The folding transition state theory in simple model systems

2008 J. Phys.: Condens. Matter 20 244134

(http://iopscience.iop.org/0953-8984/20/24/244134)

J. Phys.: Condens. Matter **20** (2008) 244134 (9pp) doi:10.1088/0953-8984/20/24/244134

The folding transition state theory in simple model systems

Szymon Niewieczerzał and Marek Cieplak

Institute of Physics, Polish Academy of Sciences, 02-668 Warsaw, Poland

Received 23 October 2007, in final form 3 November 2007
Published 29 May 2008
Online at stacks.iop.org/JPhysCM/20/244134

Abstract

We present the results of an exact analysis of several model free energy landscapes of a protein to clarify the notion of the transition state and the physical meaning of the ϕ values determined in protein engineering experiments. We argue that a proper search strategy for the transition state in more realistic models should involve identification of a common part of various methods. Two of the models considered involve explicit conformations instead of just points on the free energy axis. These models are minimalistic as they are endowed only with five or 36 states to enumerate folding paths and to identify the transition state easily. Even though they display much of the two-state behavior, the ϕ values are found not to correspond to the conformation of the transition state.

1. Introduction

Folding kinetics of simple globular proteins is often found to display a two-state behavior with a single characteristic timescale [1–6]. This behavior signifies that each individual protein molecule in a solution appears to be either in a folded (N, for native) or denatured (D) state and not in between. Thus folding can be viewed as a chemical reaction that proceeds between D and N along some reaction coordinate on which there is an essentially unpopulated bottleneck—the transition state, ‡. Another perspective on folding is to think of it as a nucleation process [7, 8] in a first order phase transition which requires formation of a critical sized ‘droplet’ to succeed in transforming the unfolded state into the folded state. This ‘droplet’, or a folding nucleus as it is known, is equivalent to the transition state in the chemical reaction picture. Both pictures allow for emergence of different pathways with distinct transition states. An established experimental way to probe them involves the substitution of amino acids in different positions with other amino acids and monitoring the resulting changes in the kinetics [9–11, 2, 12, 13, 3–5]. The changes in the kinetic folding rates normalized to the corresponding changes in the protein stability define the so called folding ϕ-values. In simple situations, the values of ϕs range between zero and one. A value that is close to one suggests a nearly native-like structure around the site of substitution in the transition state. Thus a set of the ϕ values provides an approximate description of the transition state conformation(s).

Clearly, any protein can take an astronomical number of conformations and yet it often behaves as it was a two-state system. This indicates that, in such situations, there is a huge number of relaxation modes with timescales that are too short to be measured in standard experiments and one exponential mode that is long lasting because it involves climbing the transition state. How can one determine this state theoretically?

Many ideas have been put forward to tackle this problem. One of them proposes that transition states are all edge states [14, 15]. A definition of the edge state is that when one takes it as a starting point for a temporal evolution then one finds the system reaching the N and D states with nearly equal probabilities. Another prescription is that one should find the free energy, $G(Q)$ as a function of the fraction, Q, of the established native contacts and identify all states which correspond to the maximum in $G(Q)$ [16]. Still another [17] claims that the transition state is an eigenvector of a kinetic matrix that governs relaxation processes in the system. Daggett and collaborators [18, 19] seek transition states by performing molecular dynamics of unfolding at high temperature and then by identifying stages with rapid structural transformations. In [20] and [21] the prescription involves characterization of conformations by sets of local Q_i values, where Q_i is the fraction of established native bonds that are linked to the ith amino acid, and then finding the best match to the experimental ϕ-values. And so on.

Since the subject of transition state in folding kinetics is enshrouded in confusion it seems pertinent to consider simple and exactly solvable models that would benchmark various theoretical approaches. One of such models has been recently proposed and studied by Chang *et al* [22]. This model is a variant of a system considered by Munoz, Eaton, and their

collaborators [23–25] and it embodies the topology of the 12-amino-acid β-hairpin. The model is discrete and it is endowed with an explicit free energy landscape. The variant studied by Chang *et al* has been formulated in terms of Ising spins taking values 0 and 1, as discussed in section 2. The value of 1 signifies native-like placement of a bond between successive amino acids and 0—a non-native like placement. Non-zero values of multiple spin products correspond to a native-like establishment of a fragment of the hairpin, together with the establishment of the contacts within it. The model is endowed with the spin-flip kinetics that is described by a master equation for the probability density for various possible conformations. Enumeration of all directed paths in the free energy landscape—a procedure that cannot be accomplished in any realistic system—leads to an identification of two degenerate conformations that are the transition states for the system. Essentially all of the other proposed theoretical prescriptions that we have tested, are consistent with this identification. However, they would also select, incorrectly, several other conformations as transition states even though their free energies were higher. Interestingly, the sets of putative transition states depended on the prescription used but there remained a common part that agreed with the exact enumeration result. In other words, these approximate approaches contain 'grains of truth' and the lesson learned is that a successful search strategy should involve seeking for the common 'grain'.

In this paper, we consider the same issues but in a context of three other simple, discrete, and exactly solvable models with the kinetics described by the master equation. The first of these models has been introduced by Merlo *et al* [26] and we reconsider it in section 3. The model involves just four states: N, D, and two connecting states A and B of which A is the transition state. The transition rates between these four states are defined in terms of the free energy differences, but no actual conformational changes are considered. We comment on the relevance of the eigenvectors of the kinetic matrix and show that the eigenvector corresponding to the longest relaxation time identifies state A as the transition state only after the contributions of the D and N states are discarded.

The other two models, corresponding to five and 36 conformations on the square lattice, are studied in sections 4 and 5 respectively. These minimalist models go one step further toward realism, compared to the other models discussed here. First of all, the conformations in these models correspond to specific shapes of a polymer on the lattice (as opposed to the states being defined merely in terms of which segments are native-like, or not having a specific geometry as in the case of the model proposed by Merlo *et al*). This means, in particular, that the kinetic moves correspond to well defined local adjustments in the conformations. Secondly, the free energies of conformations are not declared. Instead, the energies of the conformation are derived from contact potentials and the entropy terms are evaluated from the equilibrium probability density matrix. The five-state model is the simplest explicit-conformation analog of the model proposed by Merlo *et al* [26] whereas the 36-state model gives one a flavor of how should one deal with more convincing models of proteins. In the 36-state model the transition from the D to N state is generally

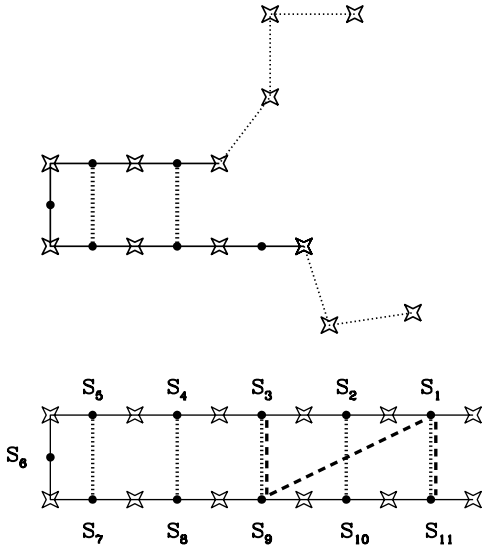

Figure 1. Bottom: the model β-hairpin system studied in [22] and in this paper. The stars denote amino acids. The spins S_n correspond to the peptide bonds between the successive amino acids. In non-native conformations only parts of the native structure are established. The dotted lines indicate presence of a hydrogen bond. The dashed lines correspond to hydrophobic bonds between hydrophobic amino acids. Top: example of a non-native conformation in the β-hairpin system. The solid line indicates established native-like backbone. The thinner dotted lines show examples of fragments which are not established in the native way.

downhill in the free energy and yet certain states act as local free-energy maxima. The lowest of these maxima is taken as the transition state.

The two minimalist models with explicit conformations are found to display many aspects of the two-state behavior but there are also deviations which may be related to an insufficient size and/or artificiality of the systems. In particular, the kinetically derived ϕ-values do not reflect the transition state conformation in an adequate way. In addition, the ϕ-values are not found to relate to the set of the local Q_i values. The conclusion is then that one should explore these issues further in larger-sized explicit-conformation models. In meantime, when dealing with more realistic models of proteins, one should use a combination of available methods when searching for transition states.

2. The Eaton-like spin model

The native state of the system we study is illustrated in the bottom panel of figure 1. The system can be described in terms of effective free energy levels which take into account their underlying microscopic degeneracies through an effective entropy term. The free energy levels are defined in terms of 11 peptide bonds which are either placed in the native fashion or not. This binary character of the bond placement allows for an Ising-like modeling and we adopt spin variables S_n which take values 1 or 0 correspondingly. An example of a non-native conformation is shown in figure 2. A part of it, near the turn in the hairpin, is set in a native fashion. All we know about

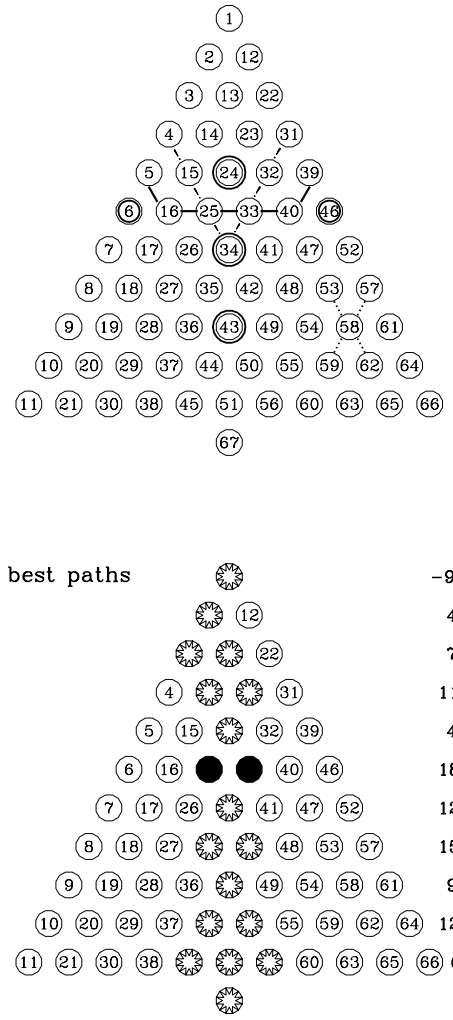

Figure 2. A triangular representation of the 67-level system. The explanations are in the main text of the paper. The values of the free energies and of the contact numbers shown on the right of the bottom panel refer only to the states along the optimal path and not to all states in each row.

the remaining parts in the model is that they are not native-like. There are many ways to have a non-native shape in the remaining segments, one of them is shown at the top of figure 2, and this richness is accounted for by introducing an entropy term which comes with a non-zero value of $(1 - S_n)$.

The free energies per mole can be written as

$$G = -J \sum_{l<m} \Delta_{lm} \prod_{n=l}^{m} S_n + T \Delta S_{\mathrm{conf}} \sum_{n=1} S_n. \qquad (1)$$

A non-zero value of the product $S_l S_{l+1} \ldots S_m$ implies that all peptide bonds between l and m are set in the native fashion which allows for the establishment of native interactions in the cluster between the bonds l and m. These interactions are either hydrophobic or due to establishment of the hydrogen bonds or both. For simplicity, we assume that the strength of the interactions, J, are the same in both cases and equal to 1000 K whereas the conformational entropy per spin, ΔS_{conf}, is taken to be 2.14 R, where R is the gas constant—in the equation above, T denotes the temperature. When writing the entropy term, we omit the constant term (the unity in $(1 - S_n)$. We choose Δ_{lm} to be 2 for $(l, m) = (1, 11)$ and $(3, 9)$, 1 for $(l, m) = (2, 10)$, $(4, 8)$, $(5, 7)$ and $(1, 9)$, and 0 otherwise. Note that the placement of the contacts brakes the symmetry between the upper and lower branches of the hairpin. Let the free energy in conformation i be denoted by G_i. The equilibrium probability to occupy this conformation, P_i is then given by

$$P_i^{\mathrm{eq}} = \frac{\mathrm{e}^{-G_i/RT}}{\sum_i \mathrm{e}^{-G_i/RT}}. \qquad (2)$$

The kinetics of the model proteins considered here are defined through the master equation

$$\frac{\mathrm{d}P(t)}{\mathrm{d}t} = -M P(t), \qquad (3)$$

where $P(t)$ is a vector whose components, $P_n(t)$, are probabilities of staying in the nth state at time t. The matrix M consists of the following elements: $M_{nm} = -w_{nm}$ for $n \neq m$ and $M_{nn} = \sum_{m \neq n} w_{mn}$. Here, w_{nm} are the transition rates for going from state m to n.

The analytical solution to the master equation has the following form:

$$P(t) = \sum_{\lambda} c_\lambda Y_\lambda \mathrm{e}^{-\lambda t}, \qquad (4)$$

where λ are the eigenvalues of the matrix M and Y_λ are the corresponding eigenvectors. The factors c_λ depend on the initial conditions. The smallest non-zero eigenvalue corresponds to the longest lasting relaxation process whereas the zero eigenvalue corresponds to the equilibrium distribution of the probability.

The matrix M describes relaxation which involves a T-controlled balance of folding and unfolding. In order to study the latter processes, we consider matrices M_{f} and M_{u} for folding and unfolding respectively [22]. In the M_{f} matrix, the N-state is considered to be a probability sink, i.e. the transition rates corresponding to leaving the N-state are set to zero (the first column in the matrix M is set equal to zero). Likewise, in the M_{u} matrix, the D-state acts as the probability sink (the last column in the M matrix is set equal to zero). The corresponding two smallest non-zero kinetic rates will be denoted by as k_{f} and k_{u}. The smallest relaxation rate will be denoted by k. In the two-state model, $k = k_{\mathrm{f}} + k_{\mathrm{u}}$ [2].

The folding ϕ-value at site i can be studied by making a small adjustment in J in contacts that involve amino acid i (which thus belongs to two consecutive bonds or spins) and then determining the resulting change, δk_{f}, in k_{f}. The definition is

$$\phi = \frac{\delta \ln (k_{\mathrm{f}})}{\delta \ln (k_{\mathrm{f}}/k_{\mathrm{u}})} = \frac{\delta k_{\mathrm{f}}}{k_{\mathrm{f}}} \Big/ \left(\frac{\delta k_{\mathrm{f}}}{k_{\mathrm{f}}} - \frac{\delta k_{\mathrm{u}}}{k_{\mathrm{u}}} \right). \qquad (5)$$

It has been demonstrated in [22] that the model just outlined has all attributes of a two-level system near the folding temperature and all other folding (unfolding, relaxation) modes than the one associated with k_{f} (k_{u}, k) last too short to matter or are associated with equilibrium.

The model comes with two variants. The first variant considers all 2^{11} spin states that are present in the system. The

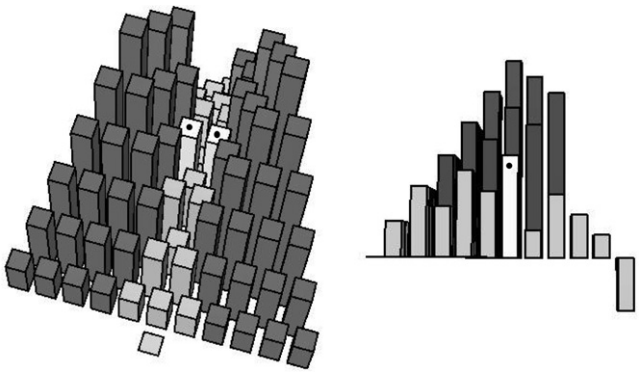

Figure 3. The left panel: the free energy landscape in a three-dimensional representation. The lighter color shows states traveled on the optimal paths. The right panel: a side view of the landscape with a half of the landscape removed to show the optimal paths against the background of other path possibilities.

second variant considers only a subset of 67 states defined by the condition that in a state described by a sequence of 0's and 1's (say 00011111000) no string of 1's is split by a 0. It turns out that this so called single sequence approximation yields all equilibrium and kinetic properties which are very close to the full space description. So in this brief summary, we confine the discussion to the 67-state system.

The free energy landscape corresponding to the 67-state system at $T = 300$ K is shown in figure 2. The single circle at the bottom represents the unfolded state (state 67). The remaining 66 states form a triangle. The row of circles at the base of the triangle represents states with one non-zero spin. The second row represents states with two contiguous non-zero spins, the third—with three, and so on. The top circle represents the single native state (state 1). The kinetic moves from the unfolded state (the bottom state) can connect to any of the single spin states (last but one row) and vice versa. In all other cases, the allowed kinetic moves are only along the diagonal directions on the triangle, as shown by the dotted lines around the 58th state. There are at most four possible moves because the single sequence condition allows for changes occurring exclusively at the interface(s) between the spin ones and the spin zeros.

The top panel of figure 2 shows the local free energy minima, in double circles. The bottom panel shows the states that are traveled on optimal paths (the decorated circles) including the two transitions states (25 and 33; black circles) which have the same free energies. One of the transition states corresponds to a conformation shown in the top panel of figure 2 and the other to its mirror image. The free energies on the optimal paths are listed on the right of the bottom panel. A three-dimensional rendition of the free energy landscape is shown in figure 3. This figure indicates clearly that the optimal paths go through up-and-down states which, however, are much lower in free energy than the remaining regions of the configuration space. The system is quite unlikely to travel into these remaining regions at 300 K.

Among the results obtained by Chang *et al* [22] for this system, we highlight the following.

(1) The reaction coordinate is a list of states that are traveled on the optimal paths: 67, followed by (any of 45, 51, 56), then (44 or 50), 43, (35 or 42), 34, (25 or 33), 24, (14 or 23), (3 or 13), 2, and finally 1.
(2) There are five edge states, 25, 33 and also 5, 16, 40, and 39. They are shown connected by a thick line in the top panel of figure 3.
(3) The free energy $G(Q)$ has a maximum at $Q = \frac{1}{4}$. In addition to states 25 and 33, there are five other states that correspond to this value of Q. These are 4, 15, 31, 32, and 34. They form a letter V on the triangular representation of the states in figure 3. These states are connected by the exclamation marks in the figure.
(4) The ϕ-values are largest for the amino acids at and near the turn in the hairpin and are zero near the terminals.
(5) There are seven states, all corresponding to $Q = \frac{1}{4}$ and thus listed in item (3), that are closest to the kinetically derived ϕ-values.
(6) Time evolution of the density matrix is smooth, i.e. rapid changes in structural measures that may occur in single trajectories (as in a Monte Carlo simulation) wash out on averaging over many trajectories.

The fourth point above suggests that the ϕ-values do indeed provide a meaningful characterization of the average conformation corresponding to the transition states 25 and 33. The remaining points, however, suggest that the theoretical approaches to identify the transition states that have been proposed so far are correct only partially. In particular, points 3 and 5 suggest that among the putative transition states there is one state, 34, which is a local energy minimum which cannot be a true transition state by definition. Nevertheless, it is clear that taking these various approaches together and identifying the common finding does agree with the correct selection of the transition states.

There is a possibility, however, that the 67-state model, though two-state, somehow is not representative of proteins so it is relevant to consider other models and benchmark the various approaches again. We now consider three other simple models.

3. The four-state model

We now consider the four-state model proposed by Merlo *et al* [26]. The allowed transitions between the states are represented schematically in figure 4. Thus one can get from state D to state N either through the state B or through state A, but not directly. Therefore, there are only two folding paths in this system. A is the transition state if its free energy is lower than that of state B.

The transition rates in this model are defined by

$$w_{nm} = \frac{1}{\tau_0}(1 + \mathrm{e}^{(G_n - G_m)/RT})^{-1}, \tag{6}$$

where G_n denotes the free energy in state n. The resulting M matrix has four eigenvalues which are given by $\lambda = 0$,

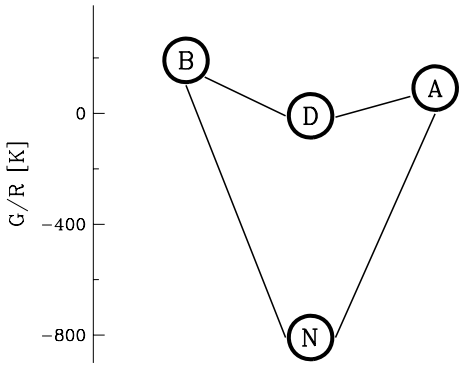

Figure 4. The 4-state model of Merlo *et al* [26]. The lines indicate the allowed kinetic transitions.

$1-q$, $1+q$, and 2, where

$$q = [1 - e^{(G_N - G_A - G_B)/RT}] \times [(1 + e^{-G_A/RT})(1 + e^{-G_B/RT}) \times (1 + e^{(G_N - G_A)/RT})(1 + e^{(G_N - G_B)/RT})]^{-\frac{1}{2}} \quad (7)$$

and $-1 < q < 1$. We require that state N must be of much lower free energy than the remaining states and that there must be a barrier to go to state D. Under these conditions, the Boltzmann factors $b_1 = e^{(G_N - G_A - G_B)/R}$, $b_2 = e^{(G_N - G_A)/R}$, $b_3 = e^{(G_N - G_B)/R}$ are negligible and $q \approx [(1+h_1)(1+h_2)]^{-1/2}$, where $h_1 = e^{(-G_A/R)}$ and $h_2 = e^{(-G_B/R)}$. The system is endowed with the two-state behavior because the eigenvalues are well separated. The second eigenvalue is small and its inverse yields the longest relaxation time. In the following, we choose $G_N/R = -800$, $G_A/R = 100$, $G_B/R = 200$ K, and $G_D = 0$ which leads to a specific heat maximum occurring at 300 K. For this choice and for $T = 300$ K, we have $h_1 = 0.72$, $h_2 = 0.51$ and the b_i factors are indeed negligible (b_1, b_2, and b_3 are 0.03, 0.05, and 0.04 respectively).

Merlo *et al* [26, 17] identify relaxation with folding and have not determined the eigenvalues of the M_f and M_u matrices. Deriving compact analytic expressions for these other matrices has turned out to be difficult and we have opted for a numerical solution. For that purpose, we take the just stated values of the free energies. The solutions for the relaxation, folding and unfolding rates are shown in figure 5. We observe that below 400 K, $k \approx k_f + k_u$ which is characteristic of two-level systems. Above this temperature, the b_i factors cease to be negligible. At very low temperatures, $k_u \approx 0$ and $k_f \approx k$. At very high temperatures, k_f approaches k_u whereas k clearly separates from $k_f + k_u$.

For the values of G_n that we consider, the transition state is clearly state A. Dill and collaborators [26, 17] have suggested that the transition state can be determined by identifying the largest components of the eigenvector corresponding to the longest lasting relaxation process in the system. The four components of the eigenvector corresponding to the longest finite relaxation time, $1/k$, are shown in figure 6 as a function of T. At each T, the weight of the state N dominates, followed by that of state D. At $T = 300$, the weights of states D, A, B, and N are 0.53, 0.16, 0.12, and -0.82 respectively. The relatively large contribution of state D is due to the fact that

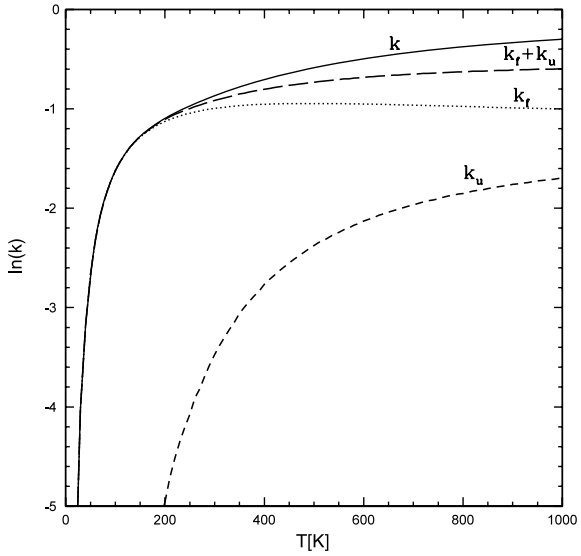

Figure 5. Logarithms of folding, unfolding and relaxation rates as functions of T in the 4-state model.

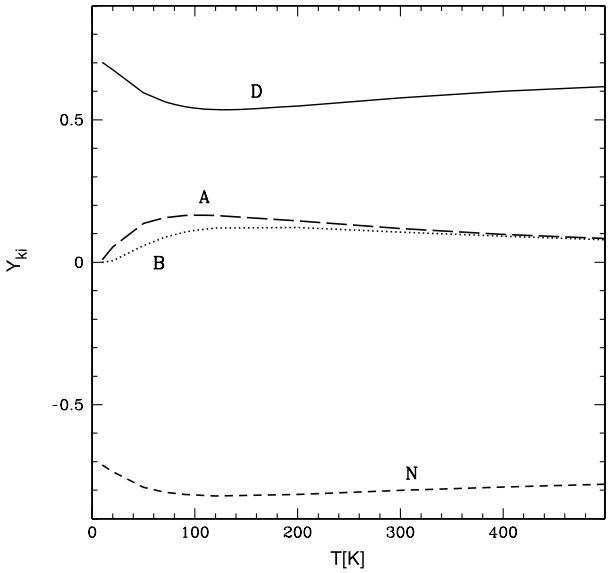

Figure 6. Components of the eigenvector corresponding to the longest finite relaxation time $1/k$ as a functions of T.

state D acts as a local energy minimum in this model. Thus, in terms of the absolute values, the transition state A comes as a third in importance. This fact indicates that the method may distinguish the transition state provided one discards the contributions of the D and N states in the eigenvector. In larger systems, all local free energy minima also need to be discarded in the eigenvector to get to a transition state in this way.

Another suggested way to identify the transition state is by determining the so called edge states for which the probabilities to flow to the N- and D-states are nearly the same [14, 15]. One can determine such states by making both D and N to be the probability sinks. In the four-state model, neither state A nor B is, strictly speaking, an edge state because of a too small number of states available. Figure 7,

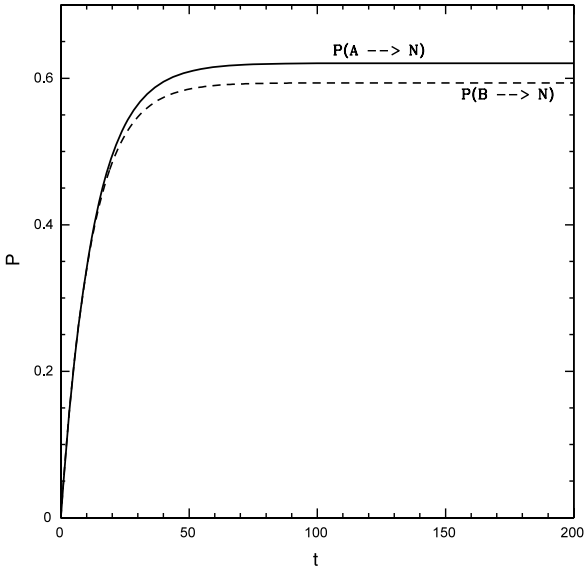

Figure 7. Probability of reaching state N, when starting from states A and B at $T = 300$ K as a function of time.

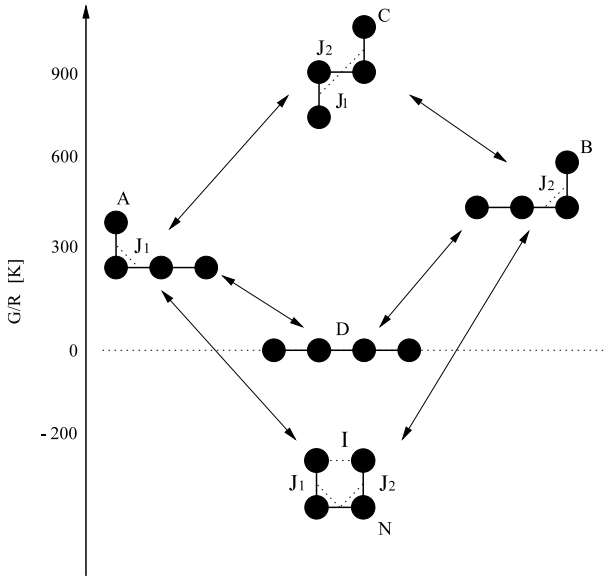

Figure 8. The five possible conformations in the five-state (or four-bead) model. The arrows indicate the allowed transitions. The energy couplings that are operational in each conformation are indicated by the dotted lines.

corresponding to $T = 300$ K, shows that when one starts in state A, then the probability to reach state N is close to 0.62. On the other hand, when one starts in state B this probability becomes close to 0.59, suggesting, quite wrongly, that B is more likely to be the transition state of this system than A because the corresponding asymptotic value is closer to 0.5. These results have been obtained by evolving the density matrix with the use of matrix M in which both N and D states act as probability sinks [22].

4. The five-state model

We now consider models with explicit conformations that come with well defined energies and not free energies. In the transition rates w_{nm} differences in free energies are now replaced by differences in energies.

The first of these models involves four beads on the square lattice. This system has only five conformations. It comprises four beads placed on the square lattice. The chain can be in five different conformations that are shown in figure 8. The kinetic moves involve turning the head or tail beads to an available lattice site. The conformations are linked kinetically as indicated in figure 8. In order to introduce some complexity into the body of four beads, we endow the system with three different couplings. The I-coupling acts between the terminal beads whenever they become neighbors on the lattice [27]. In this system this happens only in state N. In order to remove the degeneracy between the remaining four conformations, we introduce couplings J_1 and J_2 which link centers of the bonds (like a backbone–backbone hydrogen bond link in proteins). These couplings become active when the nearby bonds form a kink.

The Hamiltonian is then given by

$$H = -J_1\Delta_{123} - J_2\Delta_{234} - I\Delta_{14} \tag{8}$$

where Δ_{kl} is 1 when bead k is a lattice neighbor of bead l and 0 otherwise; Δ_{klm} is 1 if the consecutive beads k, l, and m make a 90° kink and 0 otherwise. Our choice of the parameters is as follows: $J_1/R = -300$ K, $J_2/R = -600$ K, and $I/R = 1200$ K. The linking I coupling is attractive. We make the J_l couplings repulsive to introduce some resistance to kink formation. In this way, a barrier arises when one starts in state D (the straight line conformation) and attempts to make a kinetic move. Once the barrier is crossed, the system favors establishing the strong I-bond which provides stability to state N. This choice of values of these parameters yields the specific heat with a broad maximum at 200 K. The folding temperature, at which the equilibrium occupancy of the native state is $\frac{1}{2}$, is 450 K. The system becomes two-state at temperatures below 500 K and the kinetic rates have plots similar to those shown in figure 5.

The free energy of a conformation is determined by adding the $TRP_i^{\rm eq}\ln(P_i^{\rm eq})$ term to the energy of the ith conformation, where $P_i^{\rm eq}$ is the normalized Boltzmann factor describing the equilibrium probability of occupation of this conformation. The optimal free-energy trajectories that link D with N go through state A which makes A the transition state of this system. Coupling J_1 is operational in this state.

The eigenvector corresponding to the smallest relaxation eigenvalue of the M matrix has features which are similar to those of the four-state system considered in the previous section. At $RT = 325$ K, the weights of the N, D, A, B, and C states are 0.65, −0.76, 0.07, 0.03, and 0.01 respectively. The dominating components are those which correspond to states N and D. However, the third important contribution is again associated with the transition state A like in the four-state model discussed in the previous section.

No state is a precise edge state in this system. When evolving the system from state A, B, and C, the asymptotic

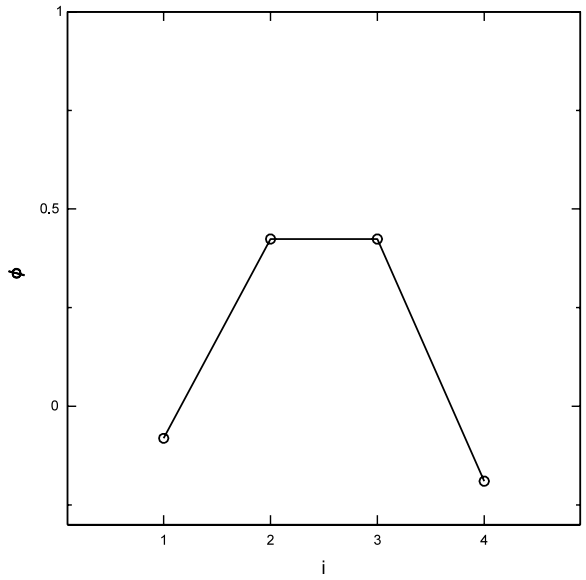

Figure 9. The ϕ values determined kinetically in the five-state model.

probabilities of arriving at state N are 0.52, 0.51, and 0.53 respectively. Thus all of these states are nearly edge-like (state B is the closest to the 0.5 criterion) and yet it is only the lowest free energy state of them is the true transition state.

We now consider the ϕ-value analysis for this system. A mutation at the ith bead is accomplished by making a 5% adjustment in all couplings that involve bead i. Each bead is linked to two possible couplings, but the pairs are all different. The resulting ϕ-values at $T = 325$ K are shown in figure 9. The maximal values of ϕ are found to be at the center beads. The terminal ϕ-values are negative. This could be due to the strong stabilization of the native state by the I bond.

Following references [20, 21], we now determine the local Q_i values. The local Q_i value for the ith bead in state S is defined as

$$Q_i(\mathrm{S}) = \frac{n_i(\mathrm{S})}{n_i(\mathrm{N})} \tag{9}$$

where $n_i(\mathrm{N})$ is the number of bonds that bead i participates in when in the native state, and $n_i(\mathrm{S})$ is the similarly defined number of bonds in state S. For instance, the transition state A makes the coupling J_1 operational and this coupling involves three beads. The resulting sets of the Q_i values are $\{\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 0\}$, $\{0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$, $\{\frac{1}{2}, 1, 1, \frac{1}{2}\}$, $\{0, 0, 0, 0\}$, and $\{1, 1, 1, 1\}$ for states A, B, C, D, and N respectively. In state C, the Q_is are bigger for the center beads and lower at the terminal beads. States A and B have asymmetric sets of the Q_i values but taken together they make the terminals to have lower Q_i than at the center. So states A, B, and C, when combined, would be consistent with the kinetically derived ϕ-values but it is only conformation A that is the transition state. This seems to suggest that the approach base on the local Q_i values may work for more realistic models of proteins but it fails for the five-state toy model.

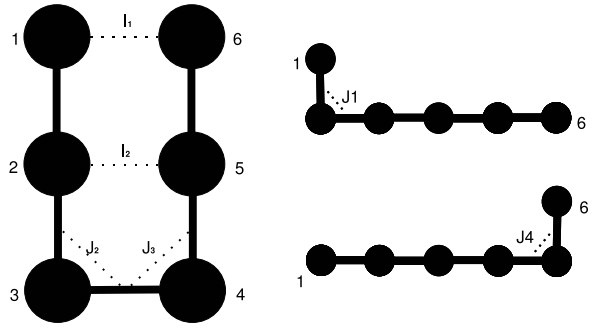

Figure 10. On the left: the native conformation in the 36-state model. The dotted lines indicate the energy couplings that are operational in the native conformation. On the right: Two conformations with one operational non-native contact each. These contacts are indicated by the dotted lines.

5. The 36-state model

In order to obtain a still richer behavior we now consider a six-bead model on the square lattice with one I-like coupling and four J-like couplings. The Hamiltonian is given by

$$H = -J_1\Delta_{123} - J_2\Delta_{234} - J_3\Delta_{345} - J_4\Delta_{456} - I_1\Delta_{16} - I_2\Delta_{25}, \tag{10}$$

where the stabilizing bead–bead interactions arise in pairs 1–6 and 2–5. The native conformation and two high energy one-kink conformations are shown in figure 10. The latter involve non-native couplings J_1 or J_4 (so the system is not Go-model-like). These non-native bonds are essential in the model since they allow for introduction of energy barrier for trajectories connecting the N and D states. We choose $J_1/R = 300$ K, $J_2/R = 100$ K, $J_3/R = 100$ K, $J_4/R = 800$ K, $I_1/R = 1600$ K and $I_2/R = 100$ K to ensure that any path from N to D must cross an energetic barrier. Unlike the previously considered system, all couplings in this model are attractive. With these couplings, the specific heat has a maximum around 200 K and the folding temperature is close to 250 K. The ϕ-values are then determined at $T = 225$ K. The kinetic rates of the 36-state model are shown in figure 11. This model is seen to have the two-state behavior ($k = k_\mathrm{f} + k_\mathrm{u}$) below 300 K. At temperatures larger than 230 K the unfolding begins to dominate over folding ($k_\mathrm{u} > k_\mathrm{f}$).

This system also displays (figure 12) chevron-like behavior when the kinetic rates are plotted (at 225 K) as a function of the concentration, x of the denaturant. One can mimic effects of the denaturant by adjusting all coupling in proportion to x [22]. Specifically, $J_i(x) = J_i(1-x)$, where $i = 1, \ldots, 4$ and $I_i(x) = I_i(1-x)$ for $i = 1, 2$.

There are many directed pathways in the free-energy space connecting the denaturated state with the native one. Some of them involve six kinetic steps and they correspond to the shortest paths as measured by the number of moves to make. The paths are shown in figure 13 together with the relevant part of the free-energy landscape. The paths correspond to a mostly downhill movement but three are several local free-energy maxima denoted as 'a', 'b', and 'c' (see also figure 14). The three states have a common L-like shape formed by the

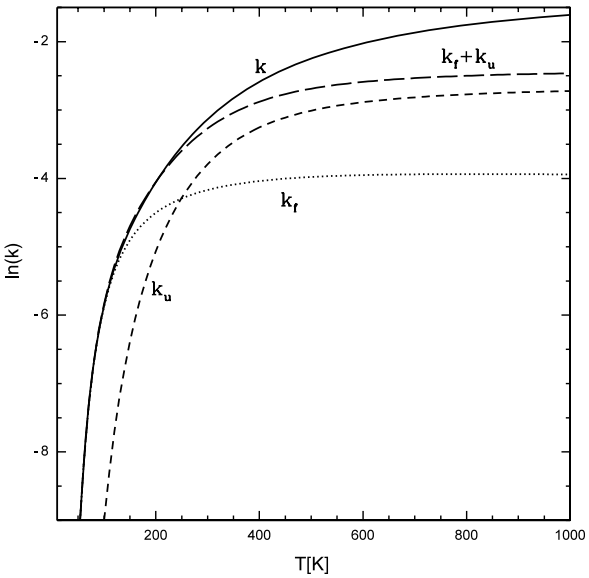

Figure 11. The logarithms of folding, unfolding and relaxation rates as a functions of T in the 36-state model.

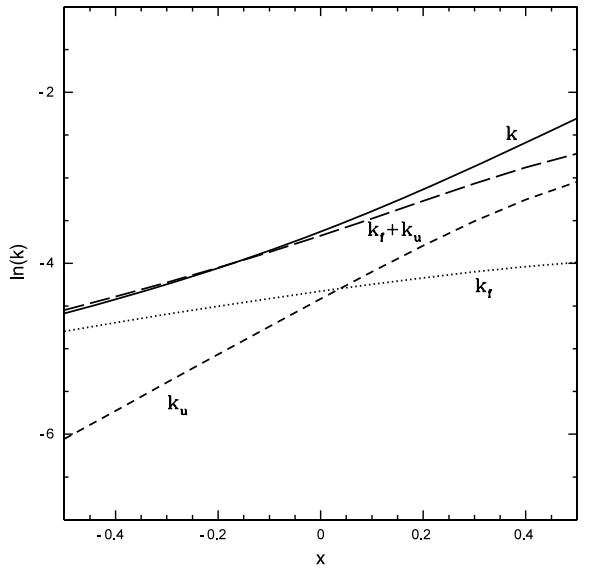

Figure 12. The logarithms of folding, unfolding and relaxation rates as a functions of the 'concentration of the denaturant', x, in the 36-state model.

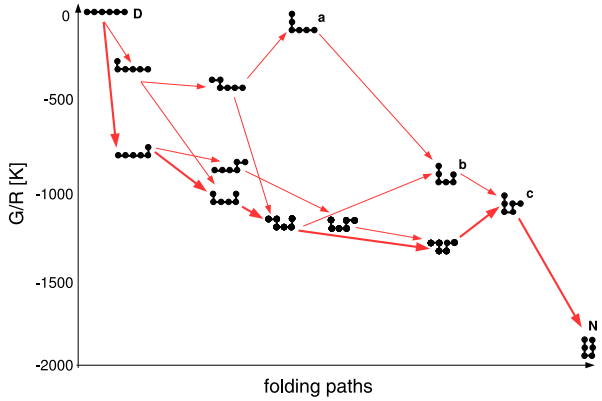

Figure 13. Some of the best folding paths in the 36-state model. The arrows indicate possible kinetic transitions and the optimal path corresponds to the thicker arrows.

(This figure is in colour only in the electronic version)

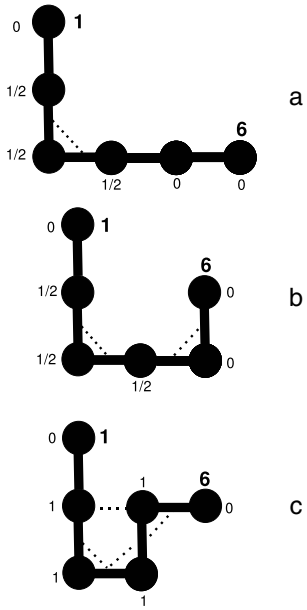

Figure 14. The local free-energy maxima states in the 36-state model. The values of Q_i are indicated at each bead. Conformation c corresponds to the transition state.

first four beads and getting to them involves breaking the same strong bond J_1. The smallest free-energy mountain to climb on going from D to N is the 'c' state so 'c' is the transition state. However, there are certain subtleties: on going down from D to N, the smallest single step increase in the free energy is associated with climbing 'a', i.e. to the highest local maximum state. Nevertheless the optimal path should be the best both for folding and for unfolding and the smallest effort on moving from N do D involves taking state 'c' at the climbing step.

The kinetically determined ϕ-values are shown in figure 15 and the values of the local parameters Q_i for states 'a' through 'c' are listed in figure 14. None of these sets of Q_i is consistent with the ϕ-values. For instance, for state 'c' the fourth and fifth beads should have the ϕ-values as large as for the second and third beads if judged by the values of Q_i. Averaging the Q_is over the three states would agree with the ϕ values better but there is no reason to make the average.

One cannot turn the J_1 and J_4 bonds off in the model, because they play the key role here. However, if we reduced the couplings I_2, J_2 and J_3 to zero, the three local maxima states would have no native bonds at all and the corresponding Q_is would be all equal to zero. In this limit, however, the ϕ-values look similar to those shown in figure 15 so the Q_i-based approach [20, 21] becomes invalid in this system.

In conclusion, the small explicit-conformation models considered here indicate existence of problems with the determination of the transition state by means other than enumeration of the optimal paths. These problems may just

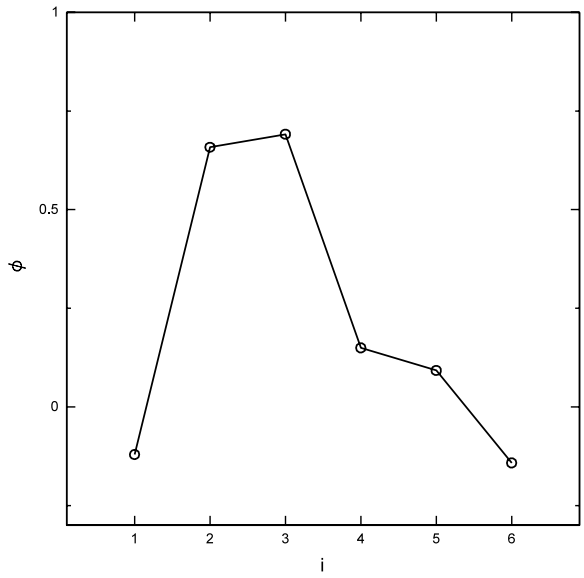

Figure 15. The solid lines connect the kinetically determined ϕ-values in the 36-state model.

be related to the remoteness of such small-sized systems from physics of real proteins. Theoretical understanding these issues requires considering less artificial and, in particular, larger systems. Perhaps the next step to do is to consider the twelve-bead model on the square lattice that comes with 15 037 conformations and yet can be solved exactly [28].

Acknowledgment

This work was funded by the MNiSW grant N202 021 31/0739.

References

[1] Eyring H and Stern A E 1939 *Chem. Rev.* **24** 253–70
[2] Fersht A R 1998 *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (New York: Freeman)
[3] Jackson S E and Fersht A R 1991 *Biochemistry* **30** 10428–35
[4] Otzen D S, ElMasry N, Jackson S E and Fersht A R 1994 *Proc. Natl Acad. Sci. USA* **91** 10422–5
[5] Itzhaki L S, Otzen D E and Fersht A R 1995 *J. Mol. Biol.* **254** 260–88
[6] Baldwin R L and Rose G D 1999 *Trends Biochem. Sci.* **24** 26–33
[7] Abkevich V I, Gutin A M and Shakhnovich E I 1994 *Biochemistry* **33** 10026–36
[8] Fersht A R 1997 *Curr. Opin. Struct. Biol.* **7** 3–9
[9] Oxender D L and Fox C F (ed) 1987 *Protein Engineering* (New York: Wiley)
[10] Robson B and Garnier J 1988 *Introduction to Proteins and Protein Engineering* (New York: Elsevier)
[11] Cleland J L and Craik C S (ed) 1996 *Protein Engineering: Principles and Practice* (New York: Wiley–Liss)
[12] Matouschek A, Kellis J T Jr, Serrano L and Fersht A R 1989 *Nature* **342** 122–6
[13] Matouschek A, Kellis J T Jr, Serrano L, Bycroft M and Fersht A R 1990 *Nature* **346** 440–5
[14] Pande V S and Rokhsar D S 1999 *Proc. Natl Acad. Sci. USA* **96** 1273–8
[15] Pande V S and Rokhsar D S 1999 *Proc. Natl Acad. Sci. USA* **96** 9062–7
[16] Clementi C, Nymeyer H and Onuchic J N 2000 *J. Mol. Biol.* **298** 937–53
[17] Ozkan S B, Bahar I and Dill K A 2001 *Nat. Struct. Biol.* **8** 765–9
[18] Li A and Daggett V 1994 *Proc. Natl Acad. Sci. USA* **91** 10430–4
[19] Kazmirski S L, Wong K-B, Freund S M V, Tan Y-J, Fersht A R and Daggett V 2001 *Proc. Natl Acad. Sci. USA* **98** 4349–54
[20] Vendruscolo M, Paci E, Dobson C M and Karplus M 2001 *Nature* **409** 641–5
[21] Paci E, Clarke J, Steward A, Vendruscolo M and Karplus M 2003 *Proc. Natl Acad. Sci. USA* **100** 394–9
[22] Chang I, Cieplak M, Banavar J R and Maritan A 2004 *Protein Sci.* **13** 2446–57
[23] Munoz V, Thompson P A, Hofrichter J and Eaton W A 1997 *Nature* **390** 196–9
[24] Munoz V, Henry E R, Hofrichter J and Eaton W A 1998 *Proc. Natl Acad. Sci.* **95** 5872–9
[25] Munoz V and Eaton W A 1999 *Proc. Natl Acad. Sci. USA* **96** 11311–6
[26] Merlo C, Dill K A and Weikl T R 2005 *Proc. Natl Acad. Sci. USA* **102** 10171–5
[27] Dill K A, Bromberg S, Yue K Z, Fiebig K M, Yee D P, Thomas P D and Chan H S 1995 *Protein Sci.* **4** 561–602
[28] Cieplak M, Henkel M, Karbowski J and Banavar J R 1998 *Phys. Rev. Lett.* **80** 3654–7