PROTEINS: Structure, Function, and Bioinformatics 67:338–349 (2007)

Understanding the Folding and Stability of a Zinc Finger-Based Full Sequence Design Protein With Replica Exchange Molecular Dynamics Simulations

Wenfei Li,[1] Jian Zhang,[1] and Wei Wang[1,2]*

[1]*National laboratory of Solid State Microstructure and Department of Physics, Nanjing University, Nanjing 210093, China*
[2]*Interdisciplinary Center of Theoretical Studies, Chinese Academy of Sciences, Beijing 100080, China*

ABSTRACT **Full sequence design protein FSD-1 is a designed protein based on the motif of zinc finger protein. In this work, its folding mechanism and thermal stability are investigated using the replica exchange molecular dynamics model with the water molecules being treated explicitly. The results show that the folding of the FSD-1 is initiated by the hydrophobic collapse, which is accompanied with the formation of the C-terminal α-helix. Then the folding proceeds with the formation of the β-hairpin and the further package of the hydrophobic core. Compared with the β-hairpin, the α-helix has much higher stability. It is also found that the N-capping motif adopted by the FSD-1 contributes to the stability of the α-helix dramatically. The hydrophobic contacts made by the side chain of Tyr3 in the native state are essential for the stabilization of the β-hairpin. It is also found that the folding of the N-terminal β-hairpin and the C-terminal α-helix exhibits weak cooperativity, which is consistent with the experimental data. Meanwhile, the folding pathway is compared between the FSD-1 and the target zinc finger peptide, and the possible role of the zinc ion on the folding pathway of zinc finger is proposed. Proteins 2007;67:338–349.**

Key words: zinc finger; FSD-1; replica exchange molecular dynamics; folding; stability

INTRODUCTION

Designing a novel protein with desired three-dimensional structure and well-defined functionalities represents the most intriguing problem for structural biologists due to its potential applications in novel machines and therapeutics.[1–7] The classical Cys_2His_2 zinc finger, due to its modular characteristic and diverse functionalities, is frequently used as the target for the structure-based protein design. As one of the most successful examples, full sequence design protein FSD-1 is designed based on the structure of the second zinc finger of Zif268 by Dahiyat and Mayo[4] [(see Fig. 5(f) for its structure)]. The zinc fingers are among the smallest natural proteins, and responsible for binding nucleic acid and regulating gene expression, as well as mediating protein–protein interactions. It consists of a N-terminal β-hairpin and a C-terminal α-helix (the so called ββα fold).[8,9] However, the proper folding of this protein depends on the binding of metal ions.[10] In comparison, despite the small size (containing 28 amino acids), the designed protein FSD-1 is capable of folding into the same structure without the help of disulfide and metal ions. Therefore, it is interesting to study the underlying factors, which contribute to the capabilities of folding and stabilization of the FSD-1. Such knowledge will be helpful in developing better strategies for *de novo* protein design, since the protein folding problem is still a hindrance for the protein design.[11,12] Meanwhile, the FSD-1 can also be considered as a model to study the folding of the ββα fold, which widely exists in natural proteins. Investigating the folding of this protein can provide a basis in understanding the folding of more complicated globular proteins.

Studying the folding mechanism of the FSD-1 is also helpful for understanding the metal ion-coupled folding of the zinc fingers. As the zinc finger protein is not capable of folding without the help of metal ions, one can not figure out the role of the metal ions on the folding process directly by comparing the folding of the apo-form (without zinc ions) peptide and holo-form (with zinc ions) peptide. The FSD-1 can fold into the same structure as the zinc fingers without the help of metal ions. Therefore, by comparing the folding processes of the FSD-1 and the zinc finger peptide, one may extract some information for the role of the metal ions on the folding of the zinc finger, which is a fundamental but not well studied problem.

In Ref. 13, Jang et al. studied the folding of the FSD-1 employing all-atom force field with the solvent being implicitly treated by the generalized Born model. They found that the folding of this protein is always initiated

Grant sponsor: National Natural Science Foundation of China; Grant numbers: 90403120, 10504012, 10021001, 10474041; Grant sponsors: Nonlinear Project (973) of NSM, China Postdoctoral Science Foundation, Jiangsu Planned Projects for Postdoctoral Research Funds.

*Correspondence to: Wei Wang, National Laboratory of Solid State Microstructure, Department of Physics and Institute of Biophysics, Nanjing University, Nanjing 210093, China.
E-mail: wangwei@nju.edu.cn

Received 17 January 2006; Revised 20 September 2006; Accepted 24 October 2006

Published online 6 February 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21312

by the folding of the C-terminal α-helix and then followed by the formation of the small loop and the N-terminal β-hairpin. In comparison, by using an optimized atomistic model, Kim et al. concluded that the folding of this protein is initiated by the collapse of the hydrophobic core, which is followed by the folding of the β-hairpin. The formation of the α-helix is the last event of folding.[14] This scenario is just opposite to that of Jang et al.[13] In Ref. 15, Lei and Duan also conducted the unfolding simulations for this protein using the all-atom molecular dynamics (MD) model with AMBER force field and explicit water molecules. Based on the assumption that folding is the reverse process of unfolding, they proposed that the folding of the individual secondary structures is prior to the native hydrophobic core formation. Therefore, contradictions still exist for the folding mechanism of the FSD-1 in literatures, and further studies employing more rigorous methods are needed.

In the present work, by using the replica exchange molecular dynamics (REMD) model, the folding mechanism of the FSD-1 is investigated. Compared with the previous work for this protein, the present work allows more efficient sampling and more realistic force field due to the use of the REMD method and the explicit treatment of solvent. We try to address the following questions: (i) What are the differences of the folding pathways between the FSD-1 and the target zinc finger peptide? (ii) What factors contribute to the capabilities of folding and stabilization of this protein? It is found based on the simulation data that the folding of the FSD-1 is initiated by the hydrophobic collapse and the formation of the α-helix, then the folding proceeds with the formation of the β-hairpin. In comparison, for the zinc finger peptide, the folding of the β-hairpin is prior to that of the α-helix. This difference indicates that the metal ions may affect the folding kinetics of the zinc fingers. By investigating the stabilities of the FSD-1, it is found that the N-capping motif adapted by the C-terminal α-helix and the contacts made by the Tyr3 in the native structure are important factors for the stability of the FSD-1. Meanwhile, weak folding cooperativity is found for the FSD-1, which is in consistence with the experimental data. These results lead to a deeper understanding on the capability of folding and stabilization of the FSD-1, and will be useful in developing better strategies for de novo protein design. Moreover, the folding mechanism studied for the FSD-1 can shed light on the general principles in the folding of the ββα fold. By comparing with the folding mechanism of the zinc finger peptide, the present work also provides some information for the role of the metal ions on the folding of the zinc fingers and other metal proteins.

MODEL AND METHODS

Replica Exchange Molecular Dynamics Model

Replica exchange method is a high efficiency sampling technique, which was first implemented in MD model by Sugita and Okamoto,[16] and was widely used in the studying of protein folding and aggregation.[17–23] Duan and coworkers proposed that the REMD can enhance the sampling efficiency by around 14.3, 35.1, and 71.5 times for temperatures of 360, 300, and 275 K, respectively, compared to the conventional MD based on the simulations for a 21-amino acid peptide with implicit solvent.[24] They observed that the simulations from two quite different initial structures start to converge within 1 ns. Sanbonmatsu and García also investigated the sampling efficiency of the REMD by performing the simulations for the peptide Met-enkephalin with explicit solvent, and found that the REMD samples five times more conformational space than the conventional MD.[25] These results demonstrate the high sampling efficiency of the REMD. In this work, we use the REMD to perform the conformational sampling.

In REMD simulations, N noninteracting replicas at N different temperatures are conducted simultaneously in parallel. After certain MD step, the exchange of replica between neighbor temperatures is attempted. The acceptance ratio of exchange is determined by the energy difference and temperature difference between the replica pair with Metropolis criterion:

$$\mathrm{P}(\mathrm{X}_i \rightarrow \mathrm{X}_j) = \min(1, \exp(-\Delta)) \quad (1)$$

where $\Delta = (\beta_i - \beta_j)(E_j - E_i)$ with β and E being the invert temperature and potential energy, respectively. With this method, the conformations at low temperatures have ability to overcome high energy barriers by being switched to high temperatures, and it provides improved sampling at lower temperatures than standard MD. Further details of the method can be found in Refs. 16 and 26.

Simulation Details

In this work, the simulations are performed using the AMBER8 MD simulation package with water molecules being treated explicitly.[27] The standard AMBER ff03 force field is used.[28] In the simulation, the nuclear magnetic resonance (NMR) structure of the FSD-1 (PDB code: 1fsd, QQYTAKIKGRTFRNEKELRDFIEKFKGR) is solvated in a TIP3P water box. The shortest distance between the amino acid atoms and the edges of the water box is 10 Å. To alleviate the boundary effects, the periodic boundary conditions are applied. In treating the long-range electrostatic interactions, the Particle Mesh Ewald (PME) summation algorithm is employed. A cutoff of 8.0 Å is used for the construction of nonbonded list. The covalent bonds involving hydrogen atoms are constrained with the SHAKE algorithm and the time step of 0.002 ps is used. We first minimize the NMR structure of the FSD-1 to reduce the improper contacts. The minimized structure is heated to 298 K in NVT ensemble, and then equilibrated in NPT ensemble at 1 atm for 20 ps to adjust the water box. From the relaxed structures, 60 standard MD are run from $T = 289$ K to

$T = 698$ K for 1 ns. The resulted structures are used as the initial structures for the REMD simulations at corresponding temperatures in NVT ensembles. For convenience, we refer to this set of initial structures as "STRUCTURE1." The temperatures are selected with an exponential law, which makes the acceptance ratio a constant and it lies between 0.15 and 0.20. The time intervals between the exchange attempts are 0.8 ps and the atomic coordinates are recorded every 0.2 ps for further analysis. Totally 18 ns are simulated for each replica, which results in 1.08 μs integration time. The structures sampled during the last 17 ns for each replica are used for analysis. In constructing the free energy landscape at certain temperature, the WHAM method is used.[29]

To test the reasonable convergence of the simulation, another REMD simulation is performed with different initial structures and temperatures. The initial structures are prepared following the method of Ref. 22. We first perform a series of standard MD simulations for 1.5 ns at $T = 275$–1000 K. The resulted structures variate between folded structure and unfolded structure. We chose 60 structures at random as the initial structures of the REMD simulation, during which the extended structures with the radius of gyration larger than 12 Å are excluded. Among the 60 initial structures, 17, 23, and 20 structures have rmsd $<$ 3.0 Å, 3.0 Å $<$ rmsd $<$ 5.0Å, and rmsd $>$ 5.0Å, respectively. We refer to this set of initial structures as "STRUCTURE2." For this REMD run, the temperatures that range from $T = 282$ K to $T = 678$ K are slightly different from the first REMD run and assigned to the replicas at random. It should be noted that among the initial structures, some of them have nativelike structures. Thus, the present simulation is not presenting a folding simulation. In fact, to fully fold a protein with 28 residues starting from fully unfolded structure in explicit water is really difficult even using REMD method. What we do in this work is to sample the whole conformational space as sufficient as possible. To achieve this, appropriate selection of the initial structures can improve the sampling efficiency and accelerate the convergence. In the above chosen initial structures, each of them locates at different position of the conformational space ranging from unfolded structure to native structure. With this set of initial structures, each replica can sample a relatively wide range of conformational space by switching to different temperatures during REMD simulation. In this way, the 60 replicas can cover a very wide range of conformational space, and the simulation is expected to be able to achieve better sampling efficiency and convergence within reasonable simulation time although the folding event is not necessarily accomplished for the individual replicas.

Since the two set of initial structures are prepared by different processes, and they represent different subspace in the conformational space although some of the structures may be close to each other, comparison of the REMD simulations started from "STRUCTURE1" and "STRUCTURE2" can eliminate the situation that the trajectories are trapped in local minimum of the energy surface, and support the reasonable convergence.

Definitions of Reaction Coordinates

Protein folding can be considered as a diffusion process among the high dimensional parameter space. To capture the main characteristics of the folding process and simplify the analysis, the high dimensional parameter space needs to be projected to limited reaction coordinates. In the present work, Q, Q_β, N_α, R_g, R_g^{core} and rmsd are used as reaction coordinates. The Q is the fraction of native contacts, which measures the similarity between any given structure and the native structure. We define the contacts are occurring whenever two amino acid side chains, which separated by at least four amino acids, have any two heavy atoms within 5.0 Å. With this definition, 42 contacts are found for the NMR structure. The Q_β is the fraction of native contacts for the N-terminal region including residues 1–14. It is used to measure formation extent of the N-terminal β-hairpin. The N_α is defined as the number of helical residues formed among residues 15–28. A residue is defined as helical residue whenever three or more consecutive residues satisfy the dihedral constraints of $-95^\circ < \phi < -25^\circ$ and $-77^\circ < \psi < -17^\circ$. The R_g is the radius of gyration and calculated for all the heavy atoms. R_g^{core} is the radius of gyration for the hydrophobic core formed by the residues Tyr3, Ala5, Ile7, Phe12, Leu18, Phe21, Ile22, and Phe25. When calculating the rmsd, we only consider the C_α atoms. The unit of the R_g, R_g^{core}, and rmsd is Angstrom (Å).

RESULTS AND DISCUSSION

Convergence Verification

One of the advantages of REMD is that it has high sampling efficiency and can give reproducible results in limited simulation time compared with the standard MD. To investigate the sampling quality, the temperature trajectories for nine selected replicas are shown in Figure 1. One can see that a wide range of temperatures are visited by the replicas, which indicates that the sampling quality is high. We also present the time series of the Q, Q_β, and N_α for three representative replicas in Figure 2 to illustrate the structural transitions, which are essential for the system to achieve the thermodynamic equilibrium. The large fluctuation of the N_α in this figure mainly resulted from the stringent definition of the α-helix residue, which demands three or more consecutive residues satisfying the dihedral constraints simultaneously. Therefore, the local structural fluctuations can result in large variation of the N_a. From Figure 2, we can observe that the component α-helix, β-hairpin, or both of them can form to large extent during the simulations. For example, for the replica in the left panels (middle panels) of Figure 2, the N_α (Q_β) increases from ∼0 up to ∼10 (from ∼0.2 up to ∼0.7) during the

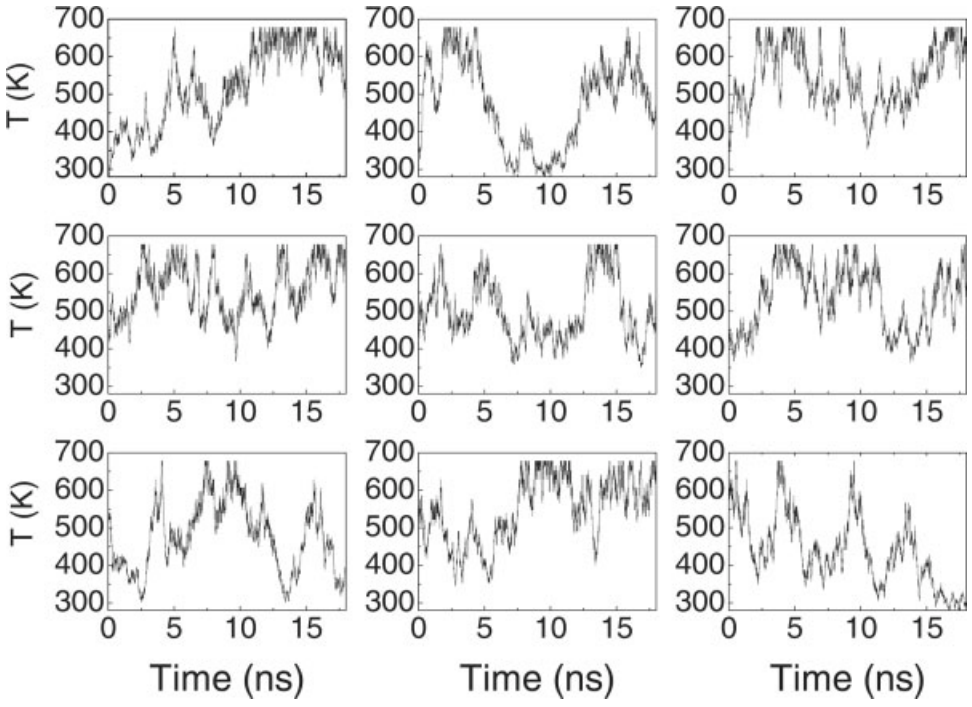

Fig. 1. Temperature trajectories of nine selected replicas.

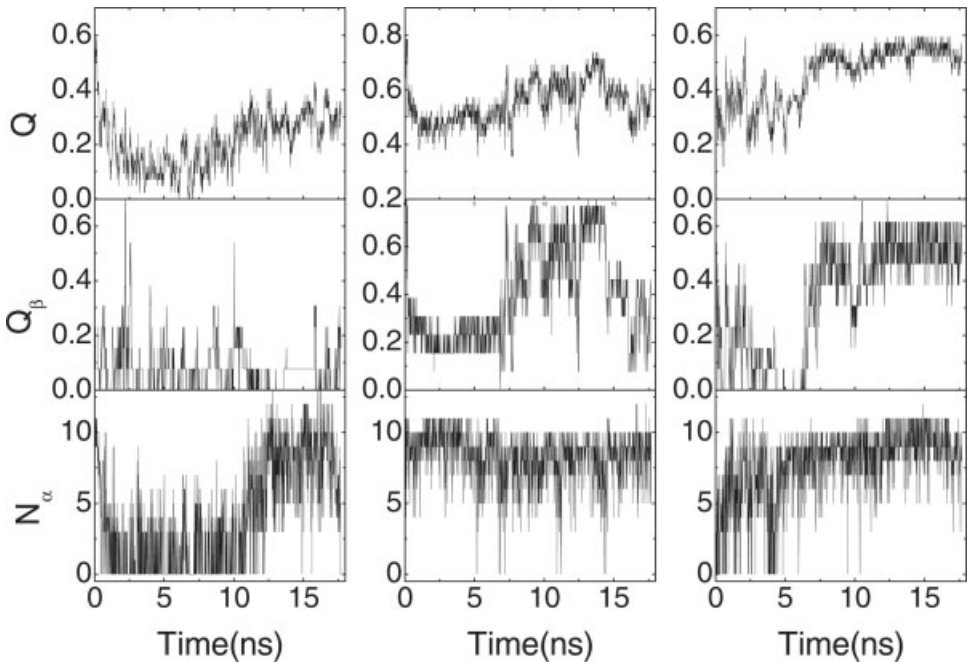

Fig. 2. Time series of the Q (upper panel), Q_β, (middle panel), and N_α (lower panel) for three representative replicas.

folding. For the replica in the right panels, the Q_β increases from ~0.0 to ~0.50, and the N_α increases from ~5 to ~10. These results demonstrate that the individual replica can sample a wide range of conformational space, although it is not necessarily a fully folding event, suggesting the high sampling quality. In Figure 2, the highest values of the Q are much less than that of the native structure. Besides the fact that the fully folding events may be not accomplished for these replicas, one possible reason is that the protein FSD-1 is not thermally stable, and the native structure can be disrupted easily. In fact, the low stability of this protein was

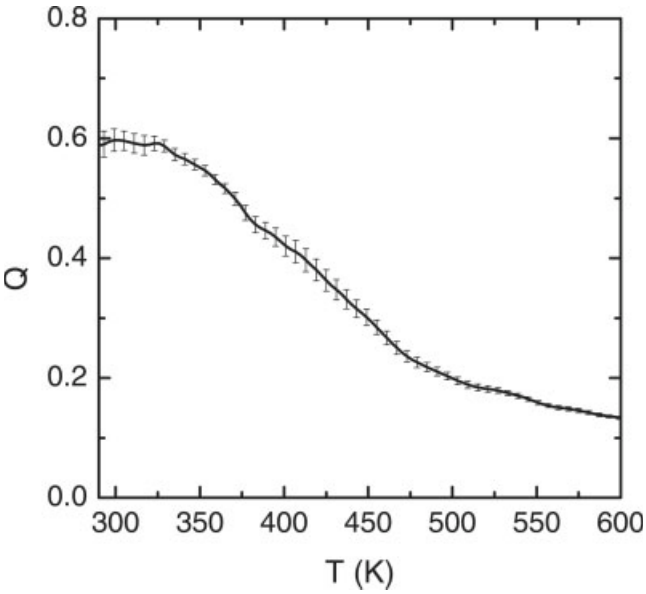

Fig. 3. Average Q as a function of temperature T. The error bars are calculated from the block average of 1 ns in length, and represent the standard deviation.

observed by Lei and Duan in Ref. 15 in which the authors found that the percentage of native contact go below 70% even at 300 K.

To further diagnose the convergence of the simulations, the average Q as a function of the temperature T is calculated and plotted in Figure 3 with the error bars (standard deviation) being calculated from the block (1 ns) averages as has been done in literature.[30] Meanwhile, we also performed a new REMD simulation with different initial structures described in the Model and Methods, and compared the free energy landscapes of two REMD simulations. It is found that the free energy landscapes are similar on the whole for the two REMD simulations, which will be discussed in the subsequent subsection [Fig. 6(a,b)]. The small uncertainties of the average Q at each temperature and the insensitivity of the free energy landscape to the initial structures suggest that the simulations are reasonably converged.

Free Energy Landscape and Folding Mechanisms

Figure 4 shows the free energy landscape projected to the reaction coordinates R_g and Q at $T = 300$ K (a) and 420 K (b). One can find that besides the extended state, two major states are mostly populated at 300 K. One is around the region of $Q = 0.8$ and $R_g = 9$ Å, which corresponds to the native state. Another one is around the region of $Q = 0.4$ and $R_g = 9.5$ Å. For convenience, we label this state as "I" state. The free energy barrier separating these two basins is about $2.0RT$. In the free energy landscape, the "I" state has higher value of Q compared to that of the extended state, which implies that the hydrophobic collapse is accompanied with the formation of the secondary structures. From the state "I" to the native state, the R_g only decreases slightly, which indicates that hydrophobic core is mostly formed in the "I" state. At 420 K, both the "I" state and the native state are still well populated. Compared with the results of 300 K, more extended state can be sampled, and the free energy landscape becomes less rugged at 420 K.

Having a close look at the structures in the state "I" is important for understanding the folding pathway of this protein. For this purpose, we make a cluster analysis[31] to the structures sampled in this basin. The structures with $0.40 < Q < 0.45$ and 9.4 Å $< R_g <$ 10.0 Å sampled at $T = 300$ K are selected for the cluster analysis. Figure 5(a–e) shows the five most populated conformations, which amount to about 64% of the total structures sampled in this state. The percentages for each of the conformations are 18, 15, 12, 10, and 9, respectively. The small percentages for each conformation indicate that the structures in this state are relatively heterogeneous. For comparison, the NMR structure of the FSD-1 is also presented in Figure 5(f). Only the side chains of Tyr3 (green), Ile7 (pink), Phe12 (blue), Leu18 (red), and Phe25 (gray) for one structure are explicitly represented for clearness. In Figure 5, the most important characteristic of the conformations is that the C-terminal α-helix is well formed. In comparison, the N-terminal β-hairpin is mostly disrupted. Therefore, the heterogeneity mainly comes from the diversity of the structures in the N-terminal region. The structure heterogeneity of the β-hairpin for the same peptide was also observed by Lei and Duan.[15] In their work, they observed that the C-terminal α-helix is more stable than the N-terminal β-hairpin. The present results based on the REMD simulations are in agreement with their work. It is also interesting to note that some α-helix may be sampled in the N-terminal region as shown by Figure 5(b,d,e). This is consistent with the observations of Ref. 13 in which the authors found that the β-hairpin moiety can undergo a transition to α-helix.

To explore the folding sequence of the secondary structures, the free energy landscape projected to the reaction coordinates N_α and Q_β is constructed at $T = 420$ K and presented in Figure 6(a). The temperature of 420 K is close to the melting temperature, which will be discussed later. For comparison, we also provide the free energy landscape calculated from the simulation with initial structures of "STRUCTURE2" in Figure 6(b). We can see that the free energy landscapes from the simulations with different initial structures are similar to each other on the whole though some details are different, which indicates that the simulation is converged to reasonable extent. In both these figures, three energy minimums can be observed, and locate at the right top corner, left top corner, and left bottom corner of the panel, respectively. The basin at the right top corner corresponds to the state with both the β-hairpin and α-helix folded. The one at the left top corner corresponds to the state with α-helix formed and the β-hairpin disrupted. The one at the left bottom corner corresponds to the state with both the α-helix and the β-hairpin unfolded.

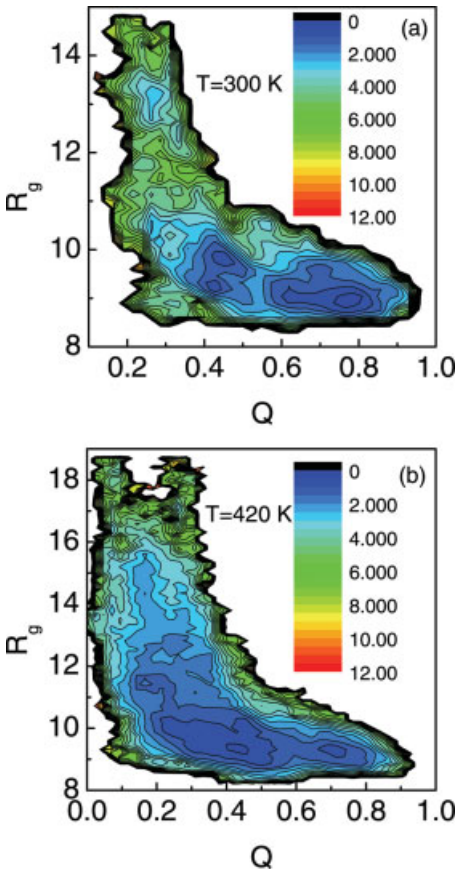

Fig. 4. Free energy landscape projected onto the reaction coordinates R_g^{core} and Q at temperatures of 300 K (**a**) and 420 K (**b**).

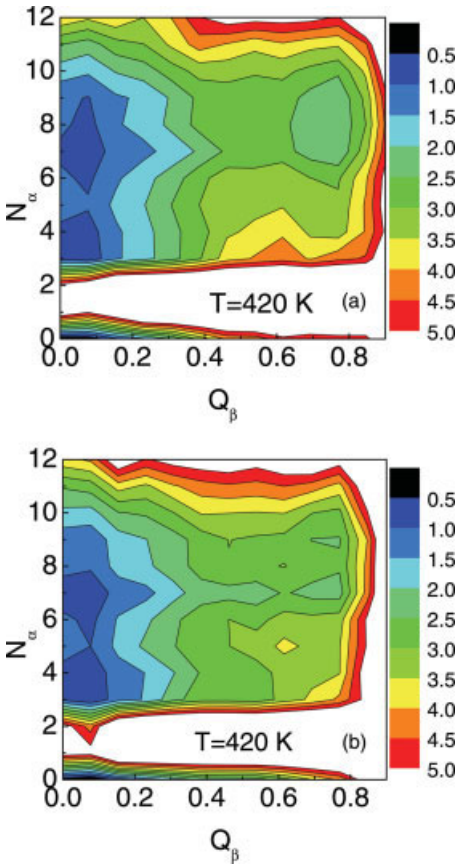

Fig. 6. Free energy landscape projected onto reaction coordinates N_α and Q_β at melting temperature of 420 K with initial structures "STRUCTURE1" (**a**) and "STRUCTURE2" (**b**).

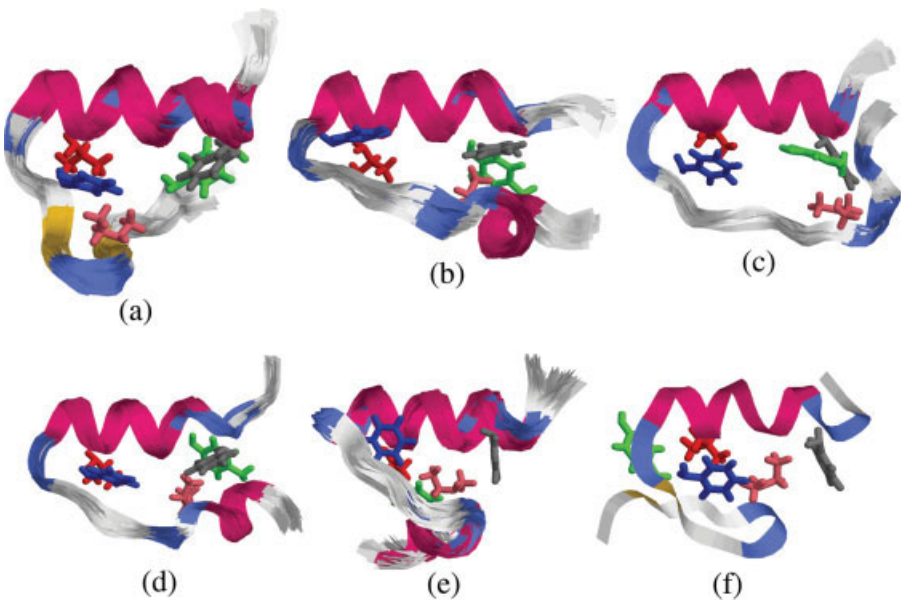

Fig. 5. The most populated conformations in the state "I" extracted based on the cluster analysis technique (**a–e**) and the NMR structure (**f**). The percentages of the conformations in (a), (b), (c), (d) and (e) are 18, 15, 12, 10, and 9, respectively. For clearness, only the side chains of Tyr3 (green), Ile7 (pink), Phe12 (blue), Leu18 (red), and Phe25 (gray) for one structure are explicitly represented.

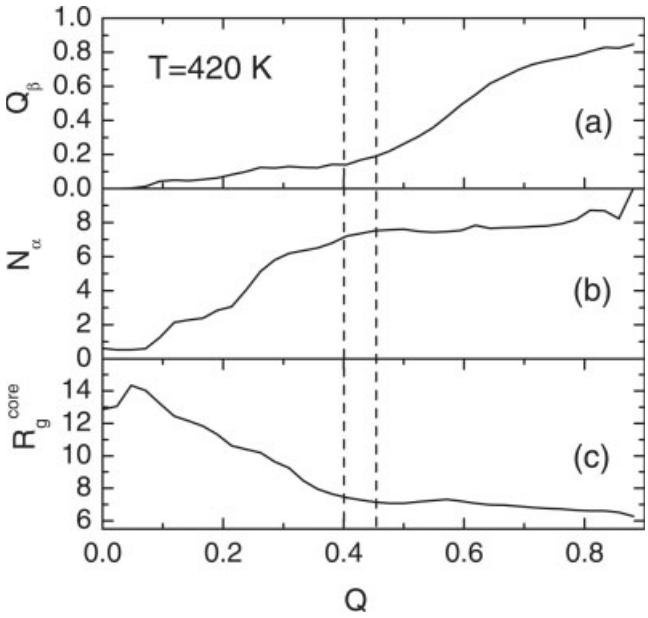

Fig. 7. Average Q_β (**a**), N_α (**b**), and R_g^{core} (**c**) as a function of Q at $T = 420$ K. The region between the dashed lines corresponds to the state where the α-helix is well folded, but the β-hairpin is mostly unstructured.

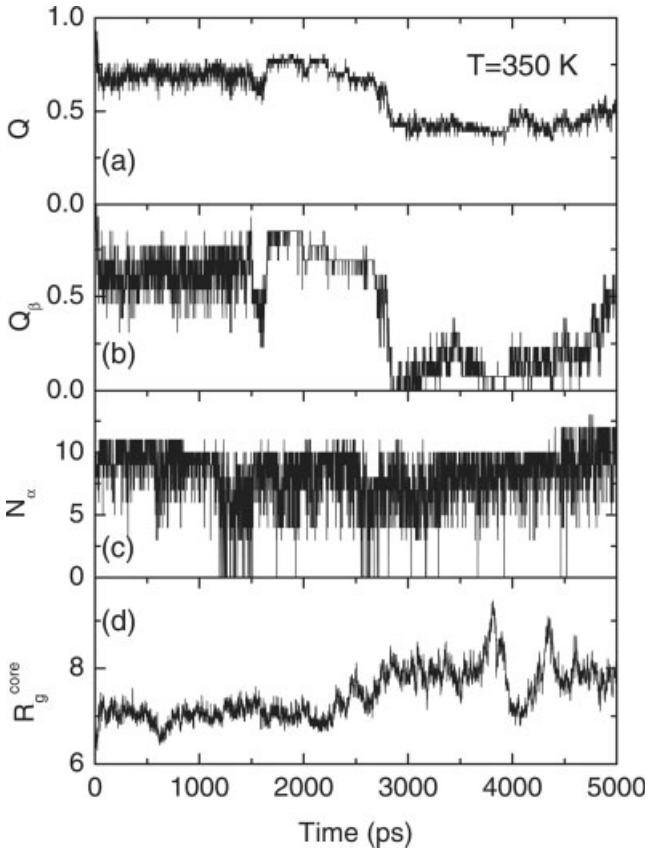

Fig. 8. Time series of Q (**a**), Q_β (**b**), N_α (**c**), and R_g^{core} (**d**) at $T = 350$ K for one representative trajectory simulated using standard MD.

The gap between 0 and 3 along the N_α axis in the left bottom of Figure 6(a,b) resulted from the definition of the helical residues, which demands three or more consecutive residues satisfying the dihedral constraints as described in the previous section. Therefore, the values of 1 and 2 cannot be appointed to N_α. From this figure, one can conclude that the α-helix is more stable compared with the β-hairpin. An intermediate exists in which the α-helix is well formed, while the β-hairpin is still unstructured. Taken together the results of Figures 4 and 6, the folding pathway in the above reaction coordinate space can be deduced: the folding is initiated by the collapse of the hydrophobic core, which is accompanied with the formation of the α-helix. After the formation of the α-helix, the β-hairpin begins to fold, which is accompanied with the further packing of the hydrophobic core.

To further confirm the above deduced folding pathway, the average Q_β, N_α, and R_g^{core} are calculated for each Q at $T = 420$ K and plotted in Figure 7(a–c), respectively. Before $Q = 0.4$, the N_α increases rapidly, which is accompanied with the decreasing of the R_g^{core}. In comparison, the Q_β keeps a low value of less than 0.2 during the above Q range. After $Q = 0.4$, the N_α reaches a saturated value. In comparison, the Q_β begins to increase dramatically with the R_g^{core} having a minor decrease. The region between the two dashed lines corresponds to the "I" state of Figure 4, where the α-helix is well folded, but the β-hairpin is mostly unstructured. These results confirm the pathway concluded from Figures 4 and 6.

The above deduced folding pathway is inferred from the reaction coordinate space. Since the time information is completely lost during the REMD simulations, one cannot obtain the folding pathway corresponding to the time sequence directly using the REMD data. To explore the time sequence of the folding events, five standard MD simulations are conducted for 5000 ps at $T = 350$ K from the native state. Figure 8 shows the results for one of the trajectories. In Figure 8(a–d), the Q, Q_β, N_α, and R_g^{core} are plotted as a function of time during the unfolding events. One can see that before 2700 ps, the Q, Q_β, N_α, and R_g^{core} all keep at certain values with fluctuations. At around 2700 ps, the Q decreases largely, which resulted from the drastic decreasing of the Q_β and the partial disruption of the hydrophobic core as indicated by the decreasing of the Q_β and the slight increasing of the R_g^{core} in Figure 8(b,d). In comparison, the N_α keeps a certain value until the end of the 5000 ps simulation. Although the 5000 ps simulation is not enough to complete the unfolding event, the unfolding sequence of the secondary structures can still be deduced unambiguously. The five trajectories give quite close results concerning the unfolding pathway. These results indicate that the α-helix is much resistant to the thermal fluctuations compared with the β-hairpin. Assuming that folding follows the reverse pathway of unfolding,[32–34] the folding of the FSD-1 should initiate with the formation of the α-helix and the collapse of the hydrophobic residues, then the

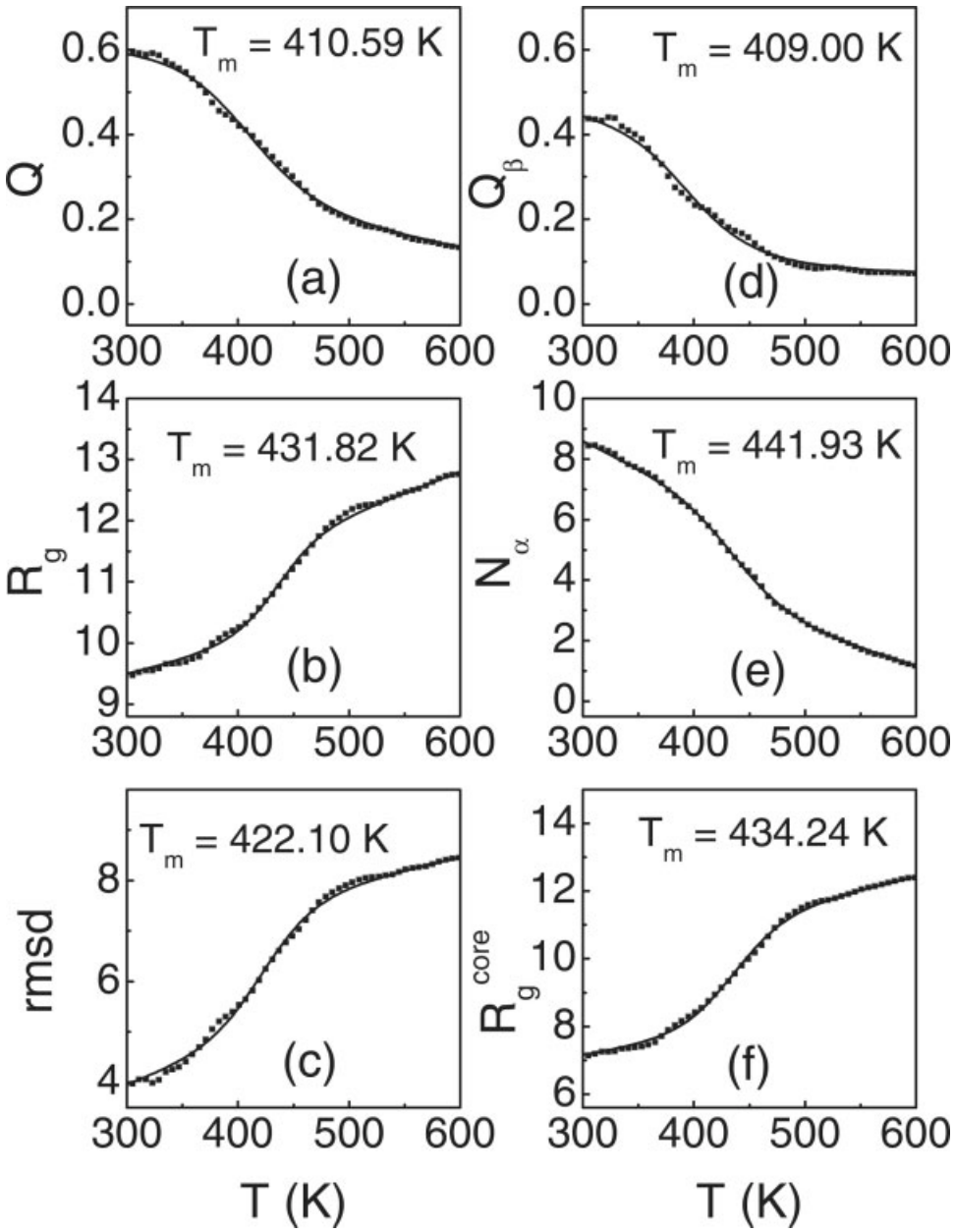

Fig. 9. Average Q (**a**), R_g (**b**), rmsd (**c**), Q_β (**d**), N_α (**e**), and R_g^{core} (**f**) as a function of temperature T (solid squares) and the two-state model fitting (solid line). The fitted melting temperature T_m is shown in the panels. The baselines are subtracted during the fitting.

folding proceeds with the formation of the β-hairpin, which is accompanied with the further packing of the hydrophobic core. This result is consistent with the folding pathway in the reaction coordinate space discussed above.

It is interesting to compare the folding pathways of the FSD-1 and the zinc finger peptide. As the zinc finger peptide is not capable of folding into the native structure without the help of metal ions, one cannot figure out the role of the metal ions on the folding pathway directly by comparing the folding of the apo-form and holo-form peptide. However, if the folding pathway is mainly determined by the native topology (a general assumption of the Go like model),[35] one may extract such information by comparing the folding pathway of the FSD-1 and the zinc finger peptide due to the same native topology adapted by them. In Ref. 36, Miura et al. by using the Raman spectrum technique studied the folding pathway of the third finger of the transcription factor Zif268. It was found that in the absence of the zinc ions, the peptide adopts β sheet structure. With the addition of the zinc ions, the cysteine residues are coordinated first, which induces the formation of the N-terminal β-hairpin and the transition of the C-terminal β strand to short

α-helix. With the further addition of the zinc ions, the histidine residues are also coordinated to the zinc ions, which results in the complete formation of the α-helix. This folding scenario is quite different from that of the FSD-1 for which the folding is initiated by the formation of the α-helix, and the formation of the β-hairpin is the last event during folding. The difference between the folding pathway of the FSD-1 and the zinc finger peptide indicates that the binding of the zinc ions not only stabilizes the fold, but may also affect the folding pathway. One possible mechanism that the metal ions affect the folding pathway is as follows: without the aid of metal ions, disulfide bonds, or other unusual residues, α-helix usually folds faster than β-hairpin because the α-helix is mainly dominated by the local contacts. However, in the zinc finger peptide, the zinc ions tend to form coordination bonds with the cysteine residues prior to the histidine residues. As the cysteine residues locate at each strand of the β-hairpin, the loop formed between the cysteine residues decreases the conformational space significantly, which fasten the folding of the β-hairpin greatly. As a result, the metal binding makes the β-hairpin fold prior to the α-helix in the zinc finger peptide. It should be noted that we are comparing the simulation data of the FSD-1 to the experimental data of the zinc finger, which has different sequence. The conclusion is drawn based on the assumption that folding pathway is mainly determined by the native topology. However, for some proteins, the sequence can also play a crucial role on the folding pathway. In addition, the comparison of the simulation data to experimental data may also result in some uncertainties. Therefore, more straightforward method is needed in further studying the role of the zinc ion on the folding pathway of zinc finger.

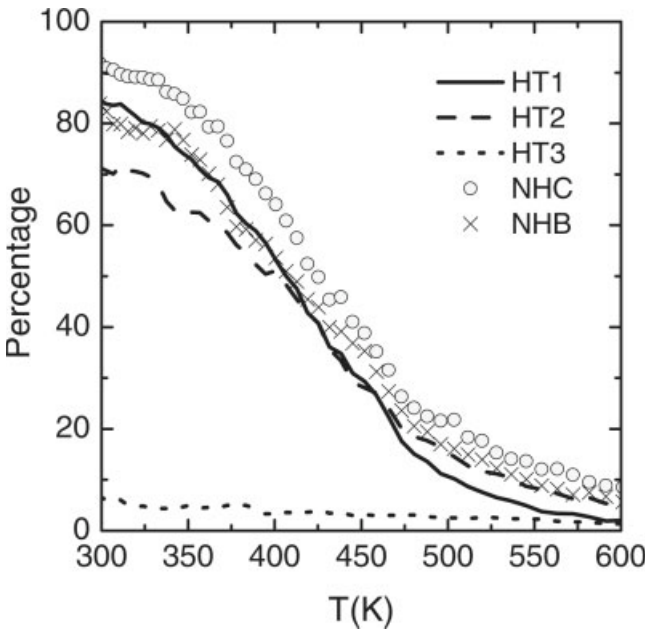

Fig. 10. Structure percentages with each of the three helical turns folded (solid line for HT1, dashed line for HT2, and dotted line for HT3) and percentages with the N-capping hydrogen bond (NHB) formed (cross) and with the N-capping hydrophobic contact (NHC) formed (open circle) as a function of temperature.

Thermal Stability of the Peptide

In the folding process of the zinc finger peptide, the binding of the zinc ions has large contributions to the overall thermodynamics.[37] For the FSD-1, the metal binding site is replaced by a larger hydrophobic core. Therefore, it is interesting to study the overall stability of the FSD-1. Such information will also be helpful to understand the successful folding of the FSD-1.

Figure 9 shows the thermal denaturation curves for the reaction coordinates Q, R_g, rmsd, Q_β, N_α, and R_g^{core} as a function of temperature T in (a), (b), (c), (d), (e), and (f), respectively, and are represented by solid squares. To obtain the melting temperatures, we fit these curves with the two-state model (solid line), and the fitted melting temperature T_m are shown in the panels. The obtained T_m for the reaction coordinates of Q, R_g, rmsd, Q_β, N_α, and R_g^{core} are 410.59, 431.82, 422.10, 409.00, 441.93, and 434.24 K, respectively. These obtained melting temperatures are much larger than 340 K obtained in experiment based on the far-UV CD spectra. The disagreement between the simulated value and the experimental data may be caused by the inaccuracy of the force field in high temperatures[30] and the use of NVT ensemble, which tends to stabilize the hydrophobic core at high temperatures.[24] This feature does not affect the discussions below. Among these melting temperatures, the one obtained from the reaction coordinate N_α (441.93 K) is much higher than that obtained from the reaction coordinate Q_β (409.00 K), which suggests that the C-terminal α-helix has higher thermal stability than the N-terminal β-hairpin.

In extracting the melting temperatures, the two-state model is used to fit the thermal denaturation curves. Despite such two-state fitting may be not physically meaningful for this protein which has weak cooperativity, it is still a reasonable way to extract and compare the transition points of the thermal denaturation curves for different reaction coordinates as done in the work of Naganathan et al. and Sadqi et al.[38,39] More detailed discussions will be given in the next subsection.

To further understand the overall stability of the FSD-1, the stabilities of the component secondary structures are discussed. In Figure 10, the population percentages for the first helical turn (HT1), the second helical turn (HT2), and the third helical turn (HT3) of the α-helix are plotted as a function of temperature T. The HT1, HT2, and HT3 are plotted with solid line, dashed line, and dotted line, respectively. We consider the HT1 as formed if residues E15, K16, E17, and L18 satisfy the dihedral constraints for α-helix. In the same way, the HT2 (or HT3) is taken as formed if residues R19, D20, F21, and I22 (or E23, K24, F25, and K26) satisfy the dihedral constraints. One can see that the percentages for the structures with the HT3 folded are less than 10%, which

implies the high flexibility of the C-terminal residues. Both the HT1 and HT2 have high populations at low temperatures, and the population decreases with the increasing of the temperatures. However, when $T < 460$ K, the HT1 has more populations compared with the HT2. When $T > 460$ K, HT2 has higher populations than the HT1. This result implies that at low temperatures, the helical turn at N-terminal of the α-helix is more stable than the one at the middle part. At high temperatures, the middle helical turn of the α-helix becomes more stable. The high stability of the HT1 at low temperatures may attribute to the N-capping motif,[40] which consists of the hydrogen bond formed between the side chain of Asn14 and the amide of Glu17 as well as the hydrophobic contact made by the side chain of Phe12 and Leu18. In Figure 10, the N-capping hydrogen bond (NHB) and N-capping hydrophobic contact (NHC) are plotted with cross and open circles, respectively. One can see that at low temperatures, both the NHB and NHC are well formed. However, when $T > 460$ K, they are disrupted mostly. Without the N-capping hydrogen bond and the N-capping hydrophobic contact, the HT1 has less steric constraints compared with the HT2, which results in the instability of the NT1 compared with the HT2 at high temperatures. The above results indicate that the N-capping motif employed in the design of the FSD-1 is an important factor, which contributes to the successful folding and stabilization of the FSD-1.

The importance of the N-capping motif in stabilizing the FSD-1 was also explored in experiment by Koscielska-Kasprzak et al.[41] In their work, the disruption of the N-capping motif by replacing the Asn14 with Arg or by replacing the Glu17 with His dramatically decreases the stabilities of the "FSD-1."[41] The present simulation work again demonstrated the importance of the N-capping motif in stabilizing the native structure of the FSD-1.

The stability and flexibility of the N-terminal β-hairpin of the FSD-1 have been discussed by Lei and Duan,[15] using the all-atom MD simulations. Prominent plasticity was found for the N-terminal β-hairpin in their work. The high flexibility of Ile7 contributes to the overall instability of the FSD-1 greatly. Meanwhile, both the simulation work and experimental data show that by replacing the Ile7 with Tyr, the stability of the β-hairpin can be improved largely due to the much larger surface of the Tyr, which can form more atomic contacts and stabilize the hydrophobic core.[15,42] In addition to the contacts made by the Ile7, the contacts involving the Tyr3 formed in the native state are also crucial for the stabilization of the β-hairpin. This result can be inferred by exploring the structures of the "I" state in Figure 5. It is found that instead of packing with the Phe12, Leu18, and the hydrophobic groups in the loop region between the N-terminal β-hairpin and the C-terminal α-helix, the Tyr3 forms contacts with the Ile7 and Phe25 in the "I" state compared with the native structure (see the position change of the green side chain compared with the NMR structure in Fig. 5), while other hydrophobic residues are mostly well packed. This result indicates that by making stronger the contacts involving the Tyr3 in the native state, one can stabilize the β-hairpin further. It is worth noting that the hydrophobic residue of this position in the classical Cys_2His_2 zinc finger motif is strictly conserved, and forms contacts with the residues located at the first helical turn of the C-terminal α-helix and with the hydrophobic groups of the loop between the β-hairpin and α-helix. It is reasonable to believe that to stabilize the β-hairpin is one possible reason for the nature to select this conservative hydrophobic residue in the zinc finger motif.

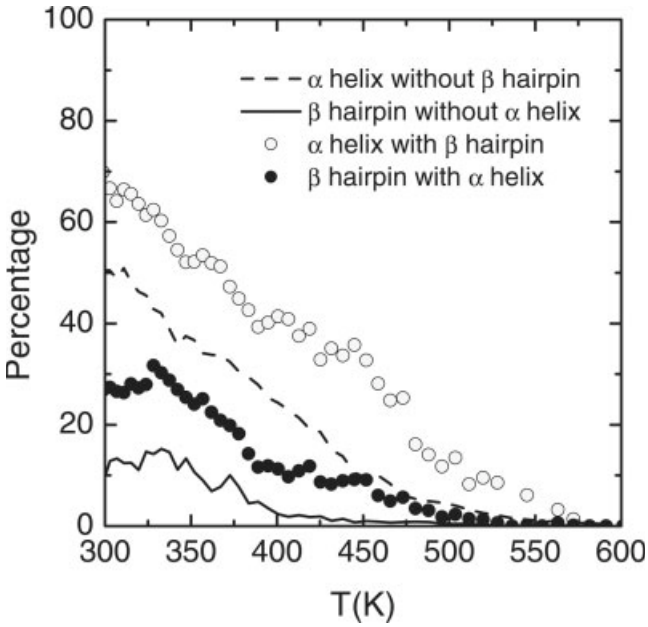

Fig. 11. Structure percentages sampled at different temperatures with certain conditions. The dashed line and open circles represent the percentages of the α-helix among the structures with the β-hairpin unfolded and folded, respectively. The solid line and close circles represent the percentages of the β-hairpin among the structures with the α-helix unfolded and folded, respectively.

Folding Cooperativity Between the Secondary Structures

The folding cooperativity between the secondary structures is widely discussed in literatures for BBA1, BBA5, and other small globular proteins.[43–47] For the FSD-1, the experimental data show that the folding of this protein only exhibits week cooperativity.[4] To analyze the cooperativity for the folding of the β-hairpin and α-helix, the conditional analysis is performed for the structures sampled at different temperatures. Figure 11 shows the structure percentages sampled at given temperatures with certain conditions. The dashed line and open circles represent the percentages of the α-helix among the structures with the β-hairpin unfolded and folded, respectively. In comparison, the solid line and close circles represent the percentages of the β-hairpin among the structures with the α-helix unfolded and folded, respectively. One can see that the percentage of the α-

helix (β-hairpin) with the β-hairpin (α-helix) folded is larger than the one with the β-hairpin (α-helix) unfolded. This result indicates that the folding of the α-helix and β-hairpin is cooperative. However, as the increments of the percentages are limited (less than 20%), the cooperativity is relatively weak.

The weak cooperativity can also be inferred from Figure 9. As mentioned in the last subsection, though the FSD-1 is not a standard two-state protein, the two-state model is still a reasonable method to extract and compare the transition points of different reaction coordinates. For the standard two-state folding with high cooperativity, the melting temperatures determined using different reaction coordinates should be very close.[38,39] However, among the melting temperatures extracted in Figure 9 from different reaction coordinates, relatively large differences can be found (the maximal difference among them is about 33K), which also indicates that the folding cooperativity is relatively weak.

In Ref. 44, the authors found that the folding of the BBA1 is noncooperative. However, the folding of the BBA5 is cooperative with the identical simulation conditions. In comparison, both the experimental data and the present simulations indicate that the folding of the FSD-1 has weak cooperativity. All of these three proteins have identical three-dimensional structures and designed based on the zinc finger motif. However, they exhibit quite different folding cooperativity, which implies the sequence dependence of the folding mechanism.

CONCLUSIONS

This work investigated the folding mechanism and thermal stability of a designed protein FSD-1 employing the REMD method and standard MD simulations with water molecules being explicitly treated. An intermediate is observed in which the α-helix is mostly populated, while the β-hairpin is deformed and disrupted. The results show that the folding of the FSD-1 is initiated by the hydrophobic collapse, which is accompanied with the formation of the α-helix. The folding further proceeds with the formation of the β-hairpin and the further package of the hydrophobic core. By comparing with the folding pathway of the target zinc finger peptide and based on the assumption that the folding pathway is mainly determined by the native topology, it was proposed that the zinc ion not only stabilizes the native structure of the zinc finger, it may also affect the folding pathway. The thermal stability of the FSD-1 is also calculated and discussed. The results show that the contacts involving the side chain of the Tyr3 in the native state is essential for the stabilization of the β-hairpin. Meanwhile, the N-capping motif improves the stability of the C-terminal α-helix dramatically. It was also found that the folding of the β-hairpin and α-helix is weak cooperative for the FSD-1. These results provide us a better understanding on the capabilities of the folding and stabilization of the designed protein FSD-1. Such knowledge can be useful in developing better strategies for de novo protein design. Meanwhile, the folding mechanism studied for the FSD-1 makes us have a deeper understanding to the general principles in the folding of the ββα fold.

ACKNOWLEDGMENTS

The authors thank Shanghai Supercomputer Center for the computational support, and the helpful discussions with Dr. Jun Wang and Dr. Meng Qin.

REFERENCES

1. Jones DT. Learning to speak the language of proteins. Science 2003;302:1347–1348.
2. DeGrado WF, Summa CM, Pavone V, Nastri F, Lombardi A. De novo design and structural characterization of proteins and metalloproteins. Ann Rev Biochem 1999;68:779–819.
3. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard B, Baker D. Design of a novel globular protein fold with atomic-level accuracy. Science 2003;302:1364–1368.
4. Dahiyat BI, Mayo SL. De Novo protein design: fully automated sequence selection. Science 1997;278:82–87.
5. Struthers MD, Cheng RP, Imperiali B. Design of a monomeric 23-residue polypeptide with defined tertiary structure. Science 1996;271:342–345.
6. Regan L, Clarke ND. A tetrahedral zinc(II)-binding site introduced into a designed protein. Biochemistry 1990;29:10878–10883.
7. Marino SF, Regan L. Secondary ligands enhance affinity at a designed metal-binding site. Chem Biol 1999;6:649–655.
8. Klug A, Schware WR. Zinc fingers. FASEB J 1995;9:597–604.
9. Berg JM, Godwin HA. Lessons from zinc-binding peptides. Annu Rev Biophys Biomol Struct 1997;26:357–371.
10. Frankel AD, Berg JM, Pabo CO. Metal-dependent folding of a single zinc finger from transcription factor IIIA. Proc Natl Acad Sci USA 1987;84:4841–4845.
11. Cox EH, McLendon GL. Zinc-dependent protein folding. Curr Opin Struct Biol 2000;4:162–165.
12. Hubner I, Deeds EJ, Shakhnovich EI. High-resolution protein folding with a transferable potential. Proc Natl Acad Sci USA 2005;102:18914–18919.
13. Jang S, Shin S, Pak Y. Molecular Dynamics Study of peptides in implicit water: Ab initio folding of β-hairpin, β-sheet, and ββα-motif. J Am Chem Soc 2002;124:4976–4977.
14. Kim SY, Lee J, Lee J. Folding simulations of small proteins. Biophys Chem 2005;115:195–200.
15. Lei H, Duan Y. The role of plastic β-hairpin and weak hydrophobic core in the stability and unfolding of a full sequence design protein. J Chem Phys 2004;121:12104–12111.
16. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 1999;314:141–151.
17. García AE, Sanbonmatsu KY. Exploring the energy landscape of a β-hairpin in explicit solvent. Proteins 2001;42:345–354.
18. Zhou R, Berne BJ, Germain R. β-hairpin folding in explicit water. Proc Natl Acad Sci USA 2001;98:14931–14936.
19. Felts AK, Narano Y, Gallicchio E, Levy RM. Free energy surface of β-hairpin and α-helical peptides generated by replica exchange molecular dynamics with the AGBNP implicit solvent model. Proteins 2004;56:310–321.
20. Zhang J, Qin M, Wang W. Folding mechanism of β-hairpins studied by replica exchange molecular simulations. Proteins 2006;62:672–685.
21. Nguyen PH, Stock G, Mittag E, Hu CK, Li MS. Free energy landscape and folding mechanism of a α-hairpin in explicit water: a replica exchange molecular dynamics study. Proteins 2005;61:795–808.
22. García AE, Onuchic JN. Folding a protein in a computer: an atomic description of the folding/unfolding of protein A. Proc Natl Acad Sci USA 2003;100:13898–13903.
23. Cecchini M, Rao F, Seeber M, Caflisch A. Replica exchange molecular dynamics simulations of amyloid peptide aggregation. J Chem Phys 2004;121:10748–10756.

24. Zhang W, Wu C, Duan Y. Convergence of replica exchange molecular dynamics. J Chem Phys 2005;123:154105.
25. Sanbonmatsu KY, García AE. Structure of met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. Proteins 2002;46:225–234.
26. Zhou R. Exploring the protein folding free energy landscape: coupling replica exchange method with P3ME/RESPA algorithm. J Mol Graph Model 2004;22:451–463.
27. Case DA, Darden TA, Cheatham TE, III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Wang B, Pearlman DA, Crowley M, Brozell S, Tui V, Gohlke H, Mongan J, Hornak V, Cui G, Beroza P, Schafmeister C, Caldwell JW, Ross WS, Kollman PA. AMBER 8, University of California, San Francisco, 2004.
28. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comp Chem 2003;24:1999–2012.
29. Ferrenberg AM, Swendsen RH. Optimized Monte Carlo data anaylsis. Phys Rev Lett 1989;63:1195–1198.
30. Zhou R. Trp-cage: Folding free energy landscape in explicit water. Proc Natl Acad Sci USA 2003;100:13280–13285.
31. Snow CD, Qiu L, Du D, Gai F, Hagen SJ, Pande VS. Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy. Proc Natl Acad Sci USA 2004;101:4077–4082.
32. Fersht AR, Daggett V. Protein folding and unfolding at atomic resolution. Cell 2002;108:573–582.
33. Li A, Daggett V. Characterization of the transition state of protein unfolding by use of molecular dynamics: chymotrypsin inhibitor 2. Proc Natl Acad Sci USA 1994;91:10430–10434.
34. Caflisch A, Karplus M. Molecular dynamics simulation of protein denaturation: solvation of the hydrophobic cores and secondary structure of barnase. Proc Natl Acad Sci USA 1994;91:1746–1750.
35. Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: what determines the structure details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. J Mol Biol 2000;298:937–953.
36. Miura T, Satoh T, Takeuchi H. Role of metal-ligand coordination in the folding pathway of zinc finger peptides. Biochim Biophys Acta 1998;1384:171–179.
37. Blasie CA, Berg JM. Structure-based thermodynamic analysis of a coupled metal binding-protein folding reaction involving a zinc finger peptide. Biochemistry 2002;41:15068–15073.
38. Naganathan AN, Perez-Jimenez R, Sanchez-Ruiz JM, Muñoz V. Robustness of downhill folding: guidelines for the analysis of equilibrium folding experiments on small proteins. Biochemistry 2005;44:7435–7449.
39. Sadqi M, Fushman D, Muñoz V. Atom-by-atom analysis of global downhill protein folding. Nature 2006;442:317–321.
40. Aurora R, Rose GD. Helix capping. Protein Sci 1998;7:21–38.
41. Koscielska-Kasprzak K, Cierpicki T, Otlewski J. Importance of α-helix N-capping motif in stabilization of ββα fold. Protein Sci 2003;12:1283–1289.
42. Sarisky CA, Mayo SL. The ββα fold: explorations in sequence space. J Mol Biol 2001;307:1411–1418.
43. Roe DR, Hornak V, Simmerling C. Folding cooperativity in a three-stranded β-sheet model. J Mol Biol 2005;352:370–381.
44. Wen EZ, Luo R. Interplay of secondary structures and side-chain contacts in the denatured state of BBA1. J Chem Phys 2004;121:2412–2421.
45. Schenck HL, Gellman SH. Use of a designed triple-stranded antiparallel β-sheet to probe β-sheet cooperativity in aqueous solution. J Am Chem Soc 1998;120:4869–4870.
46. Guo C, Cheung MS, Levine H, Kessler DA. Mechanisms of cooperativity underlying sequence-independent β-sheet formation. J Chem Phys 2002;116:4353–4365.
47. Kaya H, Chan HS. Polymer principles of protein calorimetric two-state cooperativity. Proteins 2000;40:637–661.