Multiple Folding Pathways of the SH3 Domain

Jose M. Borreguero,* Feng Ding,† Sergey V. Buldyrev,* H. Eugene Stanley,* and Nikolay V. Dokholyan†
*Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts; and †Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

ABSTRACT Experimental observations suggest that proteins follow different folding pathways under different environmental conditions. We perform molecular dynamics simulations of a model of the c-Crk SH3 domain over a broad range of temperatures, and identify distinct pathways in the folding transition. We determine the kinetic partition temperature—the temperature for which the c-Crk SH3 domain undergoes a rapid folding transition with minimal kinetic barriers—and observe that below this temperature the model protein may undergo a folding transition by multiple folding pathways via only one or two intermediates. Our findings suggest the hypothesis that the SH3 domain, a protein fold for which only two-state folding kinetics was observed in previous experiments, may exhibit intermediate states under conditions that strongly stabilize the native state.

INTRODUCTION

Recent experimental studies indicate that several proteins exhibit simultaneously a variety of intermediates and folding pathways. Kiefhaber (1995) identified at low denaturant concentration a fast pathway (50 ms) in the folding of lysozyme with no intermediates and a slow phase (420 ms) with well-populated intermediates. In other studies, authors observed the formation of a kinetic intermediate in the folding of villin 14T upon temperature decrease (Choe et al., 1998), as well as extinction of a slow pathway in the folding of the P4–P6 domain upon changes in ion concentration (Silverman et al., 2000). Kitahara and Akasaka (2003) studied a pressure-stabilized intermediate of ubiquitin, identified as an off-pathway intermediate. All these studies suggest that environmental conditions favor some folding pathways over others.

Theoretical efforts in the study of protein folding (Bryngelson and Wolynes, 1989; Eaton et al., 2000; Fersht and Daggett, 2002; Karplus and McCammon, 2002; Klimov and Thirumalai, 2002; Ozkan et al., 2002; Pande et al., 2000; Plotkin and Onuchic, 2002; Thirumalai et al., 2002; Tiana and Broglia, 2001) have focused on small, single domain proteins. It is found in experiments (Jackson, 1998) that the majority of these proteins undergo folding transition with no accumulation of kinetic intermediates in the sampled range of experimental conditions. However, kinetics studies of other two-state proteins (Bachmann and Kiefhaber, 2001; Khorasanizadeh et al., 1996) suggest the presence of short-lived intermediates that cannot be directly detected experimentally. In a recent analysis, Sanchez and Kiefhaber (2003) explained the curved Chevron plots—the nonlinear dependence of folding and unfolding rates on denaturant concentration (Fersht, 2000; Ikai and Tandford, 1973; Matouschek et al., 1990)—of 17 selected proteins by assuming the presence of an intermediate state. In addition, recent molecular dynamics studies of the SH3 domain have suggested the presence of a core-hydrated, native-like intermediate in the latest stages of the folding process (Cheung et al., 2002), as well as an intermediate state in an isolated fragment (Gnanakaran and García, 2003). Led by these studies, we hypothesize that single domain proteins may exhibit intermediates in the folding transition under suitable environmental conditions.

To test our hypothesis, we perform a molecular dynamics study of the folding pathways of the c-Crk SH3 domain (Berman et al., 2000; Branden and Tooze, 1999; Wu et al., 1995; PDB access code 1cka). The SH3 domain is a family of small globular proteins which has been extensively studied in kinetics and thermodynamics experiments (Filimonov et al., 1999; Grantcharova and Baker, 1997; Grantcharova et al., 1998; Guerois and Serrano, 2000; Knapp et al., 1998; Martinez et al., 1999; Riddle et al., 1999; Viguera et al., 1994; Villegas et al., 1995). We select c-Crk (57 residues, Fig. 1 *a*) as the SH3 domain representative, and present our results in terms of the following sequence segments: 1), N-terminal (residues 1–7); 2), RT-loop (8–20); 3), Diverging turn (21–30); 4), n-Scr loop (30–38); 5), Distal hairpin (39–50); 6), 3_{10} α-helix (51–53); and 7), C-terminus (54–57).

There is a growing body of evidence (England et al., 2003; Fersht, 2000b; Plaxco et al., 1998) supporting the hypothesis that the folding kinetics and thermodynamics of globular proteins, and in particular the SH3 domain members (Grantcharova et al., 1998; Riddle et al., 1999), are largely determined by the topology of the native state rather than by the amino acid sequence. Thus the Gō model of interactions, based on the topology of the native state, is a suitable tool to study the folding process of the c-Crk SH3 domain. Our previous thermodynamic studies (Borreguero et al., 2002) of c-Crk with this model revealed the existence of only folded and unfolded states at equilibrium conditions, in agreement with experimental results (Filimonov et al., 1999; Viguera et al., 1994; Villegas et al., 1995). Both states coexist with

Submitted January 2, 2004, and accepted for publication March 10, 2004.

Address reprint requests to J. M. Borreguero, E-mail: jmborr@bu.edu.

0006-3495/04/07/521/13 $2.00

doi: 10.1529/biophysj.104.039529

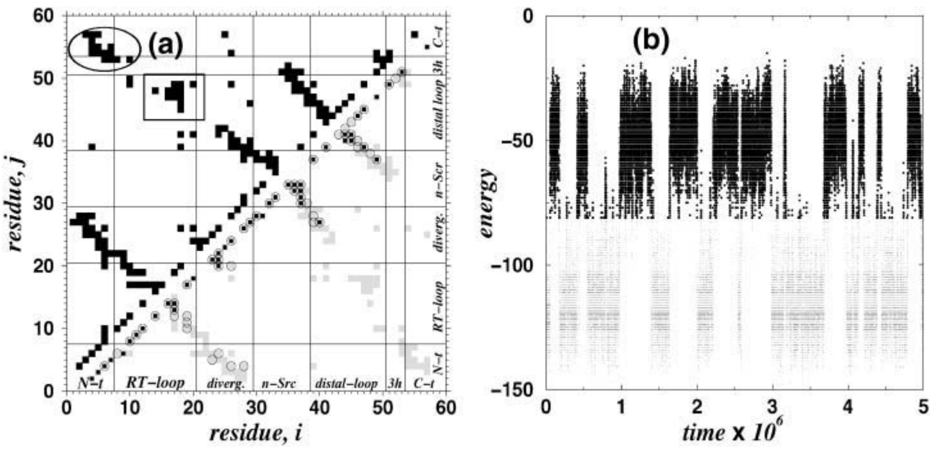

FIGURE 1 (*a*, *upper triangle*) The c-Crk SH3 domain contact map with 160 native contacts. The secondary structure elements are the clusters of contacts that are organized perpendicularly to the map diagonal. Long-range contacts between the two termini are enclosed in the circle, and long-range contacts between the RT-loop and the distal hairpin are enclosed in the square. (*a*, *lower triangle*) Contacts with a high frequency, $f \geq 0.75$ (*shaded circles*), have a sensitivity of 79% and a specificity of 65% to detect the set of native contacts (*solid squares*) with a distinctive NMR signal (Kortemme et al., 2000) in the denatured state. All other possible native contacts are depicted in shading. (*b*) Long simulation at T_F to compute the frequency map of the unfolded state. We sample the protein conformation at regular time intervals of 10 t.u. and only when the energy of the protein corresponds to an unfolded protein (*dark region*).

equal probability at the folding transition temperature $T_F = 0.626$, at which the temperature dependence of the potential energy has a sharp change, and the specific heat has a maximum (experimentally, this temperature corresponds to 67°C; Filimonov et al., 1999).

The Gō model has been successfully applied to the study of two-state proteins (Clementi et al., 2003; Ding et al., 2002; Zhou and Karplus, 1997), but there are no studies to assess the performance of the model for three-state proteins. We test the ability of our model to reproduce intermediate states in a variety of three-states proteins and apparent two-states proteins (Sanchez and Kiefhaber, 2003). We select proteins RNase (Yamasaki et al., 1995), SNAse (Walkenhorst et al., 1997), Barnase (Fersht, 2000), CheY (Lopez-Hernandez and Serrano, 1996), Im7 (Ferguson et al., 1999), and P16 (Tang et al., 1999), for which, at the experimental conditions studied, the existence of intermediates in the folding process have been observed experimentally. We also select proteins Gelsolin-WT (Isaacson et al., 1999) and U1A (Otzen et al., 1999), for which authors observed a nonlinear dependence of the observed folding/unfolding rates versus urea concentration, although evidence was not conclusive on the existence of intermediates.

In addition to the folding kinetics investigation, we address the relevance of the initial unfolded state for the subsequent evolution of the folding process. Studies suggest that the protein may retain part of the native structure even under strong denaturing conditions (García et al., 2001; Millet et al., 2002; Shortle and Ackerman, 2001; Zagrovic et al., 2002). A native-like structure of the unfolded state speeds up the conformational search to the native state that the protein must perform. In addition, a native-like structure limits the number of possible folding intermediates, and guarantees that the structure of the intermediates will share some similarities with that of the native state. We perform studies of the initial unfolded state under different temperature conditions, and compare our results in the particular temperature conditions where experiments are available (Kortemme et al., 2000).

We determine the kinetic partition temperature (Thirumalai et al., 2002), T_{KP}, below which the model c-Crk protein exhibits slow folding pathways and above which the protein undergoes a cooperative folding transition with no accumulation of intermediates. Below T_{KP}, we study the presence of one or two intermediates in the slow folding pathways and determine their structures. We find that one of the intermediates populates the folding transition for temperatures as high as T_{KP}, when the intermediate is not stabilized.

MATERIALS AND METHODS

Model protein and interactions

We adopt a coarse-grained description of the protein by which each amino acid is reduced to its C_β atom (C_α in case of Gly). Details of the model, the surrounding heat bath, and the selection of structural parameters are discussed in detail in a previous study (Borreguero et al., 2002). The selection of the set of interaction parameters among amino acids is of crucial importance for the resulting folding kinetics of the model protein (Pande et al., 2000; Plotkin and Onuchic, 2002; Thirumalai et al., 2002). We employ a variant of the Gō model of interactions (Gō and Abe, 1981)—a model based solely on the native topology—in which we prevent formation of nonnative interactions, since we are solely interested in the role that native topology and native interactions may have in the formation of intermediates.

We perform simulations and monitor the time evolution of the protein and the heat bath with the discrete molecular dynamics algorithm (DMD), which uses step potentials (Alder and Wainwright, 1959; Dokholyan et al., 1998; Rapaport, 1997; Zhou et al., 1997). The earliest molecular dynamics simulations were performed with the discrete algorithm, before the advent of continuous potentials. DMD has a higher speed performance than conventional molecular dynamics, making DMD a choice tool to simulate the folding of proteins.

Frequencies and folding simulations

To calculate the frequency map at $T = 1.0$, we probe the presence of the native contacts in each of the 1100 initially unfolded conformations. Then, we compute the probability of each native contact to be present. To calculate the frequency map at T_{target}, we select one particular folding transition and we probe the presence of the native contacts during the time interval that spans after the initial relaxation and before the simulation reaches the folding time t_F. To compute t_F, we stop the folding simulation when 90% of the native contacts form. Then, we trace back the folding trajectory and record t_F when the root mean-square deviation (RMSD), with respect to the native state, becomes smaller than 3 Å. We consider all protein conformations occurring for $t > t_F$ as belonging to the folded state and of no relevance to the folding transition.

Contact formation times

Following the kinetics of each particular native contact provides us with a detailed picture of the folding process. We compute the fraction of the 1100 folding simulations for which native contact (i, j) is present at time t and temperature T, $p_{ij}(t, T)$. For this particular contact, we estimate the characteristic contact formation time $t_{ij}(T)$ with the relation $p_{ij}(t_{ij}, T)$–$p_{ij}(0, T) = e^{-1} \times (p_{ij}(\infty, T)$–$p_{ij}(0, T))$. When $p_{ij}(t, T)$ is a single exponential distribution, then $t_{ij}(T)$ coincides with the average time of the distribution.

Similarity score function

We introduce the similarity score function, $S = (a/23)(15–b)/15$, where a is the number of native contacts belonging to the set of contacts C_1, and b is the number of native contacts belonging to set C_2 (Fig. 5 *e*). C_1 has 23 contacts and C_2 has 15 contacts. If the protein is unfolded, then $a \approx b \approx 0$, thus $S \approx 0$. Similarly, if the protein is folded, then $a \approx 23$ and $b \approx 15$, thus $S \approx 0$ again. Finally, if the protein adopts the intermediate I_1 structure, then $a \approx 23$ and $b \approx 0$, thus $S \approx 1$.

RESULTS

Unfolded state at equilibrium

To assess the ability of the protein model to reproduce the unfolded state of c-Crk at equilibrium conditions, we calculate at T_F the frequency map—the probability of two amino acids forming a contact—of the unfolded state from a long simulation of 10^6 time units (t.u.) (Fig. 1 *b*). Unfortunately, there are no detailed experimental studies on the structure of the unfolded state of the c-Crk SH3 domain. We therefore compare the computed frequency map to the experimental results on the denatured state of the homologous chicken α-spectrin SH3 domain protein (Kortemme et al., 2000). Before comparison, we first structurally align (Borreguero et al., 2002; Holm and Sander, 1996) the sequence of chicken α-spectrin SH3 domain to the sequence of c-Crk SH3 domain that we employ in our studies (sequence identity 34%, RMSD = 2.4 Å). We find that the set of native contacts with a high probability to form ($p >$ 0.75) can predict 37 native contacts out of the 47 native contacts (79% sensitivity) with a distinctive NMR signal found by Kortemme et al. (2000) (*lower triangle* of Fig. 1 *a*) in the denatured state of chicken α-spectrin. Alternatively, we find that out of the 57 predicted native contacts, 37 are correct (65% specificity). We cannot predict the set of nonnative contacts with a distinctive NMR signal, since our model does not allow us to calculate frequencies for nonnative contacts.

Relaxation of the initial unfolded state

Our initially unfolded state ensemble consists of 1100 protein conformations that we sample from a long equilibrium simulation at a very high temperature, $T_0 = 1.0$, at equal time intervals of 10^4 t.u. This time separation is long enough to ensure that the sampled conformations have low structural similarity among themselves. We calculate the frequency map of this unfolded state. At $T = 1.0$, only nearest and next-nearest contacts have high frequency, and the frequency decreases dramatically with the sequence separation between the amino acids.

When we quench the system from $T = 1.0$ to a target temperature, T_{target} (see Materials and Methods), the system relaxes in ~1500 t.u. Due to the finite size of our heat bath, the heat released by the protein upon folding increases the final temperature of the system by 0.03 energy units above T_{target}. After relaxation, the protein stays for a certain time in the unfolded state, then undergoes a folding transition. During this time interval, the protein explores unfolded conformations at equilibrium, and we calculate the frequency map of the unfolded state for different target temperatures.

At $T_{\text{target}} = 0.64$, slightly above T_F, the secondary structure is unstable (Fig. 2 *a*), with average frequency $\bar{f} = 0.24$ (see Materials and Methods). Successful folding requires the cooperative formation of contacts throughout the protein in a nucleation process (Borreguero et al., 2002; Ding et al., 2002). At $T_{\text{target}} = 0.54$, the secondary structure is more stable, although still conserving a degree of flexibility (Fig. 2 *b*, $\bar{f} = 0.50$). Then the conformational search for the native state is optimized by limiting the search to the formation of a sufficient number of long-range contacts. At $T_{\text{target}} = 0.33$, the lowest temperature studied, secondary structure elements form during the rapid collapse of the model protein in the first 1500 t.u. (Fig. 2 *c*, $\bar{f} = 0.73$). During collapse, some tertiary contacts—contacts between secondary elements—may also form. The formation of these contacts before the proper arrangement of secondary structure elements may lead the protein model to a kinetic trap. Finally, folding proceeds at this temperature through a thermally activated search for the native state.

Unfolding studies of three-state proteins

To address the ability of the model to distinguish between two- and three-state kinetics, we study a set of four different unfolding processes for several three-state and apparent two-

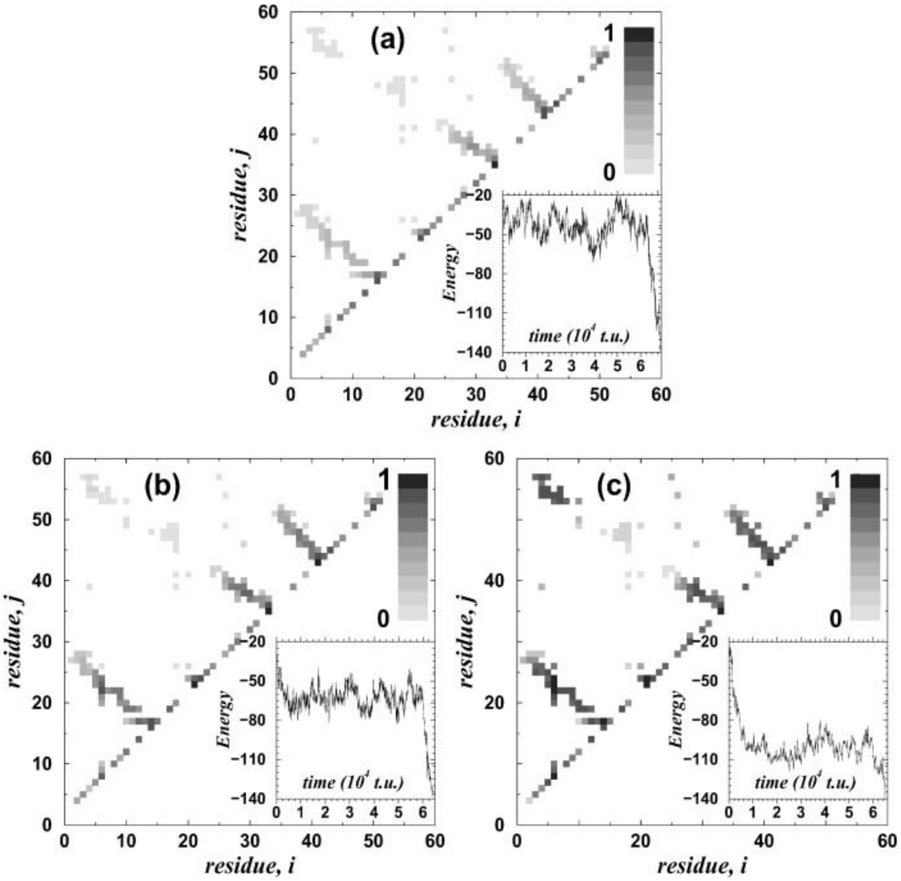

FIGURE 2 (*a*) Frequency map of one folding process at $T = 0.64$, slightly above T_F. We compute the frequencies for the particular folding transition whose potential energy trajectory we show in the inset (see Materials and Methods). Same for (*b*) $T = 0.54$ and (*c*) $T = 0.33$, the lowest temperature studied.

state proteins. Starting from a sufficiently low temperature under which the native state is stable, we steadily increase the temperature and therefore reduce the protein stability. When the temperature exceeds T_F, unfolding becomes an irreversible process. We monitor the time evolution of the potential energy and RMSD values, with respect to the native state, for a set of nine proteins (RNase, SNAse, Barnase, CheY, Im7, Im9, P16, Gelsolin-WT, and U1A), plus the c-Crk SH3 domain. We include c-Crk SH3 as the model of a two-state folder to which we can directly compare the results from the protein set.

Except for Barnase, we observe a variety of intermediate energy and RMSD values in the unfolding process of the selected proteins that do not correspond to the typical values of the native and unfolded states (Fig. 3). These values correspond to intermediate states in the unfolding process of the selected nine proteins. We observe one intermediate in rapid interconversion to the native state for Gelsolin WT. In the opposite extreme, RNase displays a long-lived intermediate before complete unfolding, whereas CheY shows a mixed behavior where the survival time of the intermediate increases with temperature. In addition, CheY exhibits a second intermediate before complete unfolding. Proteins 1UA, SNAse, and P16 show intermediates only before complete unfolding. These are on-pathway kinetic intermediates. The unfolding process of homologous proteins Im7 and Im9 is remarkably dissimilar. Whereas Im7 displays an intermediate, Im9 is a two-state protein. A similar scenario was observed in the folding process of these two proteins (Ferguson et al., 1999). Finally, we do not detect any intermediate in the unfolding process of c-Crk SH3, but only folded and unfolded states. Thus our protein model captures the essential properties that distinguish two-state from three-state proteins.

Kinetic partition temperature

To determine the temperature below which we can distinguish fast and slow folding pathways, we compute the distribution of folding times $p(t_F, T)$ (Fig. 4, *a–e*), as well as the average $\langle t_F \rangle$ (Fig. 4 *f*) and standard deviation σ_F. The ratio $r(T) \equiv \langle t_F \rangle / \sigma_F$ measures the average folding time in units of the standard deviation σ_F. This quantity characterizes the deviation of the distribution of folding times from the single exponential distribution, for which $r \equiv 1$. We expect $r \rightarrow 1$ for $T_{\mathrm{target}} > T_F$, because at these high temperatures the folding transitions become rare events and are single-exponential distributed. As we decrease T_{target}, we expect $r > 1$ just below T_F, because the folded state becomes more stable than the unfolded state, and the folding

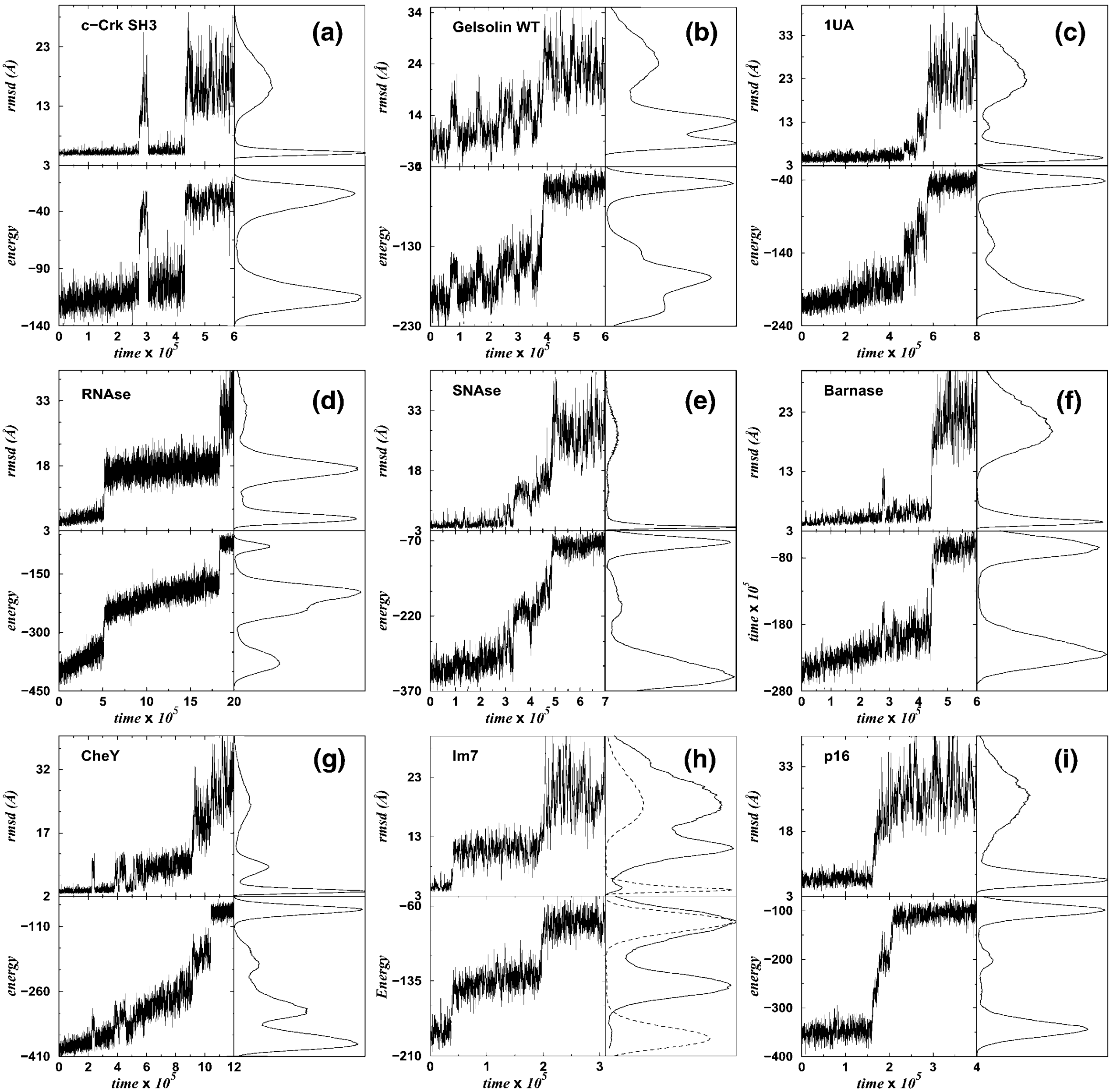

FIGURE 3 Time evolution of the energy and RMSD for one representative unfolding process. We compute the histograms with the set of four simulations for each protein. We group the proteins into 1), two-state unfolding (c-Crk SH3, barnase, Im9); 2), intermediate in rapid interconversion with the native state (Gelsolin WT); 3), intermediate with no interconversion (RNase, Im7); and 4), on-pathway kinetic intermediate (1UA, SNAse, P16). For comparison, we show histograms for Im9 as dashed lines in the box corresponding to Im7.

transitions are favored. Distributions with $r > 1$ indicate a narrow distribution centered in $\langle t_F \rangle$, so that most of the simulations undergo a folding transition for times of the order of the average folding time. However, if we continue decreasing $T_{\rm target}$, we expect some folding transitions to be kinetically trapped, and the folding time distribution will spread over several orders of magnitudes. Such distributions have $r < 1$. Thus, there is a temperature below T_F where the maximum of $r(T)$ occurs, and which signals the onset of slow folding pathways. We use the maximum of $r(T)$ to define $T_{\rm KP}$.

Fig. 4 *g* suggests that $T_{\rm KP} = 0.54$, which corresponds to a maximally compact distribution of folding times (assuming a linear relation between experimental and simulated temperatures and taking into account—Filimonov et al., 1999—that $T_F = 67°C$, we estimate $T_{\rm KP} \simeq 20°C$; see also Fig. 4 *d*). We find that r approaches 1 as we increase the temperature above $T_{\rm KP}$, and the distribution of folding times

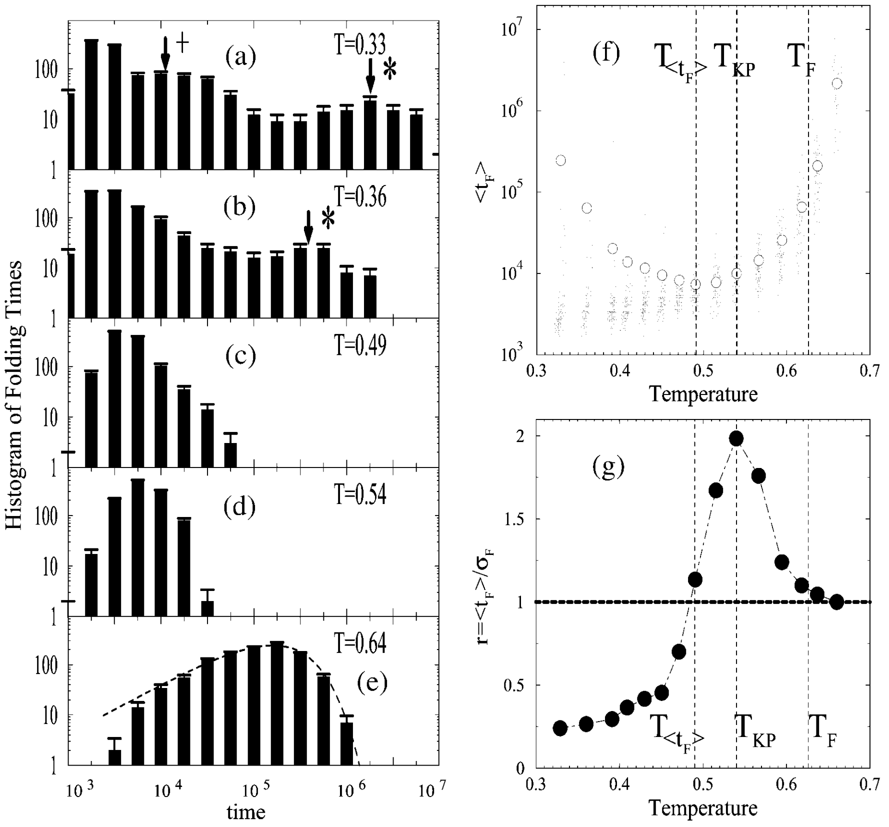

FIGURE 4 (*a–e*) Histograms of folding times for selected temperatures. At $T = 0.33$ and $T = 0.36$, the two lowest temperatures studied, histograms have a maximum for long folding times (↓ *), which suggests the existence of putative intermediates. At $T = 0.33$, a maximum in the histogram (↓ +), not present at $T = 0.36$, corresponds to short-lived kinetic traps. The distributions of folding times are unimodal at higher temperatures. At $T = 0.54$, the histogram is compact, and has no tail of long folding times. At $T = 0.64$, the histogram fits a single-exponential distribution $e^{-t_F/\langle t_F\rangle}/\langle t_F\rangle$ for times larger than the relaxation time of 1500 t.u. (*dashed line*). We estimate the errors of the histogram bars as the square-root of each bar. (*f*) Average folding time versus temperature. Each dot represents the folding time for a particular folding transition. (*g*) Ratio r of the average and the standard deviation, $r = \langle t_F\rangle/\sigma_F$, for the distribution of folding times. The ratio approaches 1 above T_{KP} and 0 below T_{KP}. The ratio is maximal at T_{KP}, indicating a compact distribution of folding times at this temperature.

approximates a single-exponential distribution. In particular, the distribution of folding times fits the single-exponential distribution $e^{-t_F/\langle t_F\rangle}/\langle t_F\rangle$ for temperatures near and above T_F. The ratio $r(T)$ decreases monotonically below T_{KP}, indicating that the distribution of folding times spreads over several orders of magnitude. This is the consequence of an increasing fraction of folding simulations kinetically trapped (Fig. 4, *a* and *b*). The average folding time $\langle t_F\rangle$ is minimal not at T_{KP}, but at a lower temperature $T_{\langle t_F\rangle} = 0.49$ (Fig. 4 *f*). At this temperature, we find that the protein becomes temporarily trapped in ~7% of the folding transitions. On the other hand, the remaining simulations undergo a folding transition much faster, thus minimizing $\langle t_F\rangle$. Interestingly, $r(T_{\langle t_F\rangle}) \simeq 1.0$, even though the distribution of folding times at this temperature is non-exponential.

Folding pathways

Below T_{KP}, an increasing fraction of the simulations undergo folding transitions that take a time up to three orders of magnitude above the minimal $\langle t_F\rangle$. In addition, $\langle t_F\rangle$ increases dramatically (Fig. 4 *f*). At the lowest temperatures studied, we distinguish between the majority of simulations that undergo a fast folding transition (the fast pathway) and the rest of the simulations that undergo folding transitions with folding times spanning three orders of magnitude (the slow pathways). At the low temperature $T = 0.33$, the potential energy of the fast pathway has on average a time evolution similar to that of all the simulations at $T_{KP} = 0.54$, indicating that there are no kinetic traps in the fast pathway.

For each folding simulation that belongs to the slow pathways, we sample the potential energy at equal time intervals of 100 t.u. until folding is finished (see Materials and Methods). Then, we collect all potential energy values and construct a distribution of potential energies. We find that below $T = 0.43$, the distribution is markedly bimodal (Fig. 5 *a*). The positions of the two peaks along the energy coordinate do not correspond to the equilibrium potential energy value of the folded state (Fig. 5 *b*). Therefore we hypothesize the existence of two intermediates in the slow pathways. We denote the two putative intermediates as I_1 and I_2 for the high energy and low energy peaks, respectively. As temperature decreases, the peaks shift to lower energies, but the energy difference between the two peaks, approximately six energy units, remains constant (Fig. 5 *b*). A constant energy difference implies that the two putative intermediates differ by a specific set of native contacts. As temperature decreases, other contacts not belonging to this set become more stable and are responsible for the overall energy decrease. At $T = 0.33$, we record the distribution of survival times for both intermediates and find

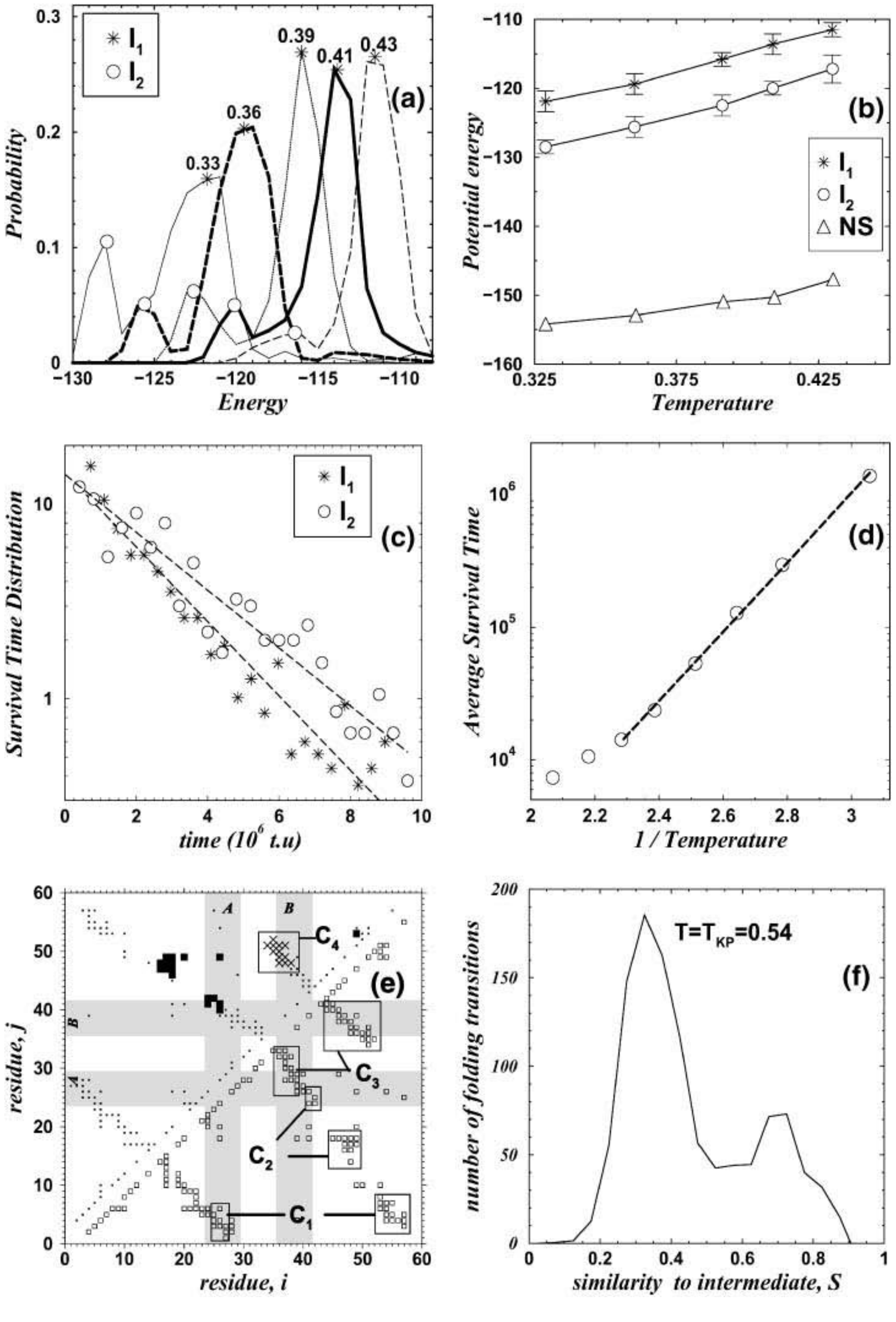

FIGURE 5 (*a*) Distributions of the potential energies of the slow folding pathways for temperatures below $T = 0.43$. The distributions are bimodal, suggesting two putative intermediates I_1 and I_2. (*b*) The potential energy of the distribution peaks (* and ○) increases with temperature, but the energy difference between peaks remains constant. The energy of the peaks is significantly larger than the equilibrium energy of the folded state (△). (*c*) Distributions of survival times at $T = 0.33$ for the high energy intermediate I_1 (*), $\langle t_F \rangle = 1.81 \times 10^6$, and $\sigma = 1.85 \times 10^6$, and the low energy intermediate I_2 (○), $\langle t_F \rangle = 2.47 \times 10^6$, and $\sigma = 2.43 \times 10^6$, fit to single-exponential distributions. (*d*) Arrhenius fit of the average survival time of intermediate I_2 below $T = 0.44$. This upper bound temperature coincides with the temperature below which the distribution of the potential energies (*a*) of the slow folding pathways becomes bimodal. (*e, upper triangle*) Absent contacts (■) and present contacts (+, C_4) in intermediate I_1. Upon the transition $I_1 \rightarrow I_2$, these contacts reverse their presence (so that the *solid squares* are the present contacts and the *crosses* are the absent contacts). There are five more squares than crosses, which roughly accounts for the difference of six energy units between the two intermediates. (*e, lower triangle*) Long-range contacts C_1 are present in intermediate I_1, and long-range contacts C_2 are absent. There are 23 contacts in C_1 and 15 contacts in C_2. We shade the positions of strands *A* and *B*. (*f*) Probability that a folding transition at $T_{KP} = 0.54$ contains a protein conformation with similarity *S* to intermediate I_1 (see Materials and Methods).

that they fit a single-exponential distribution, supporting the hypothesis that each intermediate is a local free energy minima and has a major free energy barrier (Fig. 5 *c*).

To further test the single free energy barrier hypothesis, we select a typical conformation representing intermediate I_2 and perform 200 folding simulations, each with a different set of initial velocities for a set of temperatures in the range $0.33 \leq T \leq 0.52$. For each simulation, we record the time that the protein survives in the intermediate state, and find that the average survival time fits the Arrhenius law for temperatures below $T = 0.44$ (Fig. 5 *d*). This upper bound temperature roughly coincides with the temperature $T = 0.43$ below which I_2 becomes noticeable in the histogram of potential energies (Fig. 5 *a*). This result indicates that the free energy barrier to overcome intermediate I_2 becomes independent of temperature for low temperatures, or analogously, that the same set of native contacts must form (or break) to overcome the intermediate.

Structure of the intermediates

We randomly select three conformations for each intermediate, and find that they are structurally similar, within each intermediate. Conformations belonging to intermediate I_1 have a set of long-range contacts (C_1) and a set of medium-range contacts (C_3) with high occupancy (Fig. 5 *e*). Contacts in C_1 represent a β-sheet made up by three strands: the two termini and the strand belonging to the diverging turn, which we name strand *A* (residues 24–29, Fig. 5 *e*; and I_1 in Fig. 6). For a folding transition through the slow pathway, this β-sheet stabilizes in the early events of the folding process, and strand *A* can no longer move freely. Contacts in C_3 are the contacts within the distal hairpin and with the n-Src loop. Contacts in C_3 constrain the flexibility of the strand shared by the distal hairpin and the n-Src loop, which we name strand *B* (residues 36–41, Fig. 5 *e*; and I_2 in Fig. 6). The restricted flexibility of strands *A* and *B* prevent the mutual closed packing found in the native state. Intermediate I_1 features a set of contacts (C_2) with no occupancy at all (Fig. 5 *e*) that are the result of the restricted flexibility of strands *A* and *B*.

Conformational changes leading the protein away from intermediate I_1 involve either dissociation of the β-sheet, thus breaking some contacts of C_1, or dissociation of the distal hairpin, thus breaking some contacts of C_3. We find that the latter dissociation may lead the protein conformation to intermediate I_2. Intermediate I_2 has contacts of C_1, but lacks the set of contacts (C_4) between strand *B* and the other strand of the distal hairpin (Fig. 5 *e* and *NS* in Fig. 6).

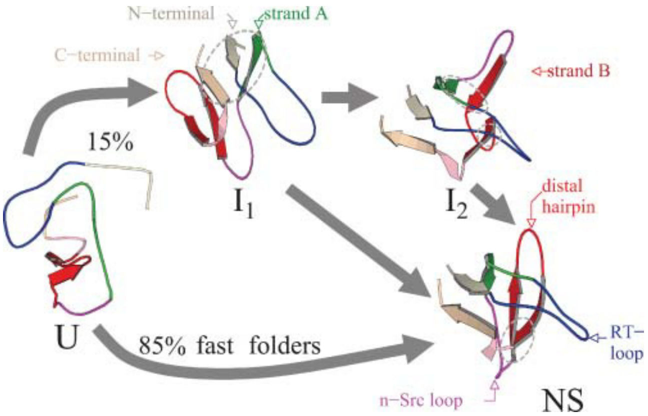

FIGURE 6 Schematic diagram of fast and slow folding pathways. At $T = 0.33$, in ~15% of the simulations, the model protein undergoes a folding transition through the slow folding pathways. We show the protein structure in I_1 and I_2 using the secondary structural elements of the native state, although some of these elements are not formed. In intermediate I_1, both termini and the strand *A* form a β-sheet (in the ellipse). The corresponding set of native contacts is C_1. Dissociation of the β-sheet leads to rearrangements of the protein conformation and successful folding to the native state. However, dissociation of the distal hairpin (in *red*) leads to more localized rearrangements that may lead the protein to intermediate I_2. Upon $I_1 \rightarrow I_2$ transition, contacts of C_2 (the two ellipses in I_2) form, but contacts of C_4 (the ellipse in native state) break.

Once we identify the structure of the intermediates, we investigate whether intermediate I_1 is present at larger temperatures when no distinction can be made concerning fast and slow folding pathways. To test this hypothesis, we sample the protein conformation during the folding transition at equal time intervals of 60 t.u. for each of the 1100 simulations, and compare these conformations to intermediate I_1 with a similarity score function (see Materials and Methods). For each folding transition, we record only the highest value of the similarity score, thus obtaining 1100 highest score values. At T_{KP}, the histogram of the highest scores is bimodal, with 25% of the folding simulations passing through intermediate I_1 (Fig. 5 *f*). Surprisingly, we find that at T_{KP}, simulations that undergo the folding transition through I_1 show kinetics of folding no different than those of the rest of simulations.

Cooperativity of the folding process

We investigate the cooperativity at T_{KP} with the time evolution of the frequency map, which we obtain with an average over the 1100 folding simulations at each moment of time. We find that different native contacts have different initial and final frequencies, as well as different time evolution. The majority of the contacts have low initial frequencies. As folding progress, the frequency increases in an exponential-like manner until it finally reaches a value close to 1, when folding is finished. Other contacts, however, do not follow this general trend but present unusual kinetics of formation (Fig. 7 *a*). In particular, some contacts have low final frequencies. We observe that these contacts are located in the surface of the protein, and can be assigned to three different categories: 1), isolated long-range contacts; 2), contacts in the base of hairpins and loops; and 3), short-range contacts whose native distance is close to the cutoff distance. Isolated long-range contacts can be easily broken by thermal fluctuations, and are difficult to form because of their long-range nature. Contacts in the base of hairpins and loops are the first to break in the transient unzipping of these structures. Finally, short-range contacts whose native distance is close to the cutoff distance can be easily broken because they undergo frequent collisions with the potential energy barrier that binds them. The more collisions a contact undergoes, the higher is the probability that this contact breaks.

We characterize the time evolution of the contact frequencies by computing the characteristic time to form a contact, $t_{ij}(T)$ (see Materials and Methods). We find that at T_{KP}, formation times correlate with sequence separation ($c = 0.75$), long-range contacts forming last (Fig. 7 *b*). We do not compute $t_{ij}(T)$ for nearest contacts because these contacts are already formed in the initial unfolded state. The histogram of formation times is bimodal, and the peak of the histogram corresponding to small formation times corresponds to contacts that form the core of loops and hairpins, located at

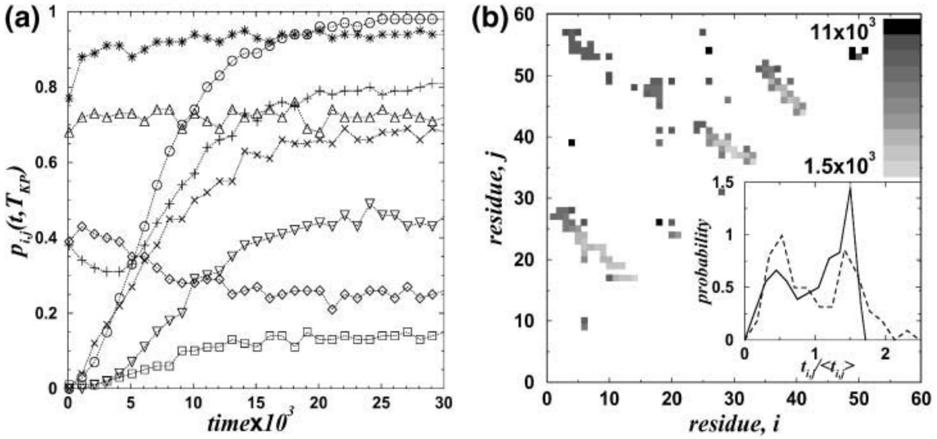

FIGURE 7 (*a*) Time evolution of the contact probability for some native contacts: □ = 20–26; ◇ = 2–4; ▽ = 26–49; × = 2–28; △ = 42–44; + = 4–6; * = 14–16; and ○ = 18–48. (*b*) Contact formation times at T_{KP} (not shown for nearest and next-nearest neighbors) are shown in shading, from light shading corresponding to 1.5×10^3 t.u. to dark shading corresponding to 11×10^3 t.u. (see *vertical bar*). The formation times increase with sequence separation. (*Inset*) Histogram of formation times at T_{KP} (*solid line*) and histogram at $T = 0.33$ for folding events through the fast folding pathway (*dashed line*).

the turns of these elements. The long-times peak corresponds to the contacts that form the base of loops and hairpins, as well as the tertiary long-range contacts between secondary structure elements. Thus, at T_{KP} the protein secondary structure is not stabilized until specific secondary structure elements interact and form the tertiary long-range contacts. This folding mechanism requires most of the native contacts to cooperate at the same time to stabilize the native state.

We also compute the time evolution of the contact frequencies at $T = 0.33$ for simulations that undergo a folding transition through the fast pathway. The histogram of formation times is bimodal, as for T_{KP} (Fig. 7 *b*). However, at $T = 0.33$ the peak corresponding to short formation times is more populated than the peak corresponding to long formation times. In fact, the long-times peak corresponds only to the tertiary long-range contacts, and has a tail for some contacts that take much longer to form. Thus, at low temperatures the secondary structure elements stabilize rapidly, independent of each other, and the folding process finishes when the formed secondary structure elements interact and form the tertiary long-range contacts. This folding mechanism only requires a cooperativity of short-range type, and is prone to be kinetically trapped.

DISCUSSION

It was shown (Borreguero et al., 2002) that the simplified protein model and interaction potentials that we use here reproduced, in a certain range of temperatures, the main experimentally determined thermodynamics characteristics of the SH3 domain (Filimonov et al., 1999). The predictive power of the model encouraged us to study the folding kinetics under initial non-equilibrium conditions in a broad range of target temperatures. The observation of intermediates in the unfolding simulations of the selected three-state and apparent two-state proteins supports the ability of our modified Gō model of interactions to distinguish between two-state and three-state kinetics. In addition, our equilibrium studies of the unfolded state at T_F show that our modified Gō model has a high sensitivity to detect the important amino acid contacts.

From our relaxation studies of the initial unfolded state, we observe that the structure of the unfolded state is highly sensitive to the target temperature, T_{target}. The role of the unfolded state in determining the folding kinetics has already been pointed out in recent experimental and theoretical studies (García et al., 2001; Plaxco and Gross, 2001; Shortle and Ackerman, 2001). We observe nucleation, folding with minimal kinetic barriers, and thermally activated mechanisms for the different observed unfolded states.

We observe that the typical formation times of secondary and tertiary contacts tend to separate from each other as we decrease T_{target} below T_F, suggesting a weakening of cooperativity between both types of contacts. In our model, a decrease in T_{target} is analogous to an increase in the stability of the native state. Thus the degree of cooperativity among the amino acids weakens under increasingly native conditions. A similar loss of cooperativity was found by Freire in extensive studies of the equilibrium fluctuations of the native state of several proteins (Luque et al., 2002). They found maximal cooperativity among the amino acids when native and denatured states had equal probability. These conditions correspond to $T = T_F$ in our study. Close to T_F, secondary structure elements stabilize only when tertiary contacts form fast after the secondary structure forms, the reason being that thermal fluctuations can disrupt the secondary structure elements when isolated. These fluctuations rapidly decrease in magnitude as temperature decreases, and secondary structure becomes stable under such conditions.

In previous studies, various methods have been developed to determine the temperature that signals the onset of multiple folding pathways. Wolynes and Onuchic's groups (Socci et al., 1996) determined a glass transition temperature, T_g, at which the average folding time is halfway between t_{min} and t_{max}, where t_{min} is the minimum average folding time and t_{max} is the total simulation time. This method is sensitive to the a priori selected t_{max}, and the authors found a 10% error in the calculation of T_g by changes of t_{max}. Also, Shakhnovich's group (Gutin et al., 1998) estimated a critical temperature, T_c, at which the temperature dependence of the equilibrium potential energy leveled off. From their results, one can evaluate a 20% error in their calculation of T_c. Both T_g and T_c are temperatures that authors use to characterize the onset of multiple folding pathways. In our study we use T_{KP}, which signals the breaking of time translational invariance of equilibrium measurements for temperatures below this value (Dokholyan et al., 2002). We estimate a 2% error in our calculation of T_{KP} from uncertainties in the location of T_{KP} in Fig. 4 *g*.

At T_{KP}, secondary structure elements are partially stable, which limits considerably the conformational search for the native state. Furthermore, T_{KP} is a relatively high temperature that prevents the stabilization of improper arrangements of the protein conformation, thus minimizing the occurrence of kinetic traps. Below T_{KP}, the model protein exhibits two intermediates with well-defined structural characteristics. The modest number of misfolded states is a direct consequence of the prevention of non-native contacts. This prevention reduces dramatically the number of protein conformations. Furthermore, since a low energy value implies that most of the native interactions have formed, there are few conformations having both low energy and structural differences with the native state (Plotkin and Onuchic, 2002).

It is found experimentally (Heidary et al., 2000; Juneja and Udgaonkar, 2002; Silverman et al., 2000; Simmons and Konermann, 2002) that proteins exhibit only a discrete set of intermediates. Even though in real proteins amino acids that do not form a native contact may still attract each other, experimental and theoretical studies confirm that native contacts have a leading role in the folding transition. Protein engineering experiments (Fersht, 1995; Grantcharova et al., 1998; Northey et al., 2002) show that transition states in two-state globular proteins are mostly stabilized by native interactions. To quantitatively determine the importance of native interactions in the folding transition, Paci et al. (2002) studied the transition states of three two-state proteins with a full-atom model. They found that on average, native interactions accounted for ~83% of the total energy of the transition states. Of relevance to our studies of the SH3 domain are the full-atom study (Shea et al., 2002) and the protein engineering experiments (Grantcharova et al., 1998; Riddle et al., 1999) showing that the transition state of the src-SH3 domain protein is largely determined by the native state. On the other hand, evidence exists that in some proteins, non-native contacts are responsible for the presence of intermediates. In a study of the homologous Im7 and Im9 proteins (Capaldi et al., 2002), authors identified a set of non-native interactions responsible for an intermediate state in the folding transition of Im7 protein. In another study (Mirny et al., 1996), authors performed Monte Carlo simulations of two different sequences with the same native state in the 3 × 3 lattice. One sequence presented a series of pathways with misfolded states due to non-native interactions.

At low temperatures, simulations that undergo folding through intermediate I_1 reveal that contacts between the two termini form earlier than the contacts belonging to the folding nucleus (Borreguero et al., 2002). This result coincides with an off-lattice study of a 36-monomer protein (Abkevich et al., 1994). In this study, the authors found an intermediate in the folding transition of their model protein. Inspection of the intermediate revealed no nucleus contacts, but a different set of long-range contacts had already formed. In addition, Serrano's group (Viguera and Serrano, 2003) engineered a variant of the α-spectrin SH3 domain, by which they increased the stability of the distal hairpin with new, stable long-range contacts. Authors observed an intermediate in the folding process when these newly introduced long-range contacts formed in the denatured state, preceding the formation of the transition state. Thus, environmental conditions that favor stabilization of long-range contacts other than the nucleus contacts may induce intermediates in the folding transition.

Alternatively, short-range contacts in key positions of the protein structure may also be responsible for slow folding pathways. In a study of the forming binding protein WW domain with the Gō model (Karanicolas and Brooks, 2003), the authors found a slow folding pathway in the model protein, and a cluster of four short-range native contacts that are responsible for this pathway. However, the authors observed that it was the absence, not the presence, of these native contacts in the unfolded state that generated biphasic folding kinetics. Thus, environmental conditions that favor destabilization of short-range contacts may promote the formation of intermediate states in the folding transition.

We also investigate the survival time of intermediate I_2, and find that the free energy barrier separating I_2 from the native state is independent of temperature. Thus, the average survival time follows Arrhenius kinetics. The value of the free energy barrier is ~5.85 energy units, indicating that approximately six native contacts break when the protein conformation reaches the transition state that separates I_2 from the native state. At the low temperatures studied, thermal fluctuations are still large enough so that the observed survival times of I_2 should be much smaller if only any six native contacts were to break. Thus we hypothesize that it is always the same set of native contacts that must break in the transition $I_2 \rightarrow$ native state. Our observations of the transition $I_1 \rightarrow I_2$ support this hypothesis. In this transition, we find that the set of contacts C_4 always breaks.

At T_{KP}, we do not detect any intermediate from kinetics measurements of the average folding time, or analogously, from the folding rate. However, with the similarity score function we detect intermediate I_1 in 25% of the folding transitions. The fact that the folding transitions populating intermediate I_1 at T_{KP} are kinetically no different than the rest suggests that thermal fluctuations are strong enough to prevent stabilization of this state. Interactions stabilizing intermediate I_1 involve but a few amino acids and therefore should not prevail over the thermal fluctuations.

In a study of protein Im9 (Gorski et al., 2001), the authors reported the existence of an intermediate in the folding transition under acidic conditions (pH = 5.5). This finding led authors to formulate the hypothesis that Im9 has an intermediate at normal conditions (pH = 7.0), but it is too unstable to be detected with current kinetic experimental techniques. Interestingly, the homologous protein Im7 (60% sequence identity) undergoes folding transition through an intermediate in all tested experimental conditions (Capaldi et al., 2002), supporting the authors' hypothesis. Similarly, in a recent report (Kamagata et al., 2003), authors observed two parallel pathways in the folding process of the proline-free staphylococcal nuclease with no accumulation of intermediates below the deadtime (4 ms) of the detection apparatus. It would be interesting to study this protein under stronger stabilizing conditions that may also stabilize any putative intermediate. Changes in both the environmental conditions and the amino acid sequence is therefore a general strategy to uncover hidden intermediates in the folding transition of a two-state protein. An alternative approach is an extensive study of the folding trajectories at T_{KP} that may reveal the hidden intermediates. This is particularly useful for computer simulations, because simulations at low temperatures, when intermediates are easily identifiable, may require several orders-of-magnitude longer than simulations at T_{KP}.

CONCLUSION

We perform analysis of the folding transition of the single domain protein c-Crk SH3 in a broad range of temperatures with molecular dynamics. At the folding transition temperature, we observe that only the folded and unfolded states are populated, in agreement with experimental results. As we decrease the temperature, we determine the kinetic partition temperature T_{KP} below which we observe two folding intermediates, I_1 and I_2. Below T_{KP}, intermediate I_1 forms when the two termini and the strand following the RT-loop form a β-sheet, before the formation of the folding nucleus. This intermediate effectively splits the folding transition into fast and slow folding pathways. Dissociation of part of the β-sheet leads the protein to either the native state or to I_2. The folding pathways of the model SH3 domain are highly sensitive to temperature, suggesting the important role of the environmental conditions in determining the folding mechanism. Since in our model a temperature decay is concomitant to a native state stability increase, our findings suggest that the SH3 domain may exhibit stable intermediates under conditions that will strongly stabilize the native state.

We acknowledge E. I. Shakhnovich for insightful discussions, R. Badzey for careful reading of the manuscript, and the Petroleum Research Fund for support. N.V.D. acknowledges support of the University of North Carolina at Chapel Hill Research Council Grant.

REFERENCES

Abkevich, V. I., A. M. Gutin, and E. I. Shakhnovich. 1994. Specific nucleus as the transition state for protein folding: evidence from the lattice model. *Biochemistry.* 33:10026–10036.

Alder, B. J., and T. E. Wainwright. 1959. Studies in molecular dynamics. I. General method. *J. Chem. Phys.* 31:459–466.

Bachmann, A., and T. Kiefhaber. 2001. Apparent two-state tendamistat folding is a sequential process along a defined route. *J. Mol. Biol.* 306:375–386.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.

Borreguero, J. M., N. V. Dokholyan, S. V. Buldyrev, E. I. Shakhnovich, and H. E. Stanley. 2002. Thermodynamics and folding kinetics analysis of the SH3 domain from discrete molecular dynamics. *J. Mol. Biol.* 318:863–876.

Branden, C., and J. Tooze. 1999. Introduction to Protein Structure. Garland Publishing, New York.

Bryngelson, J. D., and P. G. Wolynes. 1989. Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J. Phys. Chem.* 93:6902–6915.

Capaldi, A. P., C. Kleanthous, and S. E. Radford. 2002. Im7 folding mechanism: misfolding on a path to the native state. *Nat. Struct. Biol.* 9:209–216.

Cheung, M. S., A. E. García, and J. N. Onuchic. 2002. Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. USA.* 99:685–690.

Choe, S. E., P. T. Matsudaira, J. Osterhout, G. Wagner, and E. I. Shakhnovich. 1998. Folding kinetics of villin 14T, a protein domain with a central β-sheet and two hydrophobic cores. *Biochemistry.* 37:14508–14518.

Clementi, C., A. E. García, and J. N. Onuchic. 2003. Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism: all-atom representation study of protein L. *J. Mol. Biol.* 326:933–954.

Ding, F., N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich. 2002. Direct molecular dynamics observation of protein folding transition state ensemble. *Biophys. J.* 83:3525–3532.

Dokholyan, N. V., S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich. 1998. Molecular dynamics studies of folding of a protein-like model. *Fold. Des.* 3:577–587.

Dokholyan, N. V., E. Pitard, S. V. Buldyrev, and H. E. Stanley. 2002. Glassy behavior of a homopolymer from molecular dynamics simulations. *Phys. Rev. E.* 65:030801:1–030801:4.

Eaton, W. A., V. Muñoz, J. S. J. Hagen, G. S. Jas, L. J. Lapidus, E. R. Henry, and J. Hofrichter. 2000. Fast kinetics and mechanisms in protein folding. *Annu. Rev. Biophys. Biomol. Struct.* 29:327–359.

England, J. L., B. E. Shakhnovich, and E. I. Shakhnovich. 2003. Natural selection of more designable folds: a mechanism for thermophilic adaptation. *Proc. Natl. Acad. Sci. USA.* 100:8727–8731.

Ferguson, N., A. P. Capaldi, R. James, C. Kleanthous, and S. E. Radford. 1999. Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J. Mol. Biol.* 286:1597–1608.

Fersht, A. R. 1995. Characterizing transition states in protein-folding—an essential step in the puzzle. *Curr. Opin. Struct. Biol.* 5:79–84.

Fersht, A. R. 2000. A kinetically significant intermediate in the folding of barnase. *Proc. Natl. Acad. Sci. USA.* 97:14121–14126.

Fersht, A. R. 2000b. Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. *Proc. Natl. Acad. Sci. USA.* 97:1525–1529.

Fersht, A. R., and V. Daggett. 2002. Protein folding and unfolding at atomic resolution. *Cell.* 108:573–582.

Filimonov, V. V., A. I. Azuaga, A. R. Viguera, L. Serrano, and P. L. Mateo. 1999. A thermodynamic analysis of a family of small globular proteins: SH3 domains. *Biophys. Chem.* 77:195–208.

Gō, N., and H. Abe. 1981. Non-interacting local-structure model of folding and unfolding transition in globular proteins. I. Formulation. *Biopolymers.* 20:991–1011.

García, P., L. Serrano, D. Durand, M. Rico, and M. Bruix. 2001. NMR and SAXS characterization of the denatured state of the chemotactic protein CheY: implications for protein folding initiation. *Prot. Sci.* 10:1100–1112.

Gnanakaran, S., and A. E. García. 2003. Folding of a highly conserved diverging turn motif from the SH3 domain. *Biophys. J.* 84:1548–1562.

Gorski, S. A., A. P. Capaldi, C. Kleanthous, and S. E. Radford. 2001. Acidic conditions stabilise intermediates populated during the folding of Im7 and Im9. *J. Mol. Biol.* 312:849–863.

Grantcharova, V. P., and D. Baker. 1997. Folding dynamics of the Src SH3 domain. *Biochemistry.* 36:15685–15692.

Grantcharova, V. P., D. S. Riddle, J. N. Santiago, and D. Baker. 1998. Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat. Struct. Biol.* 8:714–720.

Guerois, R., and L. Serrano. 2000. The SH3-fold family: experimental evidence and prediction of variations in the folding pathways. *J. Mol. Biol.* 304:967–982.

Gutin, A., A. Sali, V. Abkevich, M. Karplus, and E. I. Shakhnovich. 1998. Temperature dependence of the folding rate in a simple protein model: search for a ''glass'' transition. *J. Chem. Phys.* 108:6466–6483.

Heidary, D. K., J. C. O'Neill, M. Roy, and P. A. Jennings. 2000. An essential intermediate in the folding of dihydrofolate reductase. *Proc. Natl. Acad. Sci. USA.* 97:5866–5870.

Holm, L., and C. Sander. 1996. Mapping the protein universe. *Science.* 273:595–602.

Ikai, A., and C. Tandford. 1973. Kinetics of unfolding and refolding of proteins. I. Mathematical analysis. *J. Mol. Biol.* 73:145–163.

Isaacson, R. L., A. G. Weedsdagger, and A. R. Fersht. 1999. Equilibria and kinetics of folding of gelsolin domain 2 and mutants involved in familial amyloidosis-Finnish type. *Proc. Natl. Acad. Sci. USA.* 96:11247–11252.

Jackson, S. E. 1998. How do small single-domain proteins fold? *Fold. Des.* 3:R81–R91.

Juneja, J., and J. B. Udgaonkar. 2002. Characterization of the unfolding of ribonuclease A by a pulsed hydrogen exchange study: evidence for competing pathways for unfolding. *Biochemistry.* 41:2641–2654.

Kamagata, K., Y. Sawano, M. Tanokura, and K. Kuwajima. 2003. Multiple parallel-pathway folding of proline-free staphylococcal nuclease. *J. Mol. Biol.* 332:1143–1153.

Karanicolas, J., and C. L. Brooks, III. 2003. The structural basis for biphasic kinetics in the folding of the WW domain from a formin-binding protein: lessons for protein design? *Proc. Natl. Acad. Sci. USA.* 100:3954–3959.

Karplus, M., and J. A. McCammon. 2002. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9:646–652.

Khorasanizadeh, S., I. D. Peters, and H. Roder. 1996. Evidence for a three-state model of protein folding from kinetics analysis of ubiquitin variants with altered core residues. *Nature Struct. Biol.* 3:193–205.

Kiefhaber, T. 1995. Kinetics traps in lysozyme folding. *Proc. Natl. Acad. Sci. USA.* 92:9029–9033.

Kitahara, R., and K. Akasaka. 2003. Close identity of a pressure-stabilized intermediate with a kinetic intermediate in protein folding. *Proc. Natl. Acad. Sci. USA.* 100:3167–3172.

Klimov, D. K., and D. Thirumalai. 2002. Stiffness of the distal loop restricts the structural heterogeneity of the transition state ensemble in SH3 domains. *J. Mol. Biol.* 317:721–737.

Knapp, S., P. T. Mattson, P. Christova, K. D. Berndt, A. Karshikoff, M. Vihinen, C. I. Smith, and R. Ladenstein. 1998. Thermal unfolding of small proteins with SH3 domain folding pattern. *PSFG.* 23:309–319.

Kortemme, T., M. J. S. Kelly, L. E. Kay, J. Forman-Kay, and L. Serrano. 2000. Similarities between the spectrin SH3 domain denatured state and its folding transition state. *J. Mol. Biol.* 297:1217–1229.

Lopez-Hernandez, E., and L. Serrano. 1996. Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein, CI-2. *Fold. Des.* 1:43–55.

Luque, I., S. A. Leavitt, and E. Freire. 2002. The linkage between protein folding and functional cooperativity: two sides of the same coin? *Annu. Rev. Biophys. Biomol. Struct.* 31:235–256.

Martinez, J. C., A. R. Viguera, R. Berisio, M. Wilmanns, P. L. Mateo, V. V. Filimonov, and L. Serrano. 1999. Thermodynamic analysis of α-spectrin SH3 and two of its circular permutants with different loop lengths: discerning the reasons for rapid folding in proteins. *Biochemistry.* 38:549–559.

Matouschek, A., J. T. Kellis, Jr., L. Serrano, M. Bycroft, and A. R. Fersht. 1990. Transient folding intermediates characterized by protein engineering. *Nature.* 346:440–445.

Millet, I. S., S. Doniach, and K. W. Plaxco. 2002. Toward a taxonomy of the denatured state: small angle scattering studies of unfolded proteins. *Adv. Prot. Chem.* 62:241–262.

Mirny, L. A., V. Abkevich, and E. I. Shakhnovich. 1996. Universality and diversity of the protein folding scenarios: a comprehensive analysis with the aid of a lattice model. *Fold. Des.* 1:103–116.

Northey, J. G. B., A. A. Di Nardo, and A. R. Davidson. 2002. Hydrophobic core packing in the SH3 domain folding transition state. *Nat. Struct. Biol.* 9:126–130.

Otzen, D. E., O. Kristensen, M. Proctor, and M. Oliveberg. 1999. Structural changes in the transition state of protein folding: alternative interpretations of curved chevron plots. *Biochemistry.* 38:6499–6511.

Ozkan, S. B., K. A. Dill, and I. Bahar. 2002. Fast-folding protein kinetics, hidden intermediates, and the sequential stabilization model. *Prot. Sci.* 11:1958–1970.

Paci, E., M. Vendruscolo, and M. Karplus. 2002. Native and non-native interactions along protein folding and unfolding pathways. *Prot. Struct. Func. Genet.* 47:379–392.

Pande, V. S., A. Y. Grosberg, and T. Tanaka. 2000. Heteropolymer freezing and design: towards physical models of protein folding. *Rev. Mod. Phys.* 72:259–314.

Plaxco, K. W., K. T. Simons, and D. Baker. 1998. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* 277:985–994.

Plaxco, K. W., and M. Gross. 2001. Unfolded, yes, but random? Never! *Nature.* 8:659–660.

Plotkin, S. S., and J. N. Onuchic. 2002. Structural and energetic heterogeneity in protein folding. I. Theory. *J. Chem. Phys.* 116:5263–5283.

Rapaport, D. C. 1997. The Art of Molecular Dynamics Simulation. Cambridge University Press, Cambridge, UK.

Riddle, D. S., V. P. Grantcharova, J. V. Santiago, A. L. M. E. Ruczinski, and D. Baker. 1999. Experiment and theory highlight role of native state topology in SH3 folding. *Nat. Struct. Biol.* 6:1016–1024.

Sanchez, I. E., and T. Kiefhaber. 2003. Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* 325:367–376.

Shea, J. E., J. N. Onuchic, and C. L. Brooks, III. 2002. Probing the folding free energy landscape of the src-SH3 protein domain. *Proc. Natl. Acad. Sci. USA.* 99:16064–16068.

Shortle, D., and M. S. Ackerman. 2001. Persistence of native-like topology in a denatured protein in 8 M urea. *Science.* 293:487–489.

Silverman, S. K., M. L. Deras, S. A. Woodson, S. A. Scaringe, and T. R. Cech. 2000. Multiple folding pathways for the P^4–P^6 RNA domain. *Biochemistry.* 39:12465–12475.

Simmons, D. A., and L. Konermann. 2002. Characterization of transient protein folding intermediates during myoglobin reconstitution by time-resolved electrospray mass spectrometry with on-line isotopic pulse labeling. *Biochemistry.* 41:1906–1914.

Socci, N. D., J. N. Onuchic, and P. G. Wolynes. 1996. Diffusive dynamics of the reaction coordinate for protein folding funnels. *J. Chem. Phys.* 15:5860–5868.

Tang, K. S., J. Guralnick, W. K. Wang, A. R. Fersht, and L. S. Itzhaki. 1999. Stability and folding of the tumour suppressor protein p^{16}. *J. Mol. Biol.* 285:1869–1886.

Thirumalai, D., D. K. Klimov, and R. I. Dima. 2002. Insights into specific problems in protein folding using simple concepts. *Adv. Chem. Phys.* 120:35–76.

Tiana, G., and R. A. Broglia. 2001. Statistical analysis of native contact formation in the folding of designed model proteins. *J. Chem. Phys.* 114:2503–2510.

Viguera, A. R., V. V. Filimonov, P. L. Mateo, and L. Serrano. 1994. Thermodynamic and kinetic analysis of the SH3 domain of spectrin shows a two-state folding transition. *Biochemistry.* 33:2142–2150.

Viguera, A. R., and L. Serrano. 2003. Hydrogen-exchange stability analysis of Bergerac-Src homology 3 variants allows the characterization of a folding intermediate in equilibrium. *Proc. Natl. Acad. Sci. USA.* 100:5730–5735.

Villegas, V., A. Azuaga, L. Catasus, D. Reverter, P. L. Mateo, F. X. Aviles, and L. Serrano. 1995. Evidence for a two-state transition in the folding process of the activation domain of human procarboxypeptidase-a^2. *Biochemistry.* 46:15105–15110.

Walkenhorst, W. F., S. M. Green, and H. Roder. 1997. Kinetic evidence for folding and unfolding intermediates in staphylococcal nuclease. *Biochemistry.* 36:5795–5805.

Wu, X. D., B. Knudsen, S. M. Feller, J. Sali, D. Cowburn, and H. Hanafusa. 1995. Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk. *Structure.* 2:215–226.

Yamasaki, K., K. Ogasahara, K. Yutani, M. Oobatake, and S. Kanaya. 1995. Folding pathway of *Escherichia coli* ribonuclease HI: a circular dichroism, fluorescence, and NMR study. *Biochemistry.* 34:16552–16562.

Zagrovic, B., C. D. Snow, S. Khaliq, M. R. Shirts, and V. S. Pande. 2002. Native-like mean structure in the unfolded ensemble of small proteins. *J. Mol. Biol.* 323:153–164.

Zhou, Y. Q., and M. Karplus. 1997. Folding thermodynamics of a model three-helix-bundle protein. *Proc. Natl. Acad. Sci. USA.* 94:14429–14432.

Zhou, Y. Q., M. Karplus, J. M. Wichert, and C. K. Hall. 1997. Equilibrium thermodynamics of homopolymers and clusters: molecular dynamics and Monte Carlo simulations of systems with square-well interactions. *J. Chem. Phys.* 107:10691–10708.