One-Dimensional Free-Energy Profiles of Complex Systems: Progress Variables that Preserve the Barriers

Sergei V. Krivov[†] **and Martin Karplus***[,†,‡]

Laboratoire de Chimie Biophysique, ISIS, Université Louis Pasteur, 67000 Strasbourg, France, and Department of Chemistry & Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

Received: January 3, 2006; In Final Form: April 26, 2006

We show that the balanced minimum-cut procedure introduced in PNAS **2004**, *101*, 14766 can be reinterpreted as a method for solving the constrained optimization problem of finding the minimum cut among the cuts with a particular value of an additive function of the nodes on either side of the cut. Such an additive function (e.g., the partition function of the reactant region) can be used as a progress coordinate to determine a one-dimensional profile (FEP) of the free-energy surface of the protein-folding reaction as well as other complex reactions. The algorithm is based on the network (obtained from an equilibrium molecular dynamics simulation) that represents the calculated reaction behavior. The resulting FEP gives the exact values of the free energy as a function of the progress coordinate; i.e., at each value of the progress coordinate, the profile is obtained from the surface with the minimal partition function among the surfaces that divide the full free-energy surface between two chosen end points. In many cases, the balanced minimum-cut procedure gives results for only a limited set of points. An approximate method based on p_{fold} is shown to provide the profile for a more complete set of values of the progress coordinate. Applications of the approach to model problems and to realistic systems (β-hairpin of protein G, LJ_{38} cluster) are presented.

1. Introduction

Considerable progress has been made in recent years in both experimental and theoretical studies of protein folding.[1,2] An understanding of this fundamental reaction in biology requires a knowledge of the free-energy surface governing the motion of the polypeptide chain as it progresses from the denatured to the native state. Recently, we have presented a method for determining the unprojected free-energy surface.[3,4] It is represented by a disconnectivity graph calculated from an equilibrium-folding trajectory with the mincut or balanced mincut procedure. The essential idea of the method is to group the coordinate sets into free-energy minima, not according to the standard geometric characteristics but rather according to the equilibrium dynamics; i.e., the trajectory is used to determine the populations of the states, which provide the relative free energies, and the rates of the transitions between the states, which provide the free-energy barriers. Application of the method to the well-studied β-hairpin of protein G[4] unmasked the fact that the free-energy surface has multiple low free-energy basins in the denatured state, in addition to the native basin. This contrasts with free-energy surfaces projected on one or two geometric degrees of freedom (such as the number of native hydrogen bonds, number of native contacts, number of native dihedral angles, rmsd from the native state, and/or the radius of gyration) are relatively simple and have a single or a few low free-energy barriers (a few kT or less). Thus, to obtain a projection with a more accurate rendition of the essential aspects of the free-energy surface, different progress coordinates are required. Projected free-energy surfaces are most useful if they preserve the barriers and minima in the order that they are met during folding/unfolding events. Here we introduce new progress coordinates that have some of the desired properties in that their use yields a projected surface that preserves the free-energy barriers on the surface; given the barriers, the minima can be determined. They are additive progress coordinates, by which we mean that they can be expressed as an integral (or sum in the case of a discrete network) of point properties. We use as progress coordinates the number of points in a given region or the partition function of that region; other choices are possible. Given such a progress coordinate, the problem of determining the free-energy profile from the reactant state, through the transition state region to the product state, can be translated into the problem of optimizing (maximizing) the free energy as a function of that coordinate. The proposed algorithm uses as input a network that accurately represents the kinetics of the system. The network is made up of nodes obtained by clustering the points on an equilibrium molecular dynamics trajectory with a sufficient number of transitions between the states of interest.

The methodology is described first, followed by application to one (1D) and two-dimensional (2D) model systems and to the β-hairpin and LJ_{38} cluster.

2. Methodology

2.1. (a) Equilibrium Kinetic Network. An equilibrium kinetic network (EKN)[3] represents the kinetics of the system as a capacitated undirected graph. The EKN can be constructed analytically, as for the 1D and 2D systems described below, or obtained from a sufficiently long molecular dynamics or MC simulation followed by clustering.[4] The edge capacity c_{ij} from node j to node i in the network is proportional to the equilibrium number of direct transitions n_{ij} made by the system from state j to state i. Detailed balance implies that at equilibrium, $n_{ij} =$

* To whom correspondence should be addressed. E-mail: marci@tammy.harvard.edu.
† Université Louis Pasteur.
‡ Harvard University.

10.1021/jp060039b CCC: $33.50
Published on Web 06/03/2006

n_{ji}; hence, $c_{ij} = c_{ji}$ and the network is undirected. The system kinetics can be reproduced on the basis of the EKN by performing MC simulation to solve the master equation with the transition probabilities $p_{ij} = c_{ij}/\sum_i c_{ij}$.

For a system with a relatively small discrete configuration space, all possible configurations and the local moves connecting them can be enumerated. On the basis of an MC scheme and the set of local moves on the potential energy surface (PES), one can determine the transition probability, p_{ji}, between the configuration. The equilibrium probabilities (P_i^e) of each point (i) can then be found through $P_i^e = \sum_j p_{ij} P_j^e$ or from the Boltzmann distribution. Finally, the capacities of the edges of the EKN are obtained as $c_{ji} = n_{ji} = p_{ji} P_i^e$, $c_{ii} = n_{ii} = p_{ii} P_i^e$, $p_{ii} = 1 - \sum_j p_{ji}$. If the PES is used as a starting point, given the set of minima and saddles, their partition functions (Z_i and Z_{ij}, respectively) can be calculated by employing the harmonic approximation.[3] The capacities of the network are then defined as $c_{ij} = Z_{ij}$.[3] If an equilibrium (MC or MD) trajectory is used, then $c_{ij} = (n_{ij} + n_{ji})/2$ and $Z_i = \sum_i n_{ij}$, where n_{ij} is the number of transition from cluster j to cluster i. The clusters, for example, can be obtained using an all-atom rmsd cutoff (or a backbone rmsd).[4] Each subsequent structure is compared with the set of clusters found so far; if the rmsd of the structure from the first configuration of all of the known clusters exceeds a given threshold, it is taken as a new cluster. Usually, in systems with a large configuration space (e.g., as in the protein-folding problem), there is a high likelihood of long sequences of clusters, each of them visited only once. In calculating the mincut values, such sequences can be substituted by just two clusters (i.e., the first and the last one in the sequence), greatly reducing the size of the problem, because a cut anywhere along the sequence always gives 1.

For each of the nodes of the EKN the partition function is $Z_i = \sum_i c_{ij}$ (if the harmonic approximation is used, the Z_i are defined through it; see above). If the nodes of the network are partitioned into two groups A and B by a cut, then the following quantities are defined: $Z_A = \sum_{i \in A} Z_i$, $Z_B = \sum_{i \in B} Z_i$, $Z_{AB} = \sum_{i \in A, j \in B} c_{ij}$, $|A| = \sum_{i \in A} 1$, $|B| = \sum_{i \in B} 1$. The free energy of the barrier is $-kT \ln(Z_{AB})$.

2.2. Bmincut Algorithm. The transition state (TS) between two points is defined as the surface with the minimal partition function that divides the FES between the points;[3,4] i.e., the partition function of the barrier is equal to the partition function of the cutting surface. The surface is isomorphic with the definition of the minimum cut between two nodes in the network.[3] To find the TS, one represents the FES as a network[3,4] and applies the minimum cut−maximum flow Ford−Fulkerson algorithm.[5]

To build the reduced free-energy profile (FEP), a parameter (progress coordinate) is introduced and the free-energy barriers for different values of the parameter are calculated; i.e., we have to find minimum cuts (TS) at each value of the progress coordinate. The minimum-cut problem thus becomes a constrained optimization problem. For additive progress coordinates, a small modification of the minimum-cut algorithm, called the balanced minimum-cut procedure, BMC,[4] solves the problem. The BMC was introduced to find a TS (a free-energy barrier) between two basins, one of which does not have a representative node (e.g., the entropic basin in ref 4). For this case, BMC introduces an "extra" node that is connected to all nodes in the graph with the same small capacity c and serves as a representative node for the basin without one. The progress coordinate is defined as "additive" if it can be written as the integral (or sum for the corresponding network) of point properties on the FES. As progress coordinates, we use the partition function of the region A (Z_A) or the number of points in the region A ($|A|$).

The FEP along the progress coordinate Z_A between two nodes A and B of the EKN is computed by the balanced mincut procedure[4] as follows. Node B is connected to each node (i) in the network by an edge with capacity equal $\lambda\omega_i$, where ω_i is the weight of node i; to use Z_A as the progress coordinate ($\omega_A = Z_A$) we set $\omega_i = Z_i$; for $\omega_A = |A|$ we set $\omega_i = 1$. The minimum cut between nodes A and B separates the nodes of the network into two groups (A and B), with the cut value $Z'_{AB} = Z_{AB} + \sum_{i \in A} \lambda\omega_i = Z_{AB} + \lambda\omega_A$, where Z_{AB} is the cut value in the original unmodified network and $\omega_A = \sum_{i \in A} \omega_i$ is the value of progress coordinate. The minimum cut finds the minimum of $Z_{AB} + \lambda\omega_A$; this corresponds to the solution of the constrained optimization problem "*min* Z_{AB} with $\omega_A = const$" by use of the Lagrange multiplier λ. Considering various λ's from 0 to some large number (for which the cut is just around the node A), one obtains the points of the FEP along the progress coordinate.

To efficiently explore the FEP in the given interval of λ, we recursively divide each subinterval into two halves until the accuracy requirements are met. We used the following accuracy indicators $\Delta\lambda < 10^{-5}$, $\Delta Z_A/Z < 10^{-3}$ (or $\Delta|A|/|A| < 10^{-3}$).

2.3. Analysis On the Basis of Reaction Probability (bpfold). The folding probability[6] has been used to separate the configuration space into two basins, assuming that the points with $p_{\mathrm{fold}} > 0.5$ and $p_{\mathrm{fold}} < 0.5$ belong to different basins; we use the terminology of protein folding here, although the approach is general and can be applied to any reaction.[7] To obtain the FEP with p_{fold}, we can again use a Lagrange multiplier. We call such a procedure balanced pfold or "bpfold".

Given the EKN and the two nodes A and B, $p_{\mathrm{fold}} = p_i$ is found as the solution of the equation $p_j = \sum_i p_{ij} p_i$, where $p_{ij} = c_{ij}/\sum_i c_{ij}$, with boundary condition $p_A = 1$ (we consider node A to be the native node) and $p_B = 0$. The equation can solved by iterative multiplication of the vector p_i by the matrix p_{ij}. To increase the speed of convergence, all diagonal entries can be reset to zero; i.e., $p_{ii} = 0$.[8] However, we found ITPACK,[9] a large sparse linear equation system iterative solver, to be more efficient then iterative multiplication.

To determine the FEP, a Lagrange multiplier is introduced similar to that used for the balanced minimum-cut method. Every node (i) in the EKN is connected with node B in the network by an edge with capacity equal to $\lambda\omega_i$, where λ is the Lagrange multiplier. To use Z_A as the progress coordinate ($\omega_A = Z_A$) we set $\omega_i = Z_i$; for $\omega_A = |A|$ we set $\omega_i = 1$. This gives different set of c_{ij} (respectively p_{ij} and p_i) compared to the original EKN ($\lambda = 0$) for each λ. The partition $p_i > 0.5$ and $p_i < 0.5$ as a function of λ yields the FEP.

2.4. Pfoldf. The procedure of finding the FEP can be simplified by introducing an approximation in the determination of p_{fold}. We assume that for small changes of the Lagrange multiplier λ in the bpfold procedure, the relative order of p_i in p_{fold} for the points with p_i around 0.5 does not change (i.e., $p'_i > p'_j$ for λ' if $p_i > p_j$ for λ), whereas the p_i themselves are changing. Then there is no need to recalculate the p_{fold} values for each new λ; instead, the p_{fold} values are calculated only for one value of λ (not necessarily $\lambda = 0$; see the β-hairpin example below).

Given the EKN and the two nodes A and B, p_{fold} is calculated as described above for a single λ. For given p_{b} in the range from 0 to 1, the nodes with $p_{\mathrm{fold}} < p_{\mathrm{b}}$ belong to the basin B and with $p_{\mathrm{fold}} > p_{\mathrm{b}}$ to the basin A. Considering various p_{b}, one

obtains the points on the FEP. By ordering the nodes according to their p_{fold} value, the computations are performed in an efficient manner.

In the pfoldf procedure, p_{fold} can be considered as the progress variable. This is similar in spirit but differs from other algorithms proposed for calculating the FEP via pfold. In ref 6, for example, the FEP is projected on p_{fold} as a reaction coordinate in the conventional way; i.e., a histogram is obtained by binning the p_{fold} reaction coordinate values obtained from an equilibrium simulation. Because p_{fold} changes in a highly nonlinear manner (i.e., it changes for almost 0 to almost 1 near the top of the barrier, and stays almost constant on the two sides) the straightforward projection reveals mainly the FEP of the top of the barrier. To overcome this, a "reverse transformation" from p_{fold} to a one-dimensional (1D) reaction coordinate has been applied.[10] The 1D FEP is constructed such that the FEP as a function of p_{fold} matches that of the original multidimensional FES. If one assumes that the motion on the FEP occurs with constant (pre-exponential) diffusion coefficient, this transformation can be evaluated analytically.[10] The "pfoldf" algorithm described here uses a different approach in that no histogram is evaluated. The nodes of the network are sorted with respect to their value of p_{fold}. The value of the FEP for a particular p is equal to $-kT \ln(n)$, where n is the total number of transitions between points with $p_{fold} > p$ and points with $p_{fold} < p$ (i.e., number of transitions across the surface $p_{fold} = p$).

Also, in the cited works, e.g.,[6] $p_{fold} = p_i$ for a configuration (i) was calculated based on a limited number of trials (MC−MD trajectories). The projection results are very sensitive to the values of the progress coordinate, and the statistical estimation of p_{fold} from a reasonable number of trajectories can introduce considerable noise. Consequently, the procedure based on such a statistical p_{fold} can produce a poorly resolved FEP. Here, we eliminate the noise by calculating the exact p_{fold} values for the EKN. However, whereas the cited works (e.g., ref 6) deal with the actual FES, here we use the EKN, which is an approximation to the original FES. We obtain essentially exact results for the EKN because the problem of calculating the kinetic properties of the network is much simpler than that of calculations based on the actual FES.

To apply the bmincut, bpfold, and pfoldf methods, one has to specify two nodes. As an initial guess, when there is no *a priori* information about the system, one can take the most visited node of the network (native state) and an "extra" node (a node that does not belong to the network[4]). When the profile is calculated between a node (A) in the network and an extra node, we call the resulting projection the "unfolding" FEP of basin A (because there is no biasing to any other node in the network and the system can go everywhere on the whole FES). This is in contrast to the case when the profile is calculated between two nodes A and B of the network, which determines the FEP between A and B. Our experience shows that to get an overall impression concerning the FES, one can calculate the unfolding profile of the "native" node with the pfoldf procedure at some small $\lambda \approx 0.01$.

3. Results

3.1. Application to Model Systems. To introduce the way the method works, we consider a simple 1D example. Although the constrained optimization problem is solved trivially in the 1D case, because every choice of Z_A determines a partition, the balanced minimum-cut approach (optimization with a Lagrange multiplier) can be illustrated by this example. In Figure 1, the free-energy surface (FES) obtained from the bmincut procedure

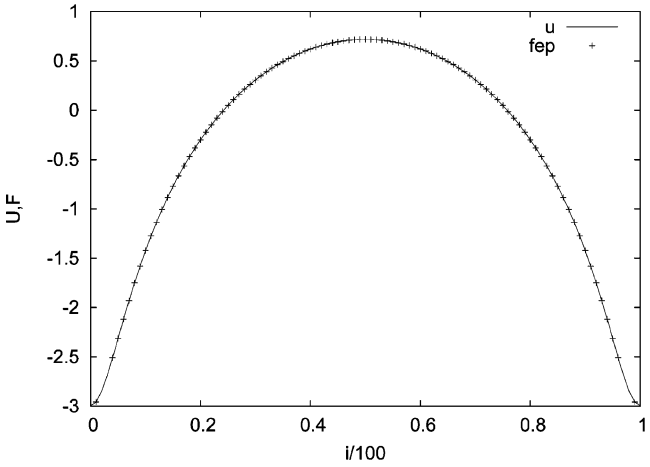

Figure 1. PES $u(x) = \log(1 - \cos(2\pi x) + 0.05)$ (line) and the FEP computed with bmincut (crosses).

is shown with the identical potential energy surface (PES). Although the bmincut procedure here produces the entire FEP, the more usual situation is that it does not. In Figure 2a we show a PES for which this is the case and outline the method. We discretize the FES into 101 points x_i between the nodes A ($x = 0$) and B ($x = 1$) to obtain the network. The c_{ij} that define the network correspond to the MC transition probabilities at $kT = 1$. Below, we consider $\omega_A = |A|$ as the progress coordinate, i.e., $\omega_i = 1$. Now we connect every node in the network with node B with capacities ω_i (Figure 2b). In this simple network, the cut is uniquely defined by its coordinate x (Figure 2b). The cut value of the modified network is $Z'_{AB} = Z_{AB} + \sum_{i \in A} \lambda\omega_i = Z_{AB} + \lambda\omega_A$, where λ is a Lagrange multiplier. The quantity Z_{AB} is the cut value of the original network, and it is equal to $c_{i\ i+1}$ for $x_i < x < x_{i+1}$. The value of the progress coordinate $\omega_A = |A|$ ($\omega_i = 1$) for the cut between x_i and x_{i+1} is equal to i. The value of the putative cut (shown by vertical line, Figure 2b) in the modified "balanced" network is equal to the sum of the cut in the original network plus the sum of the capacities of the links that connect nodes of basin A (with $x_i < 0.425$, where 0.425 is the cut position) and node B. The values of the cut versus cut position are shown for certain λ in Figure 2c. For $\lambda = 0$, the mincut corresponds to the top of the barrier in the original network ($x = 0.5$), and for $\lambda > 0$, the position of the mincut is shifted closer to node A, i.e., to smaller values of the progress coordinate as compared to that of the top of the barrier. Calculating the balanced minimum cuts at different values of parameter λ, one obtains a list of pairs of values (Z_{AB},$|A|$), which represents the FEP of the system. Figure 2 shows the final FEP ($-kT \log(Z_{AB})$ as a function of $|A|$) for this model system. As one can note, the BMC procedure reproduced the original $u(x)$ in a limited range of the x progress coordinate. To continue the FEP for $x > 0.5$, one can exchange the roles played by nodes A and B.

The cut profile for $\lambda = 0.04$ (Figure 2c) illustrates the problem that can arise in attempts to calculate the entire FEP profile. The cut profile for $\lambda = 0.04$ has two local minima ($x = 0$ and $x = 0.2$) of which only the latter is of interest. Because the minimum-cut procedure always finds the lowest minimum (here, that at $x = 0$), the algorithm works only when the local minimum of interest is also the global minimum. In the present case, the minimum at $x = 0$ is the global minimum when $\lambda < 0.0345$. With this restriction on λ, only the region $|A|/100 > 0.2$ can be sampled.

3.2. Comparison of Z_A and $|A|$ Progress Variables. The difference between two progress variables, namely $|A|$ (number

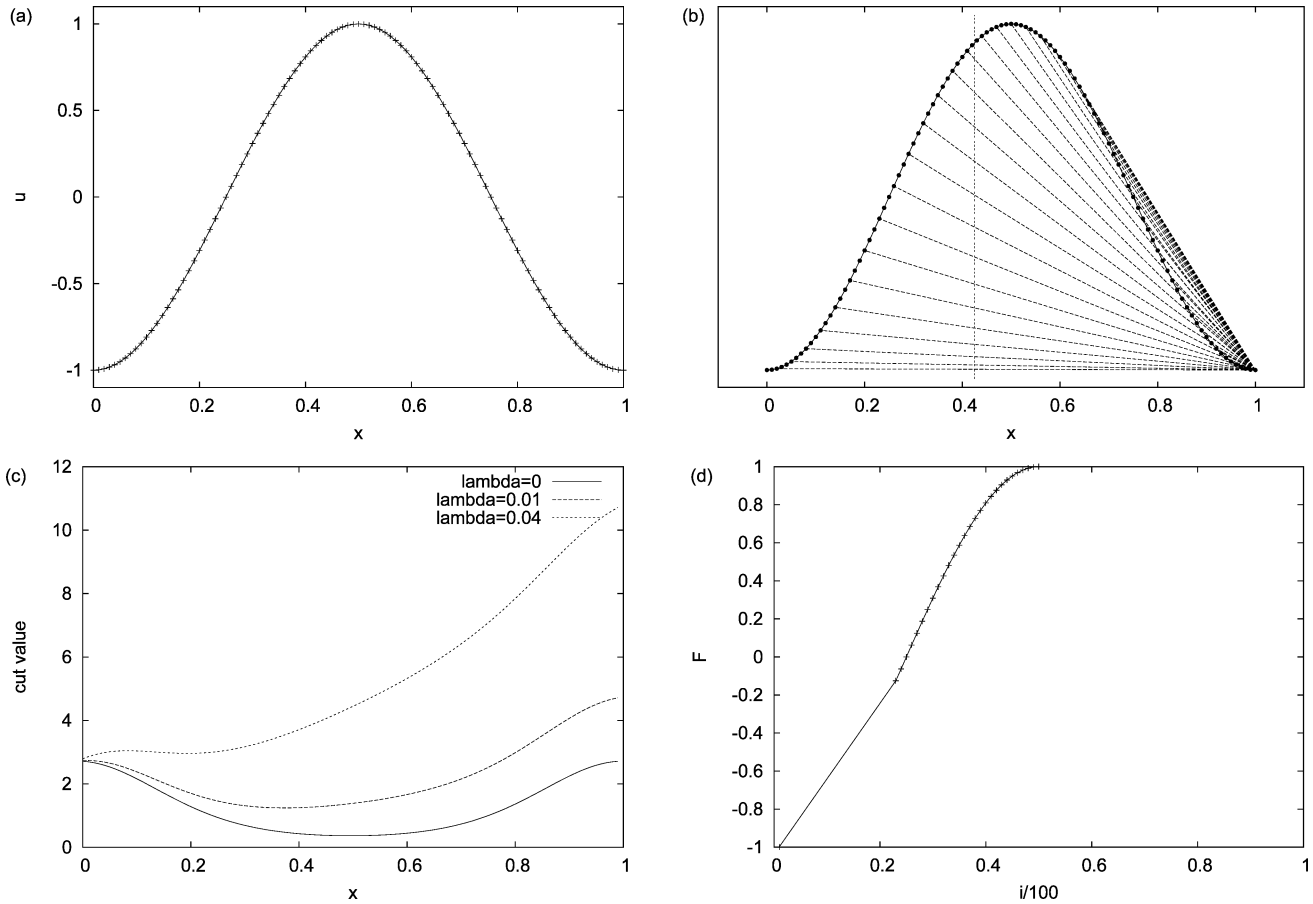

Figure 2. (a) Model PES $u(x) = -\cos(2\pi x)$. (b) Equilibrium kinetic network (capacities are not shown); dashed lines indicate the links with capacity λ (every third one is shown) that connect node B ($x = 1$) with every node of the network, as required by the balanced mincut procedure. The vertical line denotes a putative cut; ω_a for this cut is 43. (c) Cut values for different Lagrange multiplier λ. (d) FEP obtained with the balanced mincut procedure.

of configurations in basin A) and Z_A (partition function of basin A) can be illustrated with a two-pathway PES (Figure 3a); in the trivial 1D example above, Z_A and $|A|$ give identical results. The PES consists of two 1D pathways, which are joined at the two ends. Each of the 1D pathways can be considered as an idealization of a 2D pathway with infinite square well cross-section in the direction perpendicular to x.

The cuts of the PES from which the FEP is obtained with $|A|$ as the progress variable are shown in Figure 3b, whereas that with Z_A is shown in Figure 3c. As is evident, the cuts behave differently. For $|A|$ as the progress variable, the cuts propagate more readily along the lower energy path, whereas for Z_A they do so along the higher energy path; i.e., for $|A|$ as the progress variable each cut has $x_2 > x_1$, where x_1 and x_2 are the coordinates of the cut on the first and second pathway, respectively. However, both pathways have the highest cut (top of the barrier) with $x_2 = x_1 = 0.5$ as can be seen in the figure. The difference is due to the fact that the cuts are optimized while fixing different properties ($|A|$ in Figure 3b and Z_A in Figure 3c). It is similar in spirit to the differences between p_{fold} and the mean folding time for the points on the FES. The p_{fold} value does not depend on the probability of remaining in the node p_{ii}; it is determined by the transition probabilities ($p_{ij}, i \neq j$). By contrast, the mean folding time depends critically on p_{ii}. Similarly, the $|A|$ progress coordinate does not depend on Z_i because it simply counts points; by contrast, the Z_A reaction coordinate counts Z_i. Usually, the major contribution to $Z_i = \sum_j Z_{ij}$ comes from Z_{ii}.

Figure 4 shows the corresponding FEPs. Every cut shown on Figure 3 divides the paths (surface) into two parts, with the left part belonging to basin A and the right part to basin B. For each cut, the values of two progress variables ($|A|$ and Z_A) are calculated. The partition function of the cut, Z_{AB} (value of the cut), is equal to the sum of the link capacities the cut goes through. All cuts are plotted along $|A|$ in Figure 4a and Z_A in Figure 4b as progress variables. To obtain Figure 4a for every cut shown on Figure 3b, c, we calculate the progress variable $|A|$ by counting the points on the left side of the cut and plotting the value of the cut ($F = -\ln Z_{AB}$) versus the $|A|$ value. To obtain Figure 4b for every cut on Figs 3b and 3c, we calculate Z_A by calculating the total partition function of the points on the left of the cut. Figure 4 shows that each FEP is more optimal (i.e. higher) than the other along its own progress variable.

The progress variables yield similar but not identical behavior for this case. The small difference between the two FEPs on Figure 4 indicates that the FEP is not sensitive to such modifications of the progress variable. The question as to which progress variable provides a "better" 1D (one pathway) representation of the FES for this model system requires direct investigation of the master equation; this will be described in the future. The above suggests that it is possible that different progress variables are required to better preserve different quantities; e.g., that $|A|$ appears to better describe the folding probability and Z_A, the folding time.

3.3. 2D PES. We now consider the behavior of the bmincut algorithm on the 2D model PES shown in Figure 5a. From the minimum A ($x = 0, y = 0$), there are four low-energy pathways, two of which lead to two different minima (1,0) and (0,1), and for the other two, the PES stops at a saddle point (−0.5,0) and

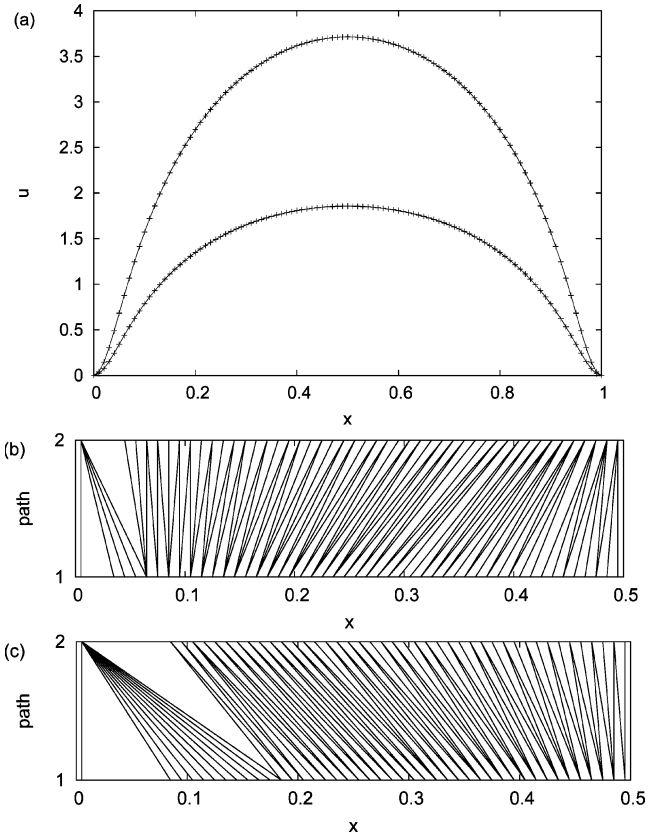

Figure 3. (a) Two pathway PES $u_i(x) = \log(1 - \cos(2\pi x) + 0.05)/i$, $i = 1, 2$. (b) Optimal cuts with $|A|$ as the progress variable. (c) Optimal cuts with Z_A as the progress variables. In (b) and (c), the cuts on path 1 are shown on the bottom and those on path 2 are on the top; the lines connect the corresponding cuts and have no other significance. The results for $x_1, x_2 \geq 0.5$ are symmetrical with those for $x_1, x_2 \leq 0.5$

(0,−0.5). Pathways along the x axis are lower in energy than the pathways along the y axis by construction. The cuts obtained with the bmincut procedure with $|A|$ as the progress coordinate are shown in Figure 5b. The cuts advance mainly along the lowest-energy pathway (along the directions (1,0) and (−1,0)), though the other pathways (along the directions (0,1) and (0,−1)) are explored as well. By decreasing the temperature, one can force the cuts to follow more closely just the lowest-energy pathway (data not shown); i.e, the bmincut can be used to find low-energy pathways. Compared to the eigenvector-following approach,[11] the present method has the advantage of not using any information about the topology of the coordinate space. It is based only on information about the transition rates between the states, so it can be readily applied to a coarse-grained version of the PES. This can be useful because the transition rates are easily to calculate for the coarse-grained FES by summing over the terms that are combined in the coarse-graining.

Figure 5c shows the FEP obtained with the bmincut procedure. Again, the bmincut procedure was able to find FEP only in the steepest part of the profile and at the TS; the relatively flat regions remain unexplored.

3.4. Comparison Between Bmincut, Bpfold, and Pfoldf Procedures. To compare the three procedures of building the FEP, we consider the model two-pathway PES shown in Figure 6a. The profiles computed with each procedure, as well as the exact profile obtained by enumerating all the possible cuts (and taking the one with lowest value) are shown in Figure 6b; the exact profile can be calculated only because the system is sufficiently simple. Although the FEP computed with bmincut procedure gives exact values, they are obtained for few points.

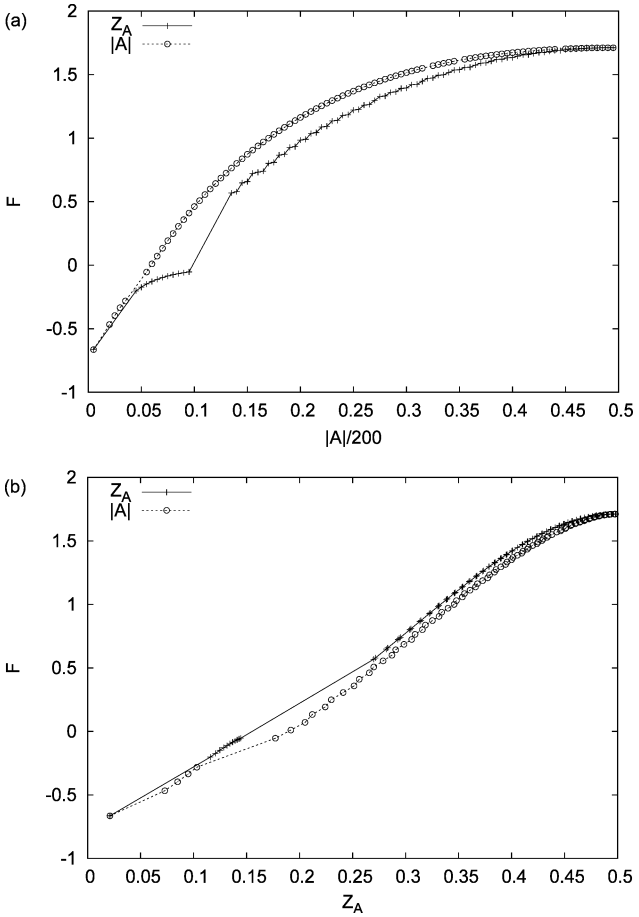

Figure 4. Free-energy profiles obtained with $|A|$ and Z_A progress variables plotted along (a) $|A|$ and (b) Z_A progress variables. Each FEP is more optimal (higher) than the other along its own progress variable. The figure again shows that the method is not applicable to all points.

However, they include important regions of the FES, such as the transition state. To label the points, we use the coordinates (x,i), where i (i=1,2) denotes the pathway and x denotes the coordinate of the point along it. The point at $|A| = 27$ in Figure 6b corresponds to the cut (with coordinates (0,1) and (27,2)) that is close to the first TS of the second pathway. The FEP computed with the bpfold procedure approximates the overall behavior of the exact FEP, though the former is notably lower than the latter. The faster version of the bpfold procedure (i.e., pfoldf) gives a very similar FEP for this case except near the transition state. There pfoldf is significantly lower than bpfold, and both are lower than bmincut, the exact value. The failure of the bpfold (pfoldf) procedure in the TS region is due to the fact that the PES and FES along the two pathways, both of which contribute, are significantly different. In particular, the first pathway has $p_{\text{fold}} = 0.5$ at the top of the barrier, and the second, due to the symmetry, has $p_{\text{fold}} = 0.5$ at the minimum between the two barriers. Thus, it is impossible to cut both pathways at the top simultaneously, i.e., by a single value of p_{fold} (see Figure 6 for the cuts). A small departure from the top for the (relatively) sharp barrier notably increases the partition function (lowers the free energy) of the cut.

These differences raise the question of the definition of the transition state. If one defines the TS as points with $p_{\text{fold}} = 0.5$,[7] then the cut at $p_{\text{fold}} = 0.5$ in the present example gives the exact result, by definition, although it is not physically meaningful. However, we define the TS as the cutting surface with the minimum value for the partition function,[3] in the spirit of variational transition state theory.[12] The cut at $p_{\text{fold}} = 0.5$ ($x_1 =$

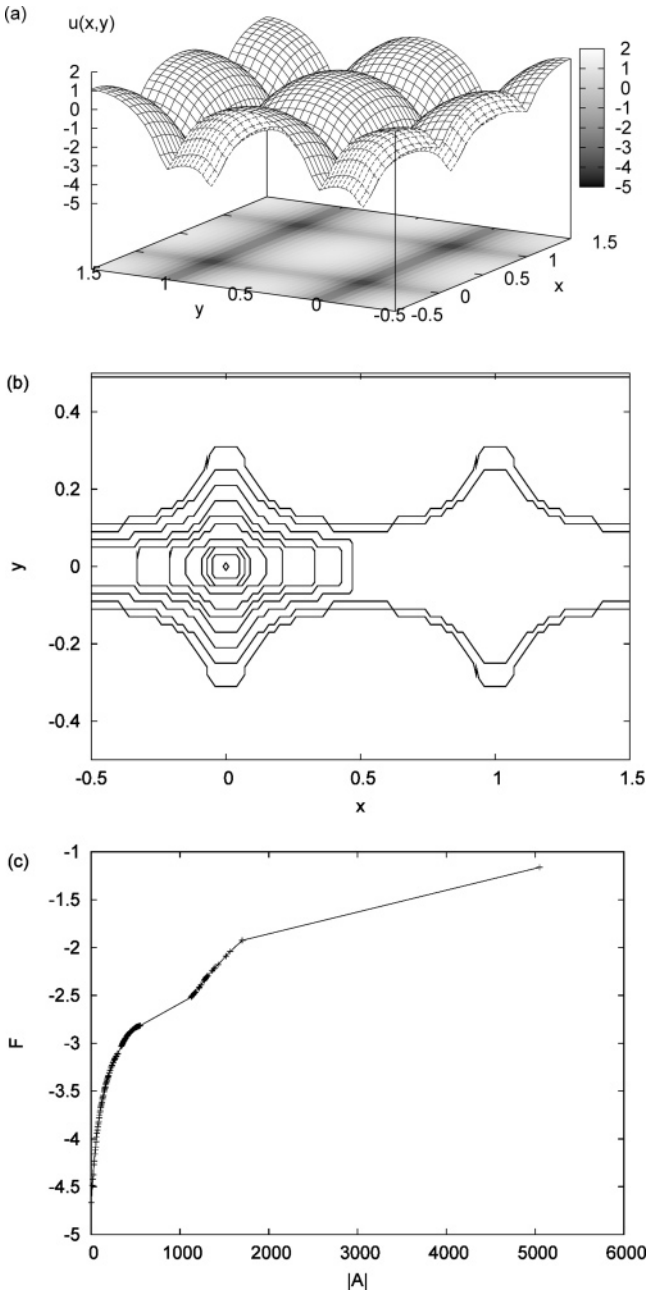

Figure 5. (a) Model PES $u(x, y) = \log(1 - \cos(2\pi x) + 0.05)/2 + \log(1 - \cos(2\pi y) + 0.05)$, $x \in [-0.5, 1.5]$, $y \in [-0.5, 1.5]$, $dx = dy = 0.02$, $kT = 0.2$. We note that the figure shows the full surface, which is truncated as indicated. (b) Cuts of FEP obtained with bmincut procedure for $|A|$ as the progress variable. (c) FEP along $|A|$ as the progress variable.

0.5, $x_2 = 0.5$, $|A| = 100$, thick line on Figure 6d) for the PES in Figure 6a gives the TS at $|A| = 100$ in Figure 6b. Clearly, this is very far from the minimum partition function at $|A| = 80$, as shown Figure 6b.

Although the model system (Figure 6a) was chosen to make a point (i.e., that problems can arise with pfold and pfoldf if used as a reaction coordinate), the system is not as artificial as it may seem. For example, the FES of proteins is believed to consist of many relatively small basins, including the region of $p_{fold} = 0.5$, where such overlapping can occur.

4. Application to Realistic Systems

4.1. β-Hairpin of Protein G. As described in the Introduction, the β-hairpin of protein G has been widely used as a simple, yet realistic, model system for protein folding.[13,14] In previous work, we analyzed the FES by constructing the free-energy disconnectivity graph (TRDG). The graph showed that the denatured state has a number of low-energy basins in addition to a large shallow entropic basin. This made clear that the FES, and the PES, which was also estimated, is not a simple funnel, but rather is a multi-minimum surface. A reduced network to calculate the kinetics of the system was constructed.[4] It takes into account only the deep free-energy minima and the transitions between them. One way to verify the validity of the reduced network is to show that the equilibration inside the basins is fast relative to the transitions between them, so that they can be considered as single entities. For this purpose, a knowledge of the correct FEP, which gives the barrier between the basins, is essential. Commonly used projection methods do not provide realistic barriers because of the overlap of different regions of configuration space, as shown in ref 4. By contrast, the present method gives the correct barriers, as demonstrated by the examples described above. The bmincut procedure gives the exact free-energy profile as a function of the progress coordinates Z_A or $|A|$, though only for a portion of the surface in most cases. The bpfold and pfoldf procedures gives meaningful results for the entire range of the progress coordinate, but they are approximate. Also, the disconnectivity graph analysis per se is not able to visualize (separate out) basins, whose free energy comes mainly from entropic terms, as in the case for a large portion of the unfolded state of the β-hairpin.[4] By contrast, the FEP with Z_A ($|A|$) as the progress coordinate is able to determine the profile of the barriers separating the localized basins.

We use here the EKN for the *β-hairpin* employed in our previous study;[4] the EKN was obtained from a 4 μs MD trajectory at $T = 330$ K and clustered with a 2.0 Å rmsd threshold. It contains 35 377 nodes and 83 331 transitions (edges) connecting them; after removing clusters that were visited just once, as described above, 10 331 nodes and 56 790 edges remain. To determine the FEP, we use the bmincut and bpfold procedures. The results are compared with these obtained with the disconnectivity graph. Figure 7 shows the unfolding FEP of the native node obtained with bmincut and bpfold; i.e., it represents the FEP between the most visited cluster, which is the native node, and an arbitrary "extra" node, outside the network. On the profile, one can identify four basins; these basins were related directly to those found in the previous analysis (Figure 2 of ref 4) by comparing the nodes of the network in the basins, because the same network has been used. The entropic basin of the denatured state is not localized in Figure 7; it is spread over the range of $0.35 < Z_A/Z < 1$, as determined by the contributing nodes. The FEP resolves a high barrier at $Z_1/Z \approx 0.35$ between basin 1 and the rest of the FES. This gives $Z_1 = 0.35Z$ in accord with Figure 5 in ref 4. The barrier is between basin 1 and the entropic basin (as checked by the node comparison). However, the FEP shows no barriers between basins 2, 3, 4 and the entropic basin. This is due to the unavoidable overlap between the different basins when they are projected on a single progress coordinate and fall into the same region. The unfolding FEP for basin 2 (Figure 8), obtained by using the most visited node in basin 2 and the same "extra" node, shows that with this projection, basin 2 is separated from the rest of the FES by a high barrier. Examination of the nodes connected by the edges making up the TS of the barrier reveals that it consists of two comparable barriers connecting basin 2 with the entropic basin and basin 3 (Figure 5 in ref 4). We note

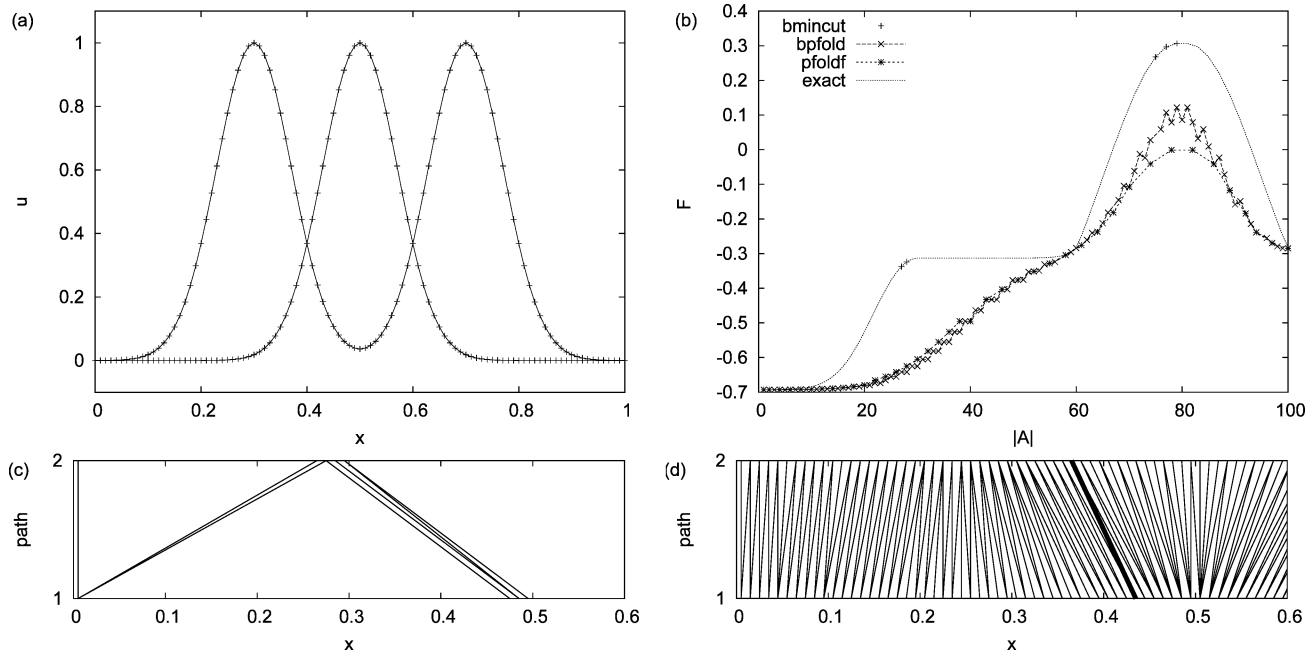

Figure 6. (a) Model PES with two pathways; first ($i = 1$) pathway $u(x) = \exp(-(x - 0.5)^2/0.01)$, second ($i = 2$) pathway $u(x) = \exp(-(x - 0.3)^2/0.01) + \exp(-(x - 0.7)^2/0.01)$. (b) FEPs computed with pfoldf, bpfold, and bmincut procedures as well as exact FEP computed by enumeration of all the cuts. Only the left halves of the FEPs are shown, because the right halves are symmetrical. (c) Cuts of the exact FEP obtained with balanced mincut procedure. (d) Cuts for the FEP obtained with pfoldf procedure. The thick lines show the cuts corresponding to the putative TS obtained here. In (c), the line goes through the top of the barriers for each of the pathways, i.e., (0.5,1) and (0.3,2); the TS corresponding to the second barrier of the second pathway is on the right part (not shown) on (b). In (d), the thick line does not go through the tops of the barriers (thus, not the optimal cut) and gives a lower free energy of the TS than the exact result.

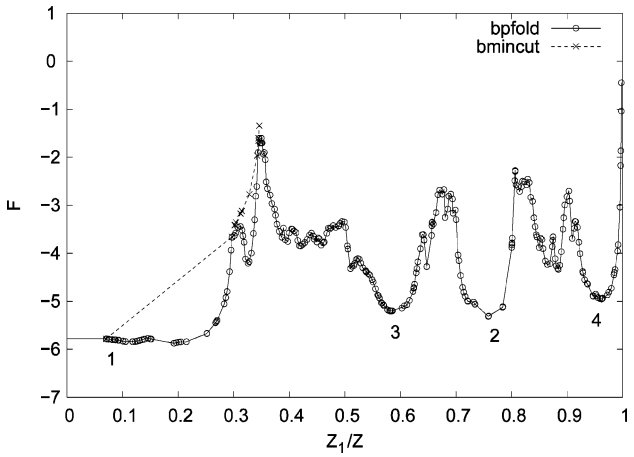

Figure 7. Unfolding FEP of the β-hairpin: bmincut (dashed line with crosses) and bpfold (solid line with circles). The symbols are the actual points obtained with the algorithms, and the lines are to guide the eye.

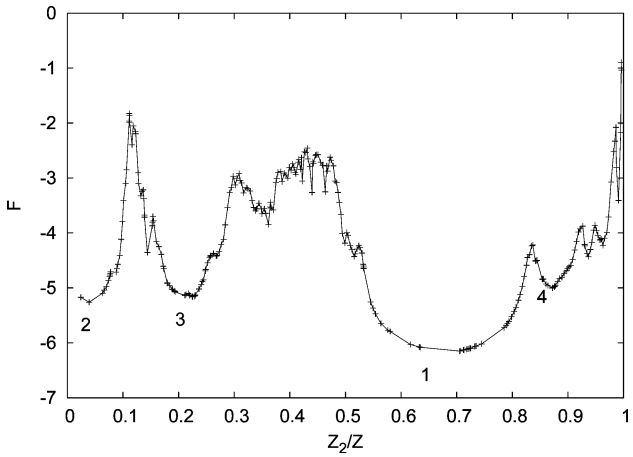

Figure 8. Unfolding FEP of the β-hairpin of basin 2 obtained with pfoldf. The symbols are the actual points obtained with the algorithm, and the line is to guide the eye.

that Z_A in Figures 7 and 8 are different; i.e., in Figure 7, Z_A measures the partition function relative to the chosen node in basin 1 and in Figure 8, relative to the chosen node in basin 2.

The shape of the barrier between basins 2 and 3 in Figure 8 is very similar to that in Figure 7 at $Z_1/Z \approx 0.67$; the latter is just a little lower. This is because in both cases, the same region of configuration space is projected, but in the latter there is a contribution from (overlapping with) FEP of denatured basin, which is quite flat and does not change the shape significantly.

The unfolding results show that all basins 1, 2, 3, 4, and entropic (for 3, 4 and the entropic basin, data are not shown) are separated from the rest of the FES by high, sharp free-energy barriers (on the order of 3−5 kcal/mol). This indicates that full equilibration takes place inside the basins, so that a simple kinetics model that considers basins as a whole (e.g., that used in ref 4) is valid. Most importantly, the present progress variables provide a clear depiction of the barriers between basins.

For comparison, we show in Figure 9 the FEP as a function of the rmsd from most visited cluster; i.e., that representing the native state. Figure 9 was calculated in a conventional way by counting histograms. It is clear that different basins overlap and the FEP is simpler than the one in Figure 7. Whereas basin 2 is clearly separated because it has a very high rmsd (∼8 Å), basins 3 and 4 are not visible because they are absorbed by the entropic basin. The putative TS between basin 1 and the denatured state at rmsd ≈ 3.5 Å is incorrect because in this region of configuration space, only configurations from basin 1 are present, as shown by curves representing the various contributions. In Figure 7, the transition state for going from the native state to

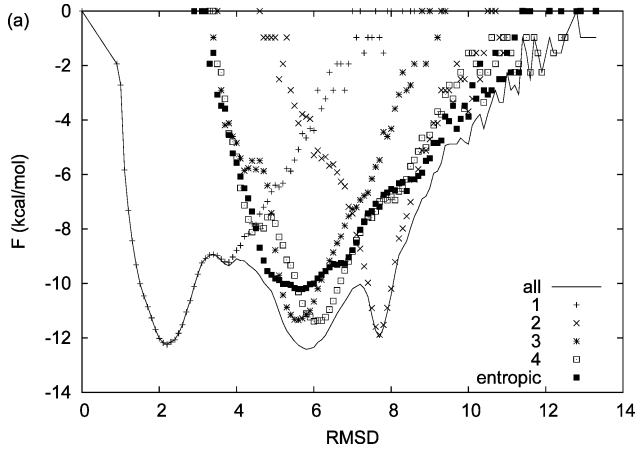

Figure 9. FEP of β-hairpin plotted as a function of the rmsd from the most visited cluster (native state). Symbols show contributions from different basins (the basins were found by bmincut procedure, ref 4).

the rest of the FES is well defined at $Z_1/Z \approx 0.35$; the rmsd from the native state of the structures of the TS is in the range 3−5.6 Å.

4.2. Lennard-Jones Cluster LJ_{38}. The second realistic example we describe is the Lennard Jones cluster LJ_{38}. It was shown by use of a potential energy disconnectivity graph[15,16] to have a two-basin PES[17] (see Figure 2 in ref 17). One basin is deep and narrow and leads to the global minimum (fcc-truncated octahedron O_h in ref 17), and the other is a wide basin leading to the next lowest (local) minimum on the PES (incomplete Mackey icosahedron C_{5v} in ref 17). The analysis in ref 17 suggested that the kinetics of the system leads to a quick descent to the second lowest (icosahedral) minimum, because of the broad funnel associated with it and its small internal barriers, and that it takes much longer to find the global (fcc) minimum.

An equilibrium Langevin MD simulation was performed at $T = 0.16\epsilon$ for 10^9 steps with a time step of 1 fs, on the LJ potential $U = 4\epsilon \sum_{i<j}[(\sigma/r_{ij})^{12} - (\sigma/r_{ij})^6]$, with $\epsilon = 1$ kcal/mol and $\sigma = 2^{-1/6}$Å. The clustering of the trajectory was performed by taking snapshots from the trajectory every 1 ps and quenching it to the current local minimum on the PES by steepest-descent minimization. This resulted in 12 496 minima and 59 533 transitions between them. After removing local minima that were visited just once, there remained 5382 minima and 51 796 transitions.

On the FEP obtained with bpfold (Figure 10a) one can identify four regions: icosahedral ($0 < Z_A/Z < 0.06$); a broad basin corresponding to minima 3 and 4 (C_s in the notation of ref 17), ($0.06 < Z_A/Z < 0.51$); a liquidlike basin ($0.51 < Z_A/Z < 0.87$) that has many minima with small partition functions, and the fcc basin ($0.87 < Z_A/Z < 1.0$). Long straight lines (e.g., 3 and 4) represent the local minima with large partition function. Due to the clustering procedure (i.e., all the points of the basin of attraction of a local minimum are combined into a single cluster), the free-energy profile inside the basins is not determined. Basin 7 and the liquidlike basin look different (i.e., they have many points inside the basins) because they consist of a large number of local minima, whereas each of the basins 3 and 4 consists only of a single local minimum. If one would use more fine-grained clustering, a basinlike FEP instead of straight lines would be obtained. The system mainly stays in the liquidlike region and in the basin around minima 3 and 4, which are connected by a low barrier, because the sum of their partition functions is about 0.8 of the partition function of the entire FES. The estimated mean times to visit either the fcc

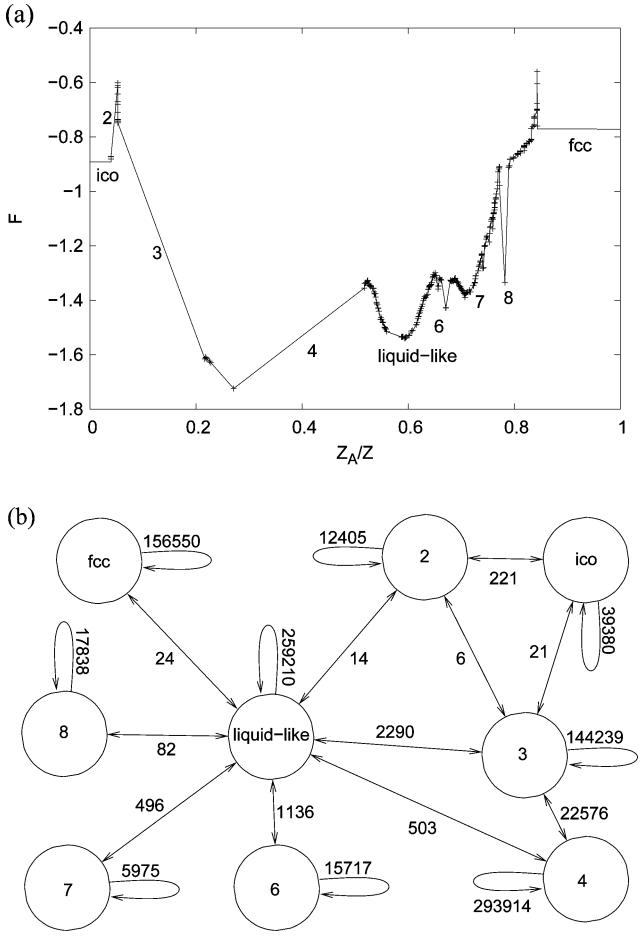

Figure 10. (a) FEP of LJ_{38} ($T = 0.16\epsilon$) obtained with bpfold. Free energy is measured in units of ϵ. (b) Simplified EKN. Numbers on the edges show the number of direct transitions in each direction between the basins; self-returning edges show the number of times the system was found in the same basin at the following quenching attempt.

basin (2.59×10^8/24 fs) or the icosahedral basin (2.59×10^8/41 fs) (i.e., the total simulation time system spent in the liquidlike state divided by the corresponding number of transitions), starting from the liquidlike basin are comparable and much longer than the meantime to go between the liquidlike basin and the basin around minima 3 and 4. The result is significantly different from the interpretation of the PES in ref 17 due to the importance of entropic contribution. In ref 17, it was stated that "it is still appropriate to say that the icosahedral minima form a single funnel because the minima are structurally very similar and because the barriers are still relatively low and so the minima give rise to a single thermodynamic state". This conclusion is not valid at $T = 0.16\epsilon$ used here. The PES graphs studied in ref 18 would became more accurate as temperature goes to zero, where the entropy does not contribute. At the present temperature, the analysis shows that, from the kinetics viewpoint, one should perhaps consider the basin around minima 3 and 4 and liquidlike state to be a single state, because the transitions are rapid. This, probably, explains why "optimization methods found the C_s minimum first; the C_{5v} minimum was only discovered relatively recently" (citing ref 17).

Figure 10b shows a simplified EKN, which has a relatively small number of nodes. It represents the basins on the FEP, as illustrated in Figure 10a. We kept only the basins that have relatively high barriers (i.e., the mean time to escape a basin, which is estimated as $Z_i/(\sum_j Z_{ji})$, is more than 10) and whose partition function is significant (i.e., $Z_i > 0.1 \sum_i Z_i$). We note

TABLE 1: Properties of Basins of EKN of LJ_{38}

basin	U_{min}[a]	F[b]	$\bar{\omega}$[c]	$\|\|$[d]
ico	−173.252378	−1.69	6.662	4
2	−172.958633	−1.51	6.627	4
3	−172.877736	−1.90	6.600	64
4	−173.134317	−2.02	6.618	262
liquidlike	−170.989860	−1.99	6.097	4956
6	−169.358251	−1.55	5.700	24
7	−170.070340	−1.39	5.772	22
8	−169.829822	−1.57	5.898	30
fcc	−173.928427	−1.91	6.795	16

[a] Potential. [b] Free energy (in units of ϵ). [c] Geometric mean of the 106 normal mode frequencies of the lowest local minimum in the basins. [d] Number of the local minima in the basin.

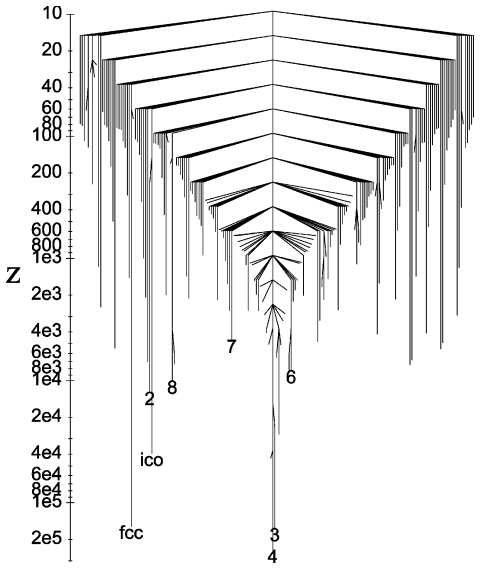

Figure 11. TRDG of LJ_{38} ($T = 0.16\epsilon$); the numbering is the same as in Figure 10a,b.

that the simplified EKN has the same internode barriers as the original EKN with 5382 nodes. The EKN shows that the barriers between the liquidlike state (it has about 5000 minima) and the basin around minima 3 and 4 are quite low, in accord with the FEP in Figure 10a. The icosahedral minimum has "direct" connection to the liquidlike state through basin 2. Also, basin 3 is much more connected with the liquidlike state than basin 4, whereas the FEP shows that basin 3 is connected to the liquidlike state through basin 4. Apart from that, the FEP is generally in accordance with the simplified network, which is an unprojected representation of the FES. Because these discrepancies involve fast kinetics, they do not change the slow kinetics of primary interest. This is because of the relative simplicity of the overall landscape; i.e., only one pathway connects the icosahedral and fcc regions of the FES. Of course, barriers "transversal" to this pathway (e.g., between the liquidlike basin and basins 6, 7, or 8) cannot be seen on the FEP. Minimum 7 is shown to illustrate the existence of the minima separated by a high barrier but having a low partition function. Table 1 lists the properties of the basins of the EKN. The second lowest basin (ico) is lower almost by kT in energy than minimum 4, but due to the entropic contribution (lower $\bar{\omega}$, as well as more local minima), the latter is much more populated.

Figure 11 shows the transition disconnectivity graph[3] (which takes into account entropic terms) of the LJ_{38} cluster, which is in agreement with the FEP and simplified network and in disagreement with PES-based disconnectivity graph shown on Figure 2 in ref 17. In particular, the FE barriers between the icosahedral basin and basin 4 and between the fcc basin and basin 4 are comparable. As stated before, disconnectivity graphs are not able to visualize an entropic basin, which here is associated with the liquidlike basin. All the local minima that constitute the liquidlike basin have small partition functions and hence are placed in the top region of the disconnectivity graph (Figure 11) and do not form a basin as in Figure 10a.

5. Concluding Discussion

One of fundamental questions in protein folding, and more generally in the reaction of complex systems, is whether the essential features of the free-energy surface can be represented in a reduced representation. The ideal would be a 1D representation that gives the free energy as a function of an appropriate progress coordinate. It is not clear, in general, whether such a 1D projection exists, though there have been many attempts to find one.

A number of approaches to construct a 1D free-energy profile are based on the commitor distribution commonly designated as p_{fold} in the protein folding reaction.[7,19,10,20] The first use of an analogous concept, referred to as "splitting probability," to discriminate product and reactant goes back to Onsager.[21] Du et al.[6] showed that conventional progress coordinates such as the number of native contacts are not adequate, in general, for protein folding because points with the same number of native contacts can have very different values of p_{fold}. They suggested that p_{fold} itself be used to build the FEP, and examples based on lattice simulation were given.[6] Because the kinetics based on this profile do not reproduce the kinetics of the systems,[10] Rhee and Pande[10] recently proposed a method to obtain a 1D FEP that reproduces the pfold values (isocommitor surfaces) of a multidimensional system in which the motion is diffusive. They illustrated their approach with a 2D model potential and commented on its utility for determining the kinetics with an assumed (constant) value for the diffusive prefactor. Best and Hummer[19] and Ma and Dinner[20] proposed approaches for constructing reaction coordinates on the basis of variational criteria by numerically finding the optimum combination from a putative set of variables. Best and Hummer[19] optimized another probabilistic quantity $p(\mathrm{TP}|r)$, i.e., the probability of being on a transition path (TP) for a given value (r) of the reaction coordinate. Ma and Dinner[20] tried to find a function of a small number of physical variables that reproduces p_{fold} by minimizing the RMS error using the GNN program.[22]

As a general remark concerning the methods using p_{fold}, we note that although it is a plausible assumption that all the points on the TS have $p_{fold} = 0.5$, the reverse (i.e., that all the points with $p_{fold} = 0.5$ belong to the TS) is not necessary true, as exemplified by the model PES in Figure 6.

The present method differs from the p_{fold}-based approaches in two ways. First, we define the transition state by use of the mincut procedure instead of by use of p_{fold}. As show above, results obtained with p_{fold} are approximate as compared with the results obtained with the mincut, although in the cases studied they are quite accurate and usually cover a larger region of the progress variable than the bmincut method. Second, the algorithm we use is based on an equilibrium kinetic network (EKN) as input, rather than the simulation results themselves. Use of the EKN simplifies the problem, but one has to verify that the EKN is a good approximation to the FES, i.e., that the equilibrium trajectory covered all of the important region of the FES and has enough transitions between them so that the statistical errors are small. Given that, the results of an analysis of the EKN correspond with high enough accuracy to the FES. Of course, the same requirement exists for any method based on the use of equilibrium trajectories.

Use of the bmincut procedure to obtain the FEP gives the exact profile in our definition because, as it was shown, the mincut in a modified (biased) network gives the solution to a constrained optimization problem for the original network with the help of a Lagrange multiplier. However, the method is restricted to certain regions of progress coordinate. This restriction arises when a minimum cut as a function of the constraint parameter has a number of local minima. The prescription for using a Lagrange multiplier would suggest that all the local minima should be checked to solve the problem. However, the minimum cut only allows determination of the lowest one. Bpfold and pfoldf can be used to obtain an approximate FEP for the entire range of progress coordinates.

The reaction coordinates we chose (Z_A and $|A|$) are among the simplest and most flexible coordinates that increase monotonically as the system goes from the initial state to the final state. If there are several well-defined pathways, these reaction coordinates will adapt their shape to them and will progress mainly along the pathways; an example is shown for the model system in Figure 5, which has two pathways, and the cutting surfaces advance mainly along the pathways. Moreover, if one wants to identify the structures associated with most important pathways, one can do this by postprocessing the obtained cuts, e.g., by considering the structural properties of the clusters between two successive cuts. The fundamental idea of the present approach is that the reaction coordinate is chosen to include any and all pathways from the initial to the final state without any prejudgment as to the geometric coordinates involved. This requires that the reaction coordinate be defined independent of structural features, as it is.

If the FEP is accurate, it describes the essence of the reaction kinetics by showing the barriers and basins on the way from the initial to the final state. Because the chosen progress coordinate is very flexible, the obtained FEP is likely to be the best way of projecting the FES onto a 1D coordinate. A posteriori, the structural features associated with the reaction can be determined. The reaction coordinates introduced in this paper are similar in spirit to the "pfold" reaction coordinate, which, also, is not associated with any "(reduced) structural description of the process", as stated by a referee, and whose relation to a reduced structural description can be obtained by postprocessing. A good example is provided by the analysis of the dipeptide reaction in ref 20.

Finally, we consider the complementary utility of free-energy profiles (FEP's; e.g., Figure 10a) and simplified equilibrium kinetic networks (EKN's; e.g., Figure 10b). A good way to represent the results of an equilibrium simulation on a complex potential energy surface is to construct the EKN, together with the FEP's for each basin (e.g., Figures 7−8); Figure 10a shows the results in a condensed way; i.e., all the profiles are shown in a single plot. The network is useful as a graphical representation of the interbasin kinetics, whereas the profiles are more useful for describing the intrabasin kinetics. For example, for the folding of a two-state protein, one is interested primarily in the intrabasin kinetics; i.e., it is the nature of the approach to the transition state from the denatured state that is given by the FEP. In the case of barrierless fast folding,[13,24] no meaningful network model can be constructed, because there are no well-defined states separated by well-defined barriers on the FES; the FEP indicates the absence of a barrier.

The simplified network (Figure 10b) shows clearly the interbasin kinetics, which is less evident in the FEP (Figure 10a). However, the validity of the simplified network is determined by a rate of the transitions relative to the rate of equilibration in the various states, which is obtained from the FEP. Also, the general procedure to isolate each basin, which is required to construct the network, is to build its FEP and cut at the top of the barrier. For example to separate the basin "ico" in Figure 10a, one has to cut at the top of the barrier around $Z_A/Z \approx 0.04$. In this case, the basin can be separated, in principle, by applying the mincut procedure between the most visited nodes in the basin and, say, basin 4. However, the mincut procedure fails in the case where the entropic contribution is significant (e.g., as in denatured basins[4]), and there, the FEP (balanced mincut[4]) is required.

We believe that the method proposed here can be of wide utility and hopefully will be used by others. The FORTRAN library of subroutines of the methods to build the FEP with a simple interface to the PYTHON language and scripts to build the model FEPs described in the paper are available on request.

References and Notes

(1) Dobson, C. M.; Sali, A.; Karplus, M. *Angew. Chem.* **1998**, *110*, 908.
(2) Kubelka, J.; Hofrichter, J.; Eaton, W. A. *Cur. Opin. Struct. Biol.* **2004**, *14*, 76−88.
(3) Krivov, S. V.; Karplus, M. *J. Chem. Phys.* **2002**, *117*, 10894−10903.
(4) Krivov, S. V.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14766−14770.
(5) Ford, L. R.; Fulkerson, D. R. *Can. J. Math.* **1956**, *8*, 399−404.
(6) Du, R.; Pande, V. S.; Grosberg, A.; Tanaka, T.; Shakhnovich, E. S. *J. Chem. Phys.* **1998**, *108*, 334.
(7) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291−318.
(8) Apaydin, M. S.; Brutlag, D. L.; Guestrin, C.; Hsu, D.; Latombe, J. C.; Varma, C. *J. Comput. Biol.* **2003**, *10*, 257−281.
(9) Young, D.; Kincaid, D. *Elliptic Problem Solvers*; Academic Press: New York, 1981; pp. 163−185.
(10) Rhee, Y.; Pande, V. *J. Phys. Chem. B* **2005**, *109*, 6780−6786.
(11) Doye, J. P. K.; Wales, D. J. *Z. Phys. D: At., Mol., Clusters.* **1997**, *40*, 194−197.
(12) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. *J. Phys. Chem.* **1996**, *100*, 12771−12800.
(13) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196−199.
(14) Dinner, A. R.; Lazaridis, T.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9068−9073.
(15) Becker, O. M.; Karplus, M. *J. Chem. Phys.* **1997**, *106*, 1495−1517.
(16) Wales, D. J.; Miller, M. A.; Walsh, T. R. *Nature* **1998**, *394*, 758−760.
(17) Doye, J. P. K.; Miller, M. A.; Wales, D. J. *J. Chem. Phys.* **1999**, *110*, 6896−6906.
(18) Doye, J. P. K.; Miller, M. A.; Wales, D. J. *J. Chem. Phys.* **1999**, *111*, 8417−8428.
(19) Best, R. B.; Hummer, G. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6732−6737.
(20) Ma, A.; Dinner, A. *J. Phys. Chem. B* **2005**, *109*, 6769−6779.
(21) Onsager, L. *Phys. Rev.* **1938**, *54*, 554−557.
(22) So, S. S.; Karplus, M. *J. Med. Chem.* **1996**, *39*, 1521−1530.
(23) Fujitsuka, Y.; Takada, S.; Luthey-Schulten, Z. A.; Wolynes, P. G. *Proteins* **2004**, *54*, 88−103.
(24) Yang, W. Y.; Gruebele, M. *Nature* **2003**, *423*, 193.