Is there an en route folding intermediate for cold shock proteins?

Lei Huang* and Eugene I. Shakhnovich*

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

Received 25 November 2011; Revised 30 January 2012; Accepted 8 February 2012
DOI: 10.1002/pro.2053
Published online 28 February 2012 proteinscience.org

Abstract: Cold shock proteins (Csps) play an important role in cold shock response of a diverse number of organisms ranging from bacteria to humans. Numerous studies of the Csp from various species showed that a two-state folding mechanism is conserved and the transition state (TS) appears to be very compact. However, the atomic details of the folding mechanism of Csp remain unclear. This study presents the folding mechanism of Csp in atomic detail using an all-atom Gō model-based simulations. Our simulations predict that there may exist an en route intermediate, in which β strands 1-2-3 are well ordered and the contacts between β1 and β4 are almost developed. Such an intermediate might be too unstable to be detected in the previous fluorescence energy transfer experiments. The transition state ensemble has been determined from the P_{fold} analysis and the TS appears even more compact than the intermediate state.

Keywords: cold shock protein; discrete molecular dynamics; folding dynamics; folding intermediate; ϕ-value analysis; transition state ensemble

Introduction

When the temperature decreases significantly, Cold shock proteins (Csps) will be expressed at high levels in response to cold shock. Csps can bind to single-stranded nucleic acids and regulate the expression of other proteins.[1,2] Owing to their important role in stress response, Csps are generally conserved from bacteria to human and share similar structures and sequences across various species. The structure of Csp from *Bacillus subtilis* is shown in Figure 1. It is a β-barrel constituted by five β sheets. Meanwhile, Csps have been known as a protein family that can fold fast in a two-state folding mechanism[3] as Schmid and coworkers showed that the folding dynamics of CspB from *B. subtilis* follows a two-state model in 1995.[4] Folding of Csps from several organisms, for example, *B. subtilis*,[3–8] *B. caldolytics*,[3,9–11], *Escherichia coli*,[12–16] and *Thermotoga maritima*,[3] have been studied intensively by ensemble experiments. With the advances in single-molecule techniques, valuable insights have been gained from recent single molecule fluorescence energy transfer (FRET) studies of the Csp from *T. maritima*.[17–21] The two-state folding dynamics has been demonstrated clearly at the single-molecule level.[17]

In many experimental studies,[3,4,12] the fluorescence of a single tryptophan is used alone to probe the folding dynamics. Vu *et al.*[15] argued that such measurements might report only on local conformational preferences and their dynamics instead of the global folding dynamics. Other than using W11 in Csp-A from *E. coli* as the probe, Y42, S52, and T68 have been substituted with tryptophan to study Csp folding.[15] Very similar rate constants of folding were found for tryptophans located in various structural contexts. Furthermore, several different methods, including the use of tryptophan or tyrosine as a probe as well as an ensemble FRET measurement, have been employed to study the folding dynamics of the Csp from *B. caldolytics*.[10,11] Most folding rates in their studies are comparable, which suggests that the folding of Csp indeed follows the kinetic mechanism of a two-state protein.

Current address: Lei Huang, Department of Biochemistry & Molecular Biology, University of Chicago, 929 East 57th Street, Chicago, Illinois 60637.

*Correspondence to: Lei Huang, Department of Biochemistry & Molecular Biology, University of Chicago, 929 East 57th Street, Chicago, IL 60637. E-mail: lhuang3@uchicago.edu or Eugene I. Shakhnovich, Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138. E-mail: shakhnovich@chemistry.harvard.edu

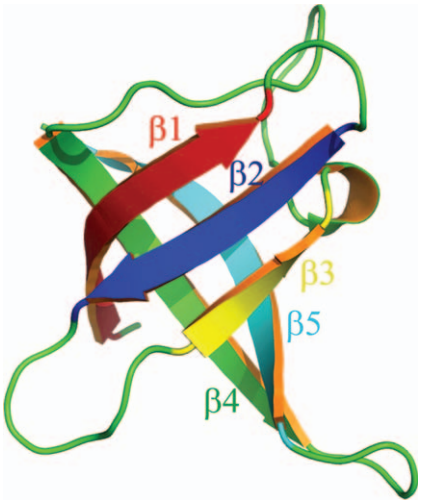

Figure 1. The structure of Csp from *B. subtilis* (Brookhaven PDB accession code: 1csp) is shown. Five β strands are rendered in different colors. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

For the two-state folding dynamics, identifying the transition state ensemble (TSE) is crucial to understand the underlying folding mechanism. As the structures in TSE have relatively high free energies and very short lifetimes,[22] it is difficult to study them directly in experiments. The dependence of rate constants of refolding and unfolding on denaturant concentration suggests that the transition state (TS) of Csp should be very compact.[3–7,9–13,15] On the other hand, ϕ-value analysis, proposed by Fersht et.al.,[23] has served as a powerful method to study the TSE. So far, ϕ-value analysis has been used to systematically study the TSE of Csp from *B. subtilis*.[7] They suggested that in the TS β1 resembles the structure in folded state, the hairpin by β1–β2 is formed, the hairpin by β2–β3 is partially formed, and native contacts between β1–β4 are partially developed. This is partially consistent with the TS suggested by Alm *et al.*,[24] who proposed that β1–β2–β3 form a well-ordered structure and the native contacts in the loop between β3 and β4 are partially developed. Other than experimental studies, Csps have been studied by computer simulations and theories.[24–28] The atomic details of folding pathway, including the TSE, are not easily obtainable from experiments alone. Ab initio computer simulations using an all-atom model might bridge this gap. However, the folding time of Csp is on the order of several milliseconds,[3] which is far beyond the time scale that most simulations using all-atom force fields today can reach. Although significant progress has been made in designing specific platforms that can accelerate molecular dynamics by two or three orders of magnitude and run milliseconds simulations with explicit solvents,[29] more powerful systems are highly desired as many independent trajectories are needed to gain good statistics on folding pathways. Other than the lack of powerful computational resources, the existing force fields might not be accurate enough for ab initio folding simulations, especially for β-proteins. For example, several commonly used force fields could be too helical.[30–32] In the recent study by Nettels *et al.*,[33] the Rg of unfolded Csp measured in simulations using AMBER ff03/TIP4P-Ew and OPLS-AA/L force field are around 14 Å, which is significantly smaller than the value inferred from single-molecule FRET experiments, 23 Å.[33] In this case, the unfolded/denatured states of Csp are not correctly modeled by these force fields. Although a knowledge-based force field has been used to fold several helical proteins succesfully,[34] it is still very challenging to fold β-proteins. Other than physical and empirical force fields, structure-centric Gō-like models have been widely used to study protein folding mechanisms,[35–45] especially when there are no other suitable force fields available. Despite the simplicity and well-known unphysical aspects, Gō-like models have been used, with some success, to study folding dynamics[35,37–42,44,45] and predict folding rates[46] as it is widely assumed that the folding mechanism is mainly determined by a protein's native structure.[40,44,47] Generally, Gō models provide a good description of the energy landscape of the folded state, the TS, and the native-like intermediates.[40,44,47–49] It is possible to extract meaningful results about the folding mechanism and the TS for those small proteins which directly collapse to a native-like TS. Gō models are known to have limitation of describing non-native-like collapsed structures, for example, those misfolded structures, and the contributions from non-native interactions.[48,49] Experiments show that the unfolded states of Csp behave like a random coil[17,20] (instead of a collapsed globule) and they directly collapse into a very compact TS.[4] Therefore, Gō models could be appropriate to study the folding mechanism of Csp.

In this study, we used an all-atom Gō potential to study the folding mechanism of the Csp from *B. subtilis*. It is the first ab initio Csp folding simulations with atomic details. High cooperativity in folding dynamics is observed in our simulations and the detailed folding mechanism is outlined. Our simulations suggest that the unfolded state essentially is a random coil except for the partially formed hairpin by β1–β2 and β2–β3 as well as partially developed native contacts within the loop between β3 and β4. Additionally, our simulations suggest that there is a short-lived (the dwell time is around 1/1000 of the folding time) intermediate although it is so unstable that it might be difficult to detect in existing experiments. Consequently, Csp will exhibit an apparent two-state folding dynamics.

Model and simulations

The all-atom Gō model in our study is similar to the models in our previous studies.[38,39] All heavy atoms

(nonhydrogen atoms) are modeled as beads having same mass. A square well potential with an infinite depth as shown in Eq. (1) is used to model a covalent bond connecting two atoms i and j. ${r_{ij}}^0$ is the bond length in the native structure and it is assumed as the equilibrium bond length.

$$u_{ij}^{\text{bond}}(r_{ij}) = \begin{cases} \infty, \ r_{ij} \le 0.9{r_{ij}}^0 \\ 0, \ 0.9r_{ij}^0 < r_{ij} < 1.1{r_{ij}}^0 \\ \infty, \ r_{ij} \ge 1.1{r_{ij}}^0 \end{cases} \quad (1)$$

In our simulations, the X-ray structure of Csp from *B. subtilis*[1] (Brookhaven PDB accession code: 1csp) is used as the native structure. The missing side-chain atoms in glutamic acid are added using the Swiss-Pdb viewer (http://spdbv.vital-it.ch/) and no collisions are detected in the reconstructed native structure. There are 520 heavy atoms in total. To define the Gō potential, distances between two nonlocal atoms (separated by at least two residues along the protein backbone), i and j, in the native structure are calculated (denoted as r_{ij}) and compared with the cutoff distance of native contact $r_{ij\_\text{cut}} = r\_cut(r_i+r_j)$, where r_i and r_j are the radii of atom i and j (radii data were adapted from Ref. [50]) and r_cut is a scaling factor. If $r_{ij} < r_{ij\_\text{cut}}$, atoms i and j could form a native contact; otherwise, they will form a nonnative contact for a distance smaller than r_{ij_cut}. The interaction energy is zero for a distance larger than r_{ij_cut}. The depth of the energy well for a native contact pair and for a nonnative contact pair is $-\varepsilon$ and 0, respectively. Here, ε is the unit of energy. The hardcore radius for each atom is defined as $0.8r_i$, where r_i is the VDW radius for atom i. The distances between any two nonbonded and nonlocal atoms cannot be smaller than $0.80(r_i+r_j)$. The total energy of any given conformation is calculated as the sum of all pairwise energies for nonlocal atoms. Discrete molecular dynamics with constant temperature[43,51] is used in both thermodynamic sampling and the studies of folding kinetics. As shown in Figure 1, residues 2–10, 14–19, 26–29, 44–53, and 58–66 are defined as strands β1–β5, respectively. The loop between β1 and β2 is labeled as loop-1; the loop between β2 and β3 is labeled as loop-2; the loop between β3 and β4 is labeled as loop-3; the loop between β4 and β5 is labeled as loop-4.

The reduced units of energy, length (r), mass, and temperature in our simulations are ε, Å, m, and ε/k_B, respectively. For convenience, reduced units are used for the simulations, that is, $E^* = E/\varepsilon$, $r^* = r/\text{Å}$, $T^* = T/(\varepsilon/k_\text{B})$ and $t^* = t/\sqrt{mr^2/\varepsilon}$.

Results and Discussion

The results of the simulations were found to be dependent on the parameters, especially r_cut. Extensive discussion about how parameters (r_cut) and temperature affect the energy landscape of Csp will be reported elsewhere. In the present study, r_cut is set as 1.10. For very small values of r_cut (<1.05), there are not enough native contacts to accurately maintain a stable folded state. Large values of r_cut (>1.20) make side-chain packing more degenerate. Simulations with 1.10 show an apparent two-state folding behavior which is consistent with the experiments. All kinetic simulations are conducted at relatively high temperatures, that is, slightly below the folding temperature.

The intermediate and unfolded state

The replica exchange method[52] is used in thermodynamic sampling. The temperatures in 64 replicas are set from 0.312 to 0.495. More replicas are allocated in the vicinity of the transition temperature to accurately determine the position of transition temperature. Four independent simulations are conducted to calculate the error bars. The weighted histogram analysis method (WHAM)[53–55] is used for data analysis by combining the data at all temperatures. The Rg, fraction of native contacts (Q), specific heat capacity (C_v/k_B), and potential energy from direct simulations as well as WHAM are shown in Figure 2. We can observe highly cooperative folding and the folding temperature (T_f, the temperature where C_v/k_B reaches its peak value) is found to be around 0.410 with a very narrow transition region. At lower or higher temperature, the protein is either folded or unfolded.

Our kinetic simulations with 1024 trajectories suggest that besides folded and unfolded states there exists an intermediate state. The traces of the time-dependent Q from four typical folding trajectories at $T = 0.402$ (corresponding to a physical temperature of 44°C considering the experimental unfolding temperature as 50°C[56]) are shown in Figure 3. The intermediate appears obligatory on folding pathway. The intermediate is highly unstable, that is, it can either fold completely or get unfolded. As the time of simulations is not long enough, the intermediate state is never reached in many trajectories, that is, only the unfolded states are observed within simulations. In our 1024 kinetic trajectories, the intermediate has been visited 599 times. Only 124 of them lead to the folded state and the rest unfold quickly. The distribution of dwell times for this intermediate can be fitted well with a single exponential with dwell time as 2.8×10^5 time units. On the other hand, the folding time is estimated to be around 3×10^8 time units by single exponential fitting of time-dependent survival ratio of the unfolded state. Compared with the folding time, the dwell time of the intermediate is 1000 times shorter and would presumably be on the order of a microsecond in experiments, considering that the folding time in experiments is on the order of a millisecond.[4] Owing to its short lifetime, the intermediate state cannot

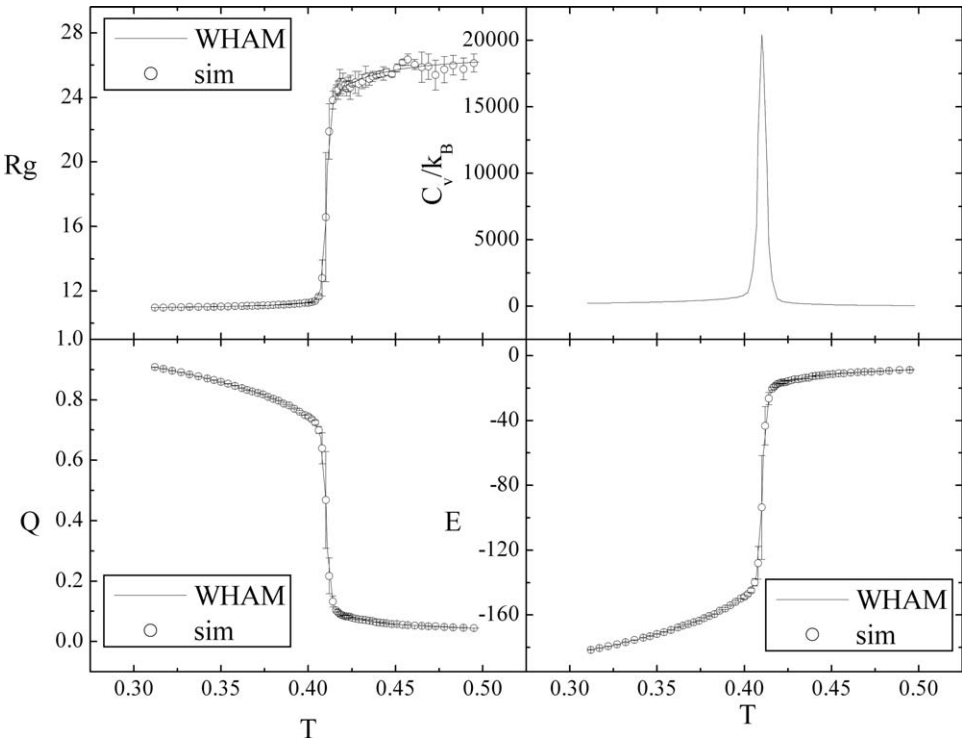

Figure 2. Temperature dependence of Rg, C_v/k_B, Q, and E. Circles represent data from direct simulations and lines represent data obtained by WHAM.

be detected in current single-molecule FRET experiments, considering the time resolution of FRET is generally around 0.1–1 ms. It would be of great interest to design single-molecule FRET experiments to check the existence of the short-lived intermediate that we observe. A dye system with relative small

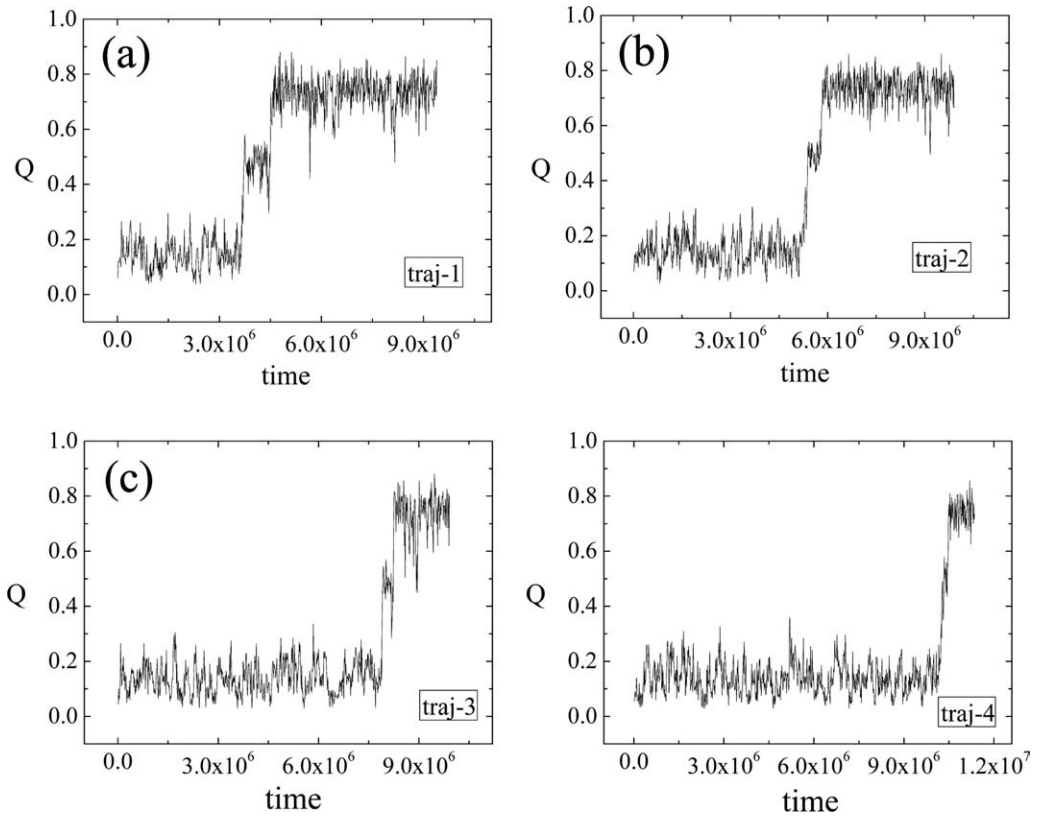

Figure 3. Four typical traces of the time-dependent Q from kinetic simulations.

Förster radius (around 30 Å) should be used to distinguish the intermediate and folded state better as the end-to-end distances in the intermediate and folded state are around 30–40 and 13 Å, respectively. As the intermediate state is not stable (i.e., there are no pronounced barriers to the unfolded and folded states), the solvent with higher viscosity might be helpful to slow down the transitions from the intermediate to unfold and fold structures as demonstrated by Eaton and coworkers.[57] It should be kept in mind, though, that adjusting solvent viscosity is still tricky as the energy landscape could be changed.

To relate the folding kinetics and free-energy landscape, one kinetic trajectory is shown in Figure 4 in the 2D free-energy profile for Rg and *Q*. The intermediate state is highlighted by a circle. As this intermediate is not stable and the simulations time in thermodynamic simulations is relatively shorter than that in kinetic simulations, the intermediate is not observed in the 2D free-energy profile. One typical snapshot of the intermediate is shown in Figure 4. It indicates that β4–β1–β2–β3 forms a well-ordered structure where only β5 is dangling around. Interestingly, the simulation based on a coarse-grained model by O'Brien *et al.*[58] also found an intermediate for the Csp from *T. maritima*. However, their structure of the intermediate is different from ours. They suggested that only β1 is unfolded and the rest is folded. One possible reason for the difference might be that different Csp proteins were used in simulations. The difference of amino acid sequences between Csp from *T. maritima* and *B. subtilis* are more than 30%.

The structures in the unfolded basin (Fig. 4) contain very small number of native contacts. Essentially, it behaves like a random coil and the average Rg is around 23 Å, which agrees very well with single-molecule FRET experiments.[33] The unfolded state has been studied by ensemble FRET experiments[10,11] in detail and they suggested that it is similar to a random coil without ordered secondary structures. The reconfiguration time of Csp in the unfolded state in our simulations is on the order of 10^4 time units (~100 ns, which is comparable with experimental value,[20,59] 50 ns). The work by Schuler and coworkers[21] suggested that around 20% of the β strands are formed after 1.3 ms dead time. This finding may be owing to the contribution of partially folded β1–β2–β3 strand and may be influenced by a contribution from a small fraction of the folded state. In our simulations, several native contacts within the β-turn area of hairpins, β1–β2 and β2–β3, are partially formed in $<10^5$ time units (<1 microsecond). The hairpin by β4–β5 is very unstable compared with the others. Several native contacts within the loop between β3 and β4 are also partially developed. The contacts between β1 and β4 are essentially not formed and the native contacts for β3–β5 are not formed.

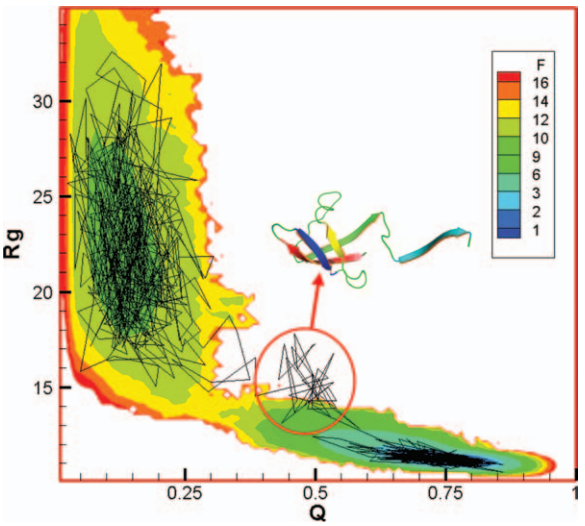

Figure 4. The 2D free-energy profile for Rg and *Q* at $T = 0.402$. Different colors represent the corresponding free energy reduced by k_BT. One typical folding trajectory from kinetic simulations is also shown in this plot. The intermediate is highlighted by a circle and one typical snapshot is shown. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

As shown in Figure 4, the free energy of the folded state is about $9k_BT$ lower than the free energy of the unfolded state. On the other hand, we can roughly estimate the free-energy difference between the unfolded state and the intermediate state with Eq. (2).

$$\Delta G = G_I - G_{\text{unfold}} \approx -k_BT \log\left(\frac{k_{\text{unfold}->\text{I}}}{k_{\text{I}->\text{unfold}}}\right) \approx 5k_BT \tag{2}$$

$k_{\text{unfold->I}}$ is the transition rate from the unfolded state to the intermediate state and $k_{\text{I->unfold}}$ is the backward transition rate. $\frac{k_{\text{unfold}->\text{I}}}{k_{\text{I}->\text{unfold}}}$ is around 1/200 in our simulations. Consequently, the free energy of the intermediate state is about 14 k_BT higher than the free energy of folded state. The potential energy decreases around 50ε (124 k_BT) from the intermediate state to the folded state. Roughly, $-T\Delta S$ is changed by 110 k_BT, which suggests that there exists a large entropy barrier in the transition from the intermediate state to the folded state.

Characterizing the TSE

As mentioned above, the P_{fold} of the intermediate is 0.2. Considering that the TS from the intermediate state to folded state has equal transition probability (0.5) to the intermediate state and folded state, its P_{fold} is around 0.6, which is close to the P_{fold} of real TS of Csp folding (0.5). Consequently, the TS of Csp folding should be close to the TS from the

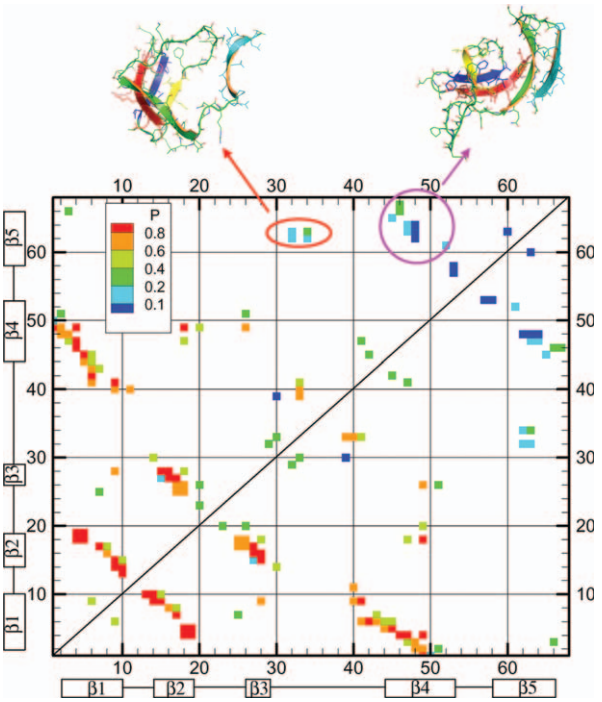

Figure 5. The average residue–residue contact map is shown for 20 transition state structures. Different colors represent the corresponding probability of native contacts. Two representative TS structures are shown in the top. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

intermediate to the folded state. All structures in the free-energy landscape shown in Figure 4 around the transition region from the intermediate to the folded state, specifically with $Q \in (0.35,0.65)$, $Rg \in (12,14)$ Å, and CαRMSD ∈ (5,12)Å, are selected as putative TS conformations. In total, 588 conformations are found satisfying such criteria and the real TS is identified using the P_{fold} analysis,[34] for example, $P_{fold} \in [0.4,0.6]$ in 64 trajectories. The simulation time for the short trajectories used in the P_{fold} analysis is 8.6×10^5 time units, roughly three times the dwell time in the intermediate. The distribution of P_{fold} over 588 conformations shows two peaks, that is, one around 0.2 and the other around 1.0, suggesting that 2D projection of free energy landscape Q–Rg plane does not provide an accurate kinetic description, that is that neither Q nor Rg can serve as a good reaction coordinate as was found earlier.[60–62] A similar distribution of P_{fold} was also observed by Hubner *et al.*[63] Eventually, only 20 out of 588 putative TSE structures have $P_{fold} \in [0.4,0.6]$ and they constitute the real TSE. The average residue–residue contact map for the real TSE is shown in Figure 5. Compared with the contact map of the intermediate state, the major differences are located in the two regions that are highlighted with an ellipse and a circle, respectively, in Figure 5. The representative structures are shown above the contact map. In one representative TS structure, (Fig. 5, left panel) new contacts are developed between β5 and the loop between β3 and β4. Alternatively, new contacts can be developed between β4 and β5 in an alternative representative TS structure (Fig. 5, right panel). The two different pathways are the consequence of the low stability of the β turn between β4 and β5. Following the way we used previously,[34] the ϕ-values are calculated for our TSE and compared with the experimental value[7] in Figure 6. The residues in β2 and β4 have larger ϕ-values in our simulations, which suggest β2 and β4 are more ordered in our TSE compared with experimental ϕ-values. Essentially, the ϕ-values of the residues located in β1, β3, and β5 are consistent with experiments. Most existing experiments show a very asymmetric chevron plot for the dependence of the apparent refolding rate on the concentration of denaturant. This suggests that the TS is very compact. The average SASA of the unfolded state, TSE and the folded state have been calculated by EDTSurf[64] as 9000, 6250, and 5210 Å,[2] respectively. The change of SASA from the unfolded state to TS over the change of SASA from the unfolded state to folded state is around 0.73. If we assume that the change in SASA is roughly proportional to the change in the interactions with the solvent, that is, the m-value in chevron plot, 0.73 is consistent with 0.6–0.9 in the experiments of CspA from *E. coli*.[13,15] However, the m-value for the experiments of Csp from other species[3] (from 0.86 to 0.93) is somewhat larger.

The folding mechanism of Csp

Almost all existing experiments suggest that Csp folds without a stable intermediate except that there may exist a fast collapse phase preceding the folding when Csp is transferred to folding conditions from denaturing conditions.[11] Such a fast collapse process was suggested mostly owing to nonspecific interactions[10] and involves a collapsed structure that is

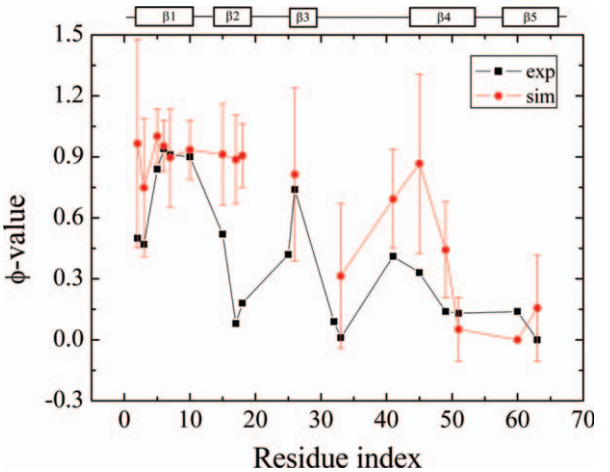

Figure 6. Comparison of ϕ-values between simulations and experiments. Error bars denote the standard deviation, σ of computed ϕ-values [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.].

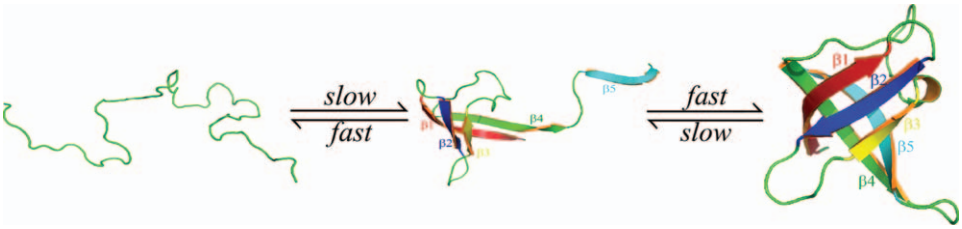

Figure 7. The proposed folding pathway for Csp. The intermediate state is very unstable. In the intermediate state, only β5 is isolated and the remaining parts resemble the native structure. The intermediate state can either unfold (with 80% probability) or fold (with 20% probability). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

essentially unfolded except some structural elements that are similar to the folded state. Furthermore, the single-molecule study by Schuler and coworkers[21] suggested that around 20% of β sheets are formed in the fast collapse phase occurring in a dead time around 1.3 ms (the folding time under such conditions is around 19 ms). Single-molecule FRET experiments[21,65] showed, not surprisingly, that the unfolded states at lower concentration of denaturant are more compact than at higher concentration of denaturant. Intuitively, when Csp is switched from a high concentration of denaturant to a low concentration of denaturant, the former unfolded state will collapse ultrafast and relax to the new unfolded state at this new lower concentration of denaturant. In fact, a similar argument was proposed by one of us in 1997.[66] Analogously, such a fast relaxation phase is also observed in the temperature jump experiments.[14] The recent single-molecule study by Schuler and colleagues[20] suggested that the collapse time of unfolded Csp is comparable with the reconfiguration time at around 50 ns. Such an ultrafast collapse phase is a purely diffusive, "downhill" process. An ultrafast collapse is often associated with the change of environment. In our simulations, the fast collapse phase is not present as the initial configurations were generated by equilibrium simulations (10^4 time units per trajectory) under the same temperature as in the kinetic simulations. Consequently, there is no such a collapse process in response to the abrupt changes in the environment (temperature, the concentration of denaturant, etc.).

The possible folding pathway for Csp suggested by our thermodynamic and kinetic simulations is shown in Figure 7. In an unfolded state, the hairpins by β1–β2 and β2–β3 are not stable. They interchange quickly between the open and the closed forms. The hairpin by β2–β3 is slightly more stable than the hairpin by β1–β2, which is consistent with the finding from a hydrogen exchange study of the Csp from *E. coli*.[16] The loop between β3 and β4 is also partially folded. It takes a long time to form a well-ordered β1–β2–β3 structure. When β4 is attached to β1 and the native contacts within the loop between β3 and β4 are fully developed at the same time, an intermediate state is formed. However, this intermediate state is not stable although β1 and β2 resemble the same state in folded structure. P_{fold} of such intermediate is 0.2, which suggests that the barrier to get folded is larger than the one in breaking the ordered strands, β4–β1–β2–β3. When the native contacts between β4 and β5 as well as β5 and loop 3 are developed, the β-barrel can be closed and the folded state is formed. Our simulations suggest that the TS is very compact and is reached at a very late stage in terms of the magnitude of protein–solvent interactions as proposed by experiments.[12,13] A well-ordered structure of β4–β1–β2–β3 is formed in the TS. Either the native contacts between β4 and β5 or between β5 and the loop connecting β3 and β4 are partially formed. Interestingly, the sequence of folding events revealed by our simulations is essentially in the reverse order of unfolding simulations at high temperature.[28] The TS is more compact than that suggested by experimental ϕ-value analysis[7] and theory.[24] The mutation study of F20 and F31 for Csp from *E. coli*[13] suggested that these two residues should be present in the TS, that is, β2 and β3 might folded. This argument is also supported by the studies of Reid *et al.*[12] that the presence of an oligonucleotide can increase the folding rate as such an oligonucleotide can bind the hydrophobic residues located in β2 and β3. Perl *et al.*[9] have proposed that the N-terminal region is folded but the C-terminal is not in the TS. Furthermore, Garcia-Mira *et al.*[7] suggested that the development of native contacts between β1 and β4 is crucial for reaching the TS. It seems that the TSE identified by this study is consistent with most experiments. However, it is still unknown how β5 interacts with other strands in experiments. It would be very interesting to design experiments to compare the TS in experiments with the TSE we propose here. For example, the change of folding rates of the mutation at residue Gln34 might shed light on the role of the interactions between the loop (residue, 30–43) and β5.

Acknowledgments

Financial support is acknowledged from the NIH (GM RO1 52126). All computer simulations were conducted on BlueGene provided by School of Engineering and Applied Sciences at Harvard University. The authors

acknowledge Dr. Muyoung Heo and Dr. Peter Kutchukian for discussions. The authors also appreciate the help of Ryan R. Cheng for reading the manuscript and helpful comments.

References

1. Schindelin H, Marahiel MA, Heinemann U (1993) Universal nucleic acid-binding domain revealed by crystal-structure of the bacillus-subtilis major cold-shock protein. Nature 364:164–168.
2. Horn G, Hofweber R, Kremer W, Kalbitzer HR (2007) Structure and function of bacterial cold shock proteins. Cell Mol Life Sci 64:1457–1470.
3. Perl D, Welker C, Schindler T, Schroder K, Marahiel MA, Jaenicke R, Schmid FX (1998) Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. Nat Struct Biol 5:229–235.
4. Schindler T, Herrler M, Marahiel MA, Schmid FX (1995) Extremely rapid protein-folding in the absence of intermediates. Nat Struct Biol 2:663–673.
5. Schindler T, Schmid FX (1996) Thermodynamic properties of an extremely rapid protein folding reaction. Biochemistry 35:16833–16842.
6. Jacob M, Holtermann G, Perl D, Reinstein J, Schindler T, Geeves MA, Schmid FX (1999) Microsecond folding of the cold shock protein measured by a pressure-jump technique. Biochemistry 38:2882–2891.
7. Garcia-Mira MM, Boehringer D, Schmid FX (2004) The folding transition state of the cold shock protein is strongly polarized. J Mol Biol 339:555–569.
8. Garcia-Mira MM, Schmid FX (2006) Key role of coulombic interactions for the folding transition state of the cold shock protein. J Mol Biol 364:458–468.
9. Perl D, Holtermann G, Schmid FX (2001) Role of the chain termini for the folding transition state of the cold shock protein. Biochemistry 40:15501–15511.
10. Magg C, Kubelka J, Holtermann G, Haas E, Schmid FX (2006) Specificity of the initial collapse in the folding of the cold shock protein. J Mol Biol 360: 1067–1080.
11. Magg C, Schmid FX (2004) Rapid collapse precedes the fast two-state folding of the cold shock protein. J Mol Biol 335:1309–1323.
12. Reid KL, Rodriguez HM, Hillier BJ, Gregoret LM (1998) Stability and folding properties of a model beta-sheet protein, *Escherichia coli* CspA. Protein Sci 7: 470–479.
13. Rodriguez HM, Vu DM, Gregoret LM (2000) Role of a solvent-exposed aromatic cluster in the folding of *Escherichia coli* CspA. Protein Sci 9:1993–2000.
14. Leeson DT, Gai F, Rodriguez HM, Gregoret LM, Dyer RB (2000) Protein folding and unfolding on a complex energy landscape. Proc Natl Acad Sci USA 97: 2527–2532.
15. Vu DM, Reid KL, Rodriguez HM, Gregoret LM (2001) Examination of the folding of *E-coli* CspA through tryptophan substitutions. Protein Sci 10:2028–2036.
16. Rodriguez HM, Robertson AD, Gregoret LM (2002) Native state EX2 and EX1 hydrogen exchange of *Escherichia coli* CspA, a small beta-sheet protein. Biochemistry 41:2140–2148.
17. Rhoades E, Cohen M, Schuler B, Haran G (2004) Two-state folding observed in individual protein molecules. J Am Chem Soc 126:14686–14687.
18. Schuler B, Lipman EA, Eaton WA (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. Nature 419:743–747.
19. Merchant KA, Best RB, Louis JM, Gopich IV, Eaton WA (2007) Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. Proc Natl Acad Sci USA 104: 1528–1533.
20. Nettels D, Gopich IV, Hoffmann A, Schuler B (2007) Ultrafast dynamics of protein collapse from single-molecule photon statistics. Proc Natl Acad Sci USA 104: 2655–2660.
21. Hoffmann A, Kane A, Nettels D, Hertzog DE, Baumgartel P, Lengefeld J, Reichardt G, Horsley DA, Seckler R, Bakajin O, Schuler B (2007) Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. Proc Natl Acad Sci USA 104:105–110.
22. Chung HS, Louis JM, Eaton WA (2009) Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. Proc Natl Acad Sci USA 106: 11837–11844.
23. Fersht AR, Matouschek A, Serrano L (1992) The folding of an enzyme 1. Theory of protein engineering analysis of stability and pathway of protein folding. J Mol Biol 224:771–782.
24. Alm E, Morozov AV, Kortemme T, Baker D (2002) Simple physical models connect theory and experiment in protein folding kinetics. J Mol Biol 322:463–476.
25. Shea JE, Brooks CL (2001) From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. Annu Rev Phys Chem 52:499–535.
26. Karanicolas J, Brooks CL (2003) The importance of explicit chain representation in protein folding models: an examination of Ising-like models. Proteins 53: 740–747.
27. Morra G, Hodoscek M, Knapp EW (2003) Unfolding of the cold shock protein studied with biased molecular dynamics. Proteins 53:597–606.
28. Huang XQ, Zhou HX (2006) Similarity and difference in the unfolding of thermophilic and mesophilic cold shock proteins studied by molecular dynamics. Biophys J 91:2451–2463.
29. Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, Young C, Batson B, Bowers KJ, Chao JC, Eastwood MP, Gagliardo J, Grossman JP, Ho CR, Ierardi DJ, Kolossváry I, Klepeis JL, Layman T, McLeavey C, Moraes MA, Mueller R, Priest EC, Shan Y, Spengler J, Theobald M, Towles B, Wang SC (2008) Anton, a special-purpose machine for molecular dynamics simulation. Commun ACM 51:91–97.
30. Best RB, Buchete NV, Hummer G (2008) Are current molecular dynamics force fields too helical? Biophys J 95:L7–L9.
31. Yoda T, Sugita Y, Okamoto Y (2004) Comparisons of force fields for proteins by generalized-ensemble simulations. Chem Phys Lett 386:460–467.
32. Freddolino PL, Liu F, Gruebele M, Schulten K (2008) Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. Biophys J 94:L75–L77.
33. Nettels D, Muller-Spath S, Kuster F, Hofmann H, Haenni D, Ruegger S, Reymond L, Hoffmann A, Kubelka J, Heinz B, Gast K, Best RB, Schuler B (2009) Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. Proc Natl Acad Sci USA 106:20740–20745.
34. Yang JS, Wallin S, Shakhnovich EI (2008) Universality and diversity of folding mechanics for three-helix bundle proteins. Proc Natl Acad Sci USA 105:895–900.

35. Zhou YQ, Karplus M (1997) Folding thermodynamics of a model three-helix-bundle protein. Proc Natl Acad Sci USA 94:14429–14432.
36. Zhou YQ, Karplus M (1999) Interpreting the folding kinetics of helical proteins. Nature 401:400–403.
37. Linhananta A, Zhou YQ (2002) The role of sidechain packing and native contact interactions in folding: discontinuous molecular dynamics folding simulations of an all-atom G(o)over-bar model of fragment B of Staphylococcal protein A. J Chem Phys 117:8983–8995.
38. Shimada J, Kussell EL, Shakhnovich EI (2001) The folding thermodynamics and kinetics of crambin using an all-atom Monte Carlo simulation. J Mol Biol 308: 79–95.
39. Shimada J, Shakhnovich EI (2002) The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation. Proc Natl Acad Sci USA 99:11175–11180.
40. Clementi C, Nymeyer H, Onuchic JN (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. J Mol Biol 298:937–953.
41. Luo ZL, Ding JD, Zhou YQ (2007) Temperature-dependent folding pathways of pin1 WW domain: an all-atom molecular dynamics simulation of a G(o)over-bar model. Biophys J 93:2152–2161.
42. Luo ZL, Ding JD, Zhou YQ (2008) Folding mechanisms of individual beta-hairpins in a Go model of Pin1 WW domain by all-atom molecular dynamics simulations. J Chem Phys 128:225103.
43. Ding F, Guo WH, Dokholyan NV, Shakhnovich EI, Shea JE (2005) Reconstruction of the src-SH3 protein domain transition state ensemble using multiscale molecular dynamics simulations. J Mol Biol 350: 1035–1050.
44. Hills RD, Brooks CL (2009) Insights from coarse-grained go models for protein folding and dynamics. Int J Mol Sci 10:889–905.
45. Klimov DK, Thirumalai D (2000) Mechanisms and kinetics of beta-hairpin formation. Proc Natl Acad Sci USA 97:2544–2549.
46. Chavez LL, Onuchic JN, Clementi C (2004) Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. J Am Chem Soc 126:8426–8432.
47. Dokholyan NV, Li L, Ding F, Shakhnovich EI (2002) Topological determinants of protein folding. Proc Natl Acad Sci USA 99:8637–8641.
48. Paci E, Vendruscolo M, Karplus M (2002) Validity of Go models: comparison with a solvent-shielded empirical energy decomposition. Biophys J 83:3032–3038.
49. Paci E, Vendruscolo M, Karplus M (2002) Native and non-native interactions along protein folding and unfolding pathways. Proteins 47:379–392.
50. Tsai J, Taylor R, Chothia C, Gerstein M (1999) The packing density in proteins: standard radii and volumes. J Mol Biol 290:253–266.
51. Zhou YQ, Karplus M, Wichert JM, Hall CK (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.
52. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 314:141–151.
53. Ferrenberg AM, Swendsen RH (1989) Optimized Monte Carlo data-analysis. Phys Rev Lett 63:1195–1198.
54. Frenkel D, Smit B (2001) Understanding Molecular Simulation. 2nd ed. San Diego: Academic Press.
55. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1995) Multidimensional free-energy calculations using the weighted histogram analysis method. J Comput Chem 16:1339–1350.
56. Makhatadze GI, Marahiel MA (1994) Effect of pH and phosphate ions on self-association properties of the major cold-shock protein from *Bacillus-Subtilis*. Protein Sci 3:2144–2147.
57. Cellmer T, Henry ER, Hofrichter J, Eaton WA (2008) Measuring internal friction of an ultrafast-folding protein. Proc Natl Acad Sci USA 105:18320–18325.
58. O'Brien EP, Ziv G, Haran G, Brooks BR, Thirumalai D (2008) Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. Proc Natl Acad Sci USA 105:13403–13408.
59. Nettels D, Hoffmann A, Schuler B (2008) Unfolded protein and peptide dynamics investigated with single-molecule FRET and correlation spectroscopy from picoseconds to seconds. J Phys Chem B 112:6137–6146.
60. Du R, Pande V, Grosberg A, Tanaka T, Shakhnovich EI (1998) On the transition coordinate for protein folding. J Chem Phys 108:334–350.
61. Shakhnovich EI (2006) Protein folding thermodynamics and dynamics: where physics, chemistry, and biology meet. Chem Rev 106:1559–1588.
62. Muff S, Caflisch A (2009) Identification of the protein folding transition state from molecular dynamics trajectories. J Chem Phys 130:125104.
63. Hubner IA, Shimada J, Shakhnovich EI (2004) Commitment and nucleation in the protein G transition state. J Mol Biol 336:745–761.
64. Xu D, Zhang Y (2009) Generating triangulated macromolecular surfaces by Euclidean distance transform. Plos One 4:e8140.
65. Lipman EA, Schuler B, Bakajin O, Eaton WA (2003) Single-molecule measurement of protein folding kinetics. Science 301:1233–1235.
66. Shakhnovich EI (1997) Theoretical studies of protein-folding thermodynamics and kinetics. Curr Opin Struct Biol 7:29–40.