Effects of Disulfide Bonds on Folding Behavior and Mechanism of the β-Sheet Protein Tendamistat

Meng Qin,* Jian Zhang,* and Wei Wang*†
*National Laboratory of Solid State Microstructure and Department of Physics, Nanjing University, Nanjing 210093, China; and †Interdisciplinary Center of Theoretical Studies, Chinese Academy of Science, Beijing 100080, China

ABSTRACT Tendamistat, a small disulfide-bonded β-sheet protein, and its three single/double-disulfide mutants are investigated by using a modified Gō-like model, aiming to understand the folding mechanism of disulfide-bonded protein as well as the effects of removal of disulfide bond on the folding process. Our simulations show that tendamistat and its two single-disulfide mutants are all two-state folders, consistent with the experimental observations. It is found that the disulfide bonds as well as three hydrogen bonds between the N-terminal loop-0 and strand-6 are of significant importance for the folding of tendamistat. Without these interactions, their two-state behaviors become unstable and the predictions of the model are inconsistent with experiments. In addition, the effect of disulfide bonds on the folding process are studied by comparing the wild-type tendamistat and its two mutants; it is found that the removal of either of the C11-C27 or C45-C73 disulfide bond leads to a large decrease in the thermodynamical stability and loss of structure in the unfolded state, and the effect of the former is stronger than that of the later. These simulation results are in good agreement with experiments and, thus, validate our model. Based on the same model, the detailed folding pathways of the wild-type tendamistat and two mutants are studied, and the effect of disulfide bonds on the folding kinetics are discussed. The obtained results provide a detailed folding picture of these proteins and complement experimental findings. Finally, the folding nuclei predicted to be existent in this protein tendamistat as well as its mutants are firstly identified in this work. The positions of the nucleus are consistent with those argued in experimental studies. Therefore, a nucleation/growth folding mechanism that can explain the two-state folding manner is clearly characterized. Moreover, the effect by the removal of each disulfide bond on the folding thermodynamics and dynamics can also be well interpreted from their influence on the folding nucleus. The implementation of this work indicates that the modified Gō-like model really describes the folding behavior of protein tendamistat and could be used to study the folding of other disulfide-bonded proteins.

INTRODUCTION

Protein folding, one of the most fundamental events in cells, is very complex and difficult to be characterized in detail. Although much progress, especially the energy landscape theory (1,2), has been made based on studies on some small and single domain proteins (3–5), this important problem has not been completely solved for more than 30 years. The folding processes of proteins involve a number of noncovalent weak interactions that result in contacts between the residues. However, many proteins, such as some membrane and secreted proteins in both bacteria and eukaryotes, fold into their native structures requiring the formation of disulfide bonds. Disulfide bonds are covalent interactions between residues, and are also vital for the stability and activities of the proteins (3–11). They can be formed in two kinds of compartments, namely the eukaryotic endoplasmic reticulum (ER) and the bacterial periplasm. When two sulfhydryl groups are close enough to each other in these two oxidative environmental compartments, oxidation of the thiol groups yields a disulfide bond. If the disulfide bond is of correct pairing, it can help to maintain the tertiary structure of the protein. Otherwise, incorrectly paired disulfide bonds can be recognized and broken up by certain special enzymes in ER or bacterial periplasm. This redox process involves the reaction between the thiol and the disulfide bond (namely thiol-disulfide bond exchange process) with the help of some special enzymes.

The folding process involving the formation of disulfide bonds is very complex. In the refolding experiment of the ribonuclease in vitro by Anfinsen and co-workers, different additive orders of oxidants and buffer lead to completely different final structures and activities (3). This indicates the delicate role of disulfide bonds in the folding process. Later, experiments showed that the folding process of some disulfide-contained proteins in vivo is much more complex because it requires not only the participation of oxidants but also the help of many special enzymes (12–14). For example, DsbA in bacterial periplasm can help to form disulfide bonds without ability of recognizing correct or incorrect pairing, and DsbC can recognize the incorrect pairing and cut it off by a thiol-disulfide exchange reaction.

Thus, it is difficult to explore such a complex folding process experimentally and theoretically. So far, people mostly focus on the refolding process without involving the rupture and reformation of disulfide bonds. If the denatured conformations of a protein are obtained by unfolding from native state in solvent without presence of reductant, the disulfide bonds remain intact. These denatured conformations act as the initial conformations of the following refolding process, and the disulfide bonds in these initial conformations are

Submitted March 23, 2005, and accepted for publication September 14, 2005.
Address reprint requests to Wei Wang, E-mail: wangwei@nju.edu.cn.

0006-3495/06/01/272/15 $2.00
doi: 10.1529/biophysj.105.063552

termed as preformed disulfide bonds. Thus the refolding only involves noncovalent interactions, and the formation process of disulfide bonds is avoided. The folding process becomes simple and could be clearly studied experimentally and theoretically. In fact, the folding of proteins with preformed disulfide bonds occurs in the unfolding-refolding experiments in the absence of reductant and oxidant. Such process is meaningful because it is shown experimentally that the formation of disulfide bonds is fast and the later folding process is mainly determined by the noncovalent interactions. As a matter of fact, in many previous works proteins with the preformed disulfide bonds were used, and many aspects related to the disulfide bonds have been obtained (15–21).

It has been shown in many experimental and theoretical studies that the preformed disulfide bonds can significantly increase the stability of the native structures of proteins. However, to the folding kinetics, there are two different opinions on the effects of the preformed disulfide bonds. One assumed that the preformed disulfide bonds can increase the folding rate because the number of possible conformations is decreased significantly (22,23), whereas the other argued that they can lead to kinetically trapped intermediates during the folding process and the folding becomes slow (24). Thus, which picture is more close to the reality is still an open question. Moreover, the folding processes of proteins with disulfide bonds mostly involve multiple pathways. The physical origins and biological behaviors of these pathways still need to be clarified.

To investigate the effects of disulfide bonds on the stability and the folding process of proteins, tendamistat, an α-amylase inhibitor from *Streptomyces tenda*, is a good model system. It is an all-β-sheet protein with 74 amino acids and has two disulfide bonds (25,26). It has been well studied in a number of experiments (16,17,19,20,27–30). It was found that protein tendamistat exhibits a two-state folding behavior that has not been found in any other disulfide-bonded proteins up to now. Interestingly, this two-state folding behavior cannot be eliminated by the removal of either disulfide bond (17). It is more intriguing that the removal of the *C*11–*C*27 disulfide bond that only circles 17 residues can lead to larger effects on the stability than the removal of the *C*45–*C*73 one that circles 29 residues (17,27). The transition state has also been studied. It is found that the first β-hairpin (the *C*11–*C*27 disulfide bond locating at its base) is formed in the transition state, and the regions of the outer strands near the *C*45–*C*73 disulfide bond are formed in the later folding process. Besides, the folding mechanism of protein tendamistat was investigated and a nucleation/growth mechanism was speculated by Kiefhaber et al. (17). Such a folding mechanism was also mentioned by Capaldi and co-workers (31). However, this mechanism has not been evidently proved yet. Although much insight has been obtained, there are still many questions to be answered. For example, i), why is protein tendamistat so special as to present a two-state folding behavior compared to other disulfide-bonded proteins? ii), How do we explain the specific effects of each disulfide bond on the thermodynamic stability and folding pathways? iii), Can the nucleation/growth mechanism describe the special folding process clearly as previously predicted?

Computer simulations, as a powerful theoretical method, can complement the experiments and provide detailed folding information related to the above-mentioned questions. The combination of experiments and computer simulations is essential for understanding the detailed folding mechanism and the roles played by disulfide bonds in the folding process. Computer simulations vary from simplified models to all atomic-level models. Although much detailed information of the folding process can be provided, the all atomic-level simulation is limited with timescale less than 1 μs, which is too short compared to the folding timescale (32,33). Only few studies concerning the effect of the disulfide bonds on the folding process are focused in the local fragments with preformed disulfide bond (34). Therefore, simplified models might be workable to simulate the folding of disulfide-bonded proteins for understanding the folding mechanism. Very recently, Cho and co-workers showed that a modified Gō-like model can be used to investigate the dimerization and aggregation with preformed disulfide bonds (35).

In this work, the detailed folding kinetics, pathway, and mechanism, as well as the effects of each disulfide bond on the thermodynamic stability and folding kinetics for the two-state folder tendamistat are studied using molecular dynamics simulations. A coarse-grained model with modified Gō-type interactions is used. The two-state folding behaviors of tendamistat and its two single-disulfide mutants are well studied. The folding transition of each mutant occurs at lower temperature than that of the wild type, corresponding to the decrease of the thermodynamic stability. The decrease in the thermodynamic stability of the *C*11*A*/*C*27*S* mutant is larger than that of the *C*45*A*/*C*73*A* mutant. This agrees well with the experimental observations (17,27). Furthermore, the detailed folding pathways of the wild-type tendamistat and its two mutants are found, which is consistent with the experimental findings. A nucleation/growth folding mechanism is clearly characterized, and the folding nuclei of the wild-type tendamistat and its two mutants are identified (36–38). This mechanism can well explain the two-state folding behavior. The effects of each disulfide bond on the thermodynamic stability and the folding processes are also interpreted from their influence on the folding nucleus.

The arrangement of this article is as follows. After the introduction, the model and methods employed in this article are described in ''Models and Methods''. The two-state behaviors of the wild-type tendamistat and its two mutants, and the effects of the disulfide bonds are investigated in the following section. In ''Folding Pathway and the Effects of

the Disulfide Bonds'', the folding pathways of the wild type and its mutants are described in detail. The nucleation/growth mechanism is discussed in ''The Folding Nucleus''. Finally, in ''Conclusions'', the conclusions are given.

MODEL AND METHODS

As shown in Fig. 1, a small all-β-sheet protein tendamistat is taken as the model protein in this work. It is an α-amylase inhibitor from *S. tenda* and consists of 74 amino acids (25,26). This small β-protein has two three-stranded antiparallel β-sheets named as $S_{\rm I}$ and $S_{\rm II}$, respectively. The β-sheet $S_{\rm I}$ is composed of strand-1 (residues 12–17), strand-2 (residues 20–26), and strand-5 (residues 52–57), and $S_{\rm II}$ is composed of strand-3 (residues 30–37), strand-4 (residues 41–49), and strand-6 (residues 67–73). The N-terminal loop-0 (residues 5–8) interacts tightly with $S_{\rm II}$ by three hydrogen bonds. The notation of the first β-hairpin that is composed of strand-1 and strand-2 S_{1-2} is S_{1-2}, as well as the second β-hairpin S_{3-4} and other β-sheets. There are two disulfide bridges between Cys-11–Cys-27 and Cys-45–Cys-73. The Cys-11–Cys-27 disulfide bond locates at the base of the first β-hairpin, and the Cys-45–Cys-73 one connects two outer strand-4 and strand-6 of $S_{\rm II}$. The three mutants of tendamistat are also studied in this work. These three mutants M_1 (C11A/C27S), M_2 (C45A/C73A), and M_3 (C11A/C27S and C45A/C73A) are obtained by substituting residue cysteine with residue alanine (A) or serine (S) as in the experiments. That is, in mutants M_1 and M_2, only one disulfide bond is kept, and in mutant M_3, both disulfide bonds are removed. In our theoretical study, the native structures of these three mutants are obtained by energy minimization based on the wild-type native structure; this takes into consideration in experiment that their native structures were observed to be similar to the wild type (28). The energy minimization is realized by using Amber software (version 7.0) (39) after the replacements of the related residues (17).

The folding processes of protein tendamistat and its mutants are studied using the molecular dynamics (MD) simulations based on a Gō-like model, which was successfully applied in many works (40–50). However, in this work, the potential energy is modified as follows (48,49). Each residue in the model protein chain is represented as a single bead centered in its corresponding C^{α} position. The energy $E_{\rm p}$ of a certain conformation Γ (Γ_0 as its native state) is given by the expression:

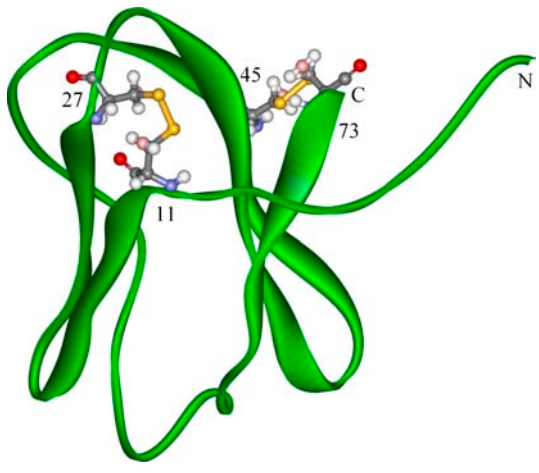

FIGURE 1 The schematic representation (backbone model) of the structure of wild-type tendamistat. The cysteines are displayed with ball-and-stick model and the disulfide bonds are shown with yellow color. The first disulfide bond (Cys-11–Cys-27) is located at the base of the first β-hairpin (composed of strand-1 and strand-2). The second disulfide bond (*C*45– *C*73) connects the two outer strands-4 and strand-6.

$$E_{\rm P}(\Gamma,\Gamma_0)=\sum_{\rm i=1}^{\rm N-1}K_{\rm r}(r_{\rm i,i+1}-r_0)^2+\sum_{\rm i=1}^{\rm N-2}K_\theta(\theta_{\rm i}-\theta_{0\rm i})^2+\sum_{\rm i=1}^{\rm N-3}[K_1(1-\cos(\phi_{\rm i}-\phi_{0\rm i}))+K_2(1-\cos(3(\phi_{\rm i}-\phi_{0\rm i})))]+\xi_1+\xi_2. \quad (1)$$

In Eq. 1, $r_{\rm i,i+1}$ denotes the distance between two consecutive residues, $\theta_{\rm i}$ and $\phi_{\rm i}$ represent the bond angle formed by three consecutive residues and dihedral angle formed by four consecutive residues at conformation Γ, respectively. The related quantities in the native state Γ_0 are represented as r_0, θ_0, and ϕ_0. Parameters $K_{\rm r}=100\varepsilon$, $K_\theta=20\varepsilon$, $K_1=\varepsilon$, and $K_2=0.5\varepsilon$ weigh the relative strength of each kind interaction. Here ε is the energy unit.

The term ξ_1 in Eq. 1 includes the native and nonnative contact interactions that take the form of Lennard-Jones potential as

$$\xi_1=\sum_{\rm i<j-3}\{\varepsilon_1(i,j)[5(\sigma_{\rm ij}/r_{\rm ij})^{12}-(\sigma_{\rm ij}/r_{\rm ij})^{10}]+\varepsilon_2(i,j)(\sigma_{\rm non_{ij}}/r_{\rm ij})^{12}\}. \quad (2)$$

Here, the native contact is defined when the distance between any pair of two heavy atoms from respective residues i and j is within 4.5 Å in the native structure of the protein. Thus, in any conformation, all contacts could be assigned as native or nonnative ones. For native contacts, we set $\varepsilon_1(i,j)=\varepsilon$ and $\varepsilon_2(i,j)=0$, and for nonnative contacts, $\varepsilon_1(i,j)=0$ and $\varepsilon_2(i,j)=\varepsilon$. In Eq. 2, the parameter $\sigma_{\rm ij}$ is the distance between residues i and j at the native state for native interactions, while $\sigma_{\rm non_{ij}}=4.0$ Å corresponds to the distance for nonnative repulsive interactions.

It is known that hydrogen bond is an important kind of interaction in proteins. Generally, in the Gō-like models, the hydrogen bonds are not taken into account explicitly but are implicitly included in the contact interactions that are approximately five times stronger than the van der Waals interactions. However, for many proteins, e.g., the protein tendamistat in this work, some hydrogen bonds are vital to the folding process and structural stability. Three specific hydrogen bonds (Ser-5–Arg-72, Leu-70–Ala-8, and Glu-6–Arg-72) existing between the N-terminal loop-0 and strand-6, are very important for the stability of protein tendamistat, especially when the *C*45–*C*73 disulfide bond is removed. It was clearly shown that strand-6 is held in place by the N-terminal arm and these hydrogen bonds keep the protein stable even after the *C*45–*C*73 disulfide bond is removed by residue replacement (27).

The roles of these three hydrogen bonds cannot be implemented by the contact interactions in the general Gō-like model because the strengths of contacts scaled by ε is only about one-fifth of the hydrogen bonds. Previous study indicated that the hydrogen bonds can also be described with the similar form of the van der Waals force but the interaction strength is ~25 times stronger (51). Thus the interactions of hydrogen bonds can be characterized with the form of Eq. 2 when the parameter $\varepsilon_1(i,j)$ and $\varepsilon_2(i,j)$ are taken to be 5ε and 0, respectively. The interaction basin of a hydrogen bond is defined when the distance between each residue of the hydrogen-bond pair is $<1.2\,\sigma_{\rm ij}$. Out of this basin, the interaction of the hydrogen-bond pair is still considered the contact potential as discussed above. It is worthy of note that the folding behaviors of the mutants are distinctly different from the experimental results when the specific role of these three hydrogen bonds is not considered (see detailed discussion in the following section).

The covalent disulfide bond is formed between two thiol groups in each of the residue cysteines, and however, can be modeled as a virtual bond between the two beads of the related cysteines. Physically, such a form of disulfide bond must have more flexibilities than the two consecutive residues of the peptide bond. Because the formation and rupture of the disulfide bonds are not considered during our simulations for the folding and unfolding processes, a modification on the Gō-like model with disulfide bonds can be made. That is, the preformed disulfide bonds existed both in the initial state and the folding processes can be taken as a harmonic oscillation potential between a pair of cysteines. This form of disulfide bonds

characterize well the strong interaction of each paired cysteines in nature as a term ξ_2 in Eq. 2

$$\xi_2 = K_s(r_{i,j} - r_{i,j}^N)^2. \quad (3)$$

Here, the interaction strength K_s is taken to be 20ε, which can describe the flexibility of the related cysteines.

The Langevin dynamics is used to simulate the folding process (52–54). The motion equation includes terms of a damping with a proper friction coefficient γ, a random force F_r to balance the energy dissipation caused by friction, and the conformation force $F_c = -\nabla_r E_P$ as follows

$$m\ddot{\vec{r}} = -\gamma\dot{\vec{r}} + F_c + F_r. \quad (4)$$

Here, $\vec{r}$ and $\dot{\vec{r}}$ are the coordinate vector and the velocity, respectively. m is the mass (the mass of each residue is assumed to be same). F_r is obtained from the Gaussian distribution with a white noise spectrum and the standard variance related to temperature T as

$$\langle F_r(0)F_r(t)\rangle = 2\gamma k_B T\delta(t), \quad (5)$$

where k_B is the Boltzmann constant, t denotes the time, and $\delta(t)$ is the Dirac δ-function. The equation of motion is numerically integrated using the velocity form of the Verlet algorithm (55,56). The integration time step is taken to be $\Delta t = 0.005\tau$, where $\tau = \sqrt{ma^2/\varepsilon}$ is a characteristic oscillatory time unit ($a = 5$ Å). The friction coefficient is taken to be $\gamma = 0.2$.

The specific heat is widely and successfully used to characterize the thermodynamic behaviors of the folding process (57–61). The curves of the specific heat versus temperature reflect the folding behaviors, i.e., a two-state folding corresponds to single transition peak and three-state folding to double transition peaks. In fact, the single transition peak is related to an overlapping of the collapse transition and the folding transition (58). The specific heat C_v and the free-energy F are calculated using the weighted histogram analysis method, which yields an optimal estimate of the density of states of the system and hence allows the calculation of relevant thermodynamic quantities (62,63). In details, the specific heat and the free energy are calculated based on 25 independent long-time runs at different temperatures (62,64).

TWO-STATE FOLDING BEHAVIORS AND THE EFFECTS OF DISULFIDE BONDS ON THERMODYNAMIC STABILITY

Previously, it has been shown that disulfide-bond contained proteins usually fold with multipathways. However, in a number of experiments, the folding and unfolding reactions of protein tendamistat (containing two disulfide bonds) were found to exhibit two-state behavior by Kiefhaber and coworkers (16,17,19–20). It was also shown that the removal of each disulfide bond has little effect on such behavior, but could lead to a large decrease in the thermodynamic stability. This decrease is larger for the removal of the $C11$–$C27$ disulfide bond than that for the $C45$–$C73$ one (17,27). In this section, some quantities (e.g., specific heat C_v, free-energy F, the structural similarity Q versus the time, unfolding time, and so on) based on the MD simulations are used to characterize the two-state folding and unfolding behaviors of the wild-type tendamistat and its mutants.

Fig. 2 shows the specific heat profile (C_v) as a function of temperature T for the wild-type tendamistat and its three mutants. There is only one main transition peak at temperature $T = 1.217$ for the wild-type tendamistat, at $T_f = 1.179$ for mutant M_1, and at $T_f = 1.189$ for mutant M_2, respectively, implying a two-state folding behavior as found in experiments (16). Here the temperature unit is ε/k_B and k_B is the Boltzmann constant. However, the peaks for the disulfide bond mutants shift obviously to the low-temperature range, which is well consistent with the experimental observations (17,27). These clearly elucidate that the removal of each disulfide bond has little effect on the two-state folding behavior, but considerably results in a large decreasing in the thermodynamic stability. Generally, this decreasing in the stability should be related to the removal of the disulfide bonds because their loop entropy, which is believed to be proportional to the loop size (65), is increased. Thus, for both mutants, the transition temperatures, i.e., $T = 1.179$ and $T = 1.189$, are smaller than that for the wild type. However, from the loop entropy, the transition temperature for M_2 should be lower than that of M_1 because the loop length is $l = 17$ for the later whereas it is $l = 29$ for the former. But this is not the case according to our simulations (Fig. 2) and the experiments. This cannot be explained simply by the loop entropy. From the detailed structure of M_2, it is known that the strand-6 in S_{II} is still held in place tightly by the N-terminal arm (loop-0, residues 5–8). Clearly, such a packing results from three hydrogen bonds between the loop-0 and strand-6, not from the $C45$–$C73$ disulfide bond that is removed for the M_2. Thus, the loss of stability due to the removal of the $C45$–$C73$ disulfide bond could be compensated and the region of S_{II} could still be kept intact, resulting in a transition temperature closer to that of the wild type.

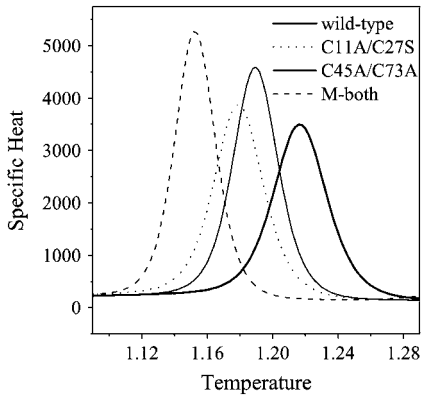

FIGURE 2 The specific heat C_v as a function of the temperature T. The transition temperatures of the wild-type tendamistat (*thick solid line*), the C11A/C27S mutant M_1 (*dotted line*), the C45A/C73A mutant M_2 (*thin solid line*), and the mutant M_3 removed both of the disulfide bonds (*dashed line*) are 1.217, 1.179, 1.189, and 1.152, respectively.

Differently, for M_1 the existence of these hydrogen bonds has no effect on the local structure S_I and the loss of the stability is mainly due to the removal of the $C11$–$C27$ disulfide bond. This disulfide bond (in the wild-type tendamistat) is located at the base of the first β-hairpin S_{1-2}, which acts as an important component for the folding nucleus. The removal of the $C11$–$C27$ disulfide bond changes the distribution of the nucleus (see the detailed discussion in the following

section), which can decrease the barrier of the transition state and leads to the large loss of the thermodynamic stability. Thus, the folding transition of M_1 occurs at a much lower temperature than that of the wild-type tendamistat, so much as lower than that of the mutant M_2.

For the mutant M_3 in which both of the disulfide bonds are removed, the folding transition temperature shifts to a much lower value as $T_f = 1.1526$. Such a transition temperature is too low and beyond the foldable temperature region of the wild type. This means that the mutant M_3 cannot fold and yield native structure when it is settled at the temperatures in the foldable temperature region of the wild type. This agrees with the experimental observation that M_3 is unstable and no native structure can be found (17). It is worthy of note that in the temperature region lower than the folding temperature $T_f = 1.1526$ of the mutant M_3, the mutant is foldable and the folded structure is native-like, exhibiting obviously a two-state folding behavior from the simulations. However, such foldability for the mutant M_3 is artificial, and is only on the theoretical base and model related.

It is worth noting that when the above-mentioned hydrogen bonds are not considered, the folding processes of the wild-type tendamistat and its two single-disulfide mutants still exhibit two-state folding behaviors. However, the folding transition of the *C45A/C73A* mutant M_2 shifts to a temperature $T_f = 1.171$, which is lower than that of the *C11A/C27S* mutant M_1, i.e., $T_f = 1.179$ (here the transition temperature of M_1 shows no changes). Though this shift seems to accord with the theory of loop entropy, it is distinctly different from the experimental results and could not depict the structural specialty between loop-0 and S_{II} in reality. Therefore, these hydrogen bonds must be taken into account particularly due to their importance.

To display the two-state folding process clearly, we monitor the reaction coordinate Q, i.e., a fraction of native contacts, as a function of time (in unit of MD steps) near the folding temperature T_f. The two-state behaviors are shown obviously in Fig. 3, *a–d*, where the ensembles of unfolded state and the folded state are equally populated and no intermediates are observed. Removal of each or both disulfide bonds cannot have any effect on this two-state folding manner, but makes the structures of unfolded state more noncompact and dissimilar to the native state. Fig. 3 *e* shows the distribution of the structural similarity Q in the unfolded state for the wild-type tendamistat and its three mutants M_1, M_2, and M_3. The value of structural similarity at the peak of the distribution of unfolded state, namely Q_u, is used to characterize the structures in the unfolded state. The value of Q_u becomes small, implying that the structures in unfolded state are less native-like due to the removal of each disulfide bond. The difference of Q_u between the wild type and the mutant M_1 (or between the wild type and the mutant M_2, M_3) is $\Delta Q_u = 0.033$ ($\Delta Q_u = 0.02, 0.048$). This exhibits that the removal of the *C11–C27* disulfide bond makes the structures in the unfolded state less native-like than those of the *C45–C73* one. This is consistent with the increasing in $\Delta m_{eq} = -0.55$ for M_1 and $\Delta m_{eq} = -0.42$ for M_2 measured experimentally. Here $\Delta m_{eq} = m_{eq}^{\text{wild type}} - m_{eq}^{\text{mutant}}$ reflects the changes in solvent-accessibility surface area that is proportional to m_{eq} upon unfolding processes. The increase of m_{eq} for mutants M_1 and M_2 indicates a loss of structure and a higher degree of solvent accessibility in the unfolded state (17). Lacking the two disulfide bonds, the unfolded state of the mutant M_3 becomes extraordinarily disordered and the value of Q at the center of the unfolded conformations is only 0.08.

Furthermore, to obtain a better understanding of the two-state folding process and the effects caused by removal of single-disulfide bond, the free-energy landscapes are investigated in detail. Fig. 4, *a–d*, show the free-energy contours using the structural similarity Q and radius gyration R_g as reaction coordinates for the wild-type tendamistat and its three mutants at their folding temperatures. The free-energy landscape is smooth and funnel-like, indicating a very stable two-state folder sliding from the unfolded state into the folded state (16,66,67). The free-energy minimum of the unfolded state for the wild-type tendamistat is located around $Q = 0.10 \sim 0.17$ and $R_g = 3.1 \sim 4.1$, while the folded state is situated around $Q = 0.64 \sim 0.90$ and $R_g = 2.1 \sim 2.5$ (Fig. 4 *a*). The free-energy barrier, located between these two minima, is responsible for the thermodynamic stability. There is no kinetic intermediate in the folding process, which is extraordinarily particular for the disulfide-bonded proteins. Moreover, the experiments found that there is no obvious chain collapse in the folding process (16), which can also be validated in the free-energy landscape. For the three mutants M_1, M_2, and M_3, the free-energy landscapes also exhibit the obvious two-state folding features, respectively. The removal of each or both disulfide bonds cannot change such folding manner (consistent with the above discussion), but can change the structural characters of the two minima of the free energy, especially for the unfolded state. The removal of the *C11–C27* disulfide bond makes the base of the first β-hairpin open, and increases the solvent accessibility of the ensemble of unfolded state. This can be shown from the position of center of the unfolded state in the free-energy plot, the values of the Q and R_g for unfolded state of the *C11–C27* mutant change to $Q = 0.1$ and $R_g = 4.1$, comparable to $Q = 0.15$ and $R_g = 3.5$ for the wide-type (Fig. 4 *b*). For the case of the removal of the *C45–C73* disulfide bond, the effects are less than that of the *C11–C27* one (Fig. 4 *c*). When both of the disulfide bonds are mutated, the value of Q and R_g change to 0.08 and 4.8, respectively, which corresponds to the extraordinary disordered unfolded state (Fig. 4 *d*).

The profiles of the probability of the unfolded state P_u versus the unfolding time (τ_u) at high temperatures $1.2T_f$ are shown in Fig. 5 for the wild-type tendamistat and the two single-disulfide mutants, respectively. From these curves, it can obviously be seen that the unfolding process also follows the two-state model. Each removal of the disulfide bond can lead to large decrease of the stability, and increases the

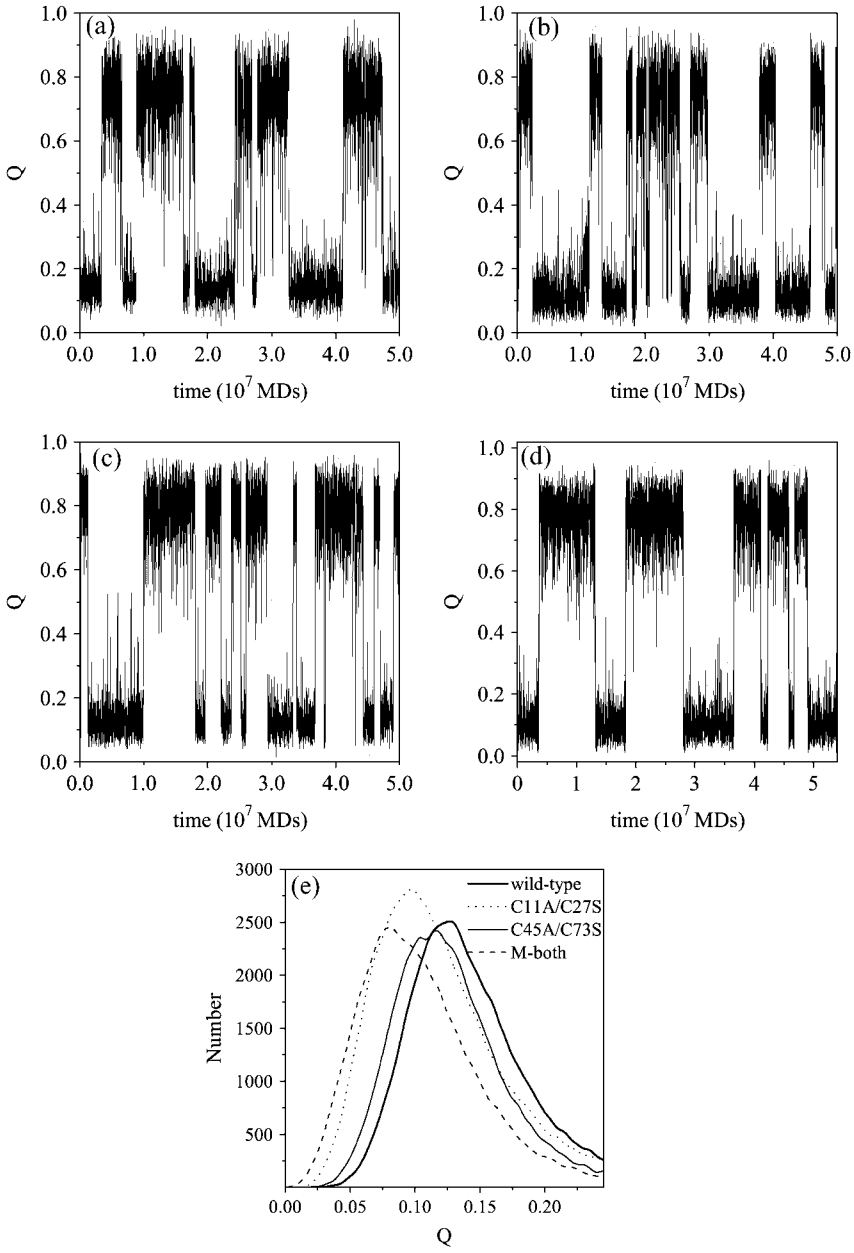

FIGURE 3 (*a–d*) The trajectories of the reaction coordinate Q as a function of time (in unit of 10^7 molecular dynamics steps), which are obtained from the typical simulations around the respective folding temperature. Each of the model protein has equal probabilities to be found in the unfolded state and in the folded state, which exhibits the two-state folding behavior clearly; (*a*) for the wild-type tendamistat, (*b*) for the mutant M_1, (*c*) for the mutant M_2, and (*d*) for the mutant M_3. Fig. 3 *e* shows the distribution of the native contact in the unfolded states for the wild-type tendamistat (*thick solid line*), the mutant M_1 (*dotted line*), the mutant M_2 (*thin solid line*), and the mutant M_3 (*dashed line*). Q_u is the value of Q at each of the peak of the distribution.

unfolding rate distinctly. This is consistent with the experimental results that the unfolding rate of the *C11A/C27S* mutant is faster than that of the *C45A/C73A* mutant and the mutant M_3 has no stable structure and unfolds completely in a very short time. The *C*11–*C*27 disulfide bond circling a loop of 17 residues leads to more effects on the stability than those by the *C*45–*C*73 disulfide bond circling a much longer loop of 29 residues.

FOLDING PATHWAY AND THE EFFECTS OF THE DISULFIDE BONDS

Experimentally, the folding behavior has been investigated intensively and many interesting features for the wild-type tendamistat and its mutants M_1 and M_2 have been shown (16,17,19,20,29). However, up to now, the detailed folding pathway and mechanism are still unclear and require theoretical interpretation. In this work, we calculate $\langle Q_i \rangle$ as a function of the structural similarity Q to show the detailed folding process. Here $\langle Q_i \rangle$ is an average formation probability of the i-th native contact over the set of all conformations with a given value Q collected in all the simulations performed at a certain temperature (68):

$$\langle Q_i(Q) \rangle_T = \left\langle \delta(Q_i - 1)\delta\left(\sum_{i=1}^{M} Q_i/M - Q\right)\right\rangle_T, \qquad (6)$$

where each value is averaged over >500 different folding trajectories. Moreover, to simplify the notation, we indicate

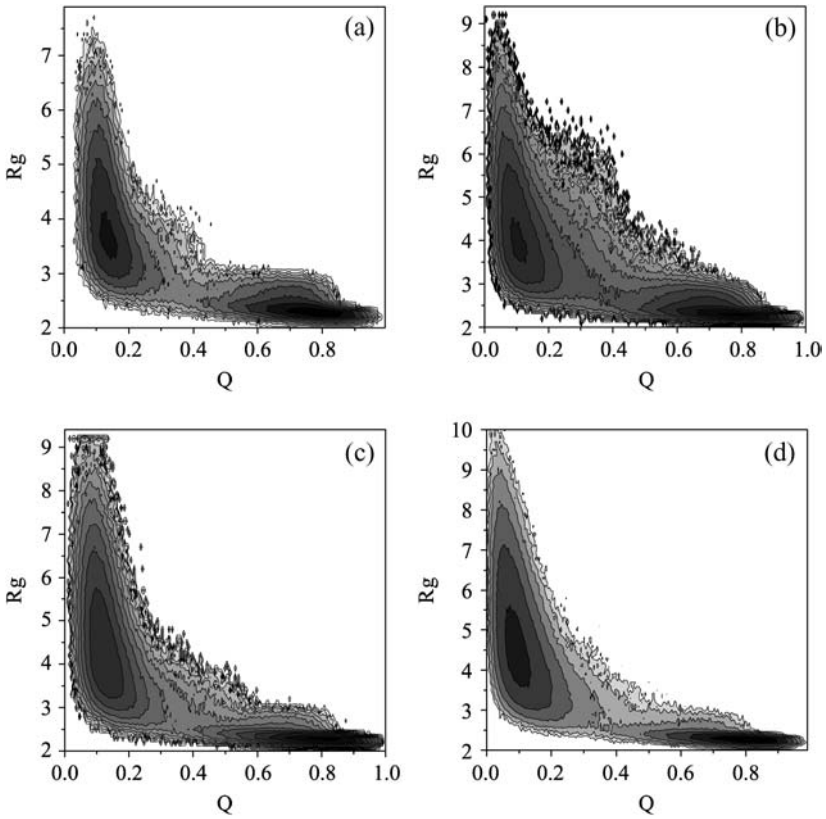

FIGURE 4 Free-energy contour maps as the function of two reaction coordinates: the structural similarity Q and the radius gyration R_g near their respective folding temperatures. The two-state folding processes can be clearly exhibited in the free-energy landscapes of the wild-type tendamistat (*a*), the mutant M_1 (*b*), the mutant M_2 (*c*), and the mutant M_3 (*d*). The removal of each single-disulfide bond does not take any effects to the two-state folding behavior (refer to the text for more detailed discussion). The free-energy difference between adjacent contour lines is 1.0 k_BT.

$\langle Q_{S1-2} \rangle$ to represent the formation probability of the total native contacts in the β-hairpin S_{1-2}, which is composed of strand-1 and strand-2. Other probabilities for S_{3-4}, S_{1-5}, S_{2-5}, S_{3-6}, S_{4-6}, and S_{0-6} are similar to $\langle Q_{S1-2} \rangle$. The formation of each local region, clearly exhibited in the curves of these probabilities, is used to characterize the whole folding pathways in detail.

Fig. 6 shows the formation probabilities of different parts of the wild-type tendamistat. The first β-hairpin S_{1-2} forms very fast and most of the native contacts (>60%) in S_{1-2} have been formed in the very early folding process ($Q < 0.1$). The formation of the other β-hairpin S_{3-4}, which is mostly formed before the transition state is similar to that of the S_{1-2}. Here, the transition state corresponds to the minimal region with $0.4 \leq Q \leq 0.5$ in the histogram of the distribution of Q values near the folding temperature. The formation processes of these two β-hairpins accord well with

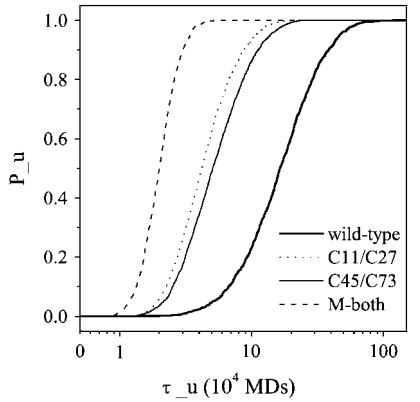

FIGURE 5 The probabilities of the unfolded state P_u as a function of the unfolding time τ_u (in unit of 10^4 molecular dynamic steps) for the wild-type tendamistat (*thick solid line*), the mutant M_1 (*dotted line*), the mutant M_2 (*thin solid line*), and the mutant M_3 (*dashed line*). Each plot is obtained from the average on >5000 unfolding simulations.

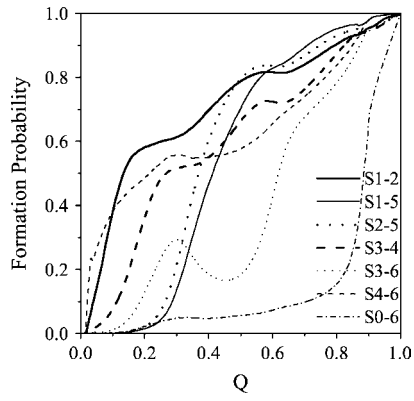

FIGURE 6 The formation probability of different parts of the wild-type tendamistat as a function of the reaction coordinate Q. Different curves correspond to the formation trajectories of different parts: the first β-hairpin *S*1–2 (*thick solid line*), β-sheet *S*1–5 (*thin solid line*), β-sheet *S*2–5 (*thick dotted line*), the second β-hairpin *S*3–4 (*thick dashed line*), β-sheet *S*3–6 (*thin dotted line*), β-sheet *S*4–6 (*thin dashed line*), and the region *S*0–6 between the N-terminal loop0 and the strand-6 (*thin dash-dotted line*).

the experimental observation that the formation of β-hairpins occurs in the earlier folding process and completes in the transition state (16,69,70). Since the $C11$–$C27$ disulfide bond tightly encloses the base of S_{1-2}, the formation of S_{1-2} is speeded up and is faster than that of S_{3-4}. It is noted that such early formation of β-hairpins also agrees with a general picture of the folding processes for the β-sheet proteins (31).

In Fig. 6, it is found that most of the contacts near the outer β-sheet S_{4-6}, which is connected by the $C45$–$C73$ disulfide bond, are formed very quickly in the early folding process. It is also found that the β-sheet S_{3-6}, an important component of S_{II}, undergoes a formation/rupture process before reaching its completely folded state. Due to the tight connection of S_{4-6} and the earlier formed β-hairpin S_{3-4}, ~30% native contacts of S_{3-6} are formed in the unfolded state, which makes S_{II} more compact. However, these formed contacts in S_{3-6} prevent the strand-5 from getting across S_{3-6} to interact with S_{1-2}. The two β-sheets S_{1-5} and S_{2-5} can be formed quickly only when the formed contacts in S_{3-6} break up gradually. (The unbroken contacts of S_{3-6} in the transition state are found to be very important for the whole folding process (will be discussed later)). This process corresponds to the rate-limiting step, namely the transition state barrier for the whole folding process.

After the time-limiting step, the rest folding process is accomplished quickly without any large barrier. The contacts in S_{0-6} between the N-terminal arm loop-0 and the strand-6, begin to form in the final stage of folding ($Q > 0.8$) when all of the other parts are almost completely formed. However, these last formed contacts play an essential role to the stability of the tendamistat because the N-terminal loop-0 situates like an arm on the outside of S_{II} and three hydrogen bonds interact strongly between each other as discussed above. The formation of this part also symbolizes the accomplishment of the total folding process.

The detailed formation probabilities of each native contact at four typical moments in the folding process are shown in Fig. 7, *a–d*. Here four moments are: 1), the earlier folding stage at $Q = 0.15$; 2), near the beginning of the transition state at $Q = 0.3$; 3), right after the transition state at $Q = 0.5$; and 4), the late folding stage at $Q = 0.8$, respectively. The folding pathway of the wild-type tendamistat can be qualitatively observed from this evolution of contact maps. In Fig. 7 *a*, most of the native contacts of S_{1-2} and S_{4-6} are formed at $Q = 0.15$, while the other regions including the β-hairpin S_{3-4} remain completely unstructured. The fast formation of S_{1-2} is mainly due to the tight connection by the $C11$–$C27$ disulfide bond. Fig. 7 *b* shows that most of the native contacts of S_{3-4} and some native contacts of S_{3-6} are formed before the transition state, while other regions remain unstructured. Fig. 7 *c* illustrates that at the moment of the end of the transition state, the β-sheets S_{1-5} and S_{2-5} are completely formed, while most of the formed contacts of S_{3-6} in Fig. 7 *b* are broken up. The wild-type tendamistat almost reaches its folded state except S_{0-6} as clearly indicated in Fig. 7 *d*.

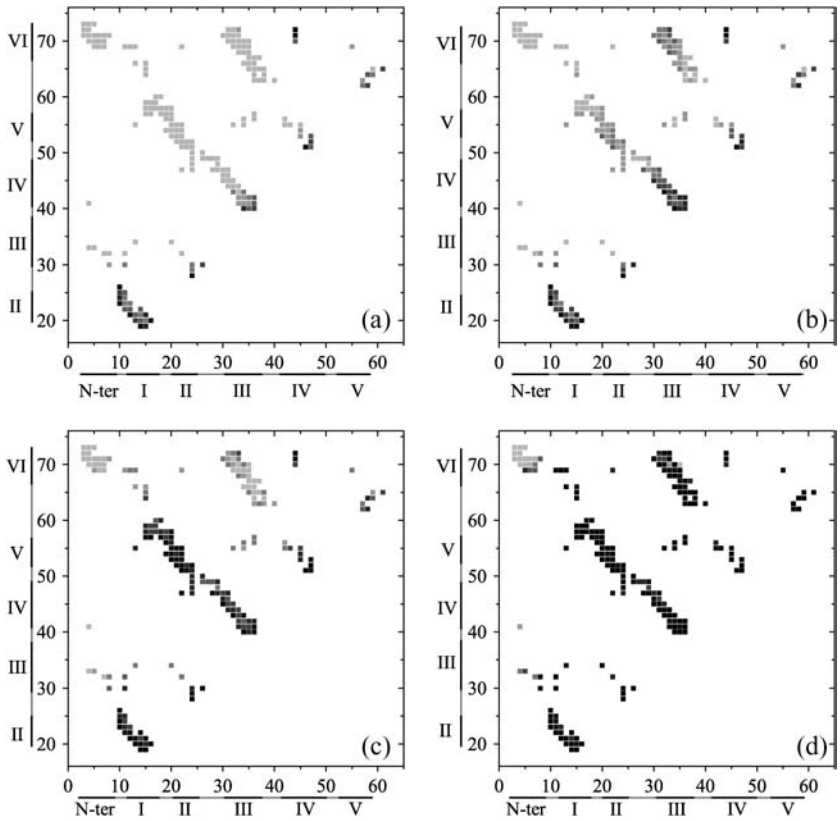

FIGURE 7 The formation probability of each native contact of the wild-type tendamistat at the four typical moments in the folding process that the moment of $Q = 0.15$ (*a*), the moment of $Q = 0.3$ (*b*), the moment of $Q = 0.5$ (*c*), and the moment of $Q = 0.8$ (*d*). Different colors indicate different values from 0 to 1, as quantified by the color scale, i.e., the darker of the color corresponds to the bigger of the value of the formation probability.

The formation probabilities of different parts versus the structural similarity Q for the wild-type tendamistat and its two single-disulfide mutants M_1 and M_2 are shown in Fig. 8, *a–g*. The removal of the *C*11–*C*27 disulfide bond leads to a large loss of the structure of S_{1-2} in the unfolded state, and only ~35% native contacts are formed in the early stage of folding at $Q = 0.15$ (see Fig. 8 *a*). Most of the contacts lost in the unfolded state due to the removal of the *C*11–*C*27 disulfide bond are found to be formed in the transition state. These contacts make S_{1-2} solvent inaccessible in the transition state. This agrees well with the results obtained from experiments (17), and also with a study on the folding behavior of a single segment of the β-hairpin S_{1-2} based on the all-atomic MD simulations (34). It is also shown in Fig. 8 *a* that the removal of the *C*45–*C*73 disulfide bond does not take any effects to the formation of this part.

For the case of the second β-hairpin S_{3-4} (Fig. 8 *b*), the removal of the *C*11–*C*27 disulfide bond enhances the formation probability thus 70% native contacts formed before the transition state. This probability is even larger than that of S_{1-2}, which is promoted by the loss of contacts in S_{1-2}. As a result, more than 40% native contacts of S_{3-6} are formed before the transition state (Fig. 8 *c*). Thus, the rupture of these formed contacts in S_{3-6}, which is favorable for strand-5 getting across this broken part to interact tightly with S_{1-2}, becomes even more difficult. As shown in Fig. 8, *d* and *e*, the removal of the *C*11–*C*27 disulfide bond has a little effect on the formation rates of the β-sheets S_{1-5} and S_{2-5} compared to that of the wild type. This makes the formation of S_{I} even more time consuming and the folding of M_1 becomes slower (17). That is, the removal of the *C*11–*C*27 disulfide bond increases the formation time of these two β-sheets, which determines the folding rate of protein tendamistat. Such an observation quantitatively accords with the experimental result that a one-eighth folding rate for M_1 with respect to that of the wild type was found (17). Therefore, the first β-hairpin S_{1-2} plays important roles not only in the thermodynamical stability but also in the folding rate.

On the contrary, there is no obvious influence of the removal of the *C*45–*C*73 disulfide bond on the early formation process of S_{3-4} for the mutant M_2. However, strand-6 no longer interacts tightly with strand-4 in the unfolded state (Fig. 8 *f*), due to the removal of this disulfide bond. The released strand-6 can even form some contacts with the N-terminal arm loop-0 in the unfolded state (Fig. 8 *g*), which does not occur for the wild-type tendamistat and the mutant M_1. As a result, there are only ~20% native contacts of S_{3-6} formed before the transition state (Fig. 8 *c*), and these formed contacts can be broken up easily in the transition state. Thus, strand-5 can get across S_{II} easier than that in the case of wild type. Therefore, the formation of S_{II} becomes more difficult.

Folding rates of small proteins are strongly dependent on a factor termed as contact order (71). In addition, a good correlation between the separations of the contacts along the sequence and the average time of the first appearances of these contacts during the folding is found (72). The average time ($\tau_{i,j}$) of the first appearance of contact between residues i and j can be used as a probe to study the folding pathways. In this work, such averaged times ($\tau_{i,j}$) are shown in the contact maps to indicate the folding pathways for the wild-type tendamistat and the two mutants M_1 and M_2, respectively (Fig. 9, *a–c*). For each contact, $\tau_{i,j}$ is obtained by averaging >1000 different simulations. Fig. 9 *a* shows that the contacts of the β-hairpins S_{1-2} and S_{3-4} in the wild-type tendamistat, distributed in the diagonal of the contact map, are formed the earliest. This is consistent with the local contacts forming faster than the nonlocal contacts, and also with the results mentioned above. It is seen that the formation rate of S_{1-2} is a little faster than that of S_{3-4} due to the existence of the *C*11–*C*27 disulfide bond. S_{3-6} seems to be formed faster than the β-sheets S_{1-5} and S_{2-5}. However, this is not the case, because the time of the first appearance of contacts does not take into consideration the rupture and reformation of these contacts. The formation of contacts between S_{1-2} and strand-5 is later than that of the hairpins. The largest sequence separations of the part S_{0-6}, located at the top left corner of the contact map, are the last formed part in our simulations. For the case of the mutant M_1 (Fig. 9 *b*), difference of ($\tau_{i,j}$) between the two β-hairpins disappears due to the removal of the *C*11–*C*27 disulfide bond. As discussed above, this mutation increases the formation probability of S_{3-6} before the transition state comparing to that of the wild type. This is obviously shown in Fig. 9 *b*. Moreover, the times of the first appearance of contacts in S_{3-6} shown in Fig. 9 *c* also supports the result of the contact formation probability in Fig. 8 *c* because the removal of the *C*45–*C*73 disulfide bond makes the formation of the *S*3–6 more difficult before the transition state.

So far, from the various features mentioned above, we can obtain a detailed folding pathway as shown in Fig. 10. The two β-hairpins S_{1-2} and S_{3-4}, situated in each of the three-stranded antiparallel β-sheets S_{I} and S_{II}, respectively, are formed in the earlier folding stage. S_{1-2} folds faster than S_{3-4}, which is attributed to the tight connection of strand-1 and strand-2 by the *C*11–*C*27 disulfide bond. The interactions between S_{1-2} and strand-5 are well established in the transition state, which occurs with most of the formed contacts in S_{3-6} breaking up at the same time. Such an establishment corresponds to the rate-limiting step for the whole folding process, and this step can be considered as the nucleation process (will be discussed in the following section). Then, after the transition state, S_{3-6} reaches its native conformation, indicating the formation of S_{II}. The last stage of the folding is the formation of the contacts between S_{I} and S_{II} with the N-terminal residues loop-0 covering tightly over the S_{II} like an arm. The folding process after the transition state occurs quickly without encountering any obvious energetic barriers.

Comparing the two single-disulfide mutants with the wild-type tendamistat, all of the results exhibit that the mutation of each disulfide bond cannot change the overall folding mechanism. This is the reason that the single mutation of each

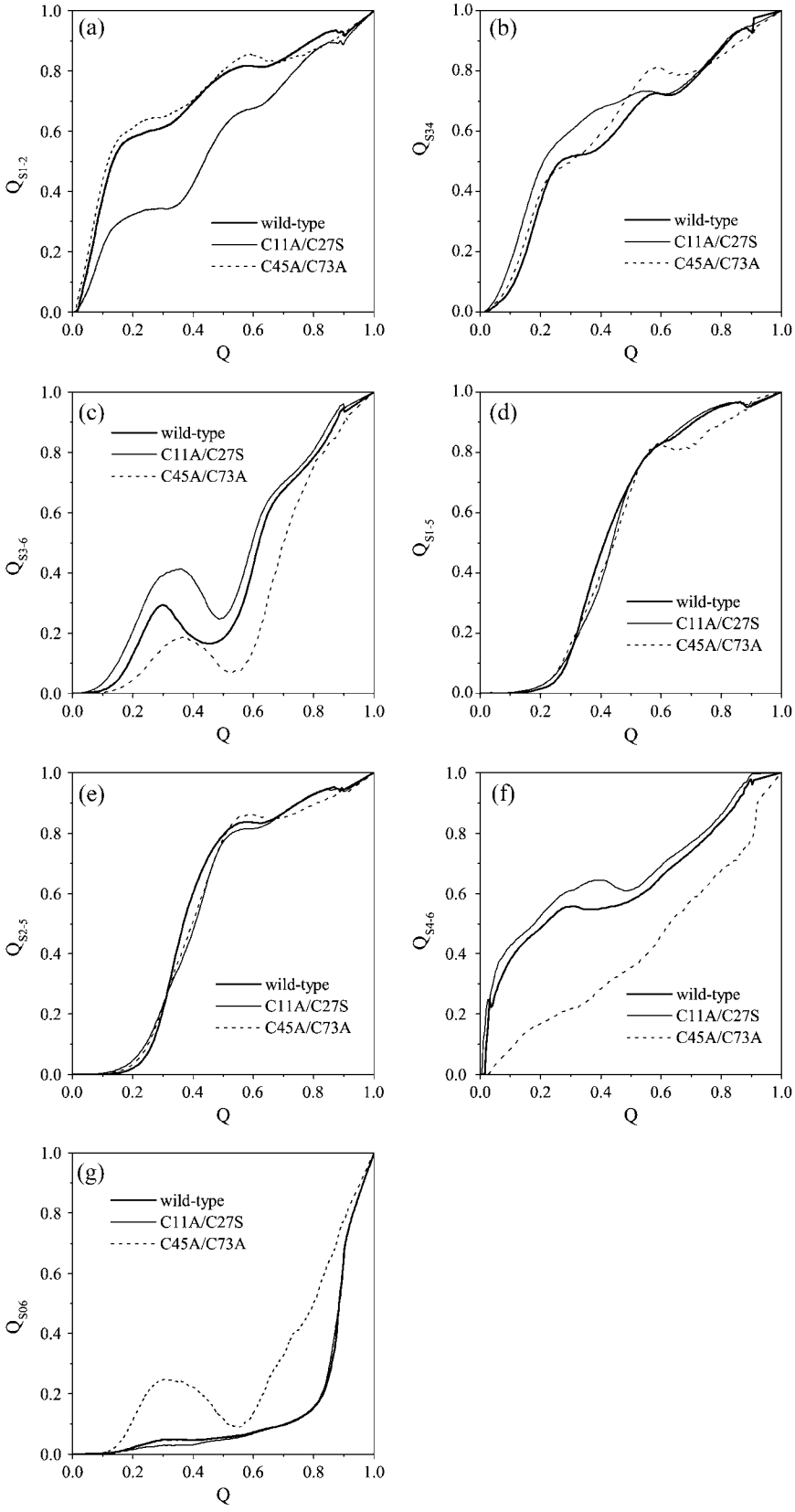

FIGURE 8 The formation probability of the different parts as a function of the structural similarity Q for the wild-type tendamistat (*thick solid line*), the mutant M_1 (*thin solid line*), and the mutant M_2 (*dashed line*). The different parts are presented, respectively, for S_{1-2} (*a*), for S_{3-4} (*b*), for S_{3-6} (*c*), for S_{1-5} (*d*), for S_{2-5} (*e*), for S_{4-6} (*f*), and for S_{0-6} (*g*). These folding processes are simulated at the temperatures that are lower than their respective folding temperatures.

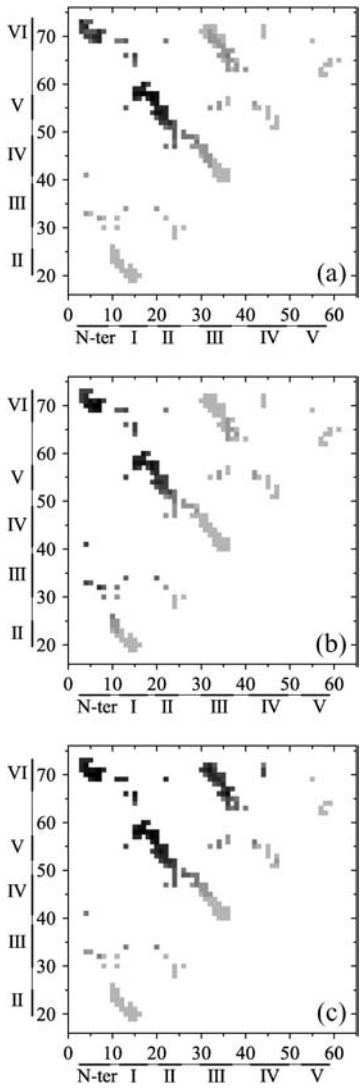

FIGURE 9 The first appearance time $\tau_{i,j}$ of the i–j native contact averaged on 2000 simulations for the wild-type tendamistat (*a*), the mutant M_1 (*b*), and the mutant M_2 (*c*), respectively. The contact color from light gray to black represents the value of the first formation time changing from minimum to maximum gradually.

disulfide bond cannot change the two-state folding behavior. However, for each mutant, there are some detailed changes in the folding pathway as discussed above and the removal of each disulfide bond can affect the main barrier caused by the formed contacts in S_{3-6} in the unfolded state. That is, the removal of the $C11$–$C27$ disulfide bond increases the barrier, while the removal of the $C45$–$C73$ disulfide bond decreases the barrier. Thus, the related rate-limiting steps for M_1 and M_2 are changed to be dependent mainly on the formation of S_{I} and S_{II}, respectively.

THE FOLDING NUCLEUS

The folding behavior of a number of small two-state proteins follows a first-order transition between the unfolded state and

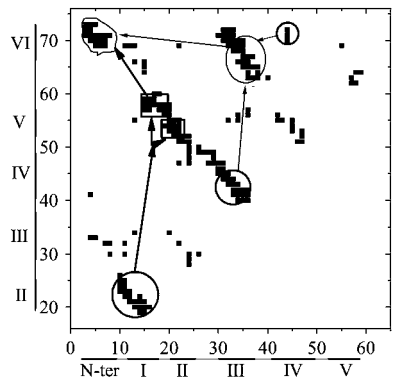

FIGURE 10 The simple description of the folding pathway for the wild-type tendamistat. The contacts inside the circle regions with thick lines are formed before the transition state; the ones inside the rectangular regions with thick line are formed in the TS; the ones inside the ellipse region with thin lines are formed in the later folding process except several key contacts are found to be formed in the TS; and the ones inside the irregular region with thin lines are formed in the last step of the whole folding process.

folded state (73–76). Mostly, the dynamic process follows a nucleation/growth mechanism, with a slow rate-limiting step and then a rapid downhill folding to the native state (77,78–80). Previous studies showed that folding nucleus play an important role in the folding process. Proteins with similar structures but different sequences could have similar folding nucleus because the nucleus is found to be topologically dependent (37,74). Because protein tendamistat and SH3 domain are rich in β-strands and have similar secondary structures, Capaldi and co-workers predicted that protein tendamistat should also follow a nucleation/growth mechanism and the β-hairpins completely formed in the transition state may participate in the formation of the nucleus (31). Our results discussed in the above section actually imply that the folding pathways of the wild-type tendamistat and the two mutants M_1 and M_2 follow the nucleation/growth mechanism. Now, let us present a detailed argument.

It is well known that for the two-state proteins, the transition state is a rate-limiting region between the denatured and native states. During the folding, the extended protein chain in the denatured state collapses directly into the native state when several key interactions between residues are formed in the transition state. These residues, which are best formed in the transition state, are termed as folding nucleus (77,78). Nucleation and overall structural formation are thus coupled, and the folding process is highly cooperative (78). Generally, in experiment, the nucleus can be identified by the method of ϕ-value analysis (78,79). Theoretically, the local fluctuation method can also be employed as an useful tool to characterize the transition state and find the folding nucleus (36–38). Here, we use this method to find the folding nucleus for the wild-type tendamistat and the two mutants M_1 and M_2.

In details, the fluctuations around the folded state and around the unfolded state near the folding temperature T_{f} are

analyzed to obtain the folding nucleus (36–38). The fluctuating events of folding $\rightleftharpoons$ unfolding are divided into four types, namely FF, UU, UF, and FU types, respectively. Here, the four events are defined as: a), the FF event that originates in and returns to the folded region without descending to the unfolded region; b), the UU event that originates in and returns to the unfolded region without ascending to the folded region; c), the UF event that originates in the unfolded region and ascends to the folded region directly without returning to the original region; and d), the FU event that comes through an absolute unfolding process from the folded region to the unfolded region (37,38). The contacts in the nucleus are assumed to form preferably in the turning points of the FF events than in the turning points of the UU events. The probabilities of the native contacts in the turning point of the FF events (P_{FF}) and that in the UU events (P_{UU}) are obtained from tens of 10^8 MD steps long-time simulations. The difference P_{FF-UU} between P_{FF} and the P_{UU} for each contact in the contact map is shown in Fig. 11, *a–c*. The folding nucleus correspond to those contacts that appear in the transition state and have positive values of P_{FF-UU}.

Two clusters of contacts for the β-hairpins S_{1-2} and S_{3-4} have high probabilities both in the FF conformations (0.5–0.7) and in the UU conformations (0.7–0.9) (Fig. 11 *a*). This shows that these two β-hairpins are highly structured in the transition state, which agrees well with the previous experimental results (16,17) and theoretical prediction (31). However, all the related values of P_{FF-UU} for S_{1-2} and S_{3-4} have negative values, indicating obviously that the formation of these two parts do not correspond to the nucleation step. The contacts in S_{0-6} (cluster-III) have quite low probabilities (<0.2) for both turning-point conformations although the values of P_{FF-UU} have positive values. Thus, these contacts are not formed in the transition state and are not included in the folding nucleus. The other two clusters (cluster-I and II) with positive values of P_{FF-UU} are also shown in Fig. 11 *a*.

Cluster-I corresponds to the region S_{1-5}, and cluster-II to the several key contacts in S_{3-6}. Clearly, there are two contacts in S_{1-5} and three contacts in S_{3-6}. These contacts are close to each other in space and have higher probabilities in both the FF turning-point conformations (0.65–0.9) and the UU turning-point conformations (0.5–0.6). Thus, they are assumed to be the folding nucleus of the wild-type tendamistat. This is consistent with the discussion in the above section that the formation of the related parts acts as the rate-limiting step in the folding process. After this step, other parts are formed quickly without encountering any energetic barriers. The nucleation mechanism of protein tendamistat can be used to interpret the special two-state folding/unfolding behavior. Whether Tendamistat is located at the folded state or unfolded state mainly depends on the formation or rupture of its folding nucleus. As a result, the formation of folding nucleus and the whole folding process are highly cooperative.

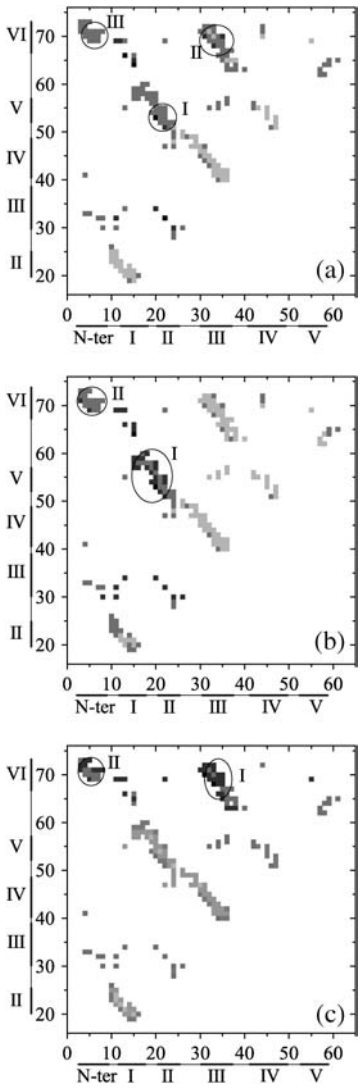

FIGURE 11 The probability difference, P_{FF-UU}, between the P_{FF} and P_{UU}, for the native contacts are shown in the contact map. Where P_{FF} and P_{UU} represent the formation probability of native contact for the structures at the turning point of the *FF* and *UU* events, respectively (see the text for the detailed definition of these two events). (*a*) For the case of the wild-type tendamistat, most of the contacts within the circles I, II, and III have positive values, while others have negative values; (*b*) for the case of the mutant M_1 (*C11A/C27S* mutant), only most of the contacts within the circles I and II have positive values; and (*c*) for the case of the mutant M_2, only the contacts within the circles I and II have positive values.

For the case of the mutant M_1 shown in Fig. 11 *b*, only the region (cluster I) composed of S_{1-2} and strand-5 has both positive values of P_{FF-UU} and high probabilities in the turning points of the FF ensemble and UU ensemble. This indicates that the residues of the nucleus are composed mainly of the residues in S_I due to the removal of the C11–C27 disulfide bond. This just validates the analysis as discussed above. The removal of the *C11–C27* disulfide bond leads to a large loss of contacts in the unfolded state for S_{1-2} and the formation is even slower than that of S_{3-4}. Thus,

more contacts in S_{3-6} are formed before the transition state and the adjustment of the positions of strands in S_{3-6}, which can favor the formation of contacts between strand-5 and S_{1-2}, becomes more difficult. Therefore, the formation of these important contacts in S_I (cluster-I) becomes even more important and rate-limited for the total folding process. However, the base of the first β-hairpin S_{1-2} is opened by the removal of the $C11$–$C27$ disulfide bond, thus S_{1-2} becomes less stable than that in the wild type. The loss of the stability of S_{1-2} is tightly relative to the loss of the stability of the folding nucleus that determines the total stability of mutant M_1. This can clearly explain why the removal of the $C11$–$C27$ disulfide bond can lead to so large a loss of the thermodynamic stability.

In Fig. 11 *c*, only cluster-I (S_{3-6}) denotes the folding nucleus region for the mutant M_2. This also interprets the result that the rate-limiting step for the mutant M_2 changes to the formation of S_{3-6} in S_{II}. As discussed above, the removal of the $C45$–$C73$ disulfide bond takes away the constraint between strand-6 and the β-hairpin S_{3-4}, which makes the contacts in S_{3-6} decrease to a half-value before the transition state compared to that of the wild-type tendamistat. Thus, the three-stranded antiparallel β-sheet S_I can reach its native conformation easier for the mutant M_1 than for the wild type. However, the formation of the part S_{3-6} in S_{II} becomes time-consuming for the whole folding. Due to the tight connections by the three hydrogen bonds, N-terminal loop-0 covers S_{II} like an arm, which can provide the required stability for the folding nucleus. Therefore, the mutant M_2 is more stable than the mutant M_1 although the related removal of the disulfide bond leads to a large loss in the thermodynamic stability.

CONCLUSION

This work is aimed to investigate the detailed folding mechanism of the disulfide-bonded protein tendamistat and the effects of the removal of each disulfide bond on the thermodynamic stability and the folding kinetics. The two-state folding behaviors for the wild-type tendamistat and its three mutants M_1, M_2, and M_3 are obviously exhibited in our simulations and agree well with the observations in experiments (16,17,19,21). Based on the modified Gō-type model, the folding pathways are also characterized in detail. Several processes, i.e., the formation of the first β-hairpin S_{1-2}, the formation of the rate-limiting step, and so on, are described. Here, the rate-limiting step corresponds to the formation of the nucleus, which is found by the local fluctuation method. The nucleation mechanism can well explain the two-state folding/unfolding behaviors for the wild-type tendamistat and the two mutants M_1 and M_2. Moreover, each removal of the disulfide bond leads to a large loss in the thermodynamic stability and the effect of the removal of the $C11$–$C27$ disulfide bond is larger than that of the $C45$–$C73$ one. This can also be clearly described by the influence on the folding nucleus. The disulfide bonds play an important role not only in the stability of the native structure but also in the kinetic process or the folding pathway.

The agreement with the experiments (16,17,19,21) indicates that our modified Gō-type model, especially, the introduction of the interactions of the disulfide bonds and the hydrogen bonds, really works. Since few theoretical studies on the folding behavior of disulfide-bonded proteins, e.g., protein tendamistat, are implemented, our work shows a successful example of study based on the coarse-grained models. The results obtained using this simplified model can provide useful insight into the folding mechanism of the disulfide-bonded proteins. In addition, this work on protein tendamistat can also help to understand the folding of β-sheet proteins, which appears to be different from that of helical and mixed α- and β-proteins (31).

For the limitation of the simplified model, we can only consider the fast folding pathway that follows a two-state process without populated intermediate states. However, for the case of the slow folding process, the rate-limiting step occurs mainly due to the *cis* to *trans* isomerization that cannot be investigated in our model. It is also noted that the formation process of the disulfide bonds is vital in the folding process, especially to the activities and functions of disulfide bonds. Although many results for the thiol-disulfide exchange process have been obtained in recent experiments, theoretically, a clear picture for the complicated formation process is still needed and more precise modeling study is required. Specifically, study based on the all-atomic model will be preferable.

This work is supported by the National Natural Science Foundation of China (Nos. 90403120, 10021001, and 10474041) and the Nonlinear Project (973) of the National Science Ministry.

REFERENCES

1. Wolynes, P. G., J. N. Onuchic, and D. Thirumalai. 1995. Navigating the folding routes. *Science.* 267:1619–1620.
2. Onuchic, J. N., and P. G. Wolynes. 2004. Theory of protein folding. *Curr. Opin. Struct. Biol.* 14:70–75.
3. Anfinsen, C. B. 1973. Principles that govern the folding of protein chains. *Science.* 181:223–230.
4. Fersht, A. R. 1999. Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. Freeman, New York, NY.
5. Daggett, V., and A. R. Fersht. 2003. The present view of the mechanism of protein folding. *Nat. Rev. Mol. Cell Biol.* 4:497–502.
6. Sevier, C. S., and C. A. Kaiser. 2002. Formation and transfer of disulfide bonds in living cells. *Nat. Rev. Mol. Cell Biol.* 3:836–847.
7. Frand, A. R., J. W. Cuozzo, and C. A. Kaiser. 2000. Pathways for protein disulphide bond formation. *Trends Cell Biol.* 10:203–210.
8. Rietsch, A., and J. Beckwith. 1998. The genetics of disulfide bond metabolism. *Annu. Rev. Genet.* 32:163–184.
9. Welker, E., W. J. Wedemeyer, and H. A. Scheraga. 2001. A role for intermolecular disulfide bonds in prion diseases. *Proc. Natl. Acad. Sci. USA.* 98:4334–4336.
10. May, B. C., C. Govaerts, S. B. Prusiner, and F. E. Cohen. 2004. Prions: so many fibers, so little infectivity. *Trends Biochem. Sci.* 29: 162–165.

11. Masip, L., J. L. Pan, S. Haldar, J. E. Penner-Hahn, M. P. Delisa, G. Georgiou, J. C. Bardwell, and J. F. Collet. 2004. An engineered pathway for the formation of protein disulfide bonds. *Science*. 303: 1185–1189.
12. Yokota, A., K. Izutani, M. Takai, Y. Kubo, Y. Noda, Y. Koumoto, H. Tachibana, and S. Segawa. 2000. The transition state in the folding-unfolding reaction of four species of three-disulfide mutant of hen lysozyme: the role of each disulfide bridge. *J. Mol. Biol.* 295:1275–1288.
13. Goldenberg, D. P. 1992. Native and non-native intermediates in the BPTI folding pathway. *Trends Biochem. Sci.* 17:257–261.
14. Robinson, A. S., and J. King. 1997. Disulphide-bonded intermediate on the folding and assembly pathway of a non-disulphide bonded protein. *Nat. Struct. Biol.* 4:450–455.
15. Matsumura, M., W. J. Becktel, M. Levitt, and B. W. Matthews. 1989. Stabilization of phage T4 lysozyme by engineered disulfide bonds. *Proc. Natl. Acad. Sci. USA*. 86:6562–6566.
16. Schönbrunner, N., K. P. Koller, and T. Kiefhaber. 1997. Folding of the disulfide-bonded β-sheet protein tendamistat: rapid two-state folding without hydrophobic collapse. *J. Mol. Biol.* 268:526–538.
17. Schönbrunner, N., G. Pappenberger, M. Scharf, J. Engels, and T. Kiefhaer. 1997. Effect of preformed correct tertiary interactions on rapid two-state tendamistat folding: evidence for hairpins as initiation sites for β-sheet formation. *Biochemistry*. 36:9057–9065.
18. Abkevich, V. I., and E. I. Shakhnovich. 2000. What can disulfide bonds tell us about protein energetics, function and folding: simulations and bioinformatics analysis. *J. Mol. Biol.* 300:975–985.
19. Bachmann, A., and T. Kiefhaber. 2001. Apparent two-state tendamistat folding is sequential process along a defined route. *J. Mol. Biol.* 306: 375–386.
20. Pappenberger, G., H. Aygün, J. W. Engels, U. Reimer, G. Fischer, and T. Kiefhaber. 2001. Nonprolyl cis peptide bonds in unfolded proteins cause complex folding kinetics. *Nat. Struct. Biol.* 8:452–458.
21. Pappenberger, G., A. Bachmann, R. Müller, H. Aygün, J. W. Engels, and T. Kiefhaber. 2003. Kinetic mechanism and catalysis of a native-state prolyl isomerization reaction. *J. Mol. Biol.* 326:235–246.
22. Flory, P. J. 1956. Theory of elastic mechanisms in fibrous proteins. *J. Am. Chem. Soc.* 78:5222–5235.
23. Camacho, C. J., and D. Thirumalai. 1995. Modeling the role of disulfide bonds in protein folding: entropic barriers and pathways. *Proteins: Struct., Funct. For. Genet.* 22:27–40.
24. Chaffotte, A. F., Y. Guillou, and M. E. Goldberg. 1992. Kinetic resolution of peptide bond and side chain far-UV circular dichroism during the folding of hen egg white lysozyme. *Biochemistry*. 31:9694–9702.
25. Kline, A. D., and K. Wuthrich. 1985. Secondary structure of the α-amylase polypeptide inhibitor tendamistat from Streptomyces tendae determined in solution by 1H nuclear magnetic resonance. *J. Mol. Biol.* 183:503–507.
26. Pflugrath, J., I. Wiegand, R. Huber, and L. Vartesy. 1986. Crystal structure determination, refinement and the molecular mode of the α-amylase inhibitor Hoe-467A. *J. Mol. Biol.* 189:383–386.
27. Volgl, T., R. Brengelmann, H. J. Hinz, M. Scharf, M. Lotzbeger, and J. W. Engels. 1995. Mechanism of protein stabilization by disulfide bridges: calorimetric unfolding studies on disulfide-deficient mutants of the α-amylase inhibitor tendamistat. *J. Mol. Biol.* 254:481–496.
28. Balbach, J., S. Seip, H. Kessler, M. Scharf, N. Kashani-Poor, and J. W. Engels. 1998. Structure and dynamic properties of the single disulfide-deficient α-amylase inhibitor [C45A/C73A] tendamistat: an NMR study. *Proteins*. 33:285–294.
29. Pappenberger, G., C. Saudan, M. Becker, A. E. Merbach, and T. Kiefhaber. 2000. Denaturant-induced movement of the transition state of protein folding revealed by high-pressure stopped-flow measurements. *Proc. Natl. Acad. Sci. USA*. 97:17–22.
30. König, V., L. Vértesy, and T. Schneider. 2003. Structure of the α-amylase inhibitor tendamistat at 0.93 Å. *Acta. Crystallogr. D Biol. Crystallogr.* D59:1737–1743.
31. Capaldi, A., P., and S. E. Radford. 1998. Kinetic studies of β-sheet protein folding. *Curr. Opin. Struct. Biol.* 8:86–92.
32. Onuchic, J. N., H. Nymeyer, A. E. Garcia, and N. D. Socci. 2000. The energy landscape theory of protein folding: insights into folding mechanisms and scenarios. *Adv. Protein Chem.* 53:87–152.
33. Daggett, V. 2000. Long timescale simulations. *Curr. Opin. Struct. Biol.* 10:160–164.
34. Bonvin, A. M., and W. F. Gunsteren. 2000. β-Hairpin stability and folding: molecular dynamics studies of the first β-hairpin of tendamistat. *J. Mol. Biol.* 296:255–268.
35. Cho, S. S., Y. Levy, J. N. Onuchic, and P. G. Wolynes. 2005. Overcoming residual frustration in domain-swapping: the roles of disulfide bonds in dimerizaiton and aggregation. *Phys. Biol.* 2:S44–S55.
36. 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.
37. Dokholyan, N. V., S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich. 2000. Identifying the protein folding nucleus using molecular dynamics. *J. Mol. Biol.* 296:1183–1188.
38. 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.
39. Case, D. A., D. A. Pearlman, J. W. Caldwell, T. E. Cheatham, III, J. Wang, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, et al. AMBER 7, 2002, University of California, San Francisco, CA.
40. Gō, N. 1983. Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.* 12:183–210.
41. Shoemaker, B. A., J. Wang, and P. G. Wolynes. 1997. Structural correlations in protein folding funnels. *Proc. Natl. Acad. Sci. USA*. 94:777–782.
42. Portman, J. J., S. Takada, and P. G. Wolynes. 1998. Variational theory for site resolved protein folding free energy surfaces. *Phys. Rev. Lett.* 81:5237–5240.
43. Alm, E., and D. Baker. 1999. Prediction of protein folding mechanisms from free-energy landscapes derived from native structures. *Proc. Natl. Acad. Sci. USA*. 96:11305–11310.
44. Muñoz, V., and W. A. Eaton. 1999. A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc. Natl. Acad. Sci. USA*. 96:11311–11316.
45. Galzitskaya, O. V., and A. V. Finkelstein. 1999. A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. *Proc. Natl. Acad. Sci. USA*. 96:11299–11304.
46. Takada, S. 1999. G(o)over-bar-ing for the prediction of protein folding mechanisms. *Proc. Natl. Acad. Sci. USA*. 96:11698–11700.
47. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: what determines the structural details of the transition state ensemble and on-route intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* 298:937–953.
48. Koga, N., and S. Takada. 2001. Poles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model. *J. Mol. Biol.* 313:171–180.
49. Karanicolas, J., and C. L. Brooks, III. 2002. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* 11:2351–2361.
50. Zhang, J., M. Qin, and W. Wang. 2005. Multiple folding mechanisms of protein Ubiquitin. *Proteins*. 59:565–579.
51. Levitt, M. 1976. A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* 104:59–107.
52. Honeycutt, J. D., and D. Thirumalai. 1992. The nature of folded states of globular proteins. *Biopolymers*. 32:695–709.

53. Guo, Z., and D. Thirumalai. 1995. Kinetics of protein folding: nucleation mechanism, time scales, and pathways. *Biopolymers*. 36:83–103.

54. Veitshans, T., D. Klimov, and D. Thirumalai. 1996. Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence dependent properties. *Fold. Des*. 2:1–22.

55. Swope, W. C., H. C. Andersen, P. H. Berens, and K. R. Wilson. 1982. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: application to small water clusters. *J. Chem. Phys*. 76:637–649.

56. Leach, A. R. 1996. Molecular Modelling Principles and Applications. Addison-Wesley, London, UK.

57. Socci, N. D., and J. N. Onuchic. 1995. Kinetic and thermodynamic analysis of protein-like heteropolymers: Monte Carlo histogram technique. *J. Chem. Phys*. 103:4732–4744.

58. Klimov, D. K., and D. Thirumalai. 1996. Criterion that determine the foldability of proteins. *Phys. Rev. Lett*. 76:4070–4073.

59. Klimov, D. K., and D. Thirumalai. 1998. Linking rates of folding in lattice models of proteins with underlying thermodynamic characteristics. *J. Chem. Phys*. 109:4119–4125.

60. Chahine, J., H. Nymeyer, V. Leite, N. Socci, and J. Onuchic. 2002. Specific and nonspecific collapse in protein folding funnels. *Phys. Rev. Lett*. 88:168101.

61. Bakk, A., J. Høye, and A. Hansen. 2001. Heat capacity of protein folding. *Biophys. J*. 81:710–714.

62. Ferrenberg, A. M., and R. H. Swendsen. 1989. Optimized Monte Carlo data analysis. *Phys. Rev. Lett*. 63:1195–1198.

63. Shea, J. E., and C. L. Brooks III. 2001. Proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem*. 52:499–535.

64. Kumar, S., J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. 1995. Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comput. Chem*. 16:1339–1350.

65. Pace, C. N., G. R. Grimsley, J. A. Thomson, and B. J. Barnett. 1988. Conformational stability and activity of ribonuclease T1 with zero, one, and two intact disulfide bonds. *J. Biol. Chem*. 263:11820–11825.

66. Dill, K. A., and H. S. Chan. 1997. From Levinthal to pathways to funnels. *Nat. Struct. Biol*. 4:10–19.

67. Wolynes, P. G. 1997. Folding funnels and energy landscapes of larger proteins within the capillarity approximation. *Proc. Natl. Acad. Sci. USA*. 94:6170–6175.

68. Clementi, C., A. E. Garcia, 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.

69. Finkelstein, A. V. 1991. Rate of beta-structure formation in polypeptides. *Proteins*. 9:23–27.

70. Blanco, F. J., M. A. Jimenez, M. Rtico, J. Santoro, J. Herranz, and J. L. Nieto. 1991. Tendamistat(12–26) fragment: NMR characterization of isolated-β-turn intermediates. *Eur. J. Biochem*. 200:345–351.

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

72. Hoang, T. X., and M. Cieplak. 2001. Sequencing of folding events in Go-type proteins. *J. Chem. Phys*. 113:8319–8328.

73. Bryngelson, J. D., and P. G. Wolynes. 1987. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA*. 84:7524–7528.

74. Martinez, J. C., M. T. Pisabarro, and L. Serrano. 1998. Obligatory steps in protein folding and the conformational diversity of the transition state. *Nat. Struct. Biol*. 5:721–729.

75. Karplus, M., and E. I. Shakhnovich. 1992. Theoretical studies of protein folding thermodynamics and dynamics. *In* Protein Folding, Chapter 4. T. Creighton, editor. W. H. Freeman and Company, New York, NY. 127–196.

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

77. Itzhaki, L. S., D. E. Otzen, and A. R. Fersht. 1995. The structure of the transition state for folding of chymotrypsin inhibitor 2 analyzed by protein engineering methods: evidence for a nucleation condensation mechanism for protein folding. *J. Mol. Biol*. 254:260–288.

78. Fersht, A. 1997. Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol*. 7:3–9.

79. Otzen, D. E., L. S. Itzhaki, N. F. Elmasry, S. E. Jackson, and A. R. Fersht. 1994. Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding. *Proc. Natl. Acad. Sci. USA*. 91:10422–10425.

80. Pande, V. S., A. Y. Grosberg, T. Tanaka, and D. Rokhsar. 1998. Pathways for protein folding: is a new view needed? *Curr. Opin. Struct. Biol*. 8:68–79.