tilde{w}/N$, where $N = 10^6$ is the total number of snapshots. In the same way, the links, which are direct transitions sampled along the MD trajectory, are weighted by n_{ij} , defined as the number of times a snapshot in node i is followed by a snapshot in node j . As mentioned above, snapshots were saved every 20 ps, which is therefore the time interval of a direct transition. The resulting equilibrium kinetic network (EKN)^{24,44} is an undirected, weighted graph where the edge capacity from node j to node i in the network, c_{ij} , is proportional to the number of direct transitions from j to i at equilibrium. Detailed balance can be “imposed”, i.e., $c_{ij} = c_{ji} = (n_{ij} + n_{ji})/2$. The transition probabilities can then be calculated as $p_{ij} = c_{ij}/\sum_k c_{kj}$.

For a node i in the EKN, the partition function is $Z_i = \sum_j c_{ij}$. If the nodes of the network are partitioned into two groups A and B, then $Z_A = \sum_{i \in A} Z_i$, $Z_B = \sum_{i \in B} Z_i$, $Z_{AB} = \sum_{i \in A, j \in B} c_{ij}$, and the free energy of the barrier between the two groups is $-kT \log(Z_{AB}/Z)$, where Z is the partition function of the full network (Figure 1).

One-Dimensional FEP. The pfoldf procedure to determine the one-dimensional FEP was published previously;²⁴ two additional procedures, pfoldt and mfpt, are introduced here.

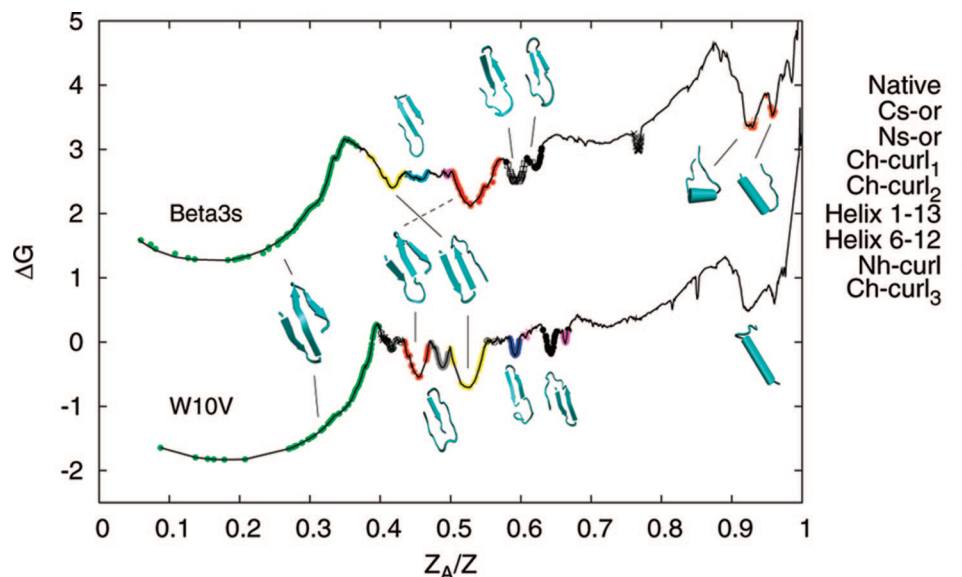


Figure 2. pfoldf-calculated FEP of Beta3s and its single-point mutant W10V, whose plot is shifted by -3 kcal/mol to avoid overlap of the curves. The progress coordinate is the relative partition function for different values of p_c (see text and Table 1). Symbols represent the secondary structure strings with ≥ 100 snapshots; they are colored according to the basins identified by the KGA.¹⁹ Conformations in the most populated basins are shown as ribbon diagrams, where β -strands are represented by arrows and helices by cylinders. Abbreviations: Ns-or, N-terminal strand out of register and folded C-terminal hairpin; Cs-or, C-terminal strand out of register and folded N-terminal hairpin; Nh-curl, curl-like conformation with folded N-terminal hairpin; Ch-curl, curl-like conformation with folded C-terminal hairpin. Note that Ns-or conformations have a topology similar to that of the folded structure but non-native orientations of the side chains in the N-terminal strand. The Z_A/Z coordinate value of the first data point (leftmost green circle) is the relative weight of the most populated secondary structure node.

TABLE 1: Overview of the Four Procedures Discussed in the Text for Determining FEPs^a

	bpfold ^b	pfoldf ^c	pfoldt ^d	mfpt ^e
for barrier to exit from a basin	yes	yes	yes	yes
for barrier between two basins	yes	yes	no	no
number of target nodes	2	2	1	1
extra node	yes	yes	no	no
system of equations	$p_i = \sum_j p_{ji} p_j$	$p_i = \sum_j p_{ji} p_j$	(see <i>Supporting Information</i>)	$\text{mfpt}_i = \Delta t + \sum(p_{ji} \times \text{mfpt}_j)$
boundary condition(s)	$p_A = 1, p_B = 0$	$p_A = 1, p_B = 0$	$p_A = 1$	$\text{mfpt}_A = 0$
progress variable	p_{fold}	p_{fold}	$p_{\text{fold}}(\tau_{\text{commit}})$	mfpt
Z_A/Z evaluation	$p_{\text{fold}} > 0.5$	$p_{\text{fold}} > p_c; 0 \leq p_c \leq 1$	$p_{\text{fold}} > p_c; 0 \leq p_c \leq 1$	$\text{mfpt} < \text{mfpt}_c; 0 \leq \text{mfpt}_c < +\infty$
number of λ values	many	1	0	0
ref	24	24	this work	this work

^a Three approaches related to bpfold (i.e., pfoldf, pfoldt, and mfpt) were used in this work. The pfoldf and mfpt procedures are available in the MD analysis package WORDOM.³⁷ ^b bpfold: The balanced pfold requires an extra node if no representative node of the unfolded state exists. This approach yields the most accurate description of the barrier to exit a basin. ^c pfoldf: The pfold fast evaluation is an approximation of bpfold that uses only one value of the Lagrange multiplier λ . ^d pfoldt: The p_{fold} between two nodes is replaced by $p_{\text{fold}}(\tau_{\text{commit}})$, which is the probability of reaching a target node within the commitment time τ_{commit} . The extra node and Lagrange multiplier are not required. pfoldf is suitable for the calculation of unfolding profiles, but cannot be employed for the detection of a barrier between two basins. ^e mfpt: Similarly to pfoldt, the mean first passage time to a target node does not require an extra node, and therefore, it applies to the same cases as pfoldt.

Table 1 lists details of all of the procedures used to calculate FEPs in this study.

pfoldf. Given the EKN and two nodes A and B, the p_{fold} value of node i , p_i , is found as the solution of the equation $p_i = \sum_j p_{ji} p_j$ with boundary conditions $p_A = 1$ and $p_B = 0$ (A is considered to be the “native node” and B the “denatured node”). The system of equations can be solved efficiently numerically by iterative multiplication of the vector p_j by the matrix p_{ji} .²⁴

To determine the FEP relative to a chosen node A, node B is considered to be the representative node of everything not belonging to the basin of A. However, in many systems, a node such as B does not exist, because there are multiple basins and/or an entropic state that cannot be represented by a single node. Both occur in the peptides investigated in this study. Thus, as in the balanced minimum-cut procedure,²² an extra node B is introduced and connected to all nodes in the network with capacity $\lambda \bar{w}$, where λ is a Lagrange multiplier. The p_{fold} calculations are performed on the EKN with the extra node,

and in the pfoldf procedure, the nodes are sorted according to their p_{fold} values using only one value of λ . The assumption in this procedure is that the order of the nodes does not change for different values of λ .²⁴ Each value p_c between 0 and 1 can then be used to cut the network into set A containing all nodes with $p_{\text{fold}} > p_c$ and set B containing the nodes with $p_{\text{fold}} < p_c$. For each cut, a point [$x = Z_A/Z$, $y = -kT \ln(Z_{AB}/Z)$] of the FEP is obtained; Z_A/Z is used as the progress coordinate, and Z_{AB} is the number of EKN transitions between the two sets. Note that p_{fold} is the progress variable used to divide the configuration space, with pfoldf evaluation based on this variable. Moreover, this procedure and those described below do not require any special treatment of low-population nodes (i.e., secondary structure strings with only one or a few snapshots) because they are automatically grouped on the same side of the barrier as the reference node A if they satisfy the condition $p_{\text{fold}} > p_c$.

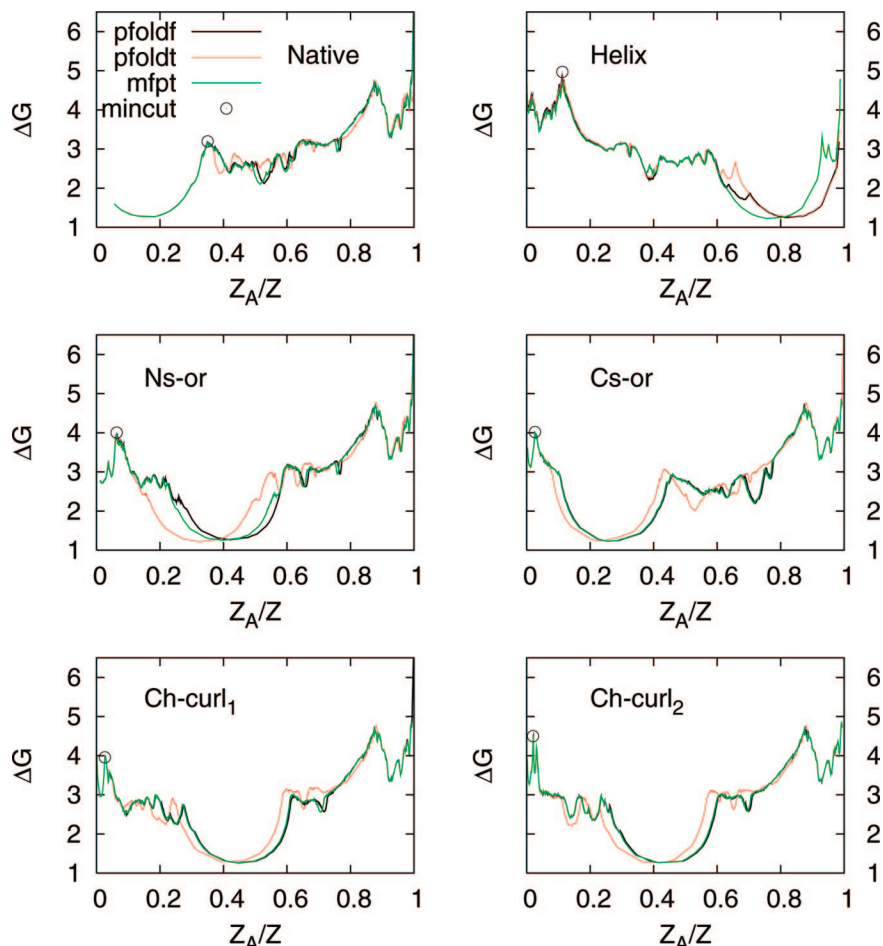


Figure 3. One-dimensional FEPs calculated using the kinetic distance from individual basins of Beta3s. Notably, the barriers separating the reference state from the rest are almost identical for pfoldf with $\lambda = 0.0001$ (black), pfoldt (red), and mfpt (green), and all three procedures yield good approximations to the exact barrier height calculated with mincut²² (open circle). pfoldt was calculated with different τ_{commit} values, namely, 20 ns (native), 40 ns (Ns-or), 100 ns (Cs-or, Ch-curl₁, Ch-curl₂), and 200 ns (helical). Except for the shift of some of the minima after the first barrier, the three FEPs are very similar, which indicates that the approach is robust upon the choice of the progress variable encoding for kinetic distance from a reference state.

pfoldt. The extra node required by the pfoldf procedure is not necessary if p_{fold} is calculated not between two representative nodes A and B, but rather with a commitment time τ_{commit} , referred to as $p_{\text{fold}}(\tau_{\text{commit}})$ and defined as the probability of reaching A within τ_{commit} .^{33–35} The calculation of $p_{\text{fold}}(\tau_{\text{commit}})$ values for all nodes in the EKN (with the initial boundary condition $p_A = 1$) is more complex than that for pfoldf; details are given in the Supporting Information. Once $p_{\text{fold}}(\tau_{\text{commit}})$ has been evaluated for all nodes, the procedure is the same as for pfoldf: The nodes are sorted according to $p_{\text{fold}}(\tau_{\text{commit}})$ values and split into sets with $p_{\text{fold}}(\tau_{\text{commit}}) > p_c$ and $p_{\text{fold}}(\tau_{\text{commit}}) < p_c$. For each p_c between 0 and 1, the pair $[Z_A/Z, -kT \ln(Z_{AB}/Z)]$ is a point on the FEP. The choice of τ_{commit} has to be long enough to assign nonzero $p_{\text{fold}}(\tau_{\text{commit}})$ values to nodes that are kinetically very far from the node under consideration (i.e., the native node or any other node of interest) to resolve all other states. For this purpose, very long commitment times are appropriate. However, it is computationally most convenient to choose τ_{commit} values as short as possible. Thus, for each basin, one starts with a short τ_{commit} and increases it until the whole profile is covered, as illustrated in Figure S1 of the Supporting Information. For instance, 20 ns is long enough if $p_{\text{fold}}(\tau_{\text{commit}})$ is calculated with respect to the native node, but $\tau_{\text{commit}} = 200$ ns is needed if the values are calculated with respect to a node in the helical region. All pfoldt profiles shown in Figure 3 below were produced with values in this range (20–200 ns). Typically, the upper limit of

τ_{commit} is on the order of the overall relaxation time ($\tau_{\text{folding}} + \tau_{\text{unfolding}}$) of the system, which is about 200 ns for Beta3s.

mfpt. Another variable used in this work to project the free energy is the mean first passage time (mfpt; see Figure 1) to node A (representative nodes of significantly populated basins were used in Figure 3, but any node can be used as a reference).⁴⁵ Given the original EKN (i.e., without an extra node), the mfpt of node i is the solution of the equation $\text{mfpt}_i = \Delta t + \sum(p_{ji} \times \text{mfpt}_j)$ with initial boundary condition $\text{mfpt}_A = 0$.⁴⁶ The time step Δt corresponds to the saving frequency of 20 ps; that is, the mfpt of a node is defined as one time step plus the weighted average of the mfpt values of its adjacent nodes. In contrast to the other progress variables, the mfpt has an explicit time dependence through the occurrence of the time step in the equations. The resulting system of linear equations differs from that of pfoldf by the Δt constant and the boundary conditions; in pfoldf, the boundary conditions are $p_A = 1$ and $p_B = 0$, whereas there is only one condition $p_A = 1$ or $\text{mfpt}_A = 0$ for pfoldt or mfpt, respectively. Therefore, both pfoldf and mfpt equations can be solved with the same efficiency by iterative multiplication. Similarly to pfoldt, mfpt does not require an extra node, because mfpt is defined not between a pair of nodes, but only with respect to one selected node. To calculate the FEP, the nodes are sorted according to their mfpt values. For any mfpt_c between 0 and $\max(\text{mfpt})$, a point $[Z_A/Z, -kT \ln(Z_{AB}/Z)]$ on the FEP can be calculated, where A is the set of

all nodes with $\text{mfpt}_i < \text{mfpt}_c$ and B is the set of nodes with $\text{mfpt}_i > \text{mfpt}_c$ (Figure 1).

Implementation. In practice, the procedure to calculate the one-dimensional FEP consists of four steps: (1) Detailed balance is imposed on the equilibrium kinetic network (EKN), i.e., $c_{ij} = c_{ji} = (n_{ij} + n_{ji})/2$, where n_{ij} is the number of direct transitions (i.e., transitions between two MD snapshots separated by the time interval Δt , which is the inverse of the MD saving frequency) from node j to node i . The transition probabilities are then calculated as $p_{ij} = c_{ij} / \sum_k c_{kj}$. (2) The system of equations with appropriate boundary condition(s) is solved numerically. (3) Nodes are sorted according to increasing values of mfpt or decreasing values of pfoldf or pfoldt ; for each value of the progress variable, the relative partition function Z_A and the cut Z_{AB} are calculated. (4) The individual points on the profile are evaluated as $[x = Z_A/Z, y = -kT \ln(Z_{AB}/Z)]$.

Identification of Basins. The kinetic grouping analysis (KGA) groups conformations according to fast relaxation at equilibrium.¹⁹ More explicitly, two coarse-grained conformations are grouped if, along the molecular dynamics trajectory, their snapshots interconvert in more than 50% of the cases within a commitment time τ_{commit} , which represents a typical relaxation time within basins of the investigated system; the value used in this study was 1 ns. The basins obtained by KGA can be compared to those isolated from FEPs. To isolate a basin with a FEP, the unfolding profile from a node in that basin (usually its most visited node) is plotted, as shown in Figure 3. In practice, the procedure is the same as that used with the native basin as the reference, except that the native node is replaced by the new node. All nodes lying on the left of the cut at the first barrier correspond to the basin. Basins lying on the right of the first barrier are potentially overlapping (Figure 1), so each basin requires a separate unfolding profile.

Transition Disconnectivity Graph (TRDG). The TRDG is a variant of the free-energy disconnectivity graph, which provides an unprojected representation of the free-energy surface.⁴⁴ The partition function of the free-energy barrier separating states i and j , Z_{ij} , is equal to the value of the mincut between the states in the network, which can be calculated by the Ford–Fulkerson algorithm.⁴⁷ After the mincuts (i.e., the free-energy barriers) between every pair of nodes have been calculated, which can be done with only $n - 1$ total mincuts for n nodes by using the Gomory–Hu algorithm,⁴⁸ the TRDG is constructed to obtain a detailed representation of the free-energy surface. Following Becker and Karplus²⁰ and using the relation $F_{ij} = -kT \ln(Z_{ij})$, one starts with the largest Z_{ij} value (smallest F_{ij} value) and successively connects states in order of decreasing Z_{ij} (increasing F_{ij}). The TRDG is useful for visualizing basins containing representative nodes (enthalpic basins), but basins that have no such nodes (entropic basins) are not visible. To resolve such basins, one has to use either the balanced mincut procedure²² or the free-energy profiles discussed above.

Results

All analyses are based on a set of 10 2.5- μs equilibrium simulations at 330 K started from the folded state. The first 0.5 μs of each run was neglected so that a total simulation time of 20 μs was sampled for each of the two peptides (see Methods). The wild-type Beta3s peptide visited 262 433 conformations (unique strings of secondary structure) with a total of 534 383 direct transitions. The W10V mutant visited 245 032 conformations with a total of 476 721 direct transitions. In Beta3s (W10V), only 62 446 (56 118) conformations were visited more than once. The most populated conformation of Beta3s and its

W10V mutant had statistical weights of 5.6% and 8.8%, respectively. It was the three-stranded antiparallel β -sheet with type II' turns at residues 6–7 and 14–15 (secondary structure string –EEEESEEEEEESSEEEE–), which corresponds to the native state determined by NMR spectroscopy.²⁷ Totals of 120 and 105 folding events (i.e., visits to the native node) were observed for Beta3s and W10V, respectively, with an average folding time of about 0.1 μs for both peptides.

FEP of Beta3s and W10V. The pfoldf FEP, based on p_{fold} values, was calculated by projecting the free energy on the relative partition function Z_A/Z , which is a progress coordinate that increases monotonically with the distance from the reference state;²⁴ see Methods. pfoldf finds approximate barriers between the reference state and the denatured state, whereas exact values can be obtained by the mincut procedure.²² An essential attribute of Z_A/Z is that it takes into account all routes from the initial state to the final state without any prejudice as to the geometric coordinates or pathways involved, so that the FEP is determined by an unbiased procedure. Figure 2 shows the results for both Beta3s and W10V. Three main regions are identified on the FEP of Beta3s when the most populated (native) node is used as reference: native state ($Z_A/Z < 0.35$), denatured state with several enthalpic subbasins ($0.35 \leq Z_A/Z \leq 0.88$), and helical basin ($Z_A/Z > 0.88$). The pfoldf FEP procedure yields an accurate description of the reference basin (native state in Figure 2), whereas there can be some overlap between different basins after the first barrier (i.e., for $Z_A/Z > 0.35$). Overlap occurs whenever nodes belonging to different basins have similar p_{fold} values (Figure 1). Unfolding profiles (see Methods) are able to fully resolve all the basins. Such unfolding profiles from representative nodes in the subbasins of Figure 2 are plotted in Figure 3 to accurately characterize each basin by eliminating overlap with other regions. This makes possible the determination of accurate barriers between the respective basins and the exact population of each basin. In each unfolding profile of Figure 3, the exact value of the barrier, calculated by mincut,²² is indicated by an open circle. The plots show that pfoldf approximates the barriers very well for the Beta3s system. (For a description of the pfoldt and mfpt results, see below.) Notably, the most populated basins isolated by this unbiased procedure²⁴ correspond to those in Table 1 of ref 19 and are quantitatively compared in Table 2 of the present work. Figure 4 qualitatively illustrates the nodes, the basins, and their connectivity in the conformational state network.³² The basins determined by pfoldf are discriminated by different colors and shapes in the network, where brown nodes belong to the entropic state. Note that the colors in the network (i.e., basins defined by pfoldf) are in good agreement with the network in Figure 1 of ref 19 for KGA results. Table 3 contains effective and free-energy values of the six basins and the entropic state. Clearly, there are low-enthalpy, low-entropy basins (native, Ns-or, Cs-or, and the two curls), as well as high-enthalpy, high-entropy basins (helical and entropic states); the origin of the high entropy of the helical basin is discussed below. Except for the helical basin, the agreement between the FEP procedures and the KGA¹⁹ is very good, and thus, the two approaches validate each other. Furthermore, essentially identical basins are isolated by either of the two procedures using secondary structure or all-atom 2.5- \AA rmsd coarse-graining (as shown in Table S-I of the Supporting Information for pfoldf).

Although the Beta3s and W10V profiles are very similar, there are subtle differences between them. Importantly, the native state is less stable in Beta3s than in W10V (35% vs 39.5%), the helical basin is slightly more populated in Beta3s than in W10V

TABLE 2: Results of the KGA and pfoldf Procedure for the Most Populated Basins of Beta3s^a

basin	name/color	weight			
		heaviest node	KGA (%)	pfoldf (%)	similarity ^b (%)
−EEEESEEEEEESSEEEE−	native		36.4	35.0	99.8
−EEEESTEEEEESSEEEE−	Ns-or		7.4	6.2	99.99
−EEEESEEEEEESSEEEE−	Cs-or		3.6	2.6	99.8
−−−SSGGG−−−EESSEETT−	Ch-curl ₁		3.6	2.8	99.5
−−−SSGGG−EESSTTTTEE−	Ch-curl ₂		2.4	2.1	98.8
−HHHHHHHHHHHHHS−−−−−	helical ^c		2.1 (11.6)	11.2	99.4 (96.4)
−EESSEEEEEESSEEEE−	cyan		1.9	1.5	97.9
−−−SSGGG−EESSESEEEE−	white rectangle		1.2	1.1	98.8
−−−SSSS−EESTT−EEE−	white diamond		0.9	0.9	96.3
−−−−STT−−−EESSEEEE−	magenta		0.8	0.5	99.7
−	entropic-1 ^d		33.8	32.7	76.0
−	entropic-2 ^e		27.2	32.7	93.0

^a The pfoldf, pfoldt, and mfpt procedures determine the basin populations by plotting the FEP from the representative (most populated) node for each basin as shown in Figure 3. The cut at the first barrier defines each basin. Note that the pfoldf procedure yields essentially identical basins using clusters obtained by secondary structure or all-atom 2.5-Å rmsd coarse-graining (see Table S-I in the Supporting Information). ^b The similarity value is calculated as the intersection of two corresponding basins (i.e., that obtained using KGA and that using pfoldf), normalized to the one with the lower population. Note that pfoldf, pfoldt and mfpt populate the investigated basins equally (similarity at least 99%, not shown). ^c Using KGA, the population of the largest helical basin is only 2.1%, while the ensemble of strings with 4, 5 or 6 consecutive G, H and I's, respectively, populates a total of 11.6%, which is very close to the helical basin population obtained by pfoldf. ^d The entropic-1 state is defined for KGA as conformations not belonging to basins with more than 1% population, and for the pfoldf procedure as conformations not belonging to one of the significant basins of the profile. ^e The KGA entropic-2 basin does not include the helical conformations.

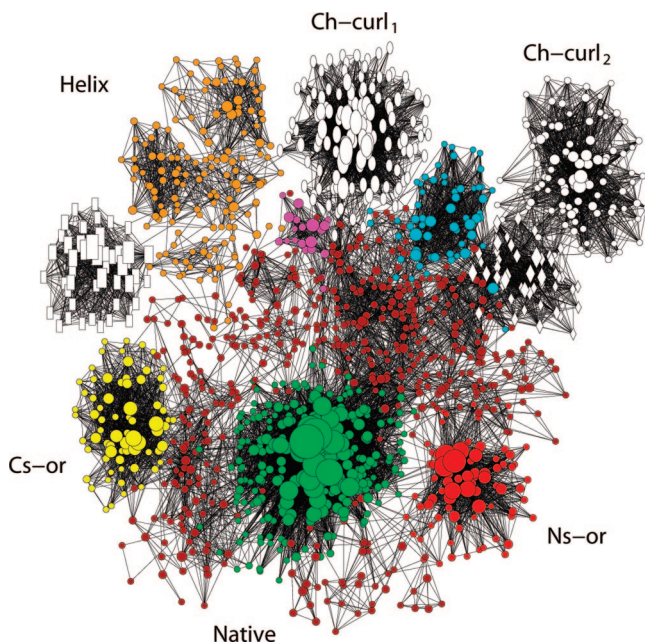


Figure 4. Conformational space network of Beta3s. Each node (i.e., conformation) of the network represents a secondary structure string. The surface of each node is proportional to its statistical weight, and only the 1430 nodes containing at least 40 snapshots are shown to avoid overcrowding.¹⁹ Nodes are colored according to the basins isolated with pfoldf FEPs (Figure 2 and Table 2). White nodes are populated significantly only in Beta3s and not in the single-point mutant W10V;¹⁹ the different white symbols are assigned according to pfoldf (Table 2). Note that the helical basin is assigned a single color when isolated according to the pfoldf FEP and multiple colors by KGA,¹⁹ as explained in the text. Brown is the color of some of the nodes in the entropic basin of which most nodes are not shown because they contain less than 40 snapshots each.

(11.2% vs 8.8%), and there is a difference in the relative stability of nonhelical misfolded species such as Ns-or (N-terminal strand out of register and folded C-terminal hairpin, basin populations of 6.2% vs 3.6%) and Cs-or (C-terminal strand out of register and folded N-terminal hairpin, basin populations of 2.6% vs 4.9%). These statistical weights were calculated by defining the

basins as described in the Methods section. Note that corresponding basins in the two systems can occur at different Z_A/Z positions in the profiles of Figure 2. Also, the FEPs reveal basins that are visited in only one of the two peptides. Such basins are illustrated in white in Figure 4.

Figure 5 shows the negative logarithm of the probability of the first passage time (fpt) to the native node (i.e., the folding time). The plot has a minimum at about 100 ns, which corresponds to the folding time. The fast folding values correspond to configurations that start in the native state (fpt < 1 ns), and the slower folding ones are configurations that start in the denatured state (fpt > 1 ns). The fpt plot shows a very simple behavior, such as is expected for a two-state system, despite the multimimum character of the free-energy surface.

FEPs Calculated with pfoldt and mfpt. The pfoldf analysis of systems without a representative node in the denatured state is based on the introduction of an extra node²⁴ that is linked with a small capacity λ (typically 0.01 or lower) to all nodes in the network. Here, p_{fold} with a commitment time of $p_{\text{fold}}(\tau_{\text{commit}})$ and the mean first passage time (mfpt) to the reference node as progress variables are introduced to plot the FEPs. The respective procedures, called pfoldt and mfpt, have the advantage that no additional node needs to be introduced.

The pfoldt- and mfpt-calculated FEPs are in good agreement with those obtained by the pfoldf procedure (Figure 3). The pfoldf, pfoldt, and mfpt barriers separating significantly populated basins are almost identical for all procedures and both peptides. Further, pfoldt and mfpt, as well as pfoldf (see above), yield good approximations to the exact barriers (open circles in Figure 3). The six basins shown in Figure 3 share more than 99% of their conformations when isolated by the three methods (not shown), indicating that the results are robust. This suggests that the choice between pfoldf, pfoldt, and mfpt can be made according to convenience. The cut between two well-defined regions with representative nodes has to be calculated by pfoldf, because it is the only procedure of the three discussed here in which two input nodes are used, but pfoldt and mfpt are the more straightforward choices if only one representative node exists, such as when an unfolding profile is calculated. Solving

TABLE 3: Energetic and Entropic Contributions to Basin Stability

pfoldf basin	weight (%)	$\langle E \rangle^a$ (kcal/mol)	$\langle \Delta E \rangle^b$ (kcal/mol)	ΔF^c (kcal/mol)	$-T\Delta S = \Delta F - \langle \Delta E \rangle$ (kcal/mol)	barrier ^d	
						secondary str (kcal/mol)	rmsd (kcal/mol)
native	35.0	-7.2	0	0	0	1.9	2.5
Ns-or	6.2	-6.4	0.8	1.1	0.3	1.2	1.7
Cs-or	2.6	-3.3	3.9	1.7	-2.2	0.8	1.6
Ch-curl ₁	2.8	-10.1	-2.9	1.7	4.6	1.0	3.2
Ch-curl ₂	2.1	-10.1	-2.9	1.9	4.8	1.2	2.6
helical ^e	11.2	2.2	9.4	0.8	-8.6	1.4	1.3
entropic ^e	32.7	2.9	10.1	0.0	-10.1		

^a Average effective energy, that is, the sum of CHARMM param19 force field³⁸ and the SAS solvation model³⁹ contributions. The average was calculated over all snapshots in the clusters (i.e., secondary structure strings) belonging to the basin determined by the pfoldf procedure. The error of the average effective energy is less than 0.5 kcal/mol as estimated by the difference of the mean values calculated on one-half of the sample, i.e., two segments of 10 μ s each. ^b Average effective energy relative to the folded state. Note that, in any force field, the absolute value of the effective energy E is arbitrary, and only ΔE values relative to a reference state are meaningful. ^c Free energy relative to the folded state calculated as $\Delta F = -kT \ln(\text{weight}/35.0)$, where 35% is the weight of the native basin as isolated by pfoldf. ^d Barrier to exit individual pfoldf basin, calculated for the secondary structure and 2.5-Å rmsd coarse-graining. ^e These basins are stabilized mainly by entropy, as indicated in the table.

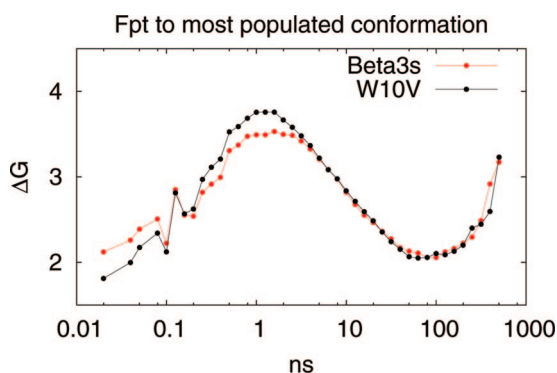


Figure 5. Profile obtained using fpt as the progress variable and calculated as $\Delta G = -k_B T \ln[P(\text{fpt})]$ on the bins, whose size increases exponentially (10 bins/decade) for better resolution of the different timescales.

the system of equations for the pfoldf and mfpt procedures is of the same complexity, because the only differences are the boundary conditions and the use of the time constant Δt that is added in the mfpt equations (see Methods). On the other hand, the pfoldf method requires some precalculations before the iterative solution of the equation can be performed and is therefore more complex (see Supporting Information). An application of the mfpt analysis is that it can be used as the progress coordinate (instead of Z_A/Z) to obtain a FEP with all basins separated from the native one by a distance in time units. The relation between the profiles projected on Z_A/Z and on mfpt is a nonlinear transformation $x \rightarrow \text{mfpt}[x(Z_A/Z)]$, where $x(Z_A/Z)$ assigns nodes to each position on Z_A/Z ; that is, the mfpt that was originally used to rank on the Z_A/Z axis is now directly assigned to the nodes. Figure 6 clearly shows mfpt values of individual basins. Interestingly, it also illustrates the origin of the single-exponential behavior (see Single-Exponential Kinetics of Folding), which is a spread of only a factor of 3 in the mfpt values (from about 60–80 ns for Ns-or and Cs-or to about 190–200 ns for Ch-curl₁, Ch-curl₂, and helical). However, as can be seen by comparing Figure 3 with Figure 6, there is more overlap of non-native basins using mfpt rather than Z_A/Z as the progress coordinate.

Helical Basin. A previous study³² suggested that the denatured state ensemble of Beta3s is highly heterogeneous and includes enthalpic traps as well as conformations with partial helical structure; the latter form the helical basin. Notably, in the FEPs of Figures 2 and 3, the entire “helical” region is

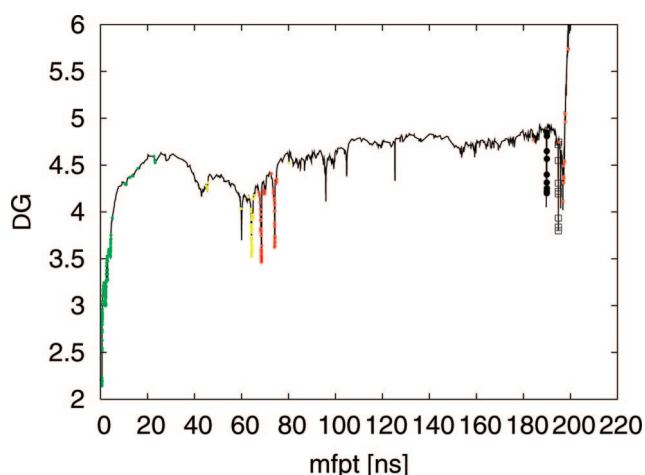


Figure 6. Beta3s unfolding FEP calculated for the directed network (see Methods) using mfpt as a progress coordinate and a progress variable, which can be obtained by the transformation $x \rightarrow \text{mfpt}[x(Z_A/Z)]$. As in Figure 3, individual basins are colored according to basins isolated from pfoldf FEPs. Interestingly, the similar mfpt values for the enthalpic traps and the spread of only about 3 between mfpt values of enthalpic traps and the helical basin are consistent with the single-exponential behavior of folding (see the Results section). The 2.5-Å rmsd clustering is used here because the corresponding FEP calculated for the secondary structure coarse-graining is affected by pseudotunneling (see text and Figure S2 in the Supporting Information).

identified and shown to be separated by a high barrier (at Z_A/Z of about 0.88) from the rest of the denatured state, which extends from 0.35 to 0.88 (see above). The helical region shows the main difference with respect to the KGA results¹⁹ and is indicated in Figure 4. KGA correctly identified the two most populated free-energy subbasins (—HHHHHHHHHH—HHS— and —TT—HHHHHHHSS—) with populations of 1.9% and 1.6%, respectively) within the helical state of Beta3s (Figure 2). The commitment time of 1 ns used in the previous work¹⁹ was too short to group all helical structures into one basin, because the helical basin is divided into various subbasins separated by barriers, as can be seen in the helical unfolding profile of Figure 3. These barriers prevent the system from rapid equilibration between all helical structures. A larger commitment time of 5 ns, however, is able to identify the entire helical basin (Supporting Information of ref 19). These results show that the definition of a basin involves the choice of “resolution”. Both the commitment time of KGA and the height of the barrier in the FEP analysis, above which

one considers a basin as separated, correspond to the “lens” with which the free-energy surface is analyzed. For each choice of a minimum barrier in the FEP procedures, there exists a commitment time for the determination of the corresponding basins with KGA. However, defining a minimum barrier height is more transparent than choosing a commitment time, which is initially extracted from the fpt plot of Figure 5 and then varied to obtain the desired resolution. In either case, there is some arbitrariness in defining a basin per se.

The helical state is the right-most basin in all five unfolding profiles from the nonhelical basins in Figure 3 (see also Figure S2 in the Supporting Information). This observation is consistent with the high barrier that has to be overcome to enter the helical basin from the rest of the conformational space, as the barriers generally appear in increasing order along the Z_A/Z progress coordinate.

Interestingly, as shown in the thermodynamic analysis presented in Table 3, the helical basin has a high energy and is entropically stabilized. This is because the strings associated with the helical basin have unstructured residues that do not make hydrogen bonds. In the helical basin, 78.4% and 7.7% of the snapshots have more than 5 and 10 unstructured residues, respectively; that is, they belong to strings with more than 5 and 10 “—” letters. The corresponding percentage values for the entropic region are 79.7% and 15.2%, respectively. As a basis of comparison, the native state has only 2.9% and 0.003% of its snapshots in strings with more than 5 and 10 unstructured residues, respectively. Moreover, the numbers of different secondary structure strings in the helical, entropic, and native states are 57 134, 193 666, and 2672, respectively.

Simplified Network (SEKN) and the Role of the Entropic State. In previous analyses of folding simulations, it has been found useful^{22,24} to construct a highly simplified network that shows only the main basins and their connectivity as a complement of the detailed conformational space network (Figure 3). Figure 7 shows such a network with the free-energy basins isolated by pfoldf. It includes the heterogeneous entropic state (dashed surface in Figure 7), which is made up of all conformations not belonging to any of the 10 basins that appear as a part of the network; the latter all have barriers higher than 0.3 kcal and significant partition function ($\geq 0.5\%$). The so-called entropic state, by contrast, is composed mainly of nodes that are visited only once, or at most a few times, so that it is not a “true” basin in the sense used for the other basins. Overall, at the chosen simulation temperature (330 K), the denatured state of Beta3s consists of the entropic state (populated at 33%), a helical basin (populated at 11%) and eight metastable enthalpic traps (populated at 0.5–6%). Although the free-energy basins were selected with the pfoldf procedure on the full EKN, the links in Figure 7 show the number of transitions, sampled along the molecular dynamics trajectory, between the most populated secondary structure string (bottom) of a given pfoldf basin i and the bottom of another basin j (or the same basin) through the entropic state only (i.e., without visiting another enthalpic basin). Each of the pfoldf profiles in Figure 7 was calculated using only the basin under consideration, the entropic state, and the native state while neglecting the other basins. In this way, barriers involved in the transitions between individual enthalpic basins and the native state are described accurately; see also the Supporting Information, section D. Note that the left-most profile in Figure 7 (i.e., pfoldf-calculated FEP from the entropic state) considers only the entropic and native states, which account for almost 70% of the total weight. All profiles show a barrier of only about 0.5 kcal/mol from the entropic state

toward the native state, which is near the limit of what has been termed barrierless or downhill folding.^{49–54}

The complexity of the SEKN, particularly for the denatured state, suggests that a detailed analysis would be useful to obtain a more complete understanding of the folding behavior. Figure 7 shows the number of “direct” transitions (i.e., without visiting the native state) between identified basins, which can be compared with the number of transitions from each of the basins to the native state. In most cases (the pair cyan and Ns-or is an exception), the direct transitions are rather rare compared to the number of transitions connecting each basin to the native state. However, the total number of transitions connecting the non-native basins to other non-native basins without passing through the native state is of the same order of magnitude as the number of transitions connecting each of the non-native basins to the native state.

If the equilibrium trajectories are followed, they make clear that direct transitions overwhelmingly go through the entropic basin from one of the defined basins to another. This is in accord with the results in Figure 7, which show that a long time is spent in the entropic basin in nearly all transitions. However, in most cases, the trajectories go through the native basin and can be diagrammed as ($i \rightarrow$ entropic \rightarrow native \rightarrow entropic $\rightarrow j$), often spending a long time in the native basin and making repeated transitions ($i \leftrightarrow$ entropic \leftrightarrow native \leftrightarrow entropic $\leftrightarrow j$). There is essentially full equilibration in the native basin with rapid sampling of different conformations; there are 2672 nodes (different secondary structure strings) in the native basin. Further, if one examines the content of native secondary structure in the last 0.2 ns before the trajectory exits from the folded to the entropic state, with the condition that it will next visit a certain enthalpic basin (e.g., Ns-or or Cs-or), a significant structural bias toward that basin is already present (see Figure 8). Thus, the fate of the trajectory is biased already in the native state. This analysis is in accord with the conclusion that the system does not stay in the entropic state long enough to equilibrate. The latter is due in part to the aforementioned low barrier between the entropic state and the native state (see the left-most profile of Figure 7) and in part to the presence of significant barriers between different parts of the entropic basin. This is shown in section E of the Supporting Information by the reduced pfoldf FEPs of pairs of non-native basins and the entropic state (e.g., Ns-or, entropic, and Cs-or in Figure S6a). However, not all basins of the denatured state are separated by barriers within the entropic state (one example is in Figure S6c for the Ns-or and the cyan basins). These observations explain the origin of the barriers in the entropic region at $0.3 \leq Z_A \leq 0.4$ for the reduced FEPs (i.e., FEPs calculated taking into account only one basin and the entropic and native states) from the Cs-or or the helical basin in Figure 7. In other words, Figures 7 and S6 are consistent because both show barriers in the entropic region mainly between Cs-or, Ns-or, and helical regions, but not among conformations with the C-terminal hairpin folded (i.e., Ns-or, Ch-curl₁, Ch-curl₂, cyan, and magenta basins in Figure 7).

The number of folding transitions from non-native basins with a structured C-terminal hairpin is larger than the corresponding number from basins with the N-terminal hairpin formed (Cs-or). This observation is consistent with the analysis of folding transition state structures¹⁹ identified by a node- $p_{\text{fold}} = 0.5$ criterion.³⁴ The two main folding pathways of Beta3s are in agreement with the diffusion-collision model⁵² in which the folding process involves the encounter of marginally stable secondary structural elements.⁵³ The SEKN also sheds light on

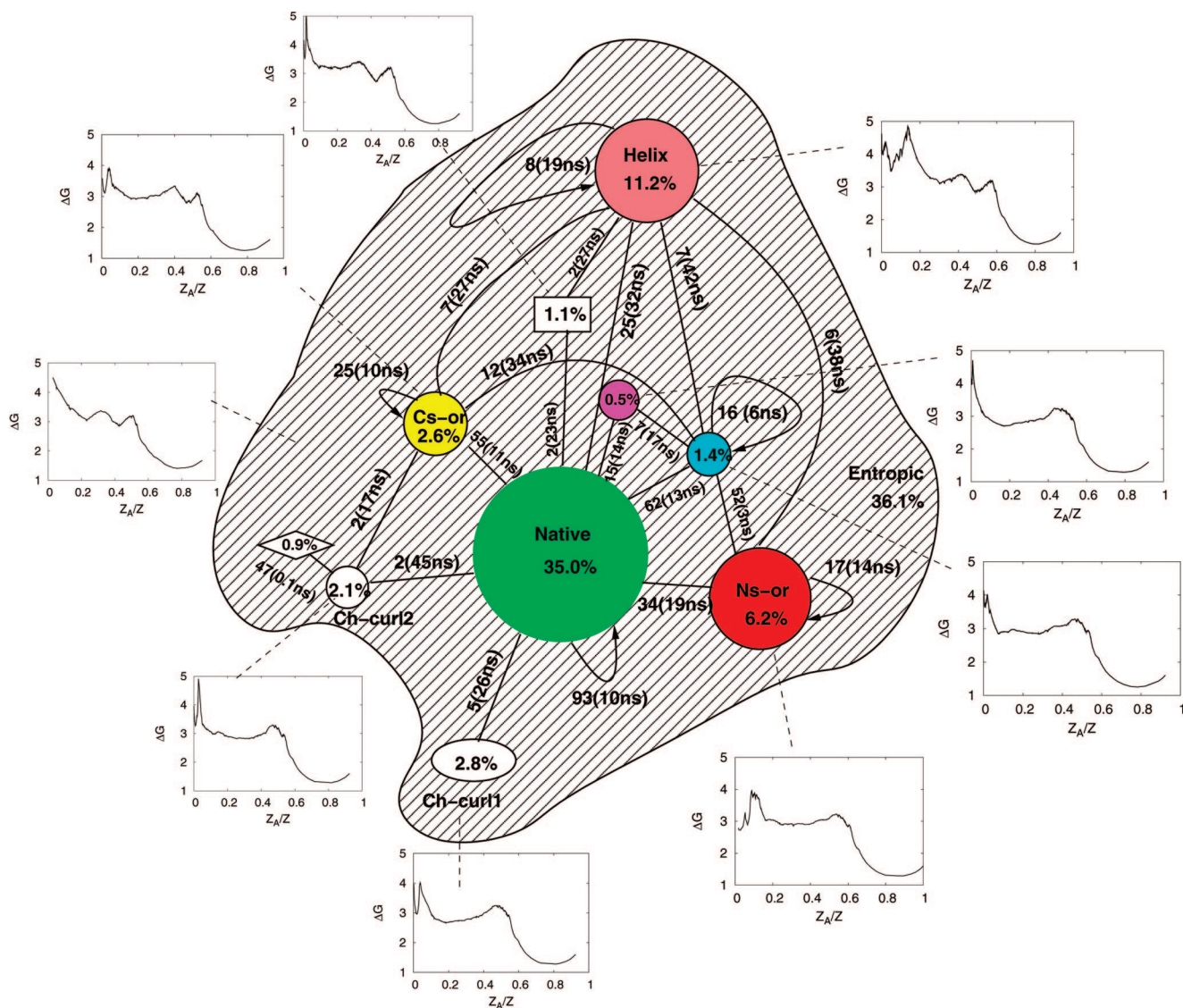


Figure 7. Simplified equilibrium kinetic network (SEKN) of Beta3s free-energy basins. The circles, ellipse, rectangle, and diamond are the 10 most populated free-energy basins of Beta3s with their respective statistical weight as isolated by pfoldf. Colors and geometrical shapes were chosen to be consistent with the basins identified by KGA.¹⁹ The dashed surface represents the entropic basin that is made up of all conformations not belonging to any of the 10 basins. Plotting the entropic basin as a circle would be incorrect given that there is no fast equilibration inside it because of the low barrier toward the native state. Each link shows the number of transitions between the bottom (most populated secondary structure string) of basin i and the bottom of basin j through the entropic basin only (i.e., without visiting any other basin bottom), and the average time spent in the entropic state is given in parentheses. Transitions from non-native enthalpic basins and from the entropic state to the native state are illustrated by pfoldf FEPs, where the representative node of the entropic state is the extra node. Each profile from a non-native basin was calculated using only the basin under consideration (dashed line), the entropic state, and the native state while neglecting the remaining eight basins. The profile from the entropic state (left middle) takes into account only the entropic and native states. Each profile gives a description of both the entropic/native barrier ($Z_A/Z \approx 0.5$) and the barrier to leave the enthalpic basin ($Z_A/Z \lesssim 0.1$), and these regions are both devoid of overlap.

the folding pathways from the helical state (Figure 7). Half of the transitions proceed via the entropic state, where the system spends considerable time, directly to the folded state. Less frequently, the trajectory visits Cs-or or Ns-or structures or even returns to the helical basin. We note that, even though the folding times from the Ns-or and Cs-or conformations are relatively large (110 and 70 ns, respectively) and the Ns-or and Cs-or conformations hardly interconvert, the difference between the most populated nodes of the three basins is only in one to three positions of the secondary structure string. This observation illustrates that a small structural change can result in a large kinetic distance. For instance, the structural change from folded to Cs-or, i.e., from two to three S letters at the second turn (see Table 2 for strings), involves a complete rearrangement of the side chains of the C-terminal strand.

For each of the basins identified in the SEKN, the analysis showed that the equilibration within the basin is fast, relative to transitions from or to the basin. However, this is not true for the entropic basin. The low barrier between the entropic state and the native state together with the high population of the former (about 33%) leads to fast transitions to the native state that prevent equilibration in the entropic basin. Therefore, plotting the entropic state as a single node would not only be misleading, but would also give an incorrect picture of the pathways. If the entropic state were replaced by one node, the native state would be connected to only the entropic node, and the picture would suggest that the entropic state acts as the hub. However, in the SEKN emerging from the pfoldf procedure, both the native and the entropic states can be considered to be hubs.

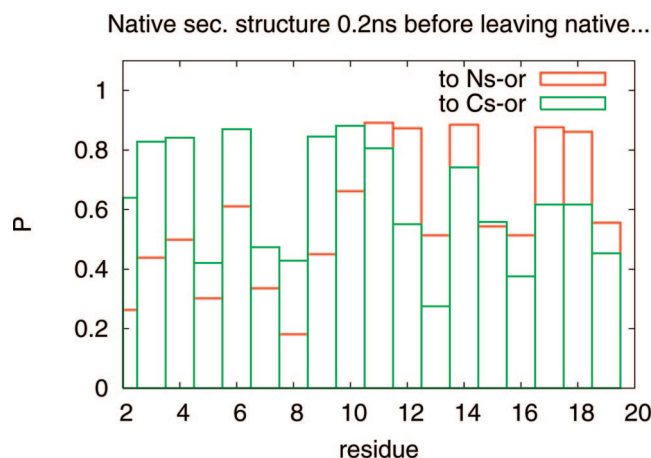


Figure 8. Native secondary structure content of the trajectory in the folded state, 10 snapshots (0.2 ns) before the system leaves toward one of the two non-native enthalpic basins Ns-or or Cs-or. Notably, if the trajectory continues to Ns-or (Cs-or), the N-terminal (C-terminal) hairpin is already disrupted before leaving the native state.

Transition Disconnectivity Graph (TRDG). The TRDG⁴⁴ of Beta3s (Figure 9) provides further evidence that the denatured state is heterogeneous and has several funnel-like basins with favorable effective energies as well as a helical basin. Disconnectivity graphs do not visualize the basins that lack representative nodes, i.e., the basins with high entropic contributions to the free energy. The split of the helical basin is consistent with the FEPs of Figures 2 and 3 and explains the longer commitment time required in KGA to isolate the complete helical basin, as discussed in the Results section. An advantage of TRDG over one-dimensional FEPs and the SEKN is that it quantitatively depicts (mainly enthalpic) minima and barriers in a single plot.

Single-Exponential Kinetics of Folding. The cumulative distribution of folding times of all configurations in the trajectory shows single-exponential behavior (Figure 10 top). This (apparently simple) behavior is consistent with the fact that the barriers to exit the individual basins are of similar heights in the FEPs from enthalpic traps or the helical basin (Figure 3 and Table 3). Together with the aforementioned low barrier from the entropic state to the native state, the similar barrier heights explain the single-exponential behavior. The similar free-energy barriers from Ns-or and Cs-or are likely to be a consequence of the high sequence identity (67%) between the N-terminal β -hairpin (residues 1–12) and the C-terminal β -hairpin (residues 9–20). On the other hand, there is no straightforward explanation for the similar barrier height from the helical basin.

Model rate calculations for a photoswitchable peptide have shown that, for an SEKN similar to the present one, a spread in the rates on the order of a factor of 9 is required to observe significant deviations from single-exponential behavior.⁵⁴

Because the entropic state cannot be represented as a single node in the SEKN (because of slow relaxation), the distribution of folding times was obtained by simulating MC kinetics on the network rather than on the SEKN. An initial point was picked arbitrarily among nodes that do not belong to the native basin with a probability proportional to the statistical weight, and MC simulations were performed until the system reached a node corresponding to the native basin. Figure 10 (bottom) shows the density function of the distribution obtained with 10^5 trajectories and a single-exponential distribution with the corresponding folding time. The curves are in reasonable agreement, which indicates that the kinetics can be approximately described as single-exponential.

Interestingly, the present analysis indicates that somewhat deceptive single-exponential behavior can emerge from completely different energy landscapes as for both Beta3s, which has a kinetically partitioned denatured state and a hub-like native state, and the β -hairpin of protein G, which shows fast equilibration within a multibasin denatured state.²²

Concluding Discussion

Considerable progress has been made recently in experimental investigations of protein folding. Particularly, for the small fast-folding proteins, which have been studied most,⁴ a two-state description (folded and unfolded) is adequate to describe the measurements. This means that little, if any, information concerning the details of the folding pathways is obtained, although mutation studies have been used to provide a coarse-grained description of the transition state.^{55,56} Also, experiments supplementing the kinetic measurements by probes sensitive to structural details^{6,54} have yielded some insights into the folding pathways, particularly when intermediates are present.⁵⁷ However, none of the experimental studies provide a detailed description of the structures that contribute significantly to the ensembles that are sampled along the folding pathways. Although one might hope for such information from future experiments, as of now, the only way to approach this problem is by computer simulations. In recent years, aided by faster, often massively parallel computational resources, the first steps toward this aim have been realized. Mostly, the studies have been limited to peptides^{22,32,42} and a few mini-proteins (e.g., Trp cage) that fold on the simulation (nanosecond) time scale. Unfolding simulations at high temperature have been interpreted in the folding direction,⁵⁸ but with few exceptions,⁵⁹ the unfolding reaction has been followed only once in the stimulations. There are very few systems for which it has been possible to do the multiple folding simulations required to obtain statistically meaningful results for analysis. The Beta3s mini-protein, with an implicit solvent model, is one such system. The present analysis is based on equilibrium simulations 20 μ s in length that show about 100 folding/unfolding events for a temperature (330 K) at which the native and denatured states are both significantly populated (35% native and 65% denatured).

One problem in using the simulation results is the difficulty of analyzing them to obtain an understanding of the folding reaction. The number of degrees of freedom (3×215 for a small system such as Beta3s in the polar hydrogen approximation, in which aliphatic and aromatic hydrogen atoms are not considered explicitly) makes a straightforward approach impossible, even though all of the details (hopefully representative of the actual folding process) are available from the trajectory. Use of the results requires a method for reducing the problem to one or only a few dimensions that are sufficient to describe the folding reaction in a meaningful way. The recently developed cut-based FEP procedures and complex network analyses have been shown to essentially solve this problem. They have demonstrated, among other conclusions, that the very simple picture of protein folding (e.g., one or at most two barriers between the denatured and native state), often obtained by projecting the free energy on an arbitrarily chosen progress variable(s), is not consistent with the complexity of the actual free-energy surface.^{18,22,32,42} Such complexity has spurred the development of more sophisticated computational procedures for determining free-energy basins and transitions among them. The essential element of both the $p_{\text{fold}}/m\text{fpt}$ -based procedures for FEP calculation²⁴ and kinetic grouping analysis¹⁹ is the

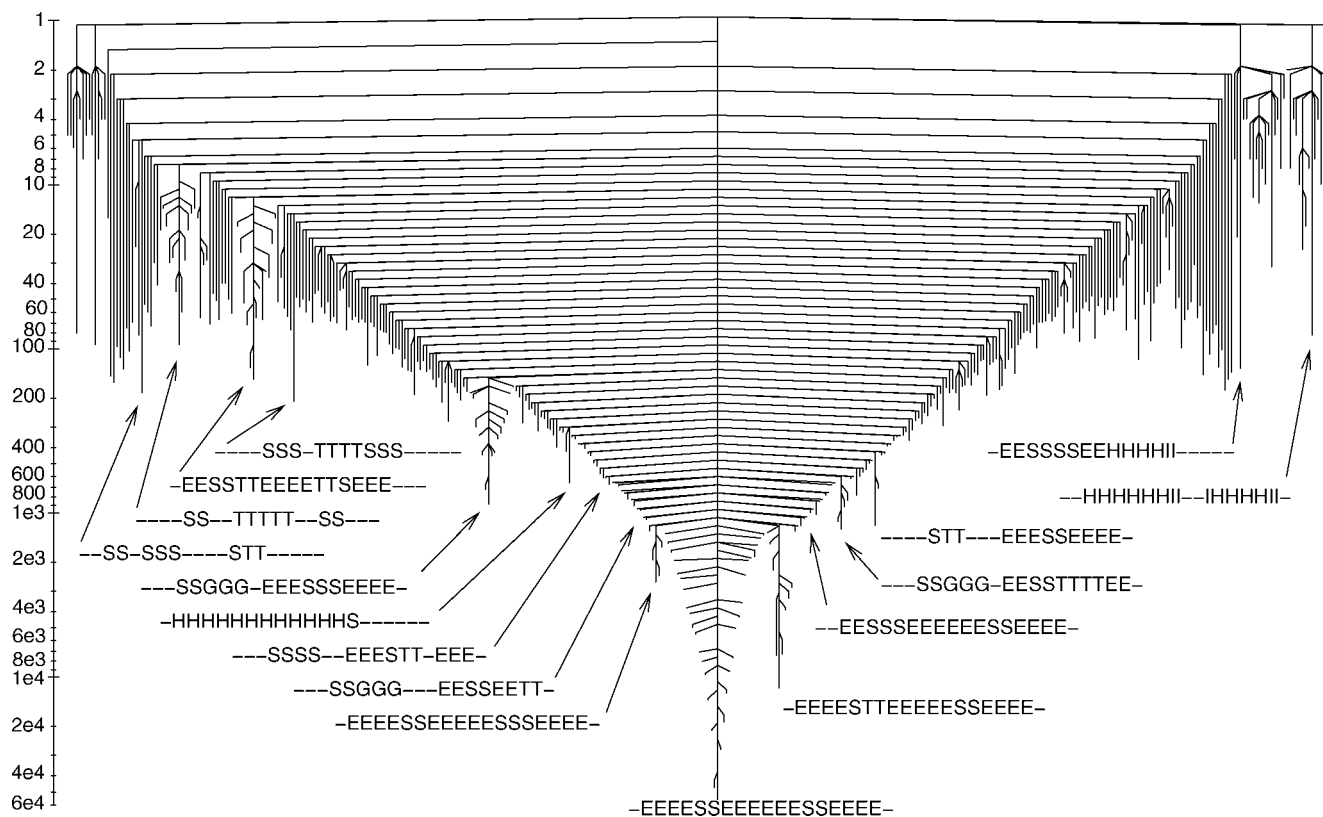


Figure 9. Free-energy (transition) disconnectivity graph of Beta3s EKN obtained with secondary structure coarse-graining. Secondary structure strings of the most populated clusters in the basins are shown. The vertical axis shows the number of times the secondary structures were visited (the bottom of each vertical line) and the number of transitions between pairs of strings (the value where the lines corresponding to two strings intersect). Information about the strings is given in Table 2.

identification of free-energy basins, not according to geometrical characteristics (such as the fraction of native contacts or rmsd from the folded structure) but rather according to the transitions that occur in folding/unfolding trajectories at equilibrium. From such an analysis, a meaningful one-dimensional projection of the free-energy surface (called FEP in this work) is obtained. It provides the basins on the surface and the barriers between them. Unlike the standard projections, which lead to overlap of the basins that smooth out the barriers and can make them disappear (as shown for the β -hairpin of protein G in ref 24), the progress variables used here do not result in such overlap if a valid clustering algorithm is used. The formulation requires a coarse-graining approach to group snapshots saved along equilibrium trajectories into nodes, so that adequate transition statistics can be obtained. Some evidence for the robustness upon changing the coarse-graining algorithm (based on rmsd or secondary structure string) was provided previously^{22,32} and is given in the Supporting Information, but this issue should be analyzed in more detail.

An important conclusion from the present study is that the $p_{\text{fold}}/\text{mfpt}$ -based procedures²⁴ find the same free-energy basins as the kinetic grouping approach described in a previous work.¹⁹ The very similar results for the free-energy surface, as well as the identification of the subtle differences between Beta3s and its W10V mutant, indicate that the two approaches are correct and complementary. One aspect of the $p_{\text{fold}}/\text{mfpt}$ -based procedures is that they are able to separate a free-energy basin consisting of an ensemble of conformations with fluctuating helical (i.e., non-native secondary structure) content. The helical basin has a population of about 11%, and from this basin, the folded state is reached through a heterogeneous entropic state. The significant statistical weight of the helical basin is surprising

if one considers that Beta3s is a peptide designed to assume a three-stranded antiparallel β -sheet fold. Because of its entropic stabilization, the helical state of Beta3s is expected to be less populated at lower temperature. Interestingly, an α -helix-rich kinetic intermediate in the refolding (by chemical denaturant dilution) of the β -sandwich protein src SH3 has been reported on the basis of circular dichroism, fluorescence, and X-ray solution scattering experiments.⁶⁰ Moreover, at pH 3, a helical equilibrium intermediate of the A45G mutant of src SH3 has been observed, and evidence has been provided that it corresponds to a kinetic intermediate.⁶¹

The differences between Beta3s and its W10V mutant are small but relevant. There is a shift of the equilibrium in the helix/ β -sheet statistical-weight ratio from 11/35 for the wild type to 9/40 for W10V. This indicates that even a single-point mutation can have an influence on the relative propensity of secondary structure formation that plays a critical role in diseases related to protein misfolding and aggregation.⁶²⁻⁶⁴

About one-third of the snapshots saved along the molecular dynamics trajectories belong to a heterogeneous entropic state that is visited during individual transitions between mainly enthalpic free-energy basins. There is no fast equilibration within the entropic state because of the low barrier of only around 0.5 kcal/mol toward the native state, but also because of barriers that split the entropic state. This explains the previously reported kinetically partitioned denatured state.¹⁹ As a consequence, the future development of the system depends strongly on the region where the system enters the entropic state or, equivalently, where it exits the native basin (Figure 8). Because the native and entropic states together make up almost 70% of the total weight, the free-energy surface of Beta3s seems to exhibit some of the features of a fast barrierless/low-barrier folder but with meta-

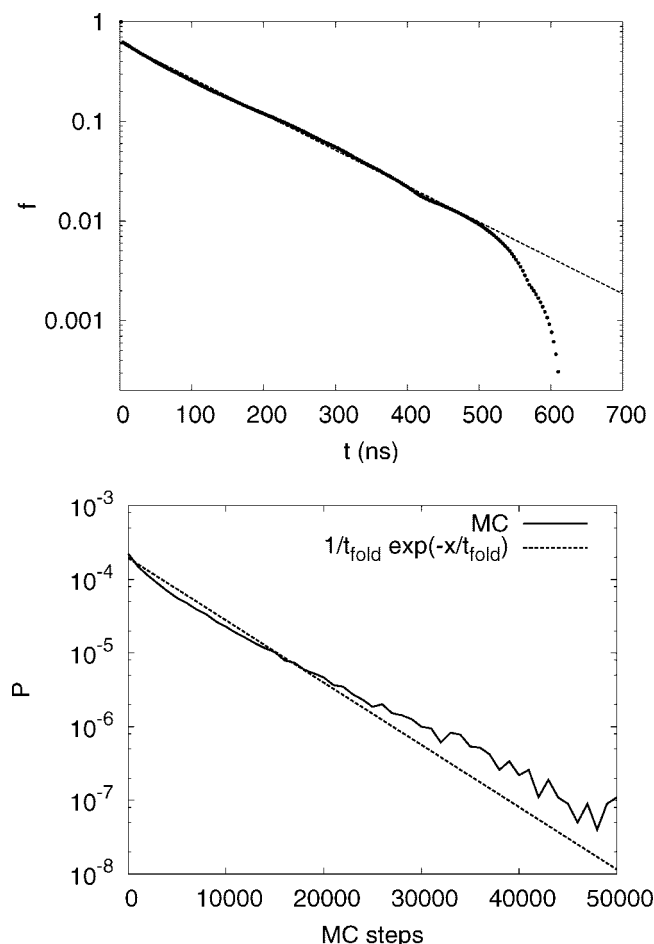


Figure 10. Single-exponential behavior of folding. (Top) Points show the cumulative distribution of the first passage time to the folded state $f(t) = \int_0^t p(\tau) d\tau$, where p is the probability distribution of the first passage time. All saved snapshots were used to calculate $f(t)$. The dashed line is the exponential fit $0.624e^{(-t/120)}$. (Bottom) Distribution of first passage times obtained from MC simulations on the directed network of 2.5-Å rmsd coarse-graining (solid line). The single-exponential approximation is also shown (dashed) with $t_{\text{fold}} = 5140$ MC steps, which corresponds to a folding time of 102.8 ns (5140×0.02 ns). The 2.5-Å rmsd clustering was used here because the corresponding distribution calculated with the secondary structure coarse-graining is affected by pseudotunneling (see text and Figure S3 in the Supporting Information).

stable enthalpic traps, each with a relatively low population and a total weight of about 20%, plus a helical region populated at about 11%.

The FEPs reveal that the barriers to exit the enthalpic traps have similar heights, and the SEKN shows that the times spent in the entropic state before reaching or after leaving the folded state are comparable. These results explain why the folding times from individual basins differ by no more than a factor of 3.¹⁹ Thus, in accord with a recent experimental analysis of a photoswitchable helical peptide,⁵⁴ the single-exponential folding behavior originates from essentially equal folding times for multiple paths. This provides another scenario, different from that of the β -hairpin of protein G,²² by which the complexity of the folding reaction can be hidden from standard kinetic experiments.

Acknowledgment. We thank F. Rao, E. Guarnera, and E. Paci for contributions to the initial stages of this work. We thank M. Seiber for help with the program WORDOM. The simulations were performed on the Matterhorn cluster of the University

of Zurich. This work was supported by a Swiss National Science Foundation grant to A.C., and the portion done at Harvard University was supported, in part, by a grant from the National Institutes of Health. S.K. was supported by the CHARMM Development Project. Procedures for calculating the FEPs have been introduced into WORDOM (<http://www.biochem-caflisch.unizh.ch/wordom>), and the program for plotting the TRDG (Figure 9) is available upon request.

Supporting Information Available: Additional calculations, figures, and tables. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) Karplus, M. *J. Phys. Chem. B* **2000**, *104*, 11–27.
- (2) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991**, *254*, 1598–1603.
- (3) Privalov, P. L.; Makhatadze, G. I. *J. Mol. Biol.* **1993**, *232*, 660–679.
- (4) Maxwell, K. L.; Wildes, D.; Zarrine-Afsar, A.; De Los Rios, M. A.; Brown, A. G.; Friel, C. T.; Hedberg, L.; Hornig, D.; Bona, J.-C.; Miller, E. J.; Vallée-Bélisle, A.; Main, E. R. G.; Bemporad, F.; Qiu, L.; Teilum, K.; Vu, N.-D.; Edwards, A. M.; Ruczinski, I.; Poulsen, F. M.; Kragelund, B. B.; Michnick, S. W.; Chiti, F.; Bai, Y.; Hagen, S. J.; Serrano, L.; Oliveberg, M.; Raleigh, D. P.; Wittung-Stafshede, P.; Radford, S. E.; Jackson, S. E.; Sosnick, T. R.; Marqusee, S.; Davidson, A. R.; Plaxco, K. W. *Protein Sci.* **2005**, *14*, 602–616.
- (5) Jackson, S. E. *Fold. Des.* **1998**, *3*, R81–91.
- (6) Dobson, C. M.; Šali, A.; Karplus, M. *Angew. Chem., Int. Ed.* **1998**, *37*, 869–893.
- (7) Dinner, A. R.; Šali, A.; Smith, L. J.; Dobson, C. M.; Karplus, M. *Trends Biochem. Sci.* **2000**, *25*, 331–339.
- (8) Mirny, L. A.; Shakhnovich, E. I. *Ann. Rev. Biophys. Biomolec. Struc.* **2001**, *30*, 361–396.
- (9) Daggett, V.; Fersht, A. R. *Nature Rev. Mol. Cell Biol.* **2003**, *4*, 497–502.
- (10) Wolynes, P. G. *Phil. Trans. R. Soc. A* **2005**, *363*, 453–467.
- (11) Chan, H. S.; Dill, K. A. *Proteins: Structure, Function, and Bioinformatics* **1998**, *30*, 2–33.
- (12) Best, R.; Hummer, G. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6732–6737.
- (13) Palyanov, A. Y.; Krivov, S. V.; Karplus, M.; Chekmarev, S. F. *J. Phys. Chem. B* **2007**, *111*, 2675–2687.
- (14) Go, N.; Abe, H. *Biopolymers* **1981**, *20*, 991–1011.
- (15) Schueler-Furman, O.; Wang, Ch.; Bradley, P.; Misura, K.; Baker, D. *Science* **2005**, *310*, 638–642.
- (16) Kussell, E.; Shimada, J.; Shakhnovich, E. I. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 5343–5348.
- (17) Paci, E.; Vendruscolo, M.; Karplus, M. *Proteins: Structure, Function, and Bioinformatics* **2002**, *47*, 379–392.
- (18) Caflisch, A. *Curr. Opin. Struct. Biol.* **2006**, *16*, 71–78.
- (19) Muff, S.; Caflisch, A. *Proteins: Structure, Function, and Bioinformatics* **2008**, *70*, 1185–1195.
- (20) Becker, O. M.; Karplus, M. *J. Chem. Phys.* **1997**, *106*, 1495–1517.
- (21) Wales, D. J. *Energy Landscapes*, Cambridge Univ. Press, Cambridge, U.K., 2003.
- (22) Krivov, S. V.; Karplus, M. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 14766–14770.
- (23) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196–199.
- (24) Krivov, S. V.; Karplus, M. *J. Phys. Chem. B* **2006**, *110*, 12689–12698.
- (25) Du, R.; Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Shakhnovich, E. I. *J. Chem. Phys.* **1998**, *108*, 334–350.
- (26) Ferrara, P.; Caflisch, A. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10780–10785.
- (27) De Alba, E.; Santoro, J.; Rico, M.; Jiménez, M. A. *Protein Sci.* **1999**, *8*, 854–865.
- (28) Rao, F.; Caflisch, A. *J. Chem. Phys.* **2003**, *119*, 4035–4042.
- (29) Cavalli, A.; Ferrara, P.; Caflisch, A. *Proteins: Structure, Function, and Bioinformatics* **2002**, *47*, 305–314.
- (30) Serrano, L.; Matouschek, A.; Fersht, A. R. *J. Mol. Biol.* **1992**, *224*, 805–818.
- (31) Settanni, G.; Rao, F.; Caflisch, A. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 628–633.
- (32) Rao, F.; Caflisch, A. *J. Mol. Biol.* **2004**, *342*, 299–306.
- (33) Hubner, I. A.; Shimada, J.; Shakhnovich, E. I. *J. Mol. Biol.* **2004**, *336*, 745–761.

- (34) Rao, F.; Settanni, G.; Guarnera, E.; Caffisch, A. *J. Chem. Phys.* **2005**, *122*, 184901.
- (35) Snow, C. D.; M Rhee, Y.; Pande, V. S. *Biophys. J.* **2006**, *91*, 14–24.
- (36) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (37) Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caffisch, A. *Bioinformatics* **2007**, *23*, 2625–2627.
- (38) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902–1921.
- (39) Ferrara, P.; Apostolakis, J.; Caffisch, A. *Proteins: Structure, Function, and Bioinformatics* **2002**, *46*, 24–33.
- (40) Ferrara, P.; Apostolakis, J.; Caffisch, A. *J. Phys. Chem. B* **2000**, *104*, 5000–5010.
- (41) Eaton, W. A.; Munoz, V.; Hagen, J.; Jas, S. G. S.; Lapidus, L. J.; Henry, E. R.; Hofrichter, J. *Ann. Rev. Biophys. Biomolec. Struct.* **2000**, *29*, 327–359.
- (42) Hubner, I. A.; Deeds, E. J.; Shakhnovich, E. I. *Proc. Natl. Acad. Sci. USA.* **2006**, *103*, 17747–17752.
- (43) Andersen, C. A. F.; Palmer, A. G.; Brunak, S.; Rost, B. *Structure* **2002**, *10*, 174–184.
- (44) Krivov, S. V.; Karplus, M. *J. Chem. Phys.* **2002**, *117*, 10894–10903.
- (45) Park, S.; Sener, M. K.; Lu, D.; Schulten, K. *J. Phys. Chem.* **2003**, *119*, 1313–1319.
- (46) Apaydin, M.; Brutlag, D.; Guesttin, C.; Hsu, D.; Latombe, J. *In International Conference on Computational Molecular Biology (RECOMB)* **2002**.
- (47) Ford, L. R.; Fulkerson, D. R. *Canadian J. of Math.* **1956**, *8*, 399–404.
- (48) Gomory, R. E.; Hu, T. C. *SIAM J. Applied Math.* **1961**, *9*, 551–570.
- (49) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins: Structure, Function, and Bioinformatics* **1995**, *21*, 167–195.
- (50) Garcia-Mira, M. M.; Sadqi, M.; Fischer, N.; Sanchez-Ruiz, J. M.; Munoz, V. *Science* **2002**, *298*, 2191–2195.
- (51) Eaton, W. A.; Munoz, V.; Thompson, P. A.; Chan, C. K.; Hofrichter, J. *Curr. Opin. Struct. Biol.* **1997**, *7*, 10–14.
- (52) Karplus, M.; Weaver, D. L. *Biopolymers* **1979**, *18*, 1421–1437.
- (53) Pappu, R. V.; Weaver, D. D. *Protein Sci.* **1998**, *7* (2), 480–490.
- (54) Ihalainen, J. A.; Bredenbeck, J.; Pfister, R.; Helbing, J.; Woolley, G. A.; Hamm, P. *Proc. Natl. Acad. Sci. USA.* **2007**, *104*, 5383–5388.
- (55) Matouschek, A.; Kellis, J. T., Jr.; Serrano, L.; Fersht, A. R. *Nature* **1989**, *340*, 122–126.
- (56) Vendruscolo, M.; Paci, E.; Dobson, C. M.; Karplus, M. *Nature* **2001**, *409*, 641–645.
- (57) Radford, S. E.; Dobson, C. M.; Evans, P. A. *Nature* **1992**, *358*, 302–307.
- (58) Fersht, A. R.; Daggett, V. *Cell* **2002**, *108*, 573–582.
- (59) Day, R.; Daggett, V. *Proc. Natl. Acad. Sci. USA.* **2005**, *102*, 13445–13450.
- (60) Li, J.; Shinjo, M.; Matsumura, Y.; Morita, M.; Baker, D.; Ikeguchi, M.; Kihara, H. *Biochemistry* **2007**, *46*, 5072–5082.
- (61) Li, J.; Matsumura, Y.; Shinjo, M.; Kojima, M.; Kihara, H. *J. Mol. Biol.* **2007**, *747*–755.
- (62) Chiti, F.; Stefani, M.; Taddei, N.; Ramponi, G.; Dobson, C. M. *Nature* **2003**, *424*, 805–808.
- (63) Tartaglia, G. G.; Cavalli, A.; Pellarin, R.; Caffisch, A. *Protein Sci.* **2004**, *13*, 1939–1941.
- (64) Tartaglia, G. G.; Cavalli, A.; Pellarin, R.; Caffisch, A. *Protein Sci.* **2005**, *14*, 2723–2734.

JP711864R