Folding Pathways and Rates for the Three-Stranded β-Sheet Peptide Beta3s using Discrete Path Sampling

Joanne M. Carr and David J. Wales*

University Chemical Laboratories, Lensfield Road, Cambridge CB2 1EW, United Kingdom

Received: February 29, 2008; Revised Manuscript Received: April 24, 2008

The discrete path sampling method was used to investigate the folding of a three-stranded antiparallel β-sheet peptide, Beta3s, described by an empirical potential and implicit solvent model. After application of a coarse-graining scheme that groups together sets of minima in local equilibrium, the calculated folding time was in reasonable agreement with other simulations and consistent with the experimental upper bound. The folding mechanism exhibited by the most significant discrete paths involves early formation of the C-terminal hairpin followed by docking of the N-terminal strand.

1. Introduction

Understanding the complexity of a protein folding reaction at an atomic level of detail is still a significant challenge for both computer simulation and experiment. From a computational point of view, progress has arisen from, for example, the use of enhanced sampling techniques,[1] coarse-graining configuration space in some way,[2] or studying not full proteins but smaller peptides.[3] Peptides are more computationally tractable, but their folding still involves the formation of secondary structure elements—a significant process in the folding of full proteins. In this study, the folding of the Beta3s peptide was investigated using discrete path sampling:[4–6] an approach based on connected sequences of stationary points on the potential energy surface.

Beta3s is a 20-residue peptide designed to adopt a three-stranded antiparallel β-sheet conformation in its native state.[7] Experimentally, Beta3s was found to be monomeric in aqueous solution,[7] with an equilibrium between the expected β-sheet geometry and unfolded structures at 283 K. The hairpin turns at Gly-6-Ser-7 and Gly-14-Ser-15 are both of type II′. An upper bound on the folding time between 4 and 14 μs at 283 K was obtained from 1-D ^{1}H NMR spectra.[7] These features made Beta3s, among other three-stranded β-sheets, an attractive target for computational studies using a variety of atomistic force fields and implicit solvation models.[8–18]

Mohanty and Hansmann constructed a free energy surface using data from parallel tempering simulations;[9] the population of the native state was underestimated compared to experiment,[7] and the potential was found to give rise to structures with overbound hydrophobic cores. A mechanism was conjectured in which the hydrogen bonds at the two turns form first in the rate-determining step, followed by zipper-like completion of the hairpins, with the one that forms first serving as a template for the second. No preference was apparent for which hairpin forms first.

Caflisch et al. studied Beta3s extensively, using the CHARMM19 potential[19] and an implicit solvation model based on solvent-accessible surface area[20] (SASA). In their early work, multiple independent molecular dynamics (MD) trajectories were run for up to 250 ns at three different temperatures.[11,13] At 360 K, the highest temperature studied, the mean folding and unfolding times were 30 and 8 ns, respectively. Free energy surfaces obtained at the three temperatures, which span 60 K, were qualitatively similar. On the dominant folding pathway, contacts between the second and the third (i.e., middle and C-terminal) strands formed first, starting close to the hairpin turn, followed by docking of the first (N-terminal) strand onto the existing hairpin. This order is in agreement with the mechanism postulated from a free energy surface constructed[3] using explicit water simulation data for another three-stranded β-sheet peptide, Betanova. Two thermodynamic transition state ensembles (TSEs) of unequal free energy were identified, each with one hairpin almost fully formed and the remaining strand unstructured; the one with the N-terminal strand unstructured was lower. The free energy barriers to formation of the first hairpin were mostly entropic, and native side chain interactions were found to be more important than hydrogen bonds in driving the folding. While Bursulaya and Brooks observed that the solvent did not play a detailed role in their explicit water simulations[3] of the folding of Betanova, complete removal of the implicit solvation energetic contribution disrupted the folding process in the simulations of Caflisch and Ferrara[11] for Beta3s.

From another set of independent MD runs using the same CHARMM19/SASA force field, but simulated for microseconds to capture reversible unfolding/folding events, Cavalli et al. concluded that while the denatured state ensemble of a small protein cannot be characterized by a few statistically relevant conformations, folding occurs via a small number of conformations per trajectory.[14] An average folding time of 85 ns was obtained at 330 K. After a clustering algorithm based on structural similarity was applied, the unfolded state was found to consist of many small clusters of configurations from the simulation. Folding occurred efficiently at 330 K, and a downhill profile was observed for the average potential energy as a function of order parameters based on fractions of native contacts for the full β-sheet and within each hairpin. In the current work, the term potential energy means the sum of intramolecular and solvation energies. This quantity is sometimes referred to elsewhere as the effective energy. At lower temperatures (below 300 K), such average potential energy profiles from replica exchange MD simulations[21] show two transition states and an intervening local minimum, which arises from the intrinsic enthalpic stabilization of the C-terminal hairpin due to the interaction of its side chains.[15] The replica exchange

* Corresponding author. E-mail: dw34@cam.ac.uk.

10.1021/jp801777p CCC: $40.75
Published on Web 06/28/2008

simulations also revealed that non-native interactions feature significantly in the unfolded state.

The MD snapshots collected and used by Cavalli et al.[14] were subsequently analyzed within a network formulation by Rao and Caflisch.[16] Snapshots grouped according to secondary structure assignments became nodes in the network. Two nodes were connected by edges (links) if a pair of snapshots, one from each node, were adjacent in the MD trajectory. The network therefore described significant free energy minima and their connectivity without the need for reaction coordinates. For the Beta3s data, the network was found to be scale-free,[22] with a few hubs that are both well populated and highly connected, while the majority of the nodes had a relatively small number of connections. Nodes with high connectivity and low statistical weight were identified as putative members of TSEs; the identification was verified via P_{fold} (folding commitment probability) calculations. The same two TSEs were observed as in ref 11, and the many heterogeneous dynamical routes to folding mapped onto two average folding pathways within the network formulation because the dynamical paths have common structural features. The denatured state was found to contain both high-entropy, high-enthalpy structures and low-enthalpy, low-entropy traps (including curl-like conformations). In contrast to a random heteropolymer without a native state, the native basin for Beta3s exhibited a hierarchical organization of conformations. The network representation was further investigated by Gfeller et al. in a study that linked various topological properties of networks with the underlying free energy surfaces.[18] Among other systems (a quadratic-well model and a random lattice heteropolymer), networks for the native-state ensemble of Beta3s were constructed and analyzed using a 10 ns MD trajectory simulated at 270 K using the CHARMM19/SASA force field. The topology of the resulting networks was found to vary significantly with the number of time steps between successive saved snapshots.

A Φ-value analysis provides structural information on unstable intermediates and rate-determining TSEs in protein folding, by assessing the effect that nondisruptive amino acid mutations have on the free energy profile.[23] Settanni et al.[17] obtained theoretical Φ-values for Beta3s from multiple, independent, near-equilibrium MD simulations at 330 K for a number of single-point mutants, again modeled with the CHARMM19/SASA potential. In the mutants for which the data are reliable, the TSE was found to be heterogeneous, supporting two possible pathways, as in the wild-type. However, the relative weights of the two paths vary with the mutant: in most cases, the wild-type pathway with the C-terminal hairpin forming first was still favored, but some mutations caused destabilization of the 2−3 strand to such an extent that the alternative route, in which the N-terminal hairpin forms first, became dominant. It was postulated that the destabilization resulted from the different steric requirements of the amino acids involved. Also, non-native interactions were found to be specifically involved in determining whether a conformation can fold or not, via loose precursors of the type II′ turn for the unstructured hairpin, shifted by one residue to the C-terminus. Consequently, structural Φ-values based on native atomic contacts (rather than all possible contacts) tend to overestimate the degree of native structure present in the TSE for Beta3s. No experimental Φ-values are available currently for comparison. Beta3s was also used in force-field refinement[10] and as a test system for new methods of analyzing existing simulation data for various purposes.[24,25]

2. Methods

The Beta3s peptide was modeled using the CHARMM19 united-atom potential[19] and the EEF1 implicit solvation model.[26] The EEF1 function is based on solvent exclusion and the assumption that the solvation free energy of a protein is a sum of group contributions; the contributions are determined from values for small model compounds.[26] The EEF1 model incorporates a modification to the standard CHARMM19 parameters whereby charged side chains and termini are neutralized as well as a particular set of options in the calculation of nonbonded interactions, including the use of cutoffs and a linear distance-dependent dielectric constant. A slightly modified version of EEF1 was employed, in which discontinuities arising from the use of cutoffs were removed.[27] The modifications[28] to the CHARMM19 force field that ensure rotamers of various residues have the same energies and geometries were also used here.

The use of CHARMM19 with the SASA[20] model would have been desirable for direct comparison with the approach of Caflisch et al.[11,13,14] However, as with the original EEF1, the use of cutoffs in SASA introduces discontinuities that render the model unsuitable for geometry optimization without modification. Furthermore, EEF1 was found to better distinguish between native and non-native conformations of Beta3s in a comparison of various implicit solvation schemes and protein models.[29]

2.1. Discrete Path Sampling. The approach employed in the current work to sample peptide folding is to characterize trajectories in terms of discrete paths: connected sequences of minima and the intervening transition states of the potential energy surface.[4,5,30] Here, a transition state is defined geometrically as a stationary point with a single negative Hessian eigenvalue.[31] The connectivity of a transition state is then defined by the two minima reached by (approximate) steepest-descent paths leaving parallel and antiparallel to the eigenvector whose eigenvalue is negative; such a minimum–transition state−minimum triplet is termed an elementary rearrangement. The number of steps in a given discrete path is the number of transition states in the sequence, which is one less than the total number of minima.

The discrete path sampling (DPS) approach,[4,5] a coarse-grained analogue of the transition path sampling method,[32–34] was employed to systematically generate ensembles of discrete paths from an initial connected path between two end points. Local minima are assigned to the end point states of interest (denoted *A* and *B*) according to the values of an appropriate order parameter. These states should each be in local equilibrium for single exponential decay from an initial nonequilibrium distribution to be observed. Any remaining minima are classed as intervening, denoted *I*. The original DPS formulation was presented in detail elsewhere,[4–6] as well as more recent developments.[35] In short, the aim is to enlarge a database of connected stationary points starting from those in the initial path by adding all the minima and transition states found during successive connection-making attempts to construct discrete paths between pairs of minima selected from the current database.

Within the steady-state approximation for the *I* minima, the phenomenological $B \leftarrow A$ rate constant can be expressed as

$$k_{BA}^{\mathrm{SS}} = \frac{1}{p_A^{\mathrm{eq}}} \sum_{b \leftarrow a} \frac{k_{bi_1} k_{i_1 i_2} \cdots k_{i_n a} p_a^{\mathrm{eq}}}{\sum_{\alpha_1} k_{\alpha_1 i_1} \sum_{\alpha_2} k_{\alpha_2 i_2} \cdots \sum_{\alpha_n} k_{\alpha_n i_n}} \qquad (1)$$

where k_{ij} is the net rate constant for all the elementary rearrangements from minimum j into minimum i, p_a^{eq} is the equilibrium occupation probability of minimum a, and p_A^{eq} is

the equilibrium occupation probability of the A group of minima, which includes a. The main sum is over all paths that start and finish on the A and B boundaries, respectively, and therefore includes direct $B \leftarrow A$ transitions as well as paths that start on the A/I boundary and pass through only intervening minima. The sums over α_j in the denominator include all geometrically distinct minima directly connected to intervening minimum i_j. Eq 1 can be written more succinctly as

$$k_{BA}^{SS} = \frac{1}{p_A^{eq}} \sum_{b \leftarrow a} P_{bi_1} P_{i_1 i_2} \cdots P_{i_{n-1} i_n} P_{i_n a} p_a^{eq} \tau_a^{-1} = \frac{1}{p_A^{eq}} \sum_{a \in A} \frac{C_a^B p_a^{eq}}{\tau_a} \tag{2}$$

where $\tau_i = 1/\sum_\alpha k_{\alpha i}$ is the mean waiting time for any transition to occur from minimum i, and $P_{ji} = k_{ji}\tau_i$ is the branching probability for the $j \leftarrow i$ transition among the set of possible elementary transitions out of minimum i. C_a^B is a committor probability: the probability that a random walk starting from a will encounter a B minimum before it returns to the A state.[36] Once the mean waiting times and branching probabilities have been calculated (section 2.3), we have all the quantities required to evaluate k_{BA}^{SS} and, analogously, k_{AB}^{SS}, as well as the contribution a particular discrete path makes to the phenomenological two-state rate constants.

Although the accuracy of k_{BA}^{SS} and k_{AB}^{SS} depends on the validity of the steady-state approximation, these rate constants provide a convenient way of selecting pairs of minima for subsequent connection-making attempts.[35] The discrete path that makes the largest contribution to the phenomenological rate constant k_{BA}^{SS} (or k_{AB}^{SS}) can be extracted from a DPS database using a network formulation[37] via Dijkstra's shortest-path algorithm.[38] Here, the edge weights are chosen to reflect the branching probabilities in eq 2. We will refer to such discrete paths as the fastest paths, although it is important to note that the analysis includes the conditional occupation probability of the reactant minimum.

Having determined the fastest path in terms of the largest contribution to k_{AB}^{SS} or k_{BA}^{SS}, we can identify the Euclidean distances (minimized with respect to overall rotation and translation) between its minima,[39] and also the largest individual potential or free energy barriers. One refinement scheme[35,40] for selecting pairs of minima for further connection attempts, referred to as SHORTCUT, chooses pairs that are the closest together in configuration space, yet separated by a minimum number of steps on the fastest path. Another scheme,[35] denoted SHORTCUT BARRIER, selects pairs of minima on either side of, and an equal number of steps away from, the largest potential or free energy barriers. The latter approach provides an efficient way of finding alternative paths that circumvent the highest barriers, while the SHORTCUT scheme yields discrete paths with a smaller number of steps.

Growing a stationary point database according to either SHORTCUT approach will generally only lead to an improvement in the contributions made to the overall rate constant by the fastest paths. It is also likely to introduce kinetic traps into the database in the form of local minima, found during the connection attempts, that lie behind high barriers. Such traps are artifacts of incomplete sampling because low-barrier paths to the rest of the database exist but have not been found. Furthermore, they lead to a reduction in the calculated two-state rate constants, which are affected by the slowest relaxation time in the database. Therefore, one must also refine the database to remove such artificial traps. The first stage in such a refinement algorithm, called FREEPAIRS, is to partition the database into groups of minima and the resulting ensembles of transition states that directly connect minima in two different groups. This partitioning can be achieved using various different criteria; the scheme employed in this work is presented in section 2.2. Only groups that lie within a chosen threshold free energy of the group of lowest free energy, and are therefore thermodynamically relevant at the prevailing temperature, are considered. The groups with the lowest values of the ratio of the free energy difference between themselves and lowest-energy product group to the effective reaction barrier height are then identified. An estimate of the barrier is made by finding the lowest free energy at which the group in question and the lowest-energy product group are mutually accessible in a superbasin analysis,[41] as used in the construction of disconnectivity graphs; the difference between that energy and the free energy of the group under consideration is the barrier height.

The ratio of a stability gap to a barrier height is a similar quantity to the Z-score[42] (the normalized difference between the energy of the native structure and the average energy of a representative set of non-native structures) and the ratio of folding temperature to glass-transition temperature[43] that are often invoked to assess the structure-seeking behavior of a system. Hence, we focus on groups whose reaction kinetics are frustrated, with similar free energies to the product but separated from it by high barriers. An additional minimum distance criterion is then employed to select pairs of local minima, one from the chosen group and one from the product group of lowest free energy, as end points for new connection-making runs. A schematic view of the discrete path sampling algorithm, with the SHORTCUT procedure used for the selection of pairs of minima to be connected, is presented in Figure 1.

The connection attempts employed our iterative Dijkstra-based approach for selecting the pairs of minima to connect; this procedure is described in detail elsewhere.[44] The required double-ended transition state searches were performed with the doubly[45] nudged elastic band[46–48] (DNEB) algorithm, and the resulting transition state candidates were tightly converged using hybrid eigenvector-following.[49] The limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algorithm of Liu and Nocedal[50,51] was used for the local minimizations. An optimization was deemed to have converged when the root-mean-square gradient fell below 10^{-6} kcal mol^{-1} Å^{-1}.

2.2. Grouping a Sample of Connected Stationary Points. The full databases of local minima from the DPS runs were partitioned into groups for coarse-grained analysis using the following recursive scheme: (1) Assign each minimum under consideration to its own free energy group. Calculate the free energy of each group, and of the intergroup transition state ensembles, using the approach of section 2.3. (2) For each intergroup transition state ensemble in turn, merge the two directly connected groups together if their forward and backward free energy barriers both lie below a fixed threshold value, $\Delta F_{barrier}$. (3) Calculate the free energy of each changed group and of the changed intergroup transition state ensembles. (4) Go to step 2 or exit if the overall group membership did not change during step 2. The aim of this algorithm is to self-consistently generate groups of minima that are in local equilibrium on a time scale determined by the choice of the barrier height, $\Delta F_{barrier}$.

2.3. Calculating Thermodynamic and Kinetic Properties. Using the superposition method,[52,53] the total partition function ($Z(T)$, where T is temperature) can be obtained as a sum of the partition functions, $Z_i(T)$, from the catchment region[54] of each local minimum, i. A catchment region, or basin of attraction, is defined as the volume of configuration space surrounding a local

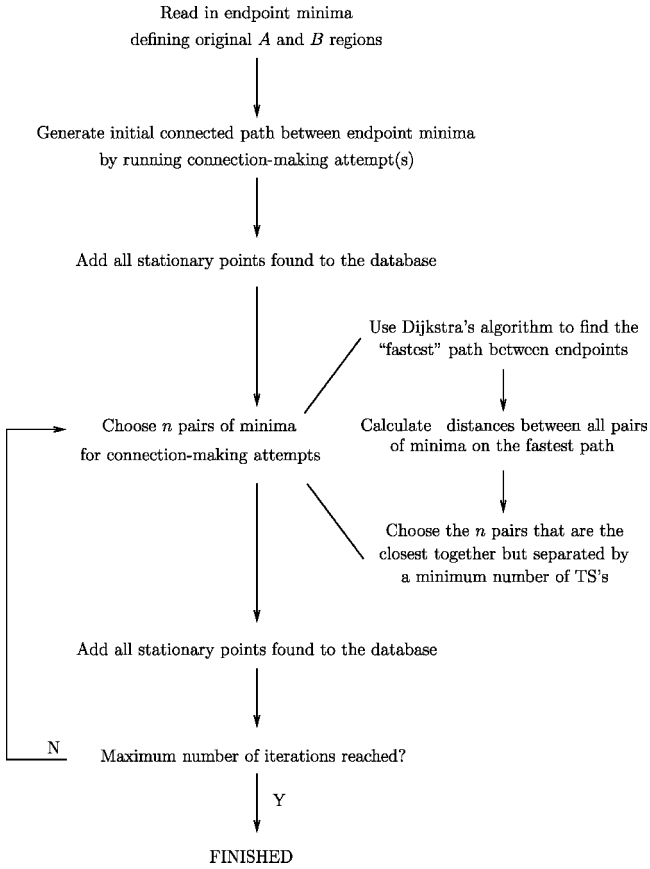

Figure 1. Schematic view of the discrete path sampling algorithm. The SHORTCUT procedure is included to illustrate the selection of pairs of minima for subsequent connection-making attempts. This procedure could be replaced by either SHORTCUT BARRIER or FREEPAIRS, as described in the text.

minimum from which steepest-descent paths lead to that minimum.[54] The equilibrium occupation probability of minimum i is then $p_i^{\mathrm{eq}}(T) = Z_i(T)/Z(T)$, and the corresponding free energy is $F_i(T) = -k_B T \ln Z_i(T)$, where k_B is the Boltzmann constant. The free energy of transition state † is $F^{\dagger}(T) = -k_B T \ln Z^{\dagger}(T)$, where the partition function, $Z^{\dagger}(T)$, does not include the reactive mode corresponding to the negative Hessian eigenvalue. We employed the harmonic approximation to provide computationally inexpensive values for the local density of vibrational states for each stationary point and hence the partition functions $Z_i(T)$ and $Z^{\dagger}(T)$.

Transition state theory[55–60] (TST) supplies the rate constant, $k_i^{\dagger}$, for an elementary rearrangement out of minimum i via transition state †. The total rate constant for direct isomerization of minimum i into minimum j, k_{ji}, is obtained by summing the $k_i^{\dagger}$ for all transition states that directly connect the two minima. The TST rate constant is often an upper bound to the true value, as no account is taken of recrossings of the dividing transition state surface: the assumption is made that a configuration at the dividing surface will proceed directly to either of the connected minima without recrossing the surface. A more accurate, but computationally expensive, approach would be to perform a full reactive flux calculation.[61–63] However, given the approximations inherent in the force fields employed, we consider harmonic densities of states and the corresponding TST formulation a reasonable approach.

For a grouped database of stationary points, the equilibrium occupation probability, p_J^{eq}, and the relative free energy, F_J, of group J are obtained from appropriate quantities for the individual members, j, as follows:

$$p_J^{\mathrm{eq}}(T) = \sum_{j \in J} p_j^{\mathrm{eq}}(T) \quad \text{and} \quad F_J(T) = -k_B T \ln \sum_{j \in J} Z_j(T) \tag{3}$$

The relative free energy of the set of transition states that connect a minimum in group J with a minimum in group L is taken as

$$F_{LJ}(T) = -k_B T \ln \sum_{(l,j)^{\dagger}} Z^{\dagger}(T) \tag{4}$$

and the corresponding intergroup rate constant from J to L, k_{LJ}, is then[40,64]

$$k_{LJ}(T) = \sum_{(l,j)^{\dagger}} \frac{p_j^{\mathrm{eq}}(T)}{p_J^{\mathrm{eq}}(T)} k_j^{\dagger}(T) \tag{5}$$

2.4. Analyzing the DPS Database of Stationary Points. We have previously found the graph transformation (GT) algorithm[36,65] to be a particularly efficient and robust method for the calculation of phenomenological rate constants k_{AB} and k_{BA}. For each reactant minimum, a, we start from a complete database and remove the intervening minima, I, in succession, and then all the other reactant minima. At each step, the branching probabilities and mean waiting times are renormalized to conserve the mean first-passage times between the A and B states. The required rate constants can then be written as

$$k_{BA}^{\mathrm{GT}} = \frac{1}{p_A^{\mathrm{eq}}} \sum_{a \in A} \frac{p_a^{\mathrm{eq}}}{\tau_a'} \sum_{b \in B} P_{ba}' \text{ and } k_{AB}^{\mathrm{GT}} = \frac{1}{p_B^{\mathrm{eq}}} \sum_{b \in B} \frac{p_b^{\mathrm{eq}}}{\tau_b'} \sum_{a \in A} P_{ab}' \tag{6}$$

where a prime denotes the final value of a renormalized quantity. In the transformed database, τ_a' is the renormalized waiting time for a transition to any of the B minima, and P_{ba}' is the transformed branching probability for the a to b path in the set of paths from a to all the B minima.

Unlike the rate constant formulation in eq 1, a GT calculation requires neither the imposition of the steady-state approximation for the I minima nor the assumption of local equilibrium within the A and B regions. However, the assumption of Markovian, or memoryless, dynamics[66] is still inherent in the GT approach. Therefore, the GT rate constants, k_{BA}^{GT} and k_{AB}^{GT}, are equivalent to the values obtainable in principle from kinetic Monte Carlo[67,68] simulations and other master equation-based[66,69] techniques. However, the GT approach does not require diagonalization of the transition matrix, and the computational cost is independent of temperature.

A GT calculation can be performed for a database of connected stationary points using a description of the kinetics in terms of transitions between either individual minima or groups of minima. The formulations for the rate constants and equilibrium occupation probabilities in both cases are given in section 2.3. For the intergroup kinetics, we assume that transitions between any two minima within a group occur instantaneously. Thus, it is important that the grouping scheme applied has indeed generated sets of minima for which there is a separation of time scales between intra- and intergroup transitions.

Mechanistic information also can be extracted from a DPS database by considering the most significant discrete paths directly,[70] using an extension to the network formulation employed to find the single path that makes the largest contribution to k^{SS} in the appropriate direction (section 2.1). Instead of using Dijkstra's shortest-path algorithm, we employ the recursive enumeration algorithm[71] (REA) to obtain the k "shortest" paths between two given end points (i.e., the k discrete paths, in rank order, that make the largest contributions to k_{AB}^{SS}

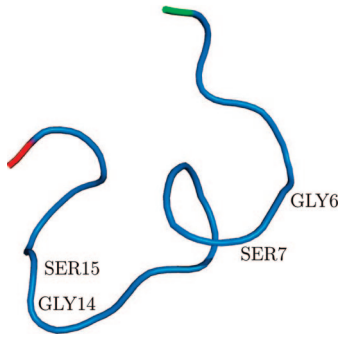

Figure 2. Structure of the starting minimum, a_{MD}, with the positions of four key residues from the two native turns marked. The N-terminus is green, and the C-terminus is red.

or k_{BA}^{SS}). Detailed descriptions of the general operation of the REA,[71] and its use in this particular context,[70] are given elsewhere.

For each of the k paths generated with the REA, we obtained the rate constant for that unbranched path, neglecting all off-path minima, via a GT calculation. These paths are then ranked according to the corresponding rate constants and analyzed. We are therefore assuming that our chosen value of k is sufficiently large that the set of paths making the largest contribution to k^{SS} (eq 1) will contain the most important paths when the steady-state condition is relaxed, even if the rank orders differ. Paths on which some barrier heights are very small compared to the thermal energy, k_BT, often give rise to sets of paths that only differ in the number of times the low-barrier transition states are fully recrossed. Such paths have the same unbranched rate constant when all possible recrossings are allowed for and do not need to be treated separately.

3. Results and Discussion

To characterize suitable initial end points in the denatured state and the folded β-sheet ensemble we first considered the results of constant temperature MD simulations. Configurations were saved every 10 ps throughout the trajectories and examined for the presence of the native hydrogen bonds and side chain−side chain contacts defined in ref 11. This definition of the native contacts was obtained from an MD simulation at 300 K with the CHARMM19/SASA model, and as such allows some comparison between our work and previous SASA-based studies.[11,13,14] Further comparison with high-resolution structural information is not possible as the conformations obtained from NMR experiments are not well defined.[7]

From a simulation run at 330 K, a configuration containing the full set of these native contacts was taken, after local minimization, as the sole initial member of the folded, B, state, and is labeled b_{MD}. An additional MD trajectory at 298 K verified that this structure is at least locally stable at the temperature used for the subsequent kinetic and thermodynamic analyses.

A reasonably short high-temperature simulation was also started from a fully extended conformation to partially explore the denatured state. A snapshot from this trajectory, possessing none of the native contacts, was used to generate the initial local minimum, denoted a_{MD}, in the A state representing the denatured ensemble. This minimum, which exhibits a loose helical turn in the region of what becomes the middle strand of the β-sheet, is shown in Figure 2.

An initial discrete path between these two end points was generated, and the resulting database of stationary points was expanded using sequential application of the SHORTCUT, SHORTCUT BARRIER, and FREEPAIRS procedures described in section 2.1. A database containing around 2 million minima and transition states was then pruned to further remove frustration, as follows. The grouping scheme of section 2.2 was applied with $\Delta F_{barrier} = 5$ kcal mol^{-1}, and the fastest path (in terms of a sequence of connected groups of free energy minima, rather than individual local potential energy minima) was obtained using Dijkstra's algorithm as discussed previously. Only the potential energy minima and transition states in the free energy groups on the fastest path were retained in a new, reduced database containing $\sim 10^5$ stationary points. The rate constants calculated for the new databases created in this way with a range of suitable values for $\Delta F_{barrier}$ were not found to differ significantly. This reduced, but less frustrated, database was subsequently expanded further by again applying the SHORTCUT and FREEPAIRS schemes in a second refinement iteration. The final database used for analysis contained $\sim$500 000 stationary points.

3.1. Calculation of Folding Times. Phenomenological rate constants for folding, k_{BA}^{GT}, were obtained using the GT approach described in section 2.4. Dijkstra's shortest-path algorithm and the recursive enumeration algorithm (REA) were used to extract the paths that make the largest contributions to the DPS nonrecrossing steady-state rate constant (k_{BA}^{SS}) as described in sections 2.1 and 2.4. All the values quoted correspond to a temperature of 298 K.

For some of the analyses, a coarse-grained approach was employed in which the kinetics were described in terms of transitions between groups of minima (section 2.3). The grouping was performed as described in section 2.2 with $\Delta F_{barrier} = 5$ kcal mol^{-1}. Similar results, in terms of both the folding rate constants and the disconnectivity graphs (section 3.3) that provide a two-dimensional representation of the energy landscape, were obtained for a meaningful range of threshold values around this chosen one. The main difference between the energy landscapes for these different groupings is that the denatured state is lowered in free energy as $\Delta F_{barrier}$ increases and classifies more minima as denatured; however, the folding transition state is similarly lowered in energy so overall the barriers and, therefore, rates remain comparable. To further justify the chosen grouping, the mean waiting time for an intergroup transition from a group, averaged over groups, was compared with the mean waiting time for an intragroup transition from a local minimum, averaged over minima that do not possess any direct connections to minima in other groups and are not dead ends. The former quantity is longer than the latter by more than 2 orders of magnitude (a factor of 460), which indicates that the assumption of a separation of time scales between inter- and intragroup motion is valid.

The folded state (B) was taken as the group containing b_{MD}, and the group containing a_{MD} became the A state. The GT folding rate constant is then 3.3×10^6 s^{-1}, which corresponds to a folding time of $\sim$300 ns. The rate constant from a GT calculation for the reverse direction was compared with the value obtained using the previous folding rate and the detailed balance relation. The relatively small difference between the two values (a factor of 10) indicates that the steady-state approximation holds reasonably well for this Beta3s energy landscape.

The $b_{MD} \leftarrow a_{MD}$ rate constant from a GT calculation on the full, ungrouped database could not be obtained on our hardware as the memory required exceeds 8 gigabytes (the peak requirement is unknown in advance). The individual fastest path between those two end points has 83 steps and makes a

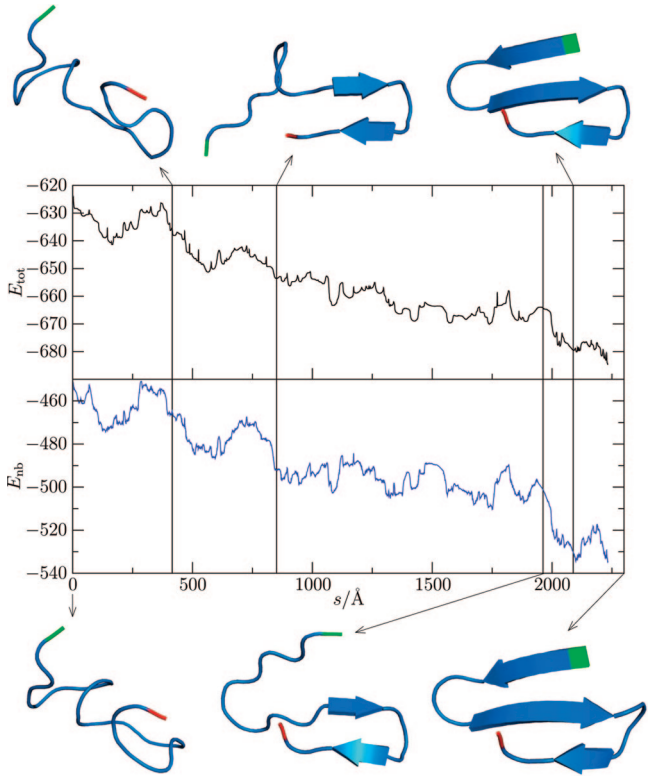

Figure 3. Total potential energy (upper) and nonbonding component (lower) as a function of the integrated path length, s, along a significant folding pathway (denoted path 47 in the text). The energy is in units of kcal mol^{-1}. The vertical lines indicate the approximate positions at which the two hairpins start to form and are completed. The corresponding four conformations and the two end point structures also are shown, using a cartoon representation. The N-termini are green, and the C-termini are red. On the pathway, the C-terminal hairpin forms first, and then the N-terminal strand docks against it to complete the β-sheet.

contribution of 5.7×10^{-82} s^{-1} to k_{BA}^{SS}. This very small value indicates that a large number of paths in the database must contribute to the folding process, as was found in previous studies of long paths.[5,40,72] This result highlights the role of entropy in terms of alternative paths, which supplements the entropic contribution to the rate constants for individual minimum-to-minimum transitions.

The 250 fastest paths found with the REA span a narrow range of values in both the contributions to k_{BA}^{SS} and the rate constants calculated in the absence of any off-path transitions and show only minor mechanistic deviations (see Figure 5 and the corresponding discussion in section 3.3). However, the rank order in terms of the two sets of rate constants differs: the path with the largest unbranched rate constant is number 47 based on its contribution to k_{BA}^{SS}. The unbranched rate constants for individual discrete paths are much larger than the corresponding contributions to k_{BA}^{SS}; a typical value was 7×10^{3} s^{-1}. This rate is still significantly smaller than the overall rate constant for folding because the analysis does not include any grouping. Hence, the rate constants that we are associating with individual paths do not allow for reassignment of some intervening minima as reactants. These quantities are simply used to provide relevant pathways for mechanistic analysis. Further discussion of the underlying energy landscape is given in section 3.3.

3.2. Sequence of Folding Events. The individual $b_{MD} \leftarrow a_{MD}$ discrete path with the largest rate constant (i.e., path 47) from the relevant folding-time analysis of the ungrouped database (section 3.1) was analyzed mechanistically. The C-terminal hairpin forms first, starting at the turn and proceeding via a zipper-like mechanism. The N-terminal strand then docks against the existing hairpin to complete the β-sheet. The total potential energy (E_{tot}, from the CHARMM19 potential plus the EEF1 solvation model) and the nonbonding contribution (E_{nb}, from the sum of the electrostatic and van der Waals terms) as a function of the integrated Cartesian path length are plotted in the upper and lower panels of Figure 3, respectively. The structures and energy values between adjacent local minima are obtained from the approximate steepest-descent paths calculated for the corresponding transition state. The four vertical lines indicate the approximate positions along the path at which the two hairpins start to form and are completed. The corresponding four conformations and the two end point structures are also shown in Figure 3. The features of the nonbonded energy profile dominate the total energy; the solvation energy tends to increase as the peptide folds and interactions with the solvent are replaced by intramolecular ones, but the effect on E_{tot} is not as significant. Furthermore, it can be seen from the nonbonded energy that the formation of the two hairpins differs. When the β-sheet is completed, the energy decreases steeply, without encountering a high uphill barrier, as the N-terminal strand approaches the existing hairpin without appreciable hindrance and with almost the correct backbone conformation. However, both the nonbonded energy and the total potential energy fall before rising and falling again overall as the first hairpin forms. The intervening energy rise is probably due to the significant conformational change required at the C-terminus before the last hydrogen bond of the hairpin can be made. Formation of the first hairpin requires more steps (transition states) than the second.

The path with the smallest unbranched rate constant from the set of 250 (number 241 according to its contribution to k_{BA}^{SS}) differs from path 47 described previously mainly in the dynamics of the N-terminal strand between the formation of the two hairpins (see section 3.3). The alternative folding pathway found by Ferrara et al.,[11,13] in which the N-terminal hairpin forms first, may be present in our database of stationary points but with a rate constant that is too small to feature in the set of paths analyzed.

3.3. Disconnectivity Graphs. To visualize the region of the energy landscape sampled, disconnectivity graphs[73] were constructed on the basis of both potential energy (PE) and free energy[74,75] (FE) at 298 K. In the latter case, the approach described in ref 75 was employed, with the required free energies obtained as discussed in section 2.3.

A PE disconnectivity graph for the ungrouped database is given in Figure 4. Only the lower half of the PE landscape was analyzed, to emphasize the features likely to be most thermodynamically relevant at the temperatures of interest. Minima from which the native state is deemed to be kinetically inaccessible, due to the presence of a very high (presumably artificial) barrier, also were excluded. The original folded minimum b_{MD} lies very low but not lowest in PE: it is surrounded by native-like minima that vary at the turns or in the docking of one of the terminal strands (usually the N-terminal one) with the other hairpin. Most of these structures have an all-atom root-mean-square deviation of less than 3 Å from b_{MD}. However, not all of these minima are grouped together with b_{MD} under the recursive scheme using $\Delta F_{barrier} = 5$ kcal mol^{-1} (section 3.1), probably because we have not sampled connections within the native state extensively due to our focus on denatured $\rightarrow$ native connections. Lines emanating from local minima that are grouped

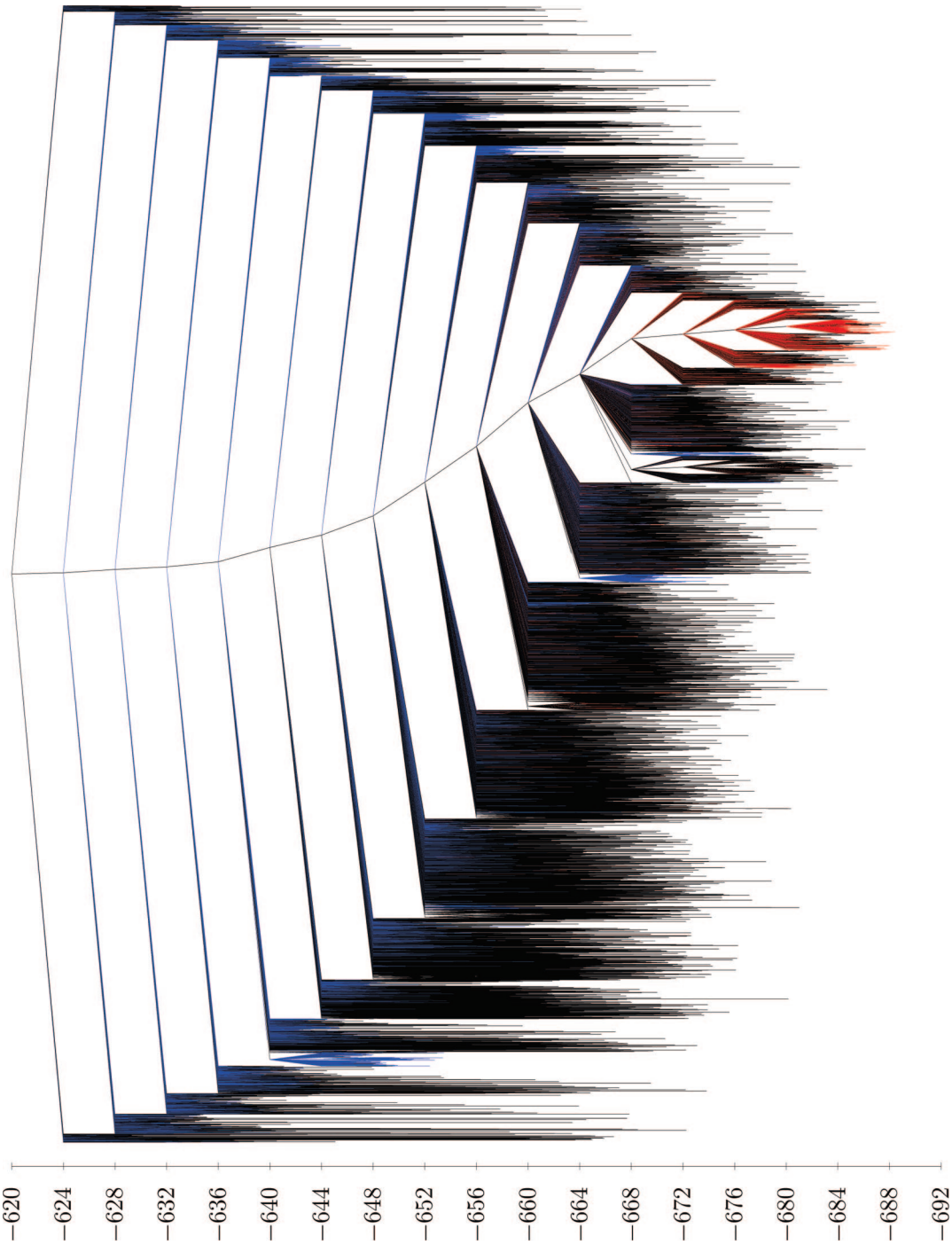

Figure 4. Potential energy (kcal mol^{-1}) disconnectivity graph for the full, ungrouped folding database. To illustrate the effect of the scheme applied to group the database for some of the analyses, the branches leading to minima grouped together with the original denatured structure a_{MD} are blue, and those leading to minima grouped with the original folded structure b_{MD} are red. In particular, the grouping significantly expanded the denatured state from a single minimum.

with b_{MD} are red in Figure 4; those from minima in the same group as a_{MD} are blue. The minima in the blue group span a wide range of PE, and some of the low-energy structures feature significant native contacts: up to 65% of the set of contacts from ref 11. The minimum in the denatured group with the largest number of native contacts has the C-terminal hairpin already formed, and the minimum of lowest energy in that group has the C-terminal hairpin fully formed and one of the native hydrogen bonds of the N-terminal hairpin present but fewer than half of the native side chain contacts. A few of the minima in the red group lie at moderate energies because one of the terminal strands is detached from the remaining hairpin. This broad distribution of the minima in the end point groups across the energy landscape is consistent with the observed two-state nature of the landscape for the grouped database (section 3.1). Alternative grouping schemes based on structural criteria are of course possible but would not guarantee that the end point groups are individually in local equilibrium, as required for two-state kinetics.

A PE disconnectivity graph plotted using only the stationary points involved in the 250 fastest $b_{MD} \leftarrow a_{MD}$ paths is presented in Figure 5. The branches leading to b_{MD} and a_{MD} are labeled, and the intervening structures from path 47 illustrated in Figure 3 are also shown here with the corre-

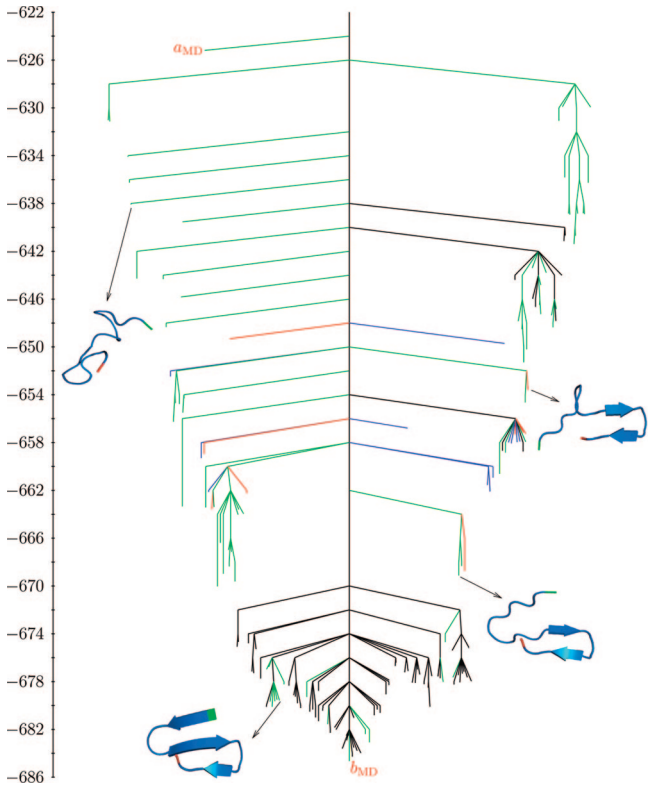

Figure 5. Potential energy (kcal mol^{-1}) disconnectivity graph constructed using only the stationary points on the 250 fastest-folding paths in the ungrouped database. The denatured and folded end points are labeled a_{MD} and b_{MD}, respectively, as in the text. To illustrate the differences within the set of paths, the branches that terminate at minima on the fastest (path 47) and slowest (path 241) of the 250 paths are colored: the green branches lead to minima present on both paths, the red ones to the minima encountered on path 47 but not 241, and the blue ones to the minima encountered on path 241 but not 47. The intervening configurations from path 47 that are illustrated in Figure 3 are also shown here with the corresponding nodes on the graph indicated. The 250 paths all exhibit a common overall mechanism, with only minor structural variations focused around the docking of the N-terminal hairpin or the region between the formation of the two hairpins.

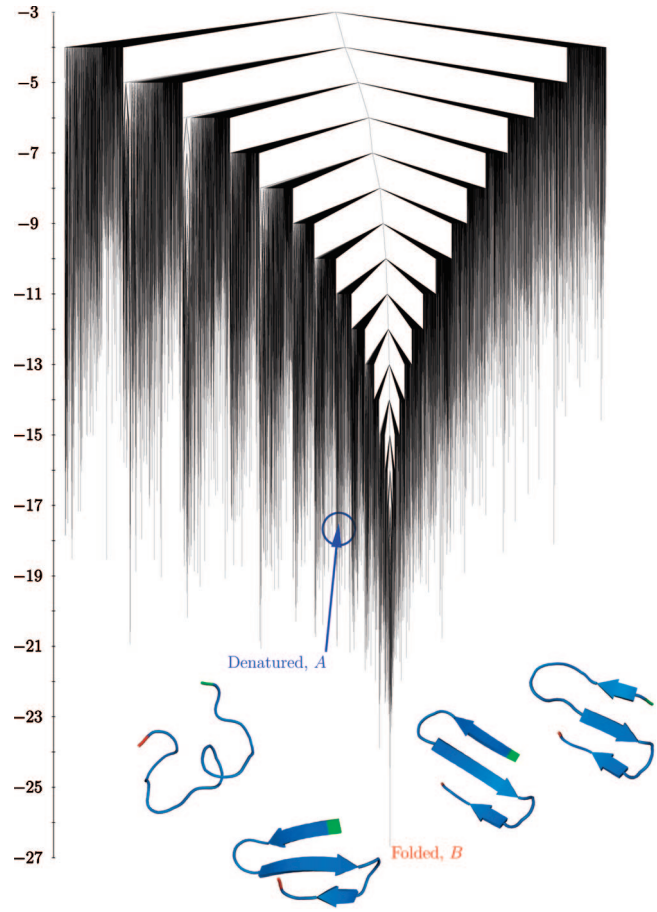

Figure 6. Free energy (kcal mol^{-1}) disconnectivity graph calculated at 298 K for the folding database, grouped using a self-consistent scheme in which two groups are merged if they can interconvert without encountering a free energy barrier higher than a chosen threshold value, as described in the text. The threshold value used here is 5 kcal mol^{-1}. The denatured and folded groups are labeled together with the structures of the defining members (a_{MD} and b_{MD}, respectively) of those groups. Two minima belonging to the lowest two groups above the folded state also are shown. The pattern of the graph is qualitatively similar to that of the ungrouped potential energy graph (Figure 4), indicating that the grouping scheme was chosen appropriately.

sponding nodes (local minima) on the graph indicated. The colored branches terminate at the minima on paths 47 and 241 (the fastest and slowest, respectively, from the set of 250 in terms of unbranched rate constants): the green branches lead to the minima present on both the paths, the red ones to the minima encountered on path 47 but not 241, and the blue ones to the minima encountered on path 241 but not 47. The minima in green are visited in the same order on the two paths. The graph can be divided vertically into three sections: −622 to −652, −652 to −670, and −670 to −686 kcal mol^{-1}, where all 250 paths visit minima only from the highest-energy section, then minima solely from the middle section, and finally minima from just the lowest part of the graph. Most of the red and blue branches lie in the middle section, which corresponds to the region between the formation of the two hairpins. Many of the branches are green, and most of the black branches (indicating minima encountered on neither path 47 nor 241) occur in the lowest-energy section, where the N-terminal strand docks with the existing C-terminal hairpin. We therefore conclude that the 250 paths all exhibit a common overall mechanism, with only minor structural variations. Although the energy does not decrease monotonically along these 250 paths, overall the graph exhibits the palm-tree pattern associated with efficient structure-seeking behavior.[73]

The coarse-grained FE graph representing the grouped database also considered in section 3.1 is presented in Figure 6. The branches for the folded group, containing b_{MD}, and the denatured group, containing a_{MD}, are labeled appropriately; the folded ensemble lies lowest of all groups. As for the full PE graph, only the lower half of the kinetically accessible landscape is shown. The graph in Figure 6 could be described as a funnelled energy landscape, with moderate downhill barriers between most groups and the well-defined global minimum. The pattern is qualitatively similar to that of the full PE graph (Figure 4), which suggests that the grouping threshold has been chosen appropriately. The local minima of lowest free energy in the groups of second- and third-lowest energy are native-like but vary from b_{MD} in the turn regions, as discussed previously in relation to the full, ungrouped PE graph. The lowest minima in these groups are illustrated in Figure 6.

4. Conclusion

In this work, we used the discrete path sampling method to investigate the folding of Beta3s, a three-stranded antiparallel

β-sheet peptide, described by an empirical potential and implicit solvation model. Within a coarse-grained formulation of the kinetics, involving transitions between groups of local minima, the folding mean first-passage time was ~300 ns at 298 K, in reasonable agreement with other simulated values obtained at similar temperatures[11,13,14] and consistent with the experimental upper bound of 4000 ns.[7] The grouping was achieved with a scheme designed to self-consistently combine sets of minima in local equilibrium on a chosen time scale. The folding mechanism exhibited by the most significant individual discrete paths on the potential energy surface involves early formation of the C-terminal hairpin followed by docking of the N-terminal strand, in agreement with previous simulations.[9,11,13,14]

The organization of the potential energy and free energy landscapes, as visualized using disconnectivity graphs, is consistent with efficient folding. In particular, the low-lying minima can be reached without the intervention of any high barriers. The thermodynamically favored region of configuration space is therefore kinetically accessible over a wide range of temperatures, as expected for a system with good structure-seeking properties.[30,76,77] This organization is also likely to result in a scale-free network if the low-lying minima are highly connected.[78] The energy landscapes of other peptides with structure-seeking characteristics, characterized in previous work,[37,74,75,79–82] may therefore share this property with Beta3s.[22]

In future work it would be interesting to compare the mechanisms and folding times for databases constructed with alternative potentials, particularly to assess the effect of the solvation model employed. The effect of our choice of initial end point structures should be mitigated by the grouping procedure, which expands the end point states to include minima both on and off significant folding pathways, thus incorporating a wider region of configuration space. The extent of this expansion relies on both the particular grouping scheme applied and the efficiency with which the discrete path sampling approach explores the energy landscape.

Acknowledgment. J.M.C. is grateful to the EPSRC for financial support. We gratefully acknowledge C. S. Whittleston for his help in preparing some of the figures.

References and Notes

(1) Garcia, A. E.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13898.
(2) Banavar, J. R.; Hoang, T. X.; Maritan, A.; Seno, F.; Trovato, A. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2004**, *70*, 041905.
(3) Bursulaya, B. D.; Brooks, C. L., III *J. Am. Chem. Soc.* **1999**, *121*, 9947.
(4) Wales, D. J. *Mol. Phys.* **2002**, *100*, 3285.
(5) Wales, D. J. *Mol. Phys.* **2004**, *102*, 891.
(6) Wales, D. J. *Int. Rev. Phys. Chem.* **2006**, *25*, 237.
(7) de Alba, E.; Santoro, J.; Rico, M.; Jiménez, M. A. *Protein Sci.* **1999**, *8*, 854.
(8) Wang, H. W.; Sung, S. S. *J. Am. Chem. Soc.* **2000**, *122*, 1999.
(9) Mohanty, S.; Hansmann, U. H. E. *Biophys. J.* **2006**, *91*, 3573.
(10) Mohanty, S.; Hansmann, U. H. E. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2007**, *76*, 12901.
(11) Ferrara, P.; Caflisch, A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10780.
(12) Ferrara, P.; Caflisch, A. *J. Mol. Biol.* **2001**, *306*, 837.
(13) Cavalli, A.; Ferrara, P.; Caflisch, A. *Proteins: Struct., Funct., Genet.* **2002**, *47*, 305.
(14) Cavalli, A.; Haberthür, U.; Paci, E.; Caflisch, A. *Protein Sci.* **2003**, *12*, 1801.
(15) Rao, F.; Caflisch, A. *J. Chem. Phys.* **2003**, *119*, 4035.
(16) Rao, F.; Caflisch, A. *J. Mol. Biol.* **2004**, *342*, 299.
(17) Settanni, G.; Rao, F.; Caflisch, A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 628.
(18) Gfeller, D.; de Lachapelle, D. M.; De Los Rios, P.; Caldarelli, G.; Rao, F. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2007**, *76*, 26113.
(19) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D. ; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(20) Ferrara, P.; Apostolakis, J.; Caflisch, A. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 24.
(21) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.
(22) Barabási, A.-L.; Albert, R. *Science (Washington, DC, U.S.)* **1999**, *286*, 509.
(23) Fersht, A. R. *Structure and Mechanism in Protein Science*; W. H. Freeman and Co.: New York, 1999.
(24) Davis, R.; Dobson, C. M.; Vendruscolo, M. *J. Chem. Phys.* **2002**, *117*, 9510.
(25) Rao, F.; Settanni, G.; Guarnera, E.; Caflisch, A. *J. Chem. Phys.* **2005**, *122*, 184901.
(26) Lazaridis, T.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 133.
(27) Bloom, J. D. Computer Simulations of Protein Aggregation. M.S. Thesis, University of Cambridge, 2002.
(28) http://www-wales.ch.cam.ac.uk/~sat39/charmm/rotamers/rotamers_in_CHARMM19.pdf.
(29) Steinbach, P. J. *Proteins: Struct., Funct., Bioinf.* **2004**, *57*, 665.
(30) Wales, D. J. *Energy Landscapes*; Cambridge University Press: Cambridge, 2003.
(31) Murrell, J. N.; Laidler, K. J. *J. Chem. Soc., Faraday Trans.* **1968**, *64*, 371.
(32) Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. *J. Chem. Phys.* **1998**, *108*, 1964.
(33) Dellago, C.; Bolhuis, P. G.; Chandler, D. *J. Chem. Phys.* **1999**, *110*, 6617.
(34) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291.
(35) Strodel, B.; Whittleston, C. S.; Wales, D. J. *J. Am. Chem. Soc.* **2007**, *129*, 16005.
(36) Trygubenko, S. A.; Wales, D. J. *J. Mol. Phys.* **2006**, *104*, 1497.
(37) Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2004**, *121*, 1080.
(38) Dijkstra, E. W. *Numerische Math.* **1959**, *1*, 269.
(39) Kearsley, S. K. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1989**, *45*, 208.
(40) Carr, J. M.; Wales, D. J. *J. Chem. Phys.* **2005**, *123*, 234901.
(41) Becker, O. M.; Karplus, M. *J. Chem. Phys.* **1997**, *106*, 1495.
(42) Godzik, A.; Koliński, A.; Skolnick, J. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 397.
(43) Bryngelson, J. D.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7524.
(44) Carr, J. M.; Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2005**, *122*, 234903.
(45) Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2004**, *120*, 2082.
(46) Henkelman, G.; Jónsson, H. *J. Chem. Phys.* **1999**, *111*, 7010.
(47) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. *J. Chem. Phys.* **2000**, *113*, 9901.
(48) Henkelman, G.; Jónsson, H. *J. Chem. Phys.* **2000**, *113*, 9978.
(49) Munro, L. J.; Wales, D. J. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 3969.
(50) Nocedal, J. *Math. Comput.* **1980**, *35*, 773.
(51) Liu, D.; Nocedal, J. *Math. Program. B* **1989**, *45*, 503.
(52) Hoare, M. R.; McInnes, J. *J. Chem. Soc., Faraday Discuss.* **1976**, *61*, 12.
(53) Hoare, M. R. *Adv. Chem. Phys.* **1979**, *40*, 49.
(54) Mezey, P. G. *Theor. Chim. Acta* **1981**, *58*, 309.
(55) Pelzer, H.; Wigner, E. *Z. Phys. Chem. B* **1932**, *15*, 445.
(56) Eyring, H. *Chem. Rev.* **1935**, *17*, 65.
(57) Evans, M. G.; Polanyi, M. *Trans. Faraday Soc.* **1935**, *31*, 875.
(58) Eyring, H. *J. Chem. Phys.* **1935**, *3*, 107.
(59) Evans, M. G.; Polanyi, M. *Trans. Faraday Soc.* **1937**, *33*, 448.
(60) Wynne-Jones, W. F. K.; Eyring, H. *J. Chem. Phys.* **1935**, *3*, 492.
(61) Anderson, J. B. *J. Chem. Phys.* **1973**, *58*, 4684.
(62) Bennett, C. H. In *Algorithms for Chemical Computations*; Christofferson, R. E., Ed.; American Chemical Society: Washington, DC, 1977; pp 63−97.
(63) Chandler, D. *J. Chem. Phys.* **1978**, *68*, 2959.
(64) Shaffer, J. S.; Chakraborty, A. K. *Macromolecules* **1993**, *26*, 1120.
(65) Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2006**, *124*, 234110.
(66) van Kampen, N. G. *Stochastic Processes in Physics and Chemistry*; North-Holland: Amsterdam, 1981.
(67) Bortz, A. B.; Kalos, M. H.; Lebowitz, J. L. *J. Comput. Phys.* **1975**, *17*, 10.
(68) Fichthorn, K. A.; Weinberg, W. H. *J. Chem. Phys.* **1991**, *95*, 1090.
(69) Kunz, R. E. *Dynamics of First-Order Phase Transitions*; Deutsch; Thun 1995.
(70) Carr J. M.; Wales, D. J. *Latest Advances in Atomic Cluster Collisions: Structure and Dynamics from the Nuclear to the Biological Scale*; Imperial College Press: London, 2008.

(71) Jiménez, V. M.; Marzal, A. In *Algorithm Engineering: 3rd International Workshop, WAE'99, London, UK, July 1999*; Vitter, J. S., Zaroliagis, C. D., Eds.; Springer: Berlin, 1999; Vol. 1668, pp 15−29.
(72) Evans, D. A. Energy Landscapes and Dynamics of Model Peptides. Ph.D. Thesis, University of Cambridge, 2003.
(73) Wales, D. J.; Miller, M. A.; Walsh, T. R. *Nature (London, U.K.)* **1998**, *394*, 758.
(74) Krivov, S. V.; Karplus, M. *J. Chem. Phys.* **2002**, *117*, 10894.
(75) Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2003**, *118*, 3891.
(76) Wales, D. J. *Philos. Trans., R. Soc. London, Ser. A* **2005**, *363*, 357.
(77) Wales, D. J.; Bogdan, T. V. *J. Phys. Chem. B* **2006**, *110*, 20765.
(78) Doye, J. P. K. *Phys. Rev. Lett.* **2002**, *88*, 238701.
(79) Mortenson, P. N.; Wales, D. J. *J. Chem. Phys.* **2001**, *114*, 6443.
(80) Mortenson, P. N.; Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2002**, *117*, 1363.
(81) Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2003**, *119*, 9947.
(82) Krivov, S. V.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14766.

JP801777P