THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

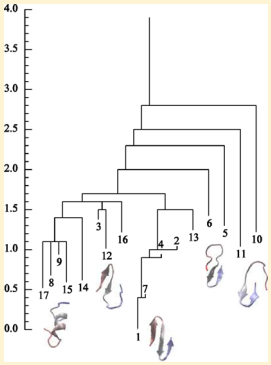
Equilibration of Protein States: A Time Dependent Free-Energy Disconnectivity Graph

Sergei F. Chekmarev*[†,‡]

[†]Institute of Thermophysics, SB RAS, 630090 Novosibirsk, Russia

[‡]Department of Physics, Novosibirsk State University, 630090 Novosibirsk, Russia

ABSTRACT: The process of equilibration of protein states in a three-stranded antiparallel β-sheet miniprotein is studied using a time-dependent free energy disconnectivity graph. To determine the rates of transitions, the molecular dynamics simulation results of a recent work (Kalgin, I. V.; et al. *J. Phys. Chem. B* **2013**, *117*, 6092) are employed. The vertices of the graph are the free energies of characteristic states of the protein, and the edges are the transition state free energies. To determine the latter, the "complete" partition function (Eyring, 1935) is used, which includes the translational partition function corresponding to the ballistic motion of the system along the reaction coordinate. The distance along the reaction coordinate that enters the translational partition function is taken to be proportional to the observation time and thus measures the number of representative points that cross the transition state surface during given time. As the time increases, the free energy barriers between the clusters of characteristic conformations (native-like, helical, and β-sheet conformations of different degree of organization) decrease and (local) equilibrium between the clusters is established. With time, these clusters are grouped into larger clusters, extending the equilibrium to a larger portion of protein states.



1. INTRODUCTION

In the course of folding, a protein passes through a number of conformations whose organization hierarchically increases on approaching the native state. On the basis of the geometric similarity of conformations,[1−3] their kinetic connectivity,[4,5]or the closeness of the representative points of the protein in a space of collective variables,[6] the conformations can be grouped into clusters of characteristic conformations. These clusters can then be considered as elements of a kinetic network, for which the folding kinetics can be described as a Markov process of transitions between the clusters.[7−10] Since the rates of the intercluster transitions are typically highly heterogeneous, detailed balance between some clusters is established much faster than between other clusters so that the probability fluxes through the corresponding segments of the network become negligible. Accordingly, the pathways that contain such segments become insignificant.[11,12] Therefore, it is important to know time scales at which different pairs of clusters come to equilibrium due to detailed balance, not only for insight into folding kinetics but also for understanding the possibility of application of different experimental techniques, whose resolution time scales vary from nanoseconds to seconds.[13]

Proteins, similar to other many-body systems such as atomic and molecular clusters,[14] have a variety of stable conformations ("inherent structures"[15]); their number increases approximately exponentially with the system size.[16] Accordingly, the energy surfaces of such systems are extremely complex. They are multidimensional, consist of a large number of catchment basins, whose minima represent the inherent structures, and correspondingly, have a large number of the transition states that connect the basins. Nevertheless, a rational characterization of such complex energy surfaces is possible; Berry and Breitengraser-Kunz[17] suggested using for this purpose one-dimensional sequences of minimum-saddle-minimum triples, and Becker and Karplus[18] proposed constructing a two-dimensional tree of the minima that are connected by saddles ("disconnectivity graph").

Initially, the disconnectivity graphs were used to visualize potential energy surfaces (PEDGs); they were applied to oligopeptides[14,18,19] and clusters.[14] One shortcoming of the PEDGs is that they do not take into account entropy effects and thus can give only limited information about kinetics. This shortcoming is remedied in the free energy disconnectivity graphs (FEDGs), in which instead of the energies of minima and saddles in the PEDGs, the free energies of the corresponding catchment basins and transition states are considered; such graphs were constructed for oligopetides[20,21] and a simplified model protein.[21] In these works, the FEDG represented the equilibrium state of the system. Concurrently, it was suggested to consider nonequilibrium, time-dependent free energy disconnectivity graphs (TD-FEDGs) (Gavrilov and Chekmarev[22]), where the heights of the free energy barriers depend on the observation time. The barriers lower with time, and different structures come to equilibrium, forming clusters of the structures; the approach was illustrated for alanine tetrapeptide.[22] A somewhat similar approach was recently proposed by Wales and Salamon,[23] where the structures were classified and clustered according to the rates of transitions

Received: May 6, 2015
Revised: June 10, 2015
Published: June 11, 2015

DOI: 10.1021/acs.jpcb.5b04336
J. Phys. Chem. B 2015, 119, 8340−8348

between the structures. A related problem is the glassy transition, in which the partition function of the system can be considered as a function of the observation time.[24]

In the present paper, we use the TD-FEDGs to examine the equilibration of protein states in a three-stranded antiparallel miniprotein (beta3s), whose equilibrium folding has been comprehensively studied by Caflisch and coauthors.[6,25−31] The protein has five characteristic conformations; they are native-like conformations, Cs-or and Ns-or conformations that have, respectively, the C-terminal and N-terminal hairpins unstructured, conformations that contain helical regions, and Ch-curl conformations that present curl-like structures with the C-terminal hairpin formed. We show that the equilibrium is first established among helical, native-like, and Cs-or conformations; later Ns-or conformations join them, and finally Ch-curl conformations come into equilibrium with the others.

The paper is organized as follows. Section 2 gives the definition of the time-dependent free energies of transitions states. Section 3 describes the system and its characterization, and section 4 contains the results and their discussion. Section 4 gives a brief summary of the results and concluding remarks.

2. DEFINITION OF FREE ENERGIES

A potential energy disconnectivity graph (PEDG) characterizes the structure of the potential energy surface (PES) in terms of the catchment basin minima and saddles that connect the basins. The minima are the vertices of the graph, and the saddle points determine the edges. These elements of the PEDG, i.e., the minima and saddles, are consistent in that each of them represents a point on the PES. A free energy disconnectivity graph (FEDG) determines the probability for the system to be in a particular basin and the probability to leave this basin for another one. In this case, the vertices of the graph are the basin free energies, and the edges are determined by the transition state free energies. The basin free energy (of basin i) is determined as

$$F_i(T) = -k_B T \ln Z_i(T) \tag{1}$$

where k_B is the Boltzmann constant, T is the temperature, and $Z_i(T)$ is the partition function of basin i. The definition of the transition state free energy is somewhat flexible and worth consideration. According to the transition state theory (TST),[33,34] the rate constant of transitions from basin i to j is

$$k_{ji}(T) = \frac{k_B T}{h}\frac{Z_{ji}^{\ddagger}(T)}{Z_i(T)} \tag{2}$$

where h is the Planck constant, and $Z_{ji}^{\ddagger}(T)$ is the partition function of the corresponding transition state. In the free energy terms, eq 2 acquires the form

$$k_{ji}(T) = \frac{k_B T}{h}\frac{e^{-F_{ji}^{\ddagger}(T)/(k_B T)}}{e^{-F_i(T)/(k_B T)}}$$

where $F_{ji}^{\ddagger}(T) = -k_B T \ln Z_{ji}^{\ddagger}(T)$ is the free energy of the transition state. The factor $k_B T/h$ is the characteristic vibrational frequency

$$\nu_{vib} = k_B T/h \tag{3}$$

At 300 K, it is equal to 0.625×10^{13} 1/s.

Typically, to construct the FEDGs, the free energies of the basin and transition state are associated with the partition functions Z_i and $Z_{ji}^{\ddagger}$, respectively. Two approaches are employed to determine them. One is to apply the harmonic approximation, using the potential energy function of the system to calculate the normal mode vibrational frequencies (Hessian eigenvalues) that determine the partition functions.[20,21,23,32,35] The other approach is to use the simulation results, counting the number of transitions between the basins (n_{ji}); the partition function of the basin is calculated as $Z_i = \sum_j n_{ij}$, and $Z_{ji}^{\ddagger}$ is estimated using eq 2, in which the rate constant k_{ji} is taken to be proportional to n_{ji}.[36] Then

$$F_{ji}^{\ddagger}(T) = F_i(T) - k_B T \ln[Z_{ji}^{\ddagger}(T)/Z_i(T)] = F_i(T) - k_B T \ln(k_{ji}/\nu_{vib}) \tag{4}$$

The partition function $Z_{ji}^{\ddagger}$ entering eq 2 does not include the mode along the reaction coordinate (the negative Hessian eigenvalue); i.e., it is related to a manifold of phase space whose dimension is one unit less than the dimension of the catchment basin. Therefore, when the free energies of the basins and transition states that are determined by eqs 1 and 4 are used to construct the FEDG, the numbers of microstates in the manifolds of different dimension are, in fact, compared. Although this problem can be easily circumvented, e.g., by a formal extension of the phase space for the transition state by multiplying $Z_{ji}^{\ddagger}$ by h, which does change its value because the phase space in eq 2 is measured in units of h (semiclassical representation), it seems conceptually proper to define the partition function of the transition state as a "complete" partition function in the early TST.[33] There it was written as

$$Z_{ji}(T) = Z_{ji}^{\ddagger}(T) Z_{ji}^{tr}(T) \tag{5}$$

where

$$Z_{ji}^{tr}(T) = (2\pi M k_B T)^{1/2} l/h \tag{6}$$

is the translational partition function for the (ballistic) motion along the reaction coordinate at the transition state (M is the effective mass of the transition state, and l is the characteristic distance along the reaction coordinate). With the partition function $Z_{ji}(T)$, the reaction rate is determined as $k_{ji}(T) = \nu_{ji}(T)\, Z_{ji}(T)/Z_i(T)$, where $\nu_{ji}(T) = k_B T/[(2\pi M k_B T)^{1/2} l]$ is the frequency to pass over the barrier corresponding to the transition state. The multiplication of Z_{ji}^{tr} and ν_{ji} results in the vibrational frequency factor $k_B T/h$ in eq 2.[33]

The definition of the transition state partition function by eqs 5 and 6 requires a specification of the distance l along the reaction coordinate. One possibility is to assume that l is proportional to the observation time, which allows us to see how equilibrium in the system is established (Gavrilov and Chekmarev[22]). This assumption has something in common with the fact that the transition region has a finite width so that the transition from reactant to product takes some time.[37−39] Here we go further and assume that the larger the l, the larger is the number of representative points that cross the transition state surface. Specifically, taking $l = v_{ji}\tau$, where $v_{ji} = [k_B T/(2\pi M)]^{1/2}$ is the average velocity of the ballistic motion through the transition state and τ is the observation time, eqs 5 and 6 give $Z_{ji}(T) = (k_B T/h) Z_{ji}^{\ddagger}(T)\tau$. Then, substituting $Z_{ji}^{\ddagger}(T)$ from this equation into eq 2, we have

$$Z_{ji}(T,\tau) = Z_i(T) k_{ji} \tau$$

or

$$F_{ji}(T, \tau) = F_i(T) - k_B T \ln(k_{ji}\tau) \tag{7}$$

DOI: 10.1021/acs.jpcb.5b04336
J. Phys. Chem. B 2015, 119, 8340−8348

Table 1. Clusters of Protein Conformations (Reproduced with Permission from Kalgin et al.[6])

cluster[a]	W_{clst} [b]	N_{str} [c]	most populated structure[d]	W_{str} [e]	cluster type
1	21.5	523	−EEEETTEEEEEETTEEEE−	38.6	native-like
			−EEEETTEEEEEETTEEE−−	37.0	
2	3.9	939	−EEEETTEEEEEETTEEEE−	16.2	
			−EEEETTEEEEEETTEE−−−	14.1	
3	2.6	2337	−EEEETTEEEEEEEEEE−−	12.3	Cs-or
			−EEEETTEEEEEEEEEEE−	9.8	
4	3.1	1173	−EEEETTEEEEE−SS−EEE−	7.2	Cs-or + native-like
			−EEEETTEEEEE−SS−EE−−	5.6	
5	3.0	773	−EEE−SSS−EEEETTEEEE−	46.1	Ns-or
			−EEEESSSEEEEETTEEEE−	5.5	
6	2.5	631	−EEE−SSS−EEEETTEEEE−	22.3	
			−EEEESSSEEEEETTEEEE−	19.8	
7	5.0	1005	−EEEETTEEEEEETTEEE−−	8.4	Ns-or + native-like
			−EE−−SSS−EEEETTEEEE−	6.6	
8	7.6	48567	−−HHHHHHHHHHT−−−−−−	0.4	helical 1
			−−−HHHHHHHHHT−−−−−−	0.2	
9	5.1	33302	−−SS−−HHHHTTT−−−−−−−	0.3	helical 2
			−−SS−−HHHHHHHSS−−−−−	0.3	
10	3.3	2347	−B−SSSSS−−EEETTEE−B−	5.6	Ch-curl 1
			−B−−SSS−−−EEETTEE−B−	4.5	
11	4.4	5758	−B−SSSSS−EEEETTTEEE−	3.3	Ch-curl 2
			−B−SSSS−−EEEETTTEEE−	3.2	
12	4.6	13206	−EEEETTEEEE−−SS−−−−−	1.5	others
			−EEEETTEEEE−SSS−−−−−	1.3	
13	3.2	3799	−EEEETTEEEEEETTEEEE−	7.1	
			−−−−BTTEEEEEETTEEEE−	3.0	
14	8.4	15590	−−−−−SS−−EEEETTEEEE−	1.5	
			−−−−SSS−−EEEETTEEEE−	1.3	
15	8.7	47727	−EE−SSS−EE−−−SS−−−B−	0.7	
			−EEE−SSS−EEEEEEEEE−−	0.4	
16	3.4	17009	−EEEETTEEE−−−SS−−−−−	0.6	
			−B−−−SSS−−−−−SSS−−B−	0.5	
17	9.7	63733	−EEETTTEEEETTTEEEE−−	0.3	
			−−−−SSS−−−−−SSS−−−−−	0.2	

[a]Cluster number. [b]Cluster weight equal to the number of representative points in the cluster relative to the total number of points (in %). [c]The number of conformations that have different secondary structure strings. [d]The secondary structure strings of the most populated conformations. [e]Weight of the given conformation in the cluster (in %).

We note that although the distance l thus determined can become unphysically large as time increases (larger than ∼10 Å, i.e., the characteristic distance to separate the reactant and product[37−39]), it does not enter the resulting eq 7; just the product of k_{ji} and τ is kept, i.e., the number of transitions from basin i to j during time τ. Comparison of eq 7 with eq 4 also shows that according to the ν_{vib} value [eq 3] the transition state free energy determined by eq 4 corresponds to the observation time $1/\nu_{vib} \approx 10^{-13}$ s.

Equation 7 suggests that the free energy barrier between the basins decreases with time, making the interbasin transitions easier. At $\tau = 1/k_{ji}$, i.e., as the time becomes equal to the mean time of passage from basin i to j, the barrier disappears, which can be interpreted as the attainment of (local) equilibrium between the structures characteristic of these basins. Correspondingly, the structures that have come into equilibrium can be grouped into clusters, which makes the FEDG time-dependent. As has been mentioned in Introduction, a somewhat similar approach was recently proposed by Wales and Salamon.[23] More specifically, with small Lennard-Jones clusters and alanine dipeptide taken as an example, they considered how the structures for these systems are clustered as the rates of transitions between them become smaller. Equating the observation time with the reciprocal of the rate constants of the transitions between the structures, the authors interpreted the FEDGs thus constructed as the graphs at different observation time scales. The present approach, in contrast, not only indicates the moment when equilibrium between the structures is established but also shows a degree of equilibration between the structures at a given time.

3. THE SYSTEM AND ITS CHARACTERIZATION

The designed three-stranded antiparallel 20-residue peptide, called beta3s (Thr1-Trp2-Ile3-Gln4-Asn5-Gly6-Ser7-Thr8-Lys9-Trp10-Tyr11-Gln12-Asn13-Gly14-Ser15-Thr16-Lys17-Ile18-Tyr19-Thr20 with charged termini[40]), is one of the few systems for which the folding reaction under equilibrium conditions has been studied in detail with an all-atom representation.[6,26−32] This study, similar to the previous ones, is based on a set of trajectories generated in the Caflisch group[27] of total length of 20 μs, during which the protein experienced on the order of 100 folding/unfolding events. The simulations were performed using the CHARMM program.[41] All heavy atoms and the hydrogen atoms bound to nitrogen or

DOI: 10.1021/acs.jpcb.5b04336
J. Phys. Chem. B 2015, 119, 8340−8348

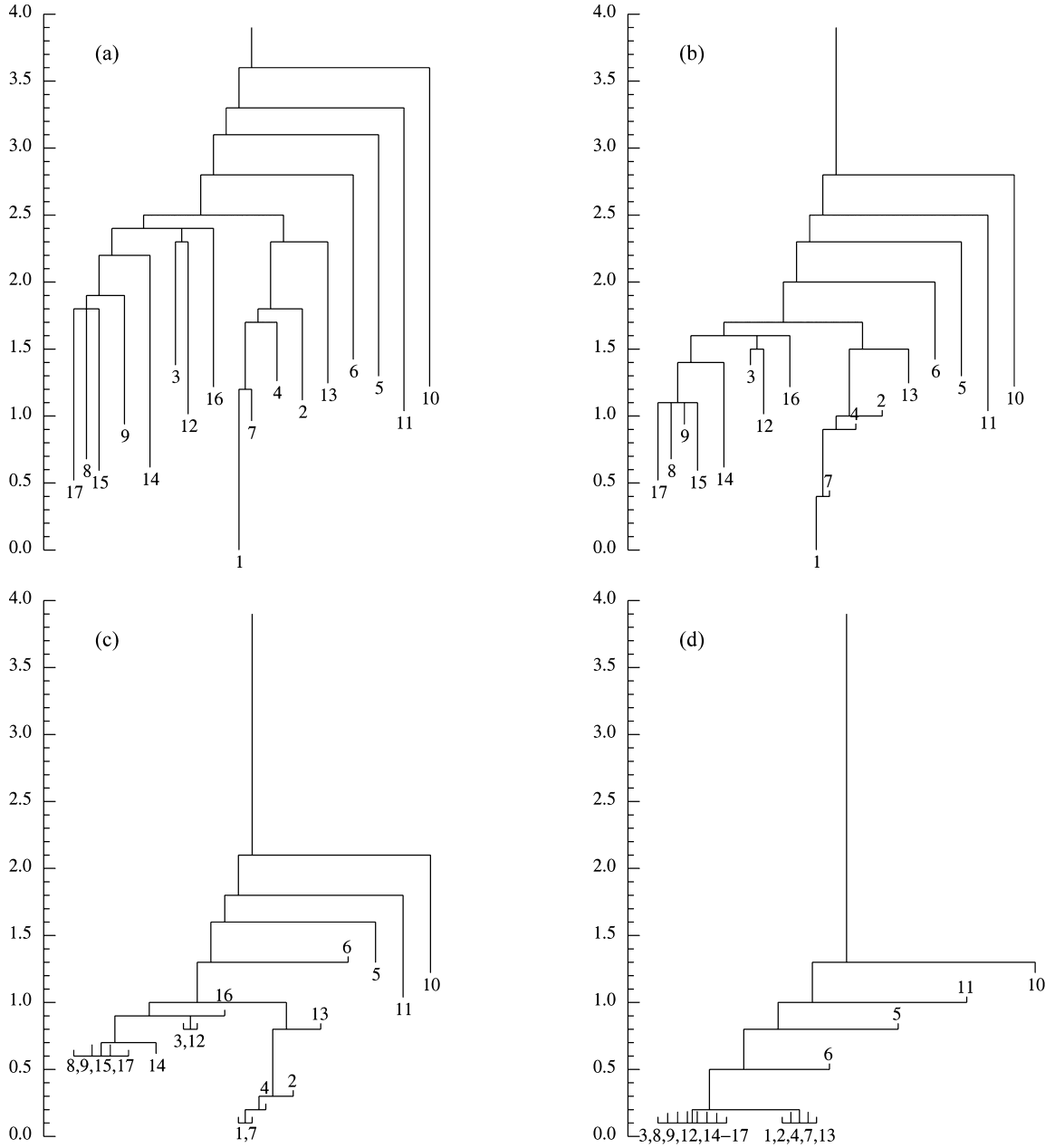


Figure 1. Time-dependent free energy disconnectivity graphs: (a) τ = 0.03 ns, (b) τ = 0.1 ns, (c) τ = 0.3 ns, and (d) τ = 1 ns. The clusters of conformations (vertices of the graph) are numbered according to Table 1. The energy is counted from the energy of the native state (cluster 1) and measured in kcal/mol.

oxygen atoms were considered explicitly; PARAM19 force field[42] and a default cutoff of 7.5 Å for the nonbonding interactions were used. A mean-field approximation based on the solvent-accessible surface (SAS) was employed to describe the main effects of the aqueous solvent.[43] Although the implicit solvent model leads to rates faster than the experimental values, 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,[44] β-hairpins in about 10 ns,[44] and three-stranded β-sheets in about 100 ns[26] compared to experimental values of ∼0.1,[46] ∼ 1,[46] and ∼10 μs,[40] respectively. The simulations were performed with the time step of 2 fs using the Berendsen thermostat (coupling constant of 5 ps) at T = 330 K. For the present protein model, this temperature is slightly above the melting temperature.[47] The atomic coordinates ("frames") were saved every 20 ps, which resulted in 10^6 snapshots.

To construct the TD-FEDG, it was found reasonable to use the rates of transitions between clusters of characteristic protein conformations rather than the rates of transitions between the conformations themselves, the number of which is very large (see Table 1). For this, the results of the previous work[6] were employed. In that work, the conformation space of beta3s in the form of the hydrogen bond distances for the formed bonds was transformed to a space of orthogonal collective variables using a principal component analysis ("hydrogen bond PCA"), and the first three modes corresponding to the largest eigenvalues were chosen to form a three-dimensional (3D) space of collective variables (g_1, g_2, g_3). The representative points of protein states in the (g_1, g_2, g_3) space were then

DOI: 10.1021/acs.jpcb.5b04336
J. Phys. Chem. B 2015, 119, 8340−8348

grouped into clusters using the MCLUST method.[48] Specifically, the collection of representative points was approximated by a set of 3D Gaussian functions with generally different covariance matrices and different weights so that each function represented a cluster of the points; the optimal number of clusters and the distribution of the points among them were determined using a maximum-likelihood estimation. Although a projection of simulation data onto a low-dimensional space can mask the complexity of the unprojected surface due to local smoothing of the folding landscape[8,36,49] (see also a recent illustration for atomic clusters[50]), the kinetic network based on the rates of intercluster transitions was found in good agreement with the "hydrodynamic" picture of the folding process[6] as well as with the results of the previous studies of beta3s folding.[26−32] Specifically, the rate of transitions from cluster i to cluster j was calculated as $k_{ji} = (N_{ji}/t_{tot})/(N_i/N_{tot})$, where N_{ji} is the number of transitions from cluster i to j (Table S7 of the Supporting Information in ref 6), t_{tot} is the total simulation time equal to 20 μs, $N_i = \sum N_{ij}$, and $N_{tot} = \sum N_i = 10^6$ is the total number of stored representative points. The ratio N_i/N_{tot} represents the weight of i cluster among the 10^6 conformations, which is given in percentage in Table 1. The resulting transition rates ranged from 1×10^{-7} 1/ps to 1×10^{-2} 1/ps,[6] which indicates that no essential transitions between the clusters were missed when the representative points (frames) were saved every 20 ps. To associate the clusters with characteristic structures, the protein conformations were discriminated according to the secondary structure strings (SSSs) encoded with the DSSP alphabet,[51] in particular, the letters H, E, B, T, S, and "−" standing for α-helix, extended, isolated β-bridge, hydrogen bonded turn, bend, and unstructured segments, respectively. With this coding, the native state is represented by the string "−EEEETTEEEEEET-TEEEE−". The other significant conformations have the following, or close to them, strings: Cs-or (−EEEETTEEEE−SSS−−), Ns-or (−EEEESSSEEEEETTEEEE−), Ch-curl (−B−SSSSS−EEEETTTEEE−), and helical (−HHHHHH-HHHHS−−). The Cs-or and Ns-or conformations differ from the native conformation in that one of the hairpins is formed and the other is unstructured ("out of register"[25]), the Ch-curl conformations have a curl-like structure with the C-terminal hairpin formed, and the conformations are called helical if they contain a well-expressed helical region. The program WORDOM[52] was used to perform the analysis.

4. RESULTS AND DISCUSSION

Figure 1 presents the TD-FEDGs for different values of the observation time τ. The vertices of the graphs are numbered according to Table 1, which relates the cluster number to the conformations characteristic of the given cluster. The relative spatial location of the clusters, which is hidden in the graphs, is shown in Figure 2; it presents a three-dimensional kinetic network for beta3s in the (g_1, g_2, g_3) space of collective variables.[6] The variable g_1 serves as a good reaction coordinate for the overall description of the process, the variable g_2 discriminates between Cs-or and Ns-or conformations, and the variable g_3 accounts for the other characteristic conformations, including Cs-or and helical conformations.

To group the clusters, a simple "kinetic connectivity" approach was used, although other approaches are also possible, such as the mincut procedure of Krivov and Karplus[36] and the regrouping scheme of Carr and Wales.[32] The present approach is somewhat similar to that of Bowman et al.,[53] which was used

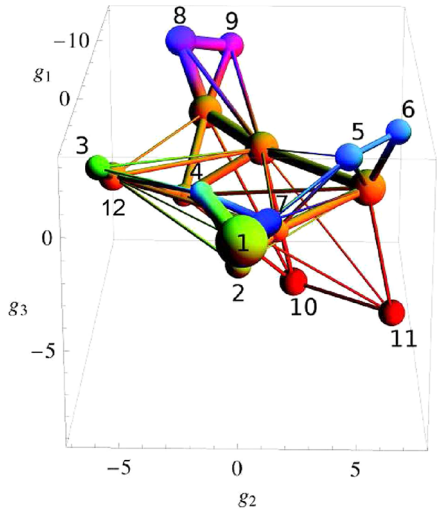


Figure 2. Spatial kinetic network in the (g_1, g_2, g_3) space of collective variables (reproduced with permission from Kalgin et al.[6]). Balls represent the clusters of characteristic conformations, which are numbered according to Table 1, and tubes represent the transitions between the clusters. Ball volumes are proportional to the number of intracluster transitions (i.e., the residence times in the clusters), and the tube cross sections are proportional to the number of intercluster transitions (the latter transitions are calculated as one-half of the total number of the forward and backward transitions between the two clusters, which were found to be very similar due to detailed balance). The g_1, g_2, and g_3 variables are in angstroms.

for automating construction of the Markov state models. Specifically, a set of the free energy thresholds F_{thr} with a certain discretization (in the present case of 0.1 kcal/mol) is introduced. Starting with the lowest threshold, two clusters (i and j) are selected for which $F_{ji}(T,\tau) \le F_{thr}$; they originate a new group. This group is then augmented according to the rule that any cluster (q) for which the condition $F_{pq}(T,\tau) \le F_{thr}$ is satisfied at least for one (p) cluster from this group is added to the group. This procedure is continued until all clusters are distributed among the groups for each threshold, and all thresholds, starting from the lowest one, are passed through. To display the connectivity of the basins in the graph, the following method is used. If $F_i(T) < F_{thr}$ and $F_j(T) < F_{thr}$, basins i and j are represented by their vertices at $F_i(T)$ and $F_j(T)$ and are connected at the level of the threshold. Otherwise, if $F_i(T) < F_{thr}$ and $F_j(T) \ge F_{thr}$, basin i is represented by a vertex at $F_i(T)$, as previously, but basin j is indicated by a short stick directed upward. In this case, the transition from basin i to j is characterized by the barrier $F_{thr} - F_i(T)$, while the opposite transition, from basin j to i, is barrierless. Finally, if $F_i(T) \ge F_{thr}$ and $F_j(T) \ge F_{thr}$, both basins i and j are indicated by short sticks directed upward, which means that the free energy barrier between the basins is absent. Correspondingly, the protein conformations in these basins are expected to be in equilibrium.

Figure 1a presents the TD-FEDG at a small observation time scale (τ = 0.03 ns), during which the probability of observing the interbasin transitions is low; accordingly, the basins are separated by large free energy barriers. Nevertheless, some ordering of the basins is present. The most notable are two groups of basins. One represents a consolidated cluster (C-cluster), which includes the native-like conformations (basins 1 and 2) and the clusters that contain mixtures of native-like and Cs-or and Ns-or conformations (basins 4 and 7, respectively); this cluster will be referred to as native C-cluster. The presence

DOI: 10.1021/acs.jpcb.5b04336
J. Phys. Chem. B 2015, 119, 8340−8348

of clusters 4 and 7 in this cluster is in accord with the conclusion of the previous work[6] that to be in agreement with the previous studies for the weights of clusters of different conformations (kinetic grouping analysis,[27,28] pfold analysis based on an equilibrium kinetic network,[28] and replica exchange molecular dynamics and constant temperature molecular dynamics[29]), these clusters should be related to the native-like conformations rather than to the Cs-or and Ns-or ones. It is also seen that cluster 13, which is determined as a cluster of unstructured conformations (Table 1), can be associated with the native C-cluster. Cluster 13 is connected to the native C-cluster better than to the others, although it is separated by a relatively large barrier, as compared to the barriers between clusters 4 and 7 and clusters 1 and 2. Both the connectivity of cluster 13 to the native C-cluster and a relatively large barrier from the native-like states can be explained by the fact that native-like conformations are present in cluster 13 but constitute a small fraction (~7%, Table 1). The other group of basins, representing a helix C-cluster, includes both the conformations that contain pronounced helical regions (basins 8 and 9) and clusters 14, 15, and 17, in which the conformations are mostly unstructured. In contrast to cluster 13, which had a noticeable fraction of (native-like) conformations that are common to this cluster and the native C-cluster, the fraction of helical conformations in clusters 14, 15, and 17 is negligible. Therefore, the occurrence of these clusters in the helix C-cluster has a different nature; it is due to the fact that the helical conformations are readily formed from unstructured ones, as it is observed in the initial stage of folding of beta3s.[11,12]

As the observation time increases, the barriers between the basins decrease, and equilibrium in the native and helix C-clusters is established (Figure 1b to Figure 1d). These clusters come to equilibrium with each other at $\tau \approx 2$ ns (results not shown). It is worthy to note that at given times, the equilibrium between different clusters is mostly local. One example, for clusters 1 and 7, is shown in Figure 3. The residence probabilities $p_i(t)$ were calculated by solving the master equation

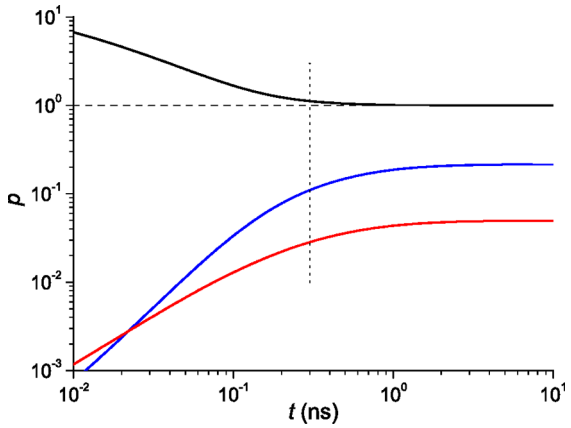


Figure 3. Residence probability distributions vs time. The blue and red lines are, respectively, for clusters 1 and 7, and the black line is for the ratio of the normalized probabilities (see the text). The dotted line marks the time $t = 0.3$ ns, at which the equilibrium between clusters 1 and 7 is indicated by the TD-FEDG (Figure 1c).

$$\frac{dp_i}{dt} = \sum_j [k_{ij}p_j(t) - k_{ji}p_i(t)]$$

with the initial conditions $p_{14}(0) = 1$ and $p_i(0) = 0$ for the other clusters; i.e., it was assumed that the process of population of the protein states started in the basin of unstructured conformations (basin 14), which are kinetically close to the helical conformations (basins 8 and 9) that are readily formed in the initial stage of folding. Along with the residence probabilities $p_1(t)$ (blue line) and $p_7(t)$ (red line), there is shown the ratio of the normalized probabilities $[(p_7(t)/p_{eq,7})]/[p_1(t)/p_{eq,1}]$ (black line), where $p_{eq,i}$ is the equilibrium probability ($i = 1, 7$). It can be seen that at $t > 0.3$ ns, i.e., as the equilibrium between clusters 1 and 7 is indicated by the TD-FEDG (Figure 1c), the ratio of the probabilities changes insignificantly (within ~10%), while $p_1(t)$ and $p_7(t)$ themselves vary as large as 2 times.

The Cs-or and Ns-or conformations, although they are structurally similar (one hairpin is formed and the other is not), come to equilibrium with the other conformations at quite different times. For the Cs-or conformations, which are represented by cluster 3, the process of equilibration consists of three stages: first, cluster 12, which has an appreciable fraction of Cs-or-like conformations (Table 1), is joined to cluster 3 ($\tau \approx 0.3$ ns, Figure 1c); then, this consolidated cluster is joined to the helix C-cluster ($\tau \approx 1$ ns, Figure 1d); and finally, as a part of the helix C-cluster, it comes to equilibrium with the native C-cluster ($\tau \approx 2$ ns). That is, although the native C-cluster contains the cluster of mixture of native-like and Cs-or conformations (cluster 4), the Cs-or conformations are mostly connected to the native C-cluster not directly but through the unstructured conformations that are in equilibrium with the helical conformations. In contrast, the Ns-or conformations, which are represented by two clusters, 5 and 6, come to equilibrium with the conformations in the native C-cluster directly; this requires a bit longer time ($\tau \approx 3$ ns) though the Ns-or conformations are formed first in the process of beta3s folding.[25,32,45] Finally, at $\tau \approx 5$ ns (results not shown), the Ch-curl conformations, which are less stable than the antiparallel β-strands because of the distortion of the hydrogen bonds[54] and more difficult to form dynamically because of the distant N- and C-terminal strands, come to equilibrium with other conformations, which completes the equilibration of the protein states. The fast equilibration among the unstructured, helical, and Cs-or conformations may be a reason why the last two are not indicated in the equilibrium disconnectivity graph of ref 32; i.e., it may happen that they are "embedded" in the pool of unstructured conformations. At the same time, the absence of the Ch-curl conformations in that graph, which are present, for example, in the disconnectivity graph of ref 28, remains unclear.

The present results shed more light on the Cs-or and Ns-or folding pathways of beta3s. Since the native C-cluster contains a number of native-like conformations (Table 1), the probability that the protein immediately attains the native state, when it comes to the native C-cluster, is low. As Figure 3 indicates, the equilibration times within the native C-cluster are finite, i.e., not negligibly small. Therefore, when visiting the native C-cluster, the protein can unfold to the conformationally close Cs-or and Ns-or states earlier than it achieves the native state. In particular, according to the previous work[6] (Table S7 of the Supporting Information of ref 6), the number of transitions

DOI: 10.1021/acs.jpcb.5b04336
J. Phys. Chem. B 2015, 119, 8340−8348

from the Cs-or and Ns-or clusters to the native C-cluster is approximately 20 times larger than the number of times the native state was achieved; i.e., the protein mostly unfolds to the Cs-or and Ns-or conformations when it comes from them to the native C-cluster. Because of this, detailed balance is established between the Cs-or and Ns-or clusters and the native C-cluster,[11] making the probability flows through the Cs-or and Ns-or clusters negligible. This is in agreement with the earlier finding that the folding pathways connect the native state with an entropic basin (unstructured conformations) mostly directly.[28] Later studies of the first-passage folding of beta3s, both the directed kinetic network[11] and the streamlines of the folding flows,[12] also show that the Cs-or and Ns-or folding pathways are close to the Cs-or and Ns-or clusters rather than go directly through them. As has been indicated earlier,[6,31] the large kinetic distances of the Ns-or and Cs-or states from the native one are due to the fact that the side chains of the N-terminal (C-terminal) strand are in the wrong orientation with respect to the rest of the three-stranded β-sheet, which requires almost complete unfolding of the N-terminal (C-terminal) hairpin for reaching the native state. The fact that the detailed balance between the Ns-or cluster and the native C-cluster is established later than that between the Cs-or cluster and the native C-cluster makes the Ns-or pathway prevalent.[25]

According to the previously mentioned difference between the rates of transitions in implicit solvent folding simulations and experiment (section 3), i.e., that the transitions in simulations are 2 orders of magnitude faster than in experiment, the time scales of equilibration of characteristic conformations in beta3s miniprotein can be expected as follows: the order of 100 ns for helical, native-like, and Cs-or conformations, several times larger for Ns-or conformations, and 1 order larger for Ch-curl conformations and the establishment of equilibrium between all states.

5. CONCLUSION

By use of a time-dependent free energy disconnectivity graph (TD-FEDG), the process of equilibration of protein states in a three-stranded antiparallel β-sheet miniprotein (betas3s) has been studied. To create a collection of protein conformations, a 20 μs equilibrium folding trajectory of beta3s generated with the CHARMM program[41] in the Caflisch group[27] was used. The protein states were represented by clusters of characteristic protein conformations, which were obtained by grouping the representative points of the protein in a reduced (three-dimensional) space of collective variables with the hydrogen bond PCA.[6] The vertices of the graph are the equilibrium free energies of the clusters, and the edges are the free energies of the transition states connecting the clusters. To determine the transition state free energy, the "complete" partition function (Eyring[33]) is used, which is equal to the product of the "standard" partition function of the transition state and the translational partition function corresponding to the ballistic motion of the system along the reaction coordinate. The distance along the reaction coordinate that enters the translational partition function is taken to be proportional to the observation time[22] so that as the time increases, a larger number of representative points cross the transition state. Accordingly, the values of the free energy barriers between the clusters decrease and (local) equilibrium between the clusters of characteristic conformations is established; the clusters are grouped into larger (consolidated) clusters, extending the equilibrium to a larger portion of the protein states until all states come into equilibrium.

It has been found that the equilibrium is first established in the native and helix consolidated clusters. It is significant that the helix consolidated cluster contains a large portion of clusters (three of six) in which the protein conformations are mostly unstructured; this is in accordance with the fact that the helical conformations are readily formed from the unstructured ones, which is observed in the initial stage of folding of beta3s.[11,12] The behavior of Cs-or and Ns-or conformations, which are structurally similar (one hairpin is formed, and the other is not), is essentially different. The Cs-or conformations join the helix and native consolidated clusters in small time after the equilibrium between these clusters is established, while the Ns-or conformations equilibrate with the native-like and helical conformations considerably later. The last conformations that come in equilibrium with the others are the Ch-curl conformations, which are less stable than the antiparallel β-strands and more difficult to form dynamically because of the distant N- and C-terminal strands. The observed hierarchy of the establishment of equilibrium between the characteristic conformations is in good agreement with the results of the previous studies[6,11,12,25,28,30−32,45] and has allowed further insight into the process of beta3s folding. One issue of particular interest is the difference between the Cs-or and Ns-or pathways, which has been indicated in the early work[25] and confirmed in the later, the first-passage folding studies:[11,12] although Cs-or and Ns-or conformations are "symmetrical" in that each of them has one hairpin formed and the other unstructured, the Ns-or pathway prevails. The TD-FEDG reveals that it is because detailed balance between the Cs-or and native-like conformations is established faster than between the Ns-or and native-like conformations, which decreases the probability flow through the Cs-or pathway and thus makes the Ns-or pathway prevalent.

AUTHOR INFORMATION

Corresponding Author

*E-mail: chekmarev@itp.nsc.ru.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

I thank Martin Karplus, Amedeo Caflisch, and Sergei Krivov for valuable discussions and comments on the manuscript. This work was performed under a grant from the Russian Science Foundation (No. 14-14-00325).

REFERENCES

(1) Unger, R.; Harel, D.; Wherland, S.; Sussman, J. L. A 3D Building Blocks Approach to Analyzing and Predicting Structure of Proteins. *Proteins: Struct., Funct., Genet.* **1989**, *5*, 355−373.

(2) Wallin, S.; Farwer, J.; Bastolla, U. Testing Similarity Measures with Continuous and Discrete Protein Models. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 144−157.

(3) Zhou, T.; Caflisch, A. Distribution of Reciprocal of Interatomic Distances: A Fast Structural Metric. *J. Chem. Theory Comput.* **2012**, *8*, 2930−2937.

(4) Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C. Automatic Discovery of Metastable States for the Construction of Markov Models of Macromolecular Conformational Dynamics. *J. Chem. Phys.* **2007**, *126*, 155101.

DOI: 10.1021/acs.jpcb.5b04336
J. Phys. Chem. B 2015, 119, 8340−8348

(5) Noé, F.; Horenko, I.; Schütte, C.; Smith, J. C. Hierarchical Analysis of Conformational Dynamics in Biomolecules: Transition Networks of Metastable States. *J. Chem. Phys.* **2007**, *126*, 155102.
(6) Kalgin, I. V.; Caflisch, A.; Chekmarev, S. F.; Karplus, M. New Insights into the Folding of a Beta-Sheet Miniprotein in a Reduced Space of Collective Hydrogen Bond Variables: Application to a Hydrodynamic Analysis of the Folding Flow. *J. Phys. Chem. B* **2013**, *117*, 6092−6105.
(7) Schutte, C.; Fischer, A.; Huisinga, W.; Deuflhard, P. A Direct Approach to Conformational Dynamics Based on Hybrid Monte Carlo. *J. Comput. Phys.* **1999**, *151*, 146−168.
(8) Noé, F.; Fischer, S. Transition Networks for Modeling the Kinetics of Conformational Change in Macromolecules. *Curr. Opin. Struct. Biol.* **2008**, *18*, 154−162.
(9) Prada-Gracia, D.; Goómez-Gardeñes, J.; Echenique, P.; Falo, F. Exploring the Free Energy Landscape: From Dynamics to Networks and Back. *PLoS Comput. Biol.* **2009**, *5*, e1000415.
(10) Bowman, G. R.; Voelz, V. A.; Pande, V. S. Taming the Complexity of Protein Folding. *Curr. Opin. Struct. Biol.* **2011**, *21*, 4−11.
(11) Kalgin, I. V.; Chekmarev, S. F.; Karplus, M. First Passage Analysis of the Folding of a Beta-Sheet Miniprotein: Is It More Realistic Than the Standard Equilibrium Approach? *J. Phys. Chem. B* **2014**, *118*, 4287−4299.
(12) Kalgin, I. V.; Chekmarev, S. F. Folding of a β-Sheet Miniprotein: Probability Fluxes, Streamlines, and the Potential for the Driving Force. *J. Phys. Chem. B* **2015**, *119*, 1380−1387.
(13) Bartlett, A. I.; Radford, S. E. An Expanding Arsenal of Experimental Methods Yields an Explosion of Insights into Protein Folding Mechanisms. *Nat. Struct. Mol. Biol.* **2009**, *16*, 582−588.
(14) Wales, D. J. *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*; Cambridge University Press: Cambridge, U.K., 2003.
(15) Stillinger, F. H.; Weber, T. A. Hidden Structure in Liquids. *Phys. Rev. A* **1982**, *25*, 978−989.
(16) Tsai, C. J.; Jordan, K. D. Use of an Eigenmode Method To Locate the Stationary Points on the Potential Energy Surfaces of Selected Argon and Water Clusters. *J. Phys. Chem.* **1993**, *97*, 11227−11237.
(17) Berry, R. S.; Breitengraser-Kunz, R. Topography and Dynamics of Multidimensional Interatomic Potential Surfaces. *Phys. Rev. Lett.* **1995**, *74*, 3951−3954.
(18) Becker, O. M.; Karplus, M. The Topology of Multidimensional Potential Energy Surfaces: Theory and Application to Peptide Structure and Kinetics. *J. Chem. Phys.* **1997**, *106*, 1495−1517.
(19) Krivov, S. V.; Chekmarev, S. F.; Karplus, M. Potential Energy Surfaces and Conformational Transitions in Biomolecules: A Successive Confinement Approach Applied to a Solvated Tetrapeptide. *Phys. Rev. Lett.* **2002**, *88*, 038101.
(20) Krivov, S. V.; Karplus, M. Free Energy Disconnectivity Graphs: Application to Peptide Models. *J. Chem. Phys.* **2002**, *117*, 10894−10903.
(21) Evans, D. A.; Wales, D. J. Free Energy Landscapes of Model Peptides and Proteins. *J. Chem. Phys.* **2003**, *118*, 3891−3897.
(22) Gavrilov, A. V.; Chekmarev, S. F. Graphic Representation of Equilibrium and Kinetics in Oligopeptides: Time-Dependent Free Energy Disconnectivity Graphs. In *Bioinformatics of Genome Regulation and Structure*; Kolchanov, N., Hofestaedt, R., Eds; Springer: New York, 2004; pp 171−178.
(23) Wales, D. J.; Salamon, P. Observation Time Scale, Free-Energy Landscapes, and Molecular Symmetry. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 617−622.
(24) Wales, D. J.; Doye, J. P. K. Dynamics and Thermodynamics of Supercooled Liquids and Glasses from a Model Energy Landscape. *Phys. Rev. B* **2001**, *63*, 214204.
(25) Rao, F.; Caflisch, A. The Protein Folding Network. *J. Mol. Biol.* **2004**, *342*, 299−306.
(26) Settanni, G.; Rao, F.; Caflisch, A. Φ-Value Analysis by Molecular Dynamics Simulations of Reversible Folding. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 628−633.
(27) Muff, S.; Caflisch, A. Kinetic Analysis of Molecular Dynamics Simulations Reveals Changes in the Denatured State and Switch of Folding Pathways upon Single-Point Mutation of a Beta-Sheet Miniprotein. *Proteins: Struct., Funct., Bioinf.* **2008**, *70*, 1185−1195.
(28) Krivov, S. V.; Muff, S.; Caflisch, A.; Karplus, M. One-Dimensional Barrier-Preserving Free-Energy Projections of a Beta-Sheet Miniprotein: New Insights into the Folding Process. *J. Phys. Chem. B* **2008**, *112*, 8701−8714.
(29) Muff, S.; Caflisch, A. ETNA: Equilibrium Transitions Network and Arrhenius Equation for Extracting Folding Kinetics from REMD Simulations. *J. Phys. Chem. B* **2009**, *113*, 3218−3226.
(30) Qi, B.; Muff, S.; Caflisch, A.; Dinner, A. R. Extracting Physically Intuitive Reaction Coordinates from Transition Networks of a Beta-Sheet Miniprotein. *J. Phys. Chem. B* **2010**, *114*, 6979−6989.
(31) Zheng, W.; Qi, B.; Rohrdanz, M. A.; Caflisch, A.; Dinner, A. R.; Clementi, C. Delineation of Folding Pathways of a β-Sheet Miniprotein. *J. Phys. Chem. B* **2011**, *115*, 13065−13074.
(32) Carr, J. M.; Wales, D. J. Folding Pathways and Rates for the Three-Stranded β-Sheet Peptide Beta3s Using Discrete Path Sampling. *J. Phys. Chem. B* **2008**, *112*, 8760−8769.
(33) Eyring, H. The Activated Complex in Chemical Reactions. *J. Chem. Phys.* **1935**, *3*, 107−115.
(34) Evans, M. G.; Polanyi, M. Some Applications of the Transition State Method to the Calculation of Reaction Velocities, Especially in Solution. *Trans. Faraday Soc.* **1935**, *31*, 875−894.
(35) Carr, J. M.; Wales, D. J. Global Optimization and Folding Pathways of Selected α-Helical Proteins. *J. Chem. Phys.* **2005**, *123*, 234901.
(36) Krivov, S. V.; Karplus, M. Hidden Complexity of Free Energy Surfaces for Peptide (Protein) Folding. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 4766−14770.
(37) Karplus, M.; Porter, N. R.; Sharma, R. D. Exchange Reactions with Activation Energy. I. Simple Barrier Potential for (H, H2). *J. Chem. Phys.* **1965**, *43*, 3259−3287.
(38) Polanyi, J. C.; Zewail, A. H. Direct Observation of the Transition State. *Acc. Chem. Res.* **1995**, *28*, 119−132.
(39) Chung, H. C.; Louis, J. M.; Eaton, W. A. Experimental Determination of Upper Bound for Transition Path Times in Protein Folding from Single-Molecule Photon-by-Photon Trajectories. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 11837−11844.
(40) De Alba, E.; Santoro, J.; Rico, M.; Jiménez, M. De Novo Design of a Monomeric Three-Stranded Antiparallel β-Sheet. *Protein Sci.* **1999**, *8*, 854−865.
(41) Brooks, B. R.; Brooks, C. L., III; MacKerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(42) Neria, E.; Fischer, S.; Karplus, M. Simulation of Activation Free Energies in Molecular Systems. *J. Chem. Phys.* **1996**, *105*, 1902−1921.
(43) Ferrara, P.; Apostolakis, J.; Caflisch, A. Evaluation of a Fast Implicit Solvent Model for Molecular Dynamics Simulations. *Proteins: Struct., Funct., Bioinf.* **2002**, *46*, 24−33.
(44) Ferrara, P.; Apostolakis, J.; Caflisch, A. Thermodynamics and Kinetics of Folding of Two Model Peptides Investigated by Molecular Dynamics Simulations. *J. Phys. Chem. B* **2000**, *104*, 5000−5010.
(45) Ferrara, P.; Caflisch, A. Folding Simulations of a Three-Stranded Antiparallel Beta-Sheet Peptide. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10780−10785.
(46) Eaton, W. A.; Muñoz, V.; Hagen, S. J.; Jas, G. S.; Lapidus, L. J.; Henry, E. R.; Hofrichter, J. Fast Kinetics and Mechanisms in Protein Folding. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 327−359.
(47) Cavalli, A.; Ferrara, P.; Caflisch, A. Weak Temperature Dependence of the Free Energy Surface and Folding Pathways of Structured Peptides. *Proteins: Struct., Funct., Bioinf.* **2002**, *47*, 305−314.

DOI: 10.1021/acs.jpcb.5b04336
J. Phys. Chem. B 2015, 119, 8340−8348

(48) Fraley, C.; Raftery, A. E. Model-Based Clustering, Discriminant Analysis, and Density Estimation. *J. Am. Stat. Assoc.* **2002**, *97*, 611−631.
(49) Wales, D. J. Energy Landscapes: Some New Horizons. *Curr. Opin. Struct. Biol.* **2010**, *20*, 3−10.
(50) Wales, D. J. Perspective: Insight into Reaction Coordinates and Dynamics from the Potential Energy Landscape. *J. Chem. Phys.* **2015**, *142*, 130901.
(51) Andersen, C. A.; Palmer, A. G.; Brunak, S.; Rost, B. Continuum Secondary Structure Captures Protein Flexibility. *Structure* **2002**, *10*, 175−184.
(52) Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caflisch, A. Wordom: A Program for Efficient Analysis of Molecular Dynamics Simulations. *Bioinformatics* **2007**, *23*, 2625−2627.
(53) Bowman, G. R.; Beauchamp, K. A.; Boxer, G.; Pande, V. S. Progress and Challenges in the Automated Construction of Markov State Models for Full Protein Systems. *J. Chem. Phys.* **2009**, *131*, 124101.
(54) Voet, D.; Voet, J. G. *Biochemistry*, 4th ed.; Wiley: Hoboken, NJ, 2011.

DOI: 10.1021/acs.jpcb.5b04336
J. Phys. Chem. B 2015, 119, 8340−8348