THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

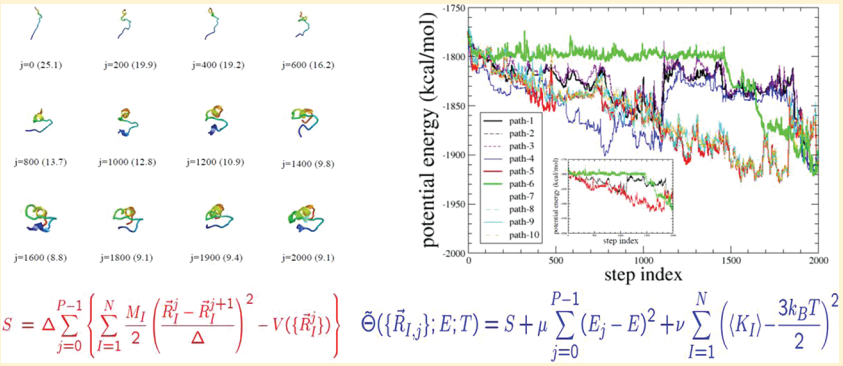
Folding Models of Mini-Protein FSD-1

In-Ho Lee,[†] Seung-Yeon Kim,[‡] and Jooyoung Lee*[,§]

[†]Korea Research Institute of Standards and Science, Daejeon 305-340, Korea

[‡]School of Liberal Arts and Sciences, Chungju National University, Chungju 380-702, Korea

[§]Korea Institute for Advanced Study, Seoul 130-722, Korea



ABSTRACT: We have carried out all-atom action-derived molecular dynamics (ADMD) folding simulations of the full-size FSD-1. FSD-1 is a designed mini-protein of 28 residues containing both α and β secondary structure elements. Multiple folding pathways are found for FSD-1, which is consistent with existing computational studies. Hydrophobic collapse is observed first, and then subsequent folding events proceeds by forming either α-helix or β-hairpin. Concurrent formation of the full tertiary structure and the secondary structure elements of α-helix and β-hairpin is observed. The folding pathway of FSD-1 elucidated by ADMD simulations does not follow the scenario of the framework model. ADMD simulations provide significant insights for the general mechanisms of protein folding and conformational changes.

INTRODUCTION

Small protein FSD-1 is one of the most successfully designed[1] mini-protein folders with a $\beta\beta\alpha$-motif (PDB code: 1fsd). FSD-1 is of 28 residues (QQYTA KIKGR TFRNE KELRD FIEKF KGR) with a well-defined hydrophobic core. FSD-1 is designed to mimic the zinc finger of Zif268, and a larger hydrophobic core consisting of Ala5, Ile7, Phe12, Leu18, Phe21, Ile22, and Phe25 is introduced for stability. In FSD-1, the Zif268 core is preserved, but the metal-chelating His residues and one of the Cys residues of Zif268 are replaced by hydrophobic Phe and Ala residues, respectively. The three dimensional structure of FSD-1 is obtained by NMR spectroscopy, and this new protein is found to be stable to fold into a similar structure as the original zinc finger without the help of metal ions and disulfide bonds. The NMR structure of FSD-1 contains secondary structure segments of both α-helix (residues 15−24) and β-hairpin (residues 3−13). Two N-terminal residues and three C-terminal residues of FSD-1 were not structurally resolved by the NMR experiment.

Recently, FSD-1 has been extensively studied experimentally as well as computationally in order to understand protein folding mechanisms and to test the accuracy of computational folding models. Knowledge about the stability and folding behavior of a zinc finger can help us to better understand its folding mechanism as well as to design new zinc fingers. Experimental data indicates that the folding of FSD-1 is weakly cooperative.[1] This is further supported by the stability and structural study by Feng et al.[2] The broad melting transition observed by circular dichroism and differential scanning calorimetry was proposed to be due to the helix-to-coil transition in the α-helical segment of FSD-1, not from the unfolding of its hydrophobic core.[2]

The mini-protein of FSD-1 containing both α and β secondary structural elements has become a rather attractive target molecule for extensive computational studies. Technically, due to its small size compared to other proteins, relatively long time simulations can be carried out.

Wu and Shea demonstrated that a version of the AMBER force field can fold FSD-1.[3] Folding of FDS-1 was observed in 2 out of 20 simulations, and the majority of the conformational population was dominated either by α-helix-containing structures or by denatured structures.[3] From the analysis of the folding trajectories, the authors observed that α-helix formed first in the early stage of folding, and the concurrent formation of β-hairpin and tertiary structure followed in a later stage. The observed broad melting transition was attributed to

Special Issue: Harold A. Scheraga Festschrift

Received: January 3, 2012
Revised: March 8, 2012
Published: March 15, 2012

 © 2012 American Chemical Society dx.doi.org/10.1021/jp300074f | J. Phys. Chem. B 2012, 116, 6916−6922

the spread of melting transitions of individual structural elements, associated with formations of helix, hydrophobic core, β-hairpin, and the tertiary fold.[3]

By performing molecular dynamics (MD) simulations of FSD-1,[4] Wang and his co-workers observed that the α-helix is formed prior to the formation of the β-hairpin. 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 subsequent consolidation of the hydrophobic core. Authors noted a difference in the folding sequence between FSD-1 and the zinc finger peptide. That is, in the folding process of the zinc finger peptide, the formation of β-hairpin precedes that of α-helix.[4]

Duan and his co-workers developed a new force field emphasizing the balance between two secondary structure elements, α-helix and β-sheet, in proteins.[5] They refitted partial charges of atoms and the main chain torsional parameters under the AMBER framework. With this force field, Lei and Duan[6] applied all-atom MD simulations to study the folding pathway of FSD-1, from which they observed that folding of secondary structures occurs prior to the formation of the native hydrophobic core. The transition state ensemble is characterized by the substantial formation of native secondary structures. This conclusion is derived from the assumption that folding follows the reverse pathway of unfolding simulations. They also found that the C-terminal helix was much more stable than the hairpin at three temperatures, 273, 300, and 330 K. In their simulations, the folding of a double mutated FSD-1 is observed with high population (64.2%) and high fidelity (C^{α} rmsd of 1.29 Å). They found that FSD-1 can fold following two pathways. In the major pathway, folding starts from the formation of the C-terminal helix. In the other minor pathway, however, the N-terminal β-hairpin folds first.[5]

For the study of FSD-1, Jang et al. used the CHARMM19[7] force field in conjunction with the GB solvation model. They found that the folding of FSD-1 always starts with the formation of the α-helix, which is followed by the formation of the N-terminal β-hairpin.[8] When the α-helix is folded, the native hydrophobic core of the protein is partially formed, and only when the β-hairpin is folded into its native position to complete the folding process, the hydrophobic core is fully consolidated.

A modified force field of AMBER was also applied to fold FSD-1, where param99 were adjusted and the generalized Born solvation model was used. The transferability of the modified force field was tested for α-helices and β-hairpins.[9]

Kim et al.[10] first optimized an atomistic potential energy function, so that with this energy function four small proteins (two α-proteins, an $\alpha\beta$ protein, and a β-hairpin) can be folded simultaneously starting from random conformations. With this optimized potential, folding simulations of FSD-1 were carried out and they found that the folding was initiated by the collapse of the hydrophobic core which was followed by the formation of secondary structures.

Mohanty and Hansmann[11] studied the folding of FSD-EY (1FME), double-mutated (Q1E and I7Y) protein of FSD-1. They found that the C-terminal helix of FSD-EY is much more stable than the N-terminal β hairpin and forms first. Authors examined correction terms for the ECEPP/3 force field,[12] and they found an increased probability of native state formation at low temperatures resulting from reduced propensity to form α helices and increased formation of β sheets.

As summarized above, existing molecular dynamics simulations on FSD-1 provide a variety of folding pathways failing to suggest a consensus folding model. This disagreement could be attributed to various force fields used in simulations. In this work, we apply a rather different method in spirit than discussed so far, namely, a trajectory sampling method for the folding study of FSD-1 by optimizing the classical action. In this approach, the protein folding problem is converted from an initial value problem to a boundary value problem. One technical difficulty in solving an initial value problem associated with rare events, as often found in conventional MD simulations, is that simulations can be trapped in local energy minima, not to mention problems arising from the inaccuracy of the force field used.

The remainder of this article is organized as follows. In the following section, we describe the computational method of action-derived molecular dynamics utilized to generate folding trajectories of the mini-protein FSD-1. In the next section, we present results by analyzing generated folding trajectories, which are discussed in relation with existing folding studies. Finally, we conclude with a summary.

METHODS

Recently, we have proposed that the action-derived molecular dynamics (ADMD)[13,14] method can serve as a useful tool to study rare events including protein folding. Our theoretical approach is rather different from the conventional molecular dynamics (MD) simulation approach which samples the conformational space of a system at a given temperature starting from a given conformation. In ADMD, for two given (initial and final) conformations, we directly search for pathways between them that approximately satisfy Newtonian trajectory conditions with a large time step.[13−15] Rather than performing an indefinite amount of MD simulation time steps for a protein to fold, which is equivalent to solving an initial value problem, we carry out transition pathway sampling to identify trajectories with optimal values of action (see below).

When searching for low-energy pathways, we expect that pathways with low potential-energy barriers are more probable than others. In this work, we refer probable transition pathway models to the trajectories with low-potential-energy barriers. We treat two given protein conformations as a "reactant" (extended structure) and a "product" (native structure) and search for the transition pathways that connect these two structures. After low potential-energy barrier trajectories are obtained, we evaluate the radius of gyration (R_g), root-mean-square deviation (rmsd) from the native structure, native contact formation, and hydrogen bond formation along the generated pathways.

The primary goal of ADMD simulations is to delineate the global temporal sequence of conformational changes during the folding process at an atomic resolution in conjunction with appropriate energy barriers. Generally speaking, this type of protein folding characterization is not possible with conventional simulation methods that typically capture only short-time dynamics. At the beginning of each ADMD simulation,[13−15] a set of random numbers is used to generate a trial atomic trajectory for each atom, connecting between initial and final conformations.

The ADMD method has been successfully applied to investigate folding pathways of α-helix (acetyl-$(Ala)_{10}$-*N*-methyl amide) and β-hairpin (residues 41−56 of protein G) systems.[16] It is shown that generated dynamic folding pathway models for

dx.doi.org/10.1021/jp300074f | J. Phys. Chem. B 2012, 116, 6916−6922

both α-helix and β-hairpin formations are consistent with experimental data, reproducing the sequences of the native hydrogen bond formations.[16] In addition, the ADMD method has been used to study the folding pathway of the 36-residue villin headpiece subdomain (HP-36).[17] The folding is initiated by hydrophobic collapse, after which concurrent formation of full tertiary structure and α-helical secondary structure is observed. The C-terminal helix forms first, followed by the N-terminal helix positioned in its native orientation. The short middle helix is shown to form last.[17]

In the current ADMD simulations, we have discretized the whole atomic trajectory into $P = 2000$ steps with the successive time increment of $\Delta = 36.28$ fs, the total simulation time being $\tau = P\Delta = 72.56$ ps. The path $\{\mathbf{q}_j\}$ represents a collection of sequential structural frames with fixed and given initial $\mathbf{q}_0$ and final $\mathbf{q}_P$. The discretized action of a trajectory can be written as

$$S = \sum_{j=0}^{P-1} L_j(\{\mathbf{q}_j\})\Delta \tag{1}$$

where the discretized Lagrangian of the jth temporal frame is defined as

$$L_j = \sum_{I=1}^{N} \frac{m_I}{2\Delta^2}(\mathbf{q}_{I,j} - \mathbf{q}_{I,j+1})^2 - V(\{\mathbf{q}_j\}) \tag{2}$$

Here, the first term is the kinetic energy and V is the potential energy. N is the total number of atoms, m_I is the mass of the Ith atom, and $\mathbf{q}_{I,j}$ is the position vector of the Ith atom at the jth frame.

The stationarity condition $\delta S = 0$ leads to a set of linear equations. However, discretized pathways generated from the minimization of eq 1 do not satisfy total-energy conservation, as discussed in the work of Passerone and Parrinello.[13] That is, accurate Verlet trajectories[18] are not guaranteed, since the action of eq 1 is not bounded.[13] Passerone and Parrinello suggested adding a constraint term to eq 1 to ensure the total-energy conservation from pathways. The modified action now becomes

$$\Theta(\{\mathbf{q}_j\}; E) = S + \mu_E \sum_{j=0}^{P-1} (E_j - E)^2 \tag{3}$$

where E is the target total-energy value to impose on the system, μ_E is an arbitrary large constant, and E_j is the total energy at the jth temporal frame defined as

$$E_j = \sum_{I=1}^{N} \frac{m_I}{2\Delta^2}(\mathbf{q}_{I,j} - \mathbf{q}_{I,j+1})^2 + V(\{\mathbf{q}_j\}) \tag{4}$$

The quality of pathways is shown to improve further[14] by introducing an additional dynamic restraint. Now the final form of the action to optimize becomes

$$\Phi(\{\mathbf{q}_j\}; E, T) = S + \mu_E \sum_{j=0}^{P-1} (E_j - E)^2 + \mu_K \sum_{I=1}^{N} \left(\langle K_I\rangle - \frac{3k_B T}{2}\right)^2 \tag{5}$$

where $\langle K_I\rangle$ is the average kinetic energy of the Ith atom along the trajectory.[14] Fictitious temperature T controls the kinetic energy of the system, μ_K is an arbitrary large constant, and k_B is the Boltzmann constant. It should be noted that T used in this work does not correspond to the physical temperature. T is only a parameter introduced to improve the quality of pathways by reducing the value of Onsager–Machlup action.[13−15] A smaller value of Onsager–Machlup action corresponds to a more Verlet-like trajectory.

Now, the task in ADMD computation is to optimize the objective function of eq 5 with $3N(P-1) = 3\,022\,488$ degrees of freedom, where the number of atoms is $N = 504$ for FSD-1. The atomic mass of each constituent atom (H, C, N, and O) is used and the total-energy conservation is taken into account. All atoms are treated as point particles. It should be noted that no additional constraints on the covalent bond lengths and bond angles other than the harmonic energy restraints of the force field used are employed. We have used the all-atom force field of amber95 with the GBSA solvent model.[19] In practice, the amber95 force field realized in the TINKER package is used.[20] The TINKER simulation package is modified and used in both protein structure preparation and energy function/force evaluations.

Preparation of the initial ($j = 0$) and final ($j = 2000$) conformations is carried out as follows. Starting from the fully extended structure and the PDB structure of FSD-1, both atomic coordinates are local energy minimized in a slow fashion, so that no dramatic structural variation can occur. Now, we need to place the final conformation (minimized native conformation) relative to the initial conformation (minimized extended conformation). For optimal superposition of the two structures, the quaternion method[21,22] is used. The first coordinate set is fixed, while the second set is translated and rotated to give the best fit.

Recently, a number of closely related variants of the ADMD method were introduced.[24−26] Here, we provide only a brief account of the method, and details of the method can be found from a review paper by Elber and his co-workers.[15]

The advantages of ADMD protein folding simulation lie in its ability to secure a reasonable degree of uniformity in temporal collection of the meaningful conformations related to the folding event. The method effectively filters out fast atomic motions of short- and intermediate-range time scales ($<\Delta$ and $\sim\Delta$). Consequently, characterization of folding pathways at an atomic resolution is possible with ADMD. It should be noted that the ADMD simulation can be executed in a highly parallel fashion.

Our method is related to the computational method popularly known as transition path sampling,[23] which aims to find transition pathways of infrequent events in both equilibrium and nonequilibrium systems. This method allows one to study rare events without requiring prior knowledge of mechanisms, reaction coordinates, and transition states. Transition path sampling conceptually follows the idea of importance sampling. That is, the ensemble of trajectories are sampled and manipulated in the manner of statistical mechanics. In general, a large number of minimum-action or minimum-energy pathways may exist, and for a more thorough understanding of the folding pathway, 10 pathways generated in this study may not be sufficient enough. However, the current ADMD approach is limited by the requirement of enormous computational resources to generate a trajectory. A more

dx.doi.org/10.1021/jp300074f | J. Phys. Chem. B 2012, 116, 6916−6922

rigorous sampling of transition pathways is beyond the scope of the current work, and it remains as a future challenge.

RESULTS AND DISCUSSION

Multiple Folding Pathways. Without the burden of fine-tuning a potential energy function for systems containing α-helices and β-sheets, we can describe the folding events of FSD-1 by ADMD, where the relative temporal ordering between secondary structure elements of α-helix and β-sheet might be of interest. We directly construct folding pathway models between given initial and final conformations. Ten independent ADMD simulations are employed to generate trajectories associated with the FSD-1 folding process. Potential energy variation of the folding pathway is shown in Figure 1 along the ADMD step index. Examining the overall pattern of the potential energy variation, 10 paths are clustered into three groups for further analysis. Three representative folding paths, denoted by path-1 (black), path-5 (red), and path-6 (green), are shown in the inset of Figure 1. Path-1 (black) represents paths 1, 2, 3, and 4, and path-5 (red) represents paths 5, 7, 8, 9, and 10, while path-6 (green) represents itself. This grouping is also consistent with the one based on the formation of secondary structure elements throughout the folding process (see below).

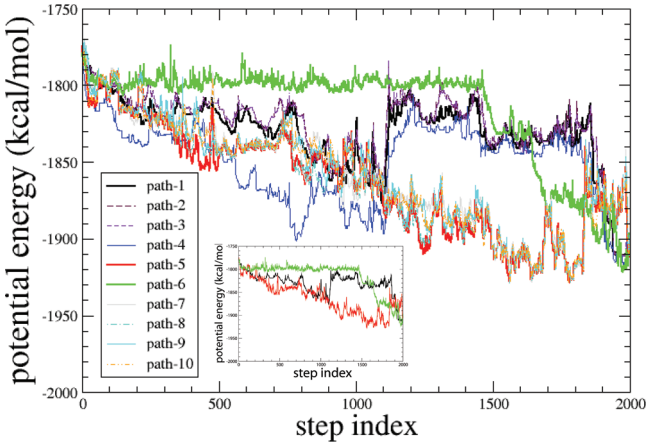


Figure 1. Ten independent ADMD simulations are carried out to generate a time series of conformational variation associated with the folding process of FSD-1 (PDB code: 1fsd). Potential energy variations are shown along the ADMD step index. Examining the profile of potential energy variation, 10 paths are clustered into three groups, which are represented by path-1 (black), path-5 (red), and path-6 (green), as shown in the inset.

The radius of gyration (R_g, solid lines) and the rmsd (dashed lines) from the NMR structure of FSD-1 are shown in Figure 2 along the ADMD step index j. Calculated linear correlation coefficients between R_g and rmsd are about 0.98 for all three trajectories. In the second half of the folding trajectories, path-1 (black) and path-5 (red) are similar to each other in the variation of R_g and rmsd. In the first half of the folding model, path-5 (red) shows larger temporal fluctuations in R_g and rmsd than path-1 (black) does. In contrast, R_g and rmsd of path-6 (green) stay almost constant until $j \sim 1500$ where drastic reduction in both R_g and rmsd is observed. Although variations of R_g and rmsd between path-1 (black) and path-5 (red) appear to be quite similar to each other, details in secondary structure formation are different from each other. The folding model of path-5 (red) shows a well formed secondary structure at $j \sim 1500$, while that of path-1 (black) shows a compact state without definite secondary structure formation. For example, at $j = 1800$, both path-1 (black) and path-5 (red) show a similar value of R_g (9.1 Å). However, the degrees of the second structure formation are rather different from each other, indicating they represent separate folding pathways.

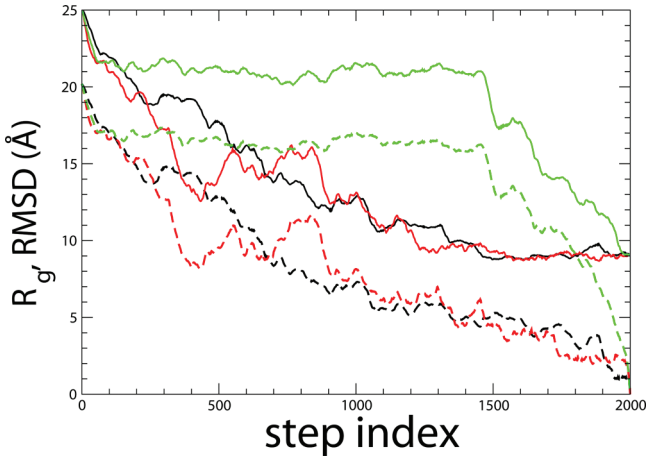


Figure 2. For conformations of path-1 (black), path-5 (red), and path-6 (green), values of radius of gyration (R_g, solid lines) and rmsd (dashed lines) are shown along the ADMD step index j. R_g and rmsd are highly correlated for all three trajectories with correlation coefficients of about 0.98. The second half of path-1 (black) is quite similar to that of path-5 (red) in R_g and rmsd, while the first half of path-5 (red) shows larger fluctuations in R_g and rmsd than that of path-1 (black). In contrast, path-6 (green) stays almost flat until $j \sim 1500$ where drastic reduction in both R_g and rmsd is observed.

Local rmsd values of α-helix (residues 15−24) and β-hairpin (residues 3−13) are measured separately, and they are shown in Figure 3 along the ADMD step index. These local rmsd values can be used to measure the degree of secondary structure formation during the folding process. Solid lines

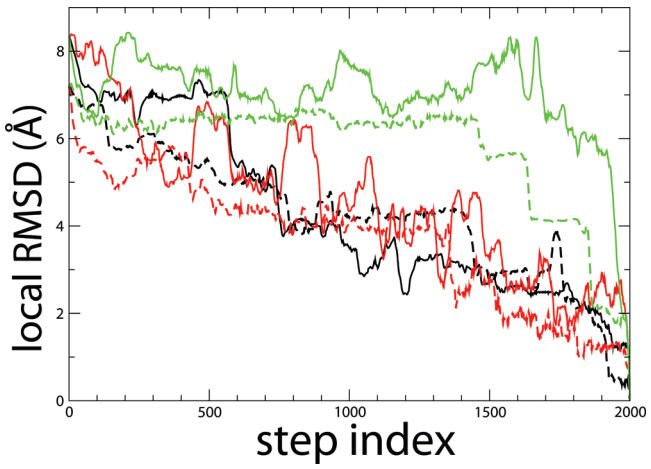


Figure 3. Local rmsd values of secondary structure elements are shown for three representative paths, path-1 (black), path-5 (red), and path-6 (green). Solid and dotted lines are for β-hairpin (residues 3−13) and α-helix (residues 15−24), respectively. Local rmsd values can be used to measure the degree of the secondary structure element formation during folding. In the folding model of path-6 (green), α-helix formation globally precedes β-sheet formation. In the folding models of path-1 (black) and path-5 (red), calculated local rmsd values decrease globally toward folding while oscillating.

dx.doi.org/10.1021/jp300074f | J. Phys. Chem. B 2012, 116, 6916−6922

(dotted lines) represent local rmsd values of β-sheet (α-helix). We observe that, in the folding model of path-6 (green), α-helix forms prior to the formation of β-sheet. In the folding models of path-1 (black) and path-6 (green), calculated local rmsd values decrease globally along the path while oscillating. The local rmsd value of the β-sheet has higher amplitude than that of the α-helix. Local rmsd variations of paths 1, 2, 3, and 4 along the pathway are observed to be similar to each other. Similarly, local rmsd variations among paths 5, 7, 8, 9, and 10 show close similarity.

In Figures 4, 5, and 6, conformational changes for the folding models of path-1 (black), path-5 (red), and path-6 (green) are

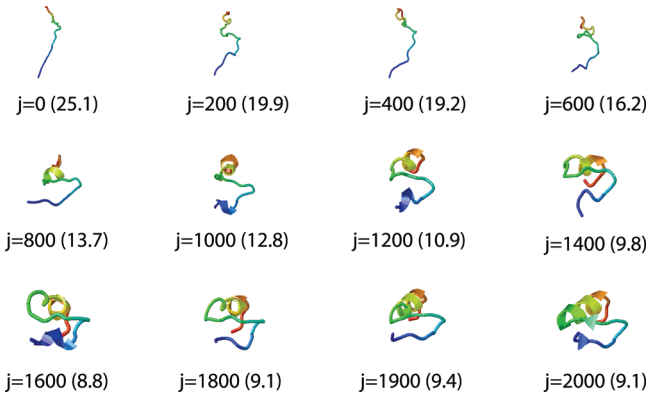


Figure 4. Snapshots for the conformational variation of path-1 are shown. ADMD step index j (radius of gyration in Å) is also indicated. FSD-1 is drawn in ribbon model. N-terminal/C-terminal residues are shown in blue/red color.

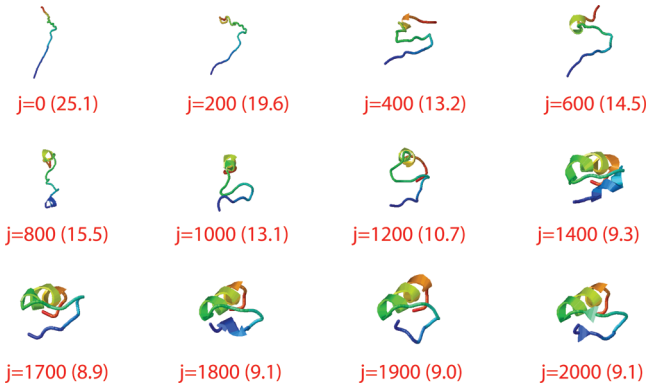


Figure 5. Snapshots for the conformational variation of path-5 are shown. See the caption of Figure 4 for additional details.

shown, respectively. Series of these conformational changes display multiple folding pathways for FSD-1. In the figures, numbers in parentheses represent values of R_g in Å. N-terminal/C-terminal residues are shown in blue/red color. The number of native contacts (responsible for tertiary structure formation) and the number of native hydrogen bonds (responsible for secondary structure formation) are also shown along the ADMD step index j in Figure 7a and b. A native contact is defined to exist between two residues (separated by more than two residues in sequence) if their native $C^{\alpha}-C^{\alpha}$ distance is less than 6.5 Å. A backbone hydrogen bond is defined to exist between a carbonyl-oxygen and an amide-hydrogen if they are separated by less than 2.5 Å and the virtual bond angle between three atoms (oxygen, nitrogen, and

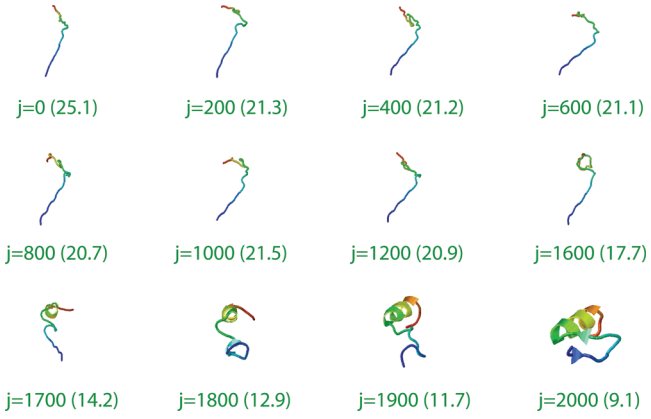


Figure 6. Snapshots for the conformational variation of path-6 are shown. See the caption of Figure 4 for additional details.

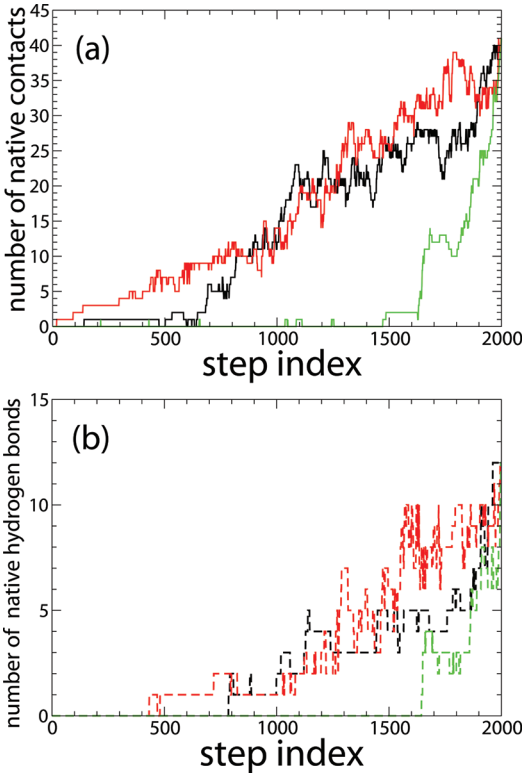


Figure 7. (a) The number of native contacts is shown for path-1 (black), path-5 (red), and path-6 (green). (b) The number of native hydrogen bonds is shown.

amide-hydrogen) is greater than 135°. For three folding models, path-1 (black), path-5 (red), and path-6 (green), the numbers of native hydrogen bonds and native contacts are highly correlated along the folding pathway with the linear correlation coefficients of r = 0.933, 0.946, and 0.977, respectively.

The folding pathway of FSD-1 suggested by the current ADMD simulations is in contrast with the framework model,[27] where secondary structures form first and then the tertiary structure is organized by packing the already established secondary structure elements. The calculated folding pathway models from this study are consistent with the concurrent formation of tertiary structure (native contacts) and secondary structures.

dx.doi.org/10.1021/jp300074f | J. Phys. Chem. B 2012, 116, 6916−6922

One of our folding models, path-6 (green), is quite close to that of Wang and his co-workers.[4] They investigated the folding mechanism by replica exchange MD simulations using the AMBER ff03 force field.[28] The folding of FSD-1 is initiated by the hydrophobic collapse with formation of the C-terminal α-helix. The folding was shown to proceed with the formation of β-hairpin and further consolidation of the hydrophobic core.[4]

The folding model, path-1 (black), is rather similar to the computational model of Kim et al.[10] Authors first optimized a force field, so that four small proteins (two α-proteins, an $\alpha\beta$ protein, and a β-hairpin) can be folded by an identical optimized force field starting from random conformations. By using this potential for the simulation of FSD-1, Kim et al.[10] suggested that the folding is initiated by the collapse of the hydrophobic core and is followed by the folding of β-hairpin and α-helix.

Lei and Duan proposed that the folding of individual secondary structures occurs prior to the formation of the native hydrophobic core.[6] Such a framework-like model[27] was not observed in the current ADMD simulations. On the other hand, the dual folding pathways[5] suggested by Duan and his co-workers are consistent in spirit with the current study. In their study, two pathways are characterized by the initial formation of either α-helix or β-hairpin. The authors[5] managed to fold FSD-1 using a force field with proper balance between α-helix and β-sheet. In contrast, it should be noted that, in the current study by ADMD, multiple folding pathways are observed without potential energy optimization.

Understanding ADMD Trajectories. Multiple folding pathways for FSD-1 are obtained by ADMD simulations. This is achieved not by applying the procedure of potential energy function optimization to stabilize the native structure under consideration or to balance the propensity between secondary structure elements, α-helix and β-hairpin. As clarified by many other authors, a successful force field optimization to fold α/β proteins does not warrant the general description of the protein folding problem.[29]

Interestingly, the folding mechanism elucidated from the current ADMD simulation is not far from the results obtained by force field optimized MD simulations. This is mainly due to the fact that temporal ordering of the structural elements of FSD-1 in the simulation is not necessarily progressive toward the native structure.

Generally speaking, the folding trajectory (kinetic information) cannot be obtained by the free energy landscape (thermodynamic information) calculation. The popular replica exchange molecular dynamics method (REMD) using generalized-ensemble algorithms can help a system to explore the free energy landscape more efficiently than conventional methods by allowing frequent escapes from local energy minima. However, this kind of method does not provide kinetic information of the folding pathway. Even worse is that sampled conformations of REMD simulations of FSD-1[4,11] are often populated away from the experimental NMR structure and the most populated conformations were centered around 2.5−4.0 Å rmsd away from the native structure. This finding may indicate that force fields used in those simulations may still need to be improved.

In the current ADMD simulation, one of the folding pathway models, path-6 (green) is in good agreement with the initial formation of the C-terminal α-helix,[4] observed by REMD simulations. In the current study, however, we find that the folding process is not necessarily limited to the model of the initial formation of the C-terminal α-helix.[4] Other possibilities are also found by direct search of the folding pathway. In general, through the equilibrium calculations such as REMD simulations, one cannot describe kinetic folding processes of a protein. This is partly due to the fact that the utilized reaction coordinates do not generally cover all possible folding pathways in practice. We observe that hydrophobic collapse usually occurs first which is followed by either sequential formation or correlated formation of secondary structures, as represented by path-1 (black) and path-5 (red). On the other hand, in path-6 (green), collapse and reduction of rmsd occur concurrently, while α-helix formation occurs at a later stage and β-hairpin is formed at the final stage of the folding.

Finally, we find that positional fluctuation of β-hairpin residues (3−13), in all 10 pathway models, is higher than that of the helical residues (15−24), which is consistent with the computational results based on unfolding simulations.[6]

Assessment of ADMD Trajectory. The ADMD method employed in this work provides us with the global nature of the protein folding kinetics, neglecting short time-scale events within and about $\Delta = 36.28$ fs. Note that the time increment of $\Delta = 36.28$ fs is much larger than that used in conventional MD simulations, where the time increment should be kept rather small (e.g., 1−2 fs) in order to maintain the properties of the Verlet trajectory.

In this section, we perform probabilistic analysis on the calculated ADMD trajectory by measuring the amount of its deviation from the Verlet trajectory formula. Deviation of an ADMD trajectory from its ideal Verlet trajectory can be estimated by the following quantity, "error variables" in a vector form,

$$\varepsilon_{I,j} = 2\mathbf{q}_{I,j} - \mathbf{q}_{I,j-1} - \mathbf{q}_{I,j+1} - \frac{\Delta^2}{m_I}\frac{\partial V(\{\mathbf{q}_j\})}{\partial \mathbf{q}_{I,j}} \quad (6)$$

which is the argument of the Onsager−Machlup action. There are $3N(P-1) = 3\,022\,488$ scalar "error variables" for each pathway representation. Since the distribution of "error variables" measures the overall deviation of an MD trajectory from its ideal Verlet trajectory, we can use it as a measure to assess the quality of the path. Recently, Elber et al.[15] also pointed out that the distribution of "error variables" represents the nature of using a discrete time step for MD simulations.

The distribution of "error variables" calculated for path-1 is shown in Figure 8, which is quite satisfactory. Obviously, with a sufficiently small value of time step Δ, generated ADMD trajectories will closely follow an ideal Verlet trajectory.[18]

To check the stability of the generated trajectories, we have repeated the ADMD simulation with new input trajectories prepared with keeping about 10% of the full degrees of freedom (3 022 488) associated with "low frequency" atomic motions. The rest of the degrees of freedom are set to zero, neglecting the "high-frequency" atomic motions in the trajectories. We find that the recalculated trajectories are virtually unchanged from the old ones for all 10 protein folding models. This indicates that the pathways generated in this study are stable against high-frequency perturbation, representing a quasi minimum-energy pathway.

CONCLUSIONS

We have used the ADMD simulation method to investigate folding events of the full-size designed sequence of mini-protein FSD-1. Multiple folding pathways are observed consistent with

dx.doi.org/10.1021/jp300074f | J. Phys. Chem. B 2012, 116, 6916−6922

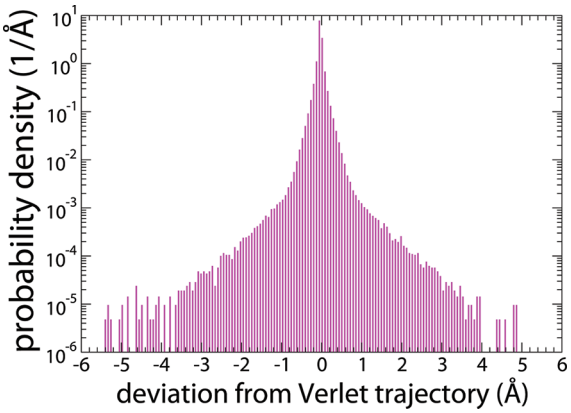


Figure 8. The distribution of "error variables" measuring the amount of deviation of an MD trajectory from its ideal Verlet trajectory is shown for path-1. Even with the relatively large time step of Δ = 36.28 fs, the probability for large deviation is shown to be low.

recent computational studies.[5] Hydrophobic collapse occurs first, and then, the folding proceeds with the formation of either α-helix or β-hairpin. Concurrent formation of the full tertiary structure and the secondary structure elements of α-helix and β-hairpin is observed.

Although we have utilized a completely different computational method in spirit relative to conventional MD methods without force field optimization, we were able to account for experimental observations on the protein folding process of FSD-1. This is due to the advantage of the ADMD method for studying pathways of rare events. By reformulating an initial value problem into a boundary value problem, we were able to explore pathways of the folding kinetics by smoothly bridging initial and final states. The folding pathway model generated by ADMD simulations may still depend on the quality of the force field employed. However, potential negative effects due to force field inaccuracy on the protein folding trajectory are rather limited in the ADMD formalism than in conventional MD simulations. Errors in conventional MD trajectories tend to become large due to the cumulative effects of folding simulations. In contrast, errors do not accumulate as much in ADMD simulations thanks to the formalism of solving boundary value problems.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jlee@kias.re.kr.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by Creative Research Initiatives (Center for *In Silico* Protein Science, 2009-0063610) of MEST/KOSEF. We thank Korea Institute for Advanced Study for providing computing resources (KIAS Center for Advanced Computation) for this work.

REFERENCES

(1) Dahiyat, B. I.; Mayo, S. L. *Science* **1997**, *278*, 82−87. DeGrado, W. F. *Science* **1997**, *278*, 80−81.
(2) Feng, J. A.; Kao, J.; Marshall, G. R. *Biophys. J.* **2009**, *97*, 2803−2810.
(3) Wu, C.; Shea, J.-E. *PLoS Comput. Biol.* **2010**, *6*, e1000998.
(4) Li, W. F.; Zhang, J.; Wang, W. *Proteins* **2007**, *67*, 338−349.
(5) Lei, H.; Wang, Z.-X.; Wu, C.; Duan, Y. *J. Chem. Phys.* **2009**, *131*, 165105.
(6) Lei, H.; Duan, Y. *J. Chem. Phys.* **2004**, *121*, 12104−12111.
(7) MacKerell, A. D., Jr.; Brooks, B.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In *CHARMM: The Encyclopedia of Computational Chemistry*; Schleyer, P. R., Ed.; John Wiley & Sons: New York, 1998; Vol. *1*, p 271.
(8) Jang, S.; Shin, S.; Pak, Y. *J. Am. Chem. Soc.* **2002**, *124*, 4976−4977.
(9) Kim, E.; Jang, S.; Pak, Y. *J. Chem. Phys.* **2009**, *131*, 195102.
(10) Kim, S.-Y.; Lee, J.; Lee, J. *Biophys. Chem.* **2005**, *115*, 195−200.
(11) Mohanty, S.; Hansmann, U. H. E. *J. Chem. Phys.* **2007**, *127*, 035102.
(12) Sipple, M. J.; Nemethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1984**, *88*, 6231−6233.
(13) Passerone, D.; Parrinello, M. *Phys. Rev. Lett.* **2001**, *87*, 108302.
(14) Lee, I.-H.; Lee, J.; Lee, S. *Phys. Rev. B* **2003**, *68*, 064303.
(15) Elber, R.; Cárdenas, A.; Ghosh, A.; Stern, H. A. *Adv. Chem. Phys.* **2003**, *126*, 93−129.
(16) Lee, I.-H.; Kim, S.-Y.; Lee, J. *Chem. Phys. Lett.* **2005**, *412*, 307−312.
(17) Lee, I.-H.; Kim, S.-Y.; Lee, J. *Comput. Chem.* **2010**, *31*, 57−65.
(18) Verlet, L. *Phys. Rev.* **1967**, *159*, 98−103.
(19) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005−3014.
(20) Ponder, J. W.; Richard, F. M. *J. Comput. Chem.* **1987**, *8*, 1016−1024 (http://dasher.wustl.edu/tinker).
(21) Kearsley, S. J. *J. Comput. Chem.* **1990**, *11*, 1187−1192.
(22) Coutsias, E. A.; Seok, C.; Dill, K. A. *J. Comput. Chem.* **2004**, *25*, 1849−1857.
(23) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291−318.
(24) Jónsson, H.; Mills, G.; Jacobsen, K. W. Nudged elastic band method for ending minimum energy paths. In *Classical and Quantum Dynamics in Condensed Phase Simulations*; Berne, G., Ciccotti, G., Coker, D., Eds.; World Scientific: Singapore, 1998; pp 385−404.
(25) E, W.; Ren, W.; Vanden-Eijnden, E. *Phys. Rev. B* **2002**, *66*, 052301.
(26) Faradjian, A.; Elber, R. *J. Chem. Phys.* **2004**, *120*, 10880−10889.
(27) Kim, P. S.; Baldwin, R. L. *Annu. Rev. Biochem.* **1990**, *59*, 631−660.
(28) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(29) Lei, H.; Dastidar, S. G.; Duan, Y. *J. Phys. Chem. B* **2006**, *110*, 22001−22008.