# Exploring the Folding Free Energy Landscape of a β-Hairpin Miniprotein, Chignolin, Using Multiscale Free Energy Landscape Calculation Method

Ryuhei Harada[†,§] and Akio Kitao[*,‡,§]

[†]Department of Physics, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan
[‡]Institute of Molecular and Cellular Bioscience, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo, 113-0032, Japan
[§]Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Tokyo, Japan

ABSTRACT: The folding process for a β-hairpin miniprotein, chignolin, was investigated by free energy landscape (FEL) calculations using the recently proposed multiscale free energy landscape calculation method (MSFEL). First, coarse-grained molecular dynamics simulations searched a broad conformational space, then multiple independent, all-atom molecular dynamics simulations with explicit solvent determined the detailed local FEL using massively distributed computing. The combination of the two models enabled efficient calculation of the free energy landscapes. The MSFEL analysis showed that chignolin has an intermediate state as well as a misfolded state. The folding process is initiated by the formation of a β-hairpin turn, followed by the formation of contacts in the hydrophobic core between Tyr2 and Trp9. Furthermore, mutation of Tyr2 shifts the population to the misfolded conformation. The results indicate that the hydrophobic core plays an important role in stabilizing the native state of chignolin.
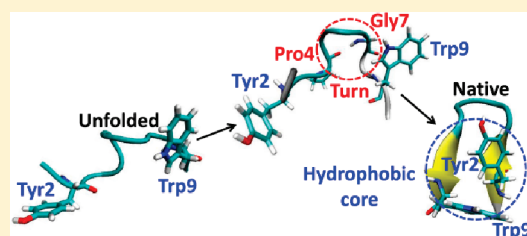
1. INTRODUCTION

Understanding protein-folding mechanisms is an attractive and challenging problem in physical chemistry. It is also of pharmaceutical importance because many diseases, such as Huntington's and Parkinson's diseases, are thought to be caused by misfolded proteins. Experimental and theoretical studies have revealed many kinetic and thermodynamic aspects of the folding process.[1−3] Recent increases in computer power have allowed molecular dynamics (MD) simulations with explicit solvent to simulate the unfolded-to-folded-state process on the order of microseconds in the case of rapidly folding miniproteins such as chignolin (10 residues),[4] Trp-cage (20 residues),[5,6] and villin headpiece subdomain (35 residues).[7−9] By tracing the trajectories, the dynamic conformational changes in the folding process can be directly detected. This process can be characterized by the free energy landscape (FEL), calculated by MD simulation. The FEL is an important thermodynamic quantity that indicates possible folding pathways as well as intermediate and misfolded states in atomic-level detail. The folding FEL has numerous local energy minimum states, and a high free energy barrier can separate each local minimum state. Therefore, systems simulated using conventional MD (CMD) are often trapped for some time in a local energy minimum state near the initial conformation. This is the so-called sampling problem in FEL calculations and increases in severity as the degrees of freedom of the system increase.

One way to overcome the sampling problem is coarse-graining of the system.[10] Some coarse-grained (CG) models simplify the atoms of each amino acid as one bead.[11−13] The coarse-grained models have been widely used in folding studies and examined if the obtained results are consistent with experiment. A typical example is a so-called Go-like model designed to give low energy value if native contact is formed and to reach the global minimum if all the native contacts are formed.[14−19] Intermediate and denatured states deduced from the Go-like models have been shown to reproduce the experimental results of the Φ-value analysis.[20−22] Therefore, this type of CG models is expected to allow sampling of a broad conformational space, including intermediate and unfolded conformations. Although CG models are suitable for efficient conformation sampling compared with full-atom CMD simulations,[23,24] problems arise from the coarse graining: the reduction in atomic detail leads to low accuracy in the FEL and results that differ from those obtained using all-atom (AA) models.

To address this problem, we recently proposed the multiscale free energy landscape calculation method (MSFEL) using both CG and AA models.[25] This multiscale approach combines the accuracy of the AA model and the efficiency of the CG model. The MSFEL consists of the following four steps: In step 1, a wide conformational space is sampled by MD simulations with the CG model (CG-MD). A C_α-based Go-like model is employed because the Go-like models are expected to allow broad conformational sampling, including intermediate and unfolding conformations, as mentioned above. However, it is still a difficult question whether CG-MD can sample sufficient conformational space. In step 2, multiple structures in the CG model are selected, and AA model structures are generated. In step 3, independent

MD simulations with the AA model (AA-MD) are performed. In step 4, the FEL is calculated from AA trajectories on the basis of the weighted histogram analysis method (WHAM).[26−28] The most time-consuming part of the MSFEL is the independent MD simulation in step 3, which is suitable for distributed computing. Relatively short MD simulations in this step can be done in parallel on distributed computational resources available without communication during MD.

In this study, the folding FEL of a β-hairpin miniprotein, chignolin (sequence: GYDPETGTWG), in explicit water is explored by the MSFEL. Chignolin is an artificially designed miniprotein consisting of 10 residues (138 atoms). This molecule folds into a stable β-hairpin structure; the structure has been determined by nuclear magnetic resonance (NMR).[29] Similar β-hairpin structures compose parts of larger proteins;[30−32] therefore, chignolin is a good target for folding studies and as a model for the folding of protein nuclei.

2. COMPUTATIONAL METHODS

2.1. Step 1: CG-MD Simulation. CG-MD simulations allow a large conformational space to be sampled at relatively low computational cost compared with AA simulations. A C_α-based CG model was used to sample conformational space around the reference structure, thereby ensuring that the global energy minimum will be similar to the model presented by Higo and Umeyama.[33] The potential energy function was defined as

$$V^{CG}\left(\vec{r}^{C_\alpha}\,|\,\vec{r}^{C_\alpha 0}\right) = \sum_{|i-j|=1,\,i<j} k_{12}\left(r_{ij}^{C_\alpha} - r_{ij}^{C_\alpha 0}\right)^2$$
$$+ \sum_{|i-j|=2,\,i<j} k_{13}\left(r_{ij}^{C_\alpha} - r_{ij}^{C_\alpha 0}\right)^2$$
$$+ \sum_{|i-j|=3,\,i<j} k_{14}\left(r_{ij}^{C_\alpha} - r_{ij}^{C_\alpha 0}\right)^2$$
$$+ \sum_{|r_{ij}|<r_c,\,|i-j|>3} k_{LJ}\left(\left(\frac{r_{ij}^{C_\alpha 0}}{r_{ij}^{C_\alpha}}\right)^{12} - \left(\frac{r_{ij}^{C_\alpha 0}}{r_{ij}^{C_\alpha}}\right)^6\right)$$
$$(1)$$

where $r_{ij}^{C_\alpha 0}$ and $r_{ij}^{C_\alpha}$ represent the distance between the ith and jth atoms of the reference and instantaneous structures, respectively. k_{12}, k_{13}, and k_{14} indicate the force constants for bond stretching, angle bending, and dihedral angle bending interactions, respectively. These interactions are local interactions that maintain the geometry of the C_α atoms of the main chain. The last term is a Lennard-Jones-type interaction considered only for the atom pairs making a nonlocal native contact within the cutoff radius r_C in the reference structure ($r_C = 6.0$ Å). One of the NMR structures[29] is used as the reference structure for the CG-potential calculation. The force constants k_{12}, k_{13}, k_{14}, and k_{LJ} are determined so as to satisfy the geometric conditions with respect to the C_α atoms of the main chain. The parameter k_{12} is first determined by the distribution of the distance between adjacent C_α atoms in the AA model; this distribution is approximately Gaussian, with a peak around 3.8 Å. k_{12} is determined to best reproduce this distribution, with the relation between the variance of the Gaussian being $k_{12} = 1/\beta\sigma^2$. The other parameters are determined by their ratios to k_{12} using $k_{13}/k_{12} = 1/5$ and $k_{14}/k_{12} = k_{LJ}/k_{12} = 1/100$.

Replica exchange molecular dynamics (REMD)[34] simulations were performed to enhance conformational space sampling. REMD simulations using the CG potentials were performed

with replicas at 10 temperatures: 200, 239, 286, 342, 404, 489, 585, 700, 853, and 1041 K. A 10^6-step production CG-MD run after a 10^3-step equilibration was performed under a canonical ensemble at each temperature with temperature controlled by a chain of two Nosé−Hoover thermostats[35] for various values of the masses $Q1 = 10$ g/mol and $Q2 = 13$ g/mol. The two replicas with neighboring temperatures were exchanged every 10^3 steps, and in total, 10^4 snapshots were recorded every 100 steps.

2.2. Step 2: Mapping from the CG Model to the AA Model. Multiple representative structures were chosen (M: the number of structures) from the CG-MD simulation results. These structures should broadly cover the conformational space sampled in the CG-MD simulations and be distributed densely enough so that each AA-MD trajectory in step 3 significantly overlaps with neighboring trajectories. From each trajectory that contains 10^4 snapshots of the C_α coordinates, 10 snapshots are selected at every 10^3 intervals. The total number of the selected C_α coordinate sets is 100 (10 snapshots × 10 replicas). The structures obtained from the CG model were used as inputs for BBQ[36] to generate the main-chain atoms (N, C′, and O) from the C_α coordinates. These were followed by SCWRL[37] to generate the side-chain atoms of chignolin. One hundred AA structures were constructed in total.

2.3. Step 3: Independent AA-MD Simulation. This step sampled local conformational space more accurately than the CG-simulation and can be conducted completely in parallel using noninteractive MD simulations. The 100 AA structures were solvated in a rectangular box (36.5 Å × 38.7 Å × 31.5 Å) containing 1421 TIP3P water molecules.[38] Two sodium ions were added to neutralize the system. The AA-MD simulations were independently performed using the PMEMD module of the Amber 9.0 software[39] with the Amber parm99SB force field.[40] The umbrella sampling method[41,42] was used for intensive local sampling. In addition, harmonic positional restraints for C_α atoms were used for the umbrella potential $V_i^{umbrella}$ for the ith simulation, as given by

$$V_i^{umbrella} = k_i\left(\vec{r}^{C_\alpha} - \vec{r}_i^{C_\alpha 0}\right)^2 \qquad (i = 1, 2, ..., M) \qquad (2)$$

where $\vec{r}_i^{C_\alpha 0}$ and $\vec{r}^{C_\alpha}$ are the reference and instantaneous coordinates of the C_α atoms, respectively, represented as $3N$-dimensional vectors (N: the number of C_α atoms), and k_i is the force constant for the ith reference structure. The original ith unbiased probability distribution $\rho_i^{unbiased}$ can be reweighted by the biased probability distribution ρ_i^{biased}.

After the CG-to-AA mappings, we performed energy minimizations for the initial structures, including water molecules, and relaxed them by short time (200 ps) MD simulations with the molecular force field that included density adjustment with an isothermal−isobaric ensemble at 300 K and 1 bar using the Berendsen method[43] with a relaxation time of 1.0 ps. Then we used the relaxed structures as references. As results of these treatments, the AA-structures have been modeled with the potential issue. The systems were equilibrated with a canonical ensemble for 100 ps at 300 K with harmonic restraints imposed on the C_α atoms (except for the N- and C-terminal residues) for the umbrella samplings ($k_i = 1.0 \times 10^{-4}$ kcal/mol/Å^2 to intensely sample local FEL around the generated initial structures. Production runs were performed for 1 ns ×100 trajectories, and each trajectory was recorded every 0.5 ps. In both the equilibration and production runs, the temperature was maintained at 300 K, and the SHAKE algorithm[44] was used to constrain the bond lengths involving hydrogen atoms and to enable the use of a long (2 fs) time step. The electrostatic interactions were calculated
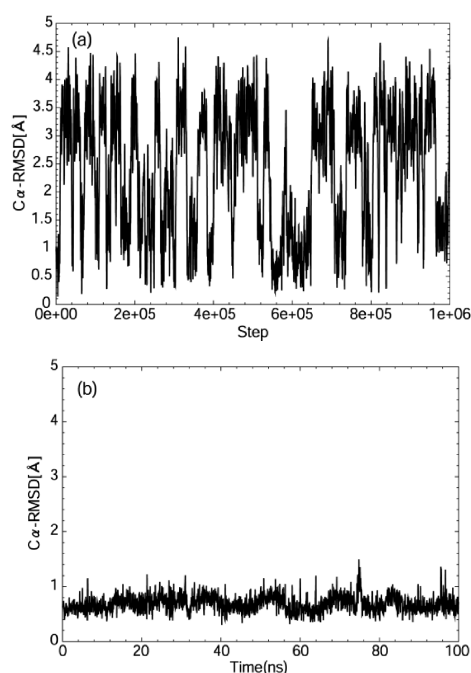
Figure 1. (a) The time series of the C_α rmsd of the CG-MD simulation using the REMD (replica 1) from the NMR structure. (b) As above, but for the 100 ns, full-atom CMD started from the native conformation.

using the particle mesh Ewald method, with a real space cutoff distance of 9 Å.[45]

2.4. Step 4: Free Energy Calculation. The probability distributions obtained in step 3 were reweighted and combined with the WHAM.[26−28] In the WHAM, the probability distributions are summed up by weighting factors, and the combined probability distribution, ρ, is represented by a linear combination of each probability distribution. The weighting factors are determined so as to minimize the statistical error using a normalization condition.

3. RESULTS AND DISCUSSION

3.1. Efficiency of CG-MD. To examine conformational sampling efficiency by the CG-MD simulation, we calculated the root-mean-square deviation (rmsd) of the C_α atoms from model 1 of the NMR structures.[29] Since the positions of the N- and C-terminal residues (Gly1 and Gly10) were poorly converged in the NMR structures, only residues 2−9 were considered. Figure 1a shows an example of the time series of the rmsd in the CG-MD simulation using the REMD (the first replica). Compared with the time series of a long, full-atom CMD simulation started from the native structure (Figure 1b), folding and unfolding events were frequently observed in the CG-MD simulation. In Figure 1b, the system was trapped around the native conformation, and unfolding was not observed. The average acceptance ratio among all 10 replicas was 0.67, which is quite high compared with that of the conventional full-atom REMD simulation (∼0.2).

During the REMD simulation, the number of folding events from unfolded states (C_α rmsd > 4.0) to the "native" state (C_α rmsd < 1.0) averaged over all replicas was 23.2.

3.2. CG-AA Mapping. The structures generated with the CG-MD simulation were mapped onto the AA model, as described in Computational Methods. To examine the accuracy of CG-AA
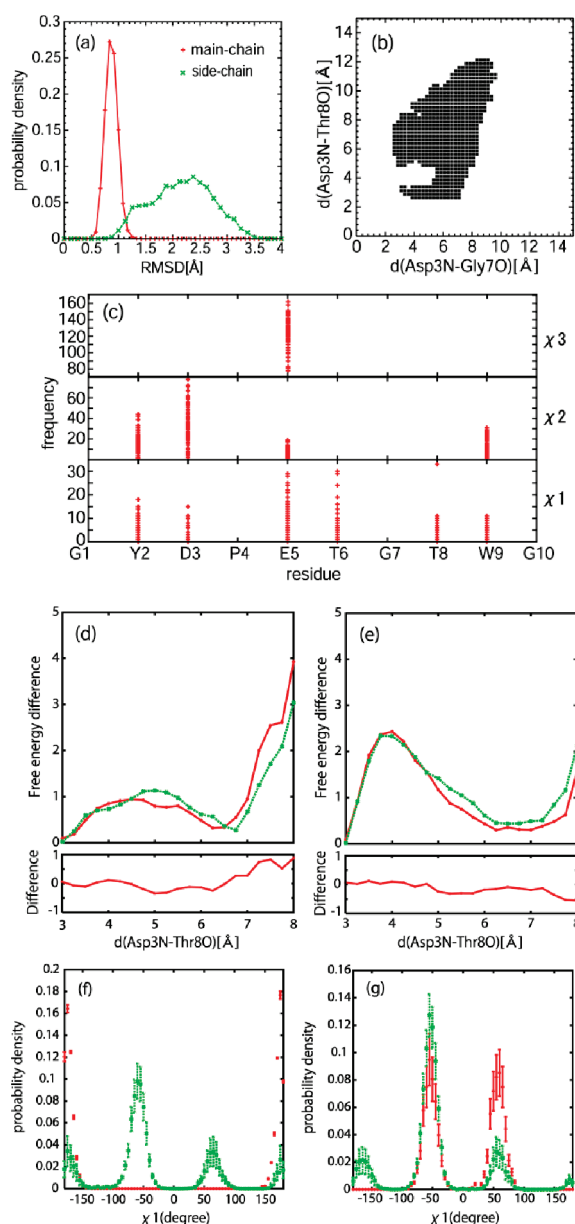

Figure 2. (a) The distributions of the rmsd between the reconstructed and original structures in the AA-CG mappings. (b) Overlapping regions among 100 distinct AA-MD trajectories depicted by solid rectangles on the projected subspace. The overlap is counted if at least two distinct trajectories visit the rectangle. The size of each rectangle is 0.25 Å × 0.25 Å. (c) Transition frequencies of chignolin side-chain dihedral angles ($\chi_1-\chi_3$) of each amino acid residue in a hundred 1 ns AA-MD simulations. (d, e) Comparison of one-dimensional FELs along the donor−acceptor distances (3N−7O, 3N−8O) calculated from the first (500 ps, red line) and second halves (500 ps, green line) of the 1 ns MD trajectories. The differences between two curves are shown in the lower panels. The unit of the ordinates is k_BT at T = 300 K. (f, g) The distributions of side-chain dihedral angles (χ_1) of Tyr2 and Trp9, respectively, with error bars. The red and green lines correspond to the native and denatured conformations, respectively.

mappings by BBQ (Backbone Building from Quadrilaterals program) and SCWRL (Side-Chains With a Rotamer Library), we prepared a test set of AA structures. We first prepared 1×10^4 AA structures by AA-MD simulations in explicit water, which contains the structures near the native state and the nonnative conformations.

This test set is considered as the answers of the CG-AA mappings. Then we extracted C_α coordinates from the test AA structures. These CG models were mapped into the AA models using BBQ and SCWRL. We calculated the rmsd between the original AA structures and reconstructed AA structures. The rmsd's were calculated using main- and side-chain heavy atoms and checked distributions of the rmsd shown in Figure 2a. The distribution of the main-chain atoms shows a sharp peak around the average (0.86 Å), with a standard deviation of 0.11 Å, which is comparable to "ultrahigh resolution" in X-ray crystallography, typically thought to be <1.0 Å. The distribution in the side chains has a broader peak, with an average of 2.15 Å and a standard deviation of 0.55 Å, indicating that the candidate structures for the side-chain arrangements are not necessarily unique for a given main-chain arrangement, although they are approximated by the geometry of the C_α coordinates. Later, we will show that side chains sample various rotamers frequently, suggesting that initial side-chain mapping is not a major problem with the protein investigated in this study.

3.3. Validation of the MSFEL. To calculate the folding FEL using the WHAM, overlaps between trajectories of the independent AA-MD simulations are essential. Figure 2b shows the overlaps of the 100 trajectories from the AA-MD simulations: black solid boxes represent the overlaps between neighboring trajectories and the 100 AA-trajectories that sufficiently overlap in a subspace spanned by donor—acceptor distances of the main chain. The results demonstrate that the MSFEL can calculate the folding FEL in this subspace.

To assess the bias given by the starting configuration of each of the 1 ns simulations in the independent AA-MD simulations, we compared free energy values calculated from the first and second half of the trajectories. We calculated one-dimensional FELs projected onto the donor—acceptor distances (3N—7O and 3N—8O) as shown in Figure 2d, e and compared the differences. The differences are less than k_BT in the range shown in the figure and less than $0.35k_BT$ for the range <7 Å, which suggests the convergence of the FEL within a relatively short sampling time.

The efficiency of sampling the side-chain rotamers is also important in the folding FEL calculation. In this method, the initial side-chain coordinates are chosen from the optimal rotamer in the rotamer library. Therefore, we investigated the sampling efficiency of the side chains to check whether other rotamers are sampled during AA-MD simulations. To examine the dependence of the efficiency of sampling on the choice of the initial side-chain rotamer, we counted the transition frequency among side-chain rotamers during 1 ns AA-MD simulations. Figure 2c shows the transition frequencies of rotamers for each residue in chignolin. Highly exposed side chains, such as Asp3, Glu5, and Thr6, show very high transition frequencies, and even the well-packed side chains of Tyr2 and Trp9 exhibit transition frequencies sufficient to allow sampling of other rotameric states. This result suggests that the calculated FEL will likely have a very weak dependence on the choice of the initial side-chain rotamers in AA-MD.

In addition to the transition frequencies, we also examined the distributions of all torsion angles on side chains. Here, we focused on distributions of two hydrophobic residues (Tyr2 and Trp9) that are important for the folding. We selected two sets of AA-MD trajectories that satisfy the two rmsd conditions: the near-native set whose C_α rmsd from the native structure is <1.0 Å and the denatured set with a C_α rmsd of >4.0 Å. Figure 2f, g are distributions of χ_1 near the native state (C_α rmsd < 1.0 Å) and the denatured state (C_α rmsd > 4.0 Å) with respect to Tyr2 and Trp9, respectively. The distributions indicate more frequent
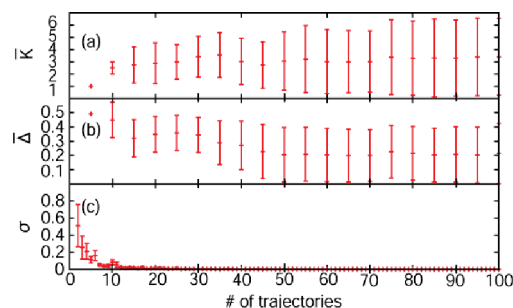

Figure 3. AA trajectory overlap and FEL convergence as a function of the number of trajectories considered, n. (a) The average number of overlapping trajectories per trajectory, . (b) Average fraction of overlapped snapshots between pairs of overlapping trajectories, $\overline{K}$. (c) The convergence of the probability density projected onto the subspace.

rotamer transitions for the side chains on the denatured conformations compared with the native ones. In the native conformations, the ranges of the torsion angles are limited because of the well-packed side-chain arrangements, which cause the sharper distributions around the native conformations compared with the broad distributions of the denatured cases.

We also calculated the number of overlapping trajectories per trajectory, K. A pair trajectories is regarded as overlapped if the C_α rmsd between the two average conformations is smaller than 1.0 Å. For each pair of overlapping trajectories, comparisons among all snapshots are made, and the fraction of overlap, Δ, is estimated with the same rmsd criterion. The results show that one trajectory overlaps with $\overline{K} = 3.4 \pm 3.2$ trajectories, and = $31.4 \pm 21.1\%$ of the snapshots overlap in each pair of overlapping trajectories on average. We also confirmed that there is no isolated trajectory and that all the snapshots are connected in conformational space. $\overline{K}$ and were also calculated as a function of the number of trajectories considered, n (Figure 3ab). $\overline{K}$ values were found to rapidly increase from 0 to 20, but the rate of increase slowed by $n = 20$ or higher. values essentially converge when 20 or more trajectories are considered. We also examined the convergence of probability distributions projected onto the subspace spanned by the above reaction coordinates by

$$\sigma = \frac{\int d\xi \left[\rho^{n+1}(\xi) - \rho^n(\xi)\right]^2}{\int d\xi \left[\rho^n(\xi)\right]^2} \quad (3)$$

As shown in Figure 3c, σ essentially converges at around $n = 20$, corresponding to the convergence of trajectory overlaps.

3.4. Folding Free Energy Landscape of Chignolin. It was earlier reported[46] that three main-chain hydrogen bonds, HB1, HB2, and HB3, play an important role in the folding process of chignolin. HB1 is formed in the native state between the 3N of Asp3 and the 8O of Thr8, HB2 is formed in the misfolded state between the 3N of Asp3 and the 7O of Gly7, and HB3 is formed in both the native and misfolded states between the 3O of Asp3 and the 7N of Gly7. These hydrogen bonds are shown in Figure 4a1 and b1. HB3 is considered to be an important trigger for the folding process. Therefore, we calculated the folding FEL projected onto the donor—acceptor distances of HB1 (native) or HB2 (misfolded) under conditions compatible with the formation of HB3.
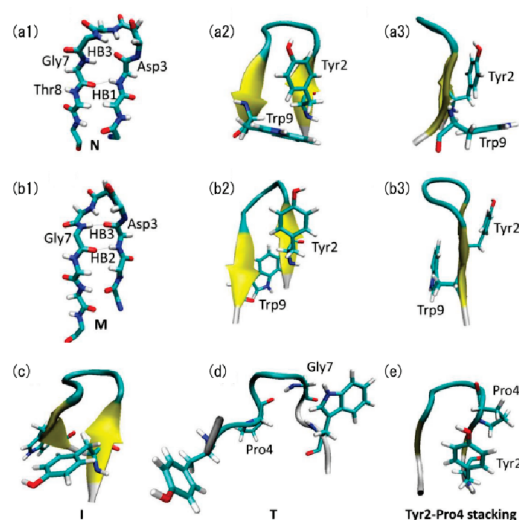
Figure 4. The representative conformations of chignolin. (a1, a2) The native state (N) and (a3) its side view, (b1, b2) misfolded states (M) and (b3) its side view, (c) intermediate state (I), (d) turn (T) and (e) Tyr2-Pro4 stacking conformation, respectively. The side chain of the hydrophobic residues and Gly7 are drawn with the licorice rendering.

The native, misfolded, and intermediate states were identified in the folding FEL of chignolin shown in Figure 5a. In the native state shown in Figure 4a1, both HB1 and HB3 are formed. The hydrophobic core is formed mainly by "edge-to-face" hydrophobic interactions between Tyr2 and Trp9, shown in Figure 4a2. This observation showed good correlation with experimental data. The misfolded state with HB2 and HB3 (Figure 4b1) has a side-chain arrangement different from that of Tyr2 and Trp9 (Figure 4b2): these side chains protrude from different sides of the β-sheet and, therefore, cannot contact directly. In addition to the two stable conformations, a novel intermediate state was identified by the MSFEL. The intermediate conformation shown in Figure 6c is characterized by the locations of Tyr2 and Trp9, with Tyr2 blocking the formation of HB1 and HB2. Compared with the folding FEL determined using the implicit water model,[46] the overstabilization of the misfolded state is suppressed in the above simulation, demonstrating the improvements obtained using explicit solvent.

3.5. Folding Process of Chignolin. To address the process by which chignolin folds, we also calculated the folding FEL projected onto the fraction of native contacts (NC) and the radius of gyration (Rg), shown in Figure 6a The folding FEL has two basins of attraction located at 0.4 < NC < 0.6 and 0.8 < NC < 1.0; the former was identified as an intermediate state, and the latter, as the native state. In the first stage of the folding process, chignolin drops into the first basin of attraction, in which a native β-turn has already been formed by Pro4, Glu5, Thr6, and Gly7 (a representative conformation is shown in Figure 4d). Next, chignolin moves to the second basin and folds into the native state by forming hydrophobic interactions between Tyr2 and Trp9. Therefore, the folding process of this β-hairpin consists of a two-step mechanism: β-turn formation and nucleation of the hydrophobic core. To further investigate this two-step mechanism, we redefined the reaction coordinates for NC1 and NC2 using the fraction of the native contacts of the β-turn (Pro4 and Gly7) and the hydrophobic core (Tyr2 and Trp9), respectively. Figure 5b shows the folding FEL projected onto the subspace defined by NC1 and NC2. The new folding FEL clearly indicates the folding pathway
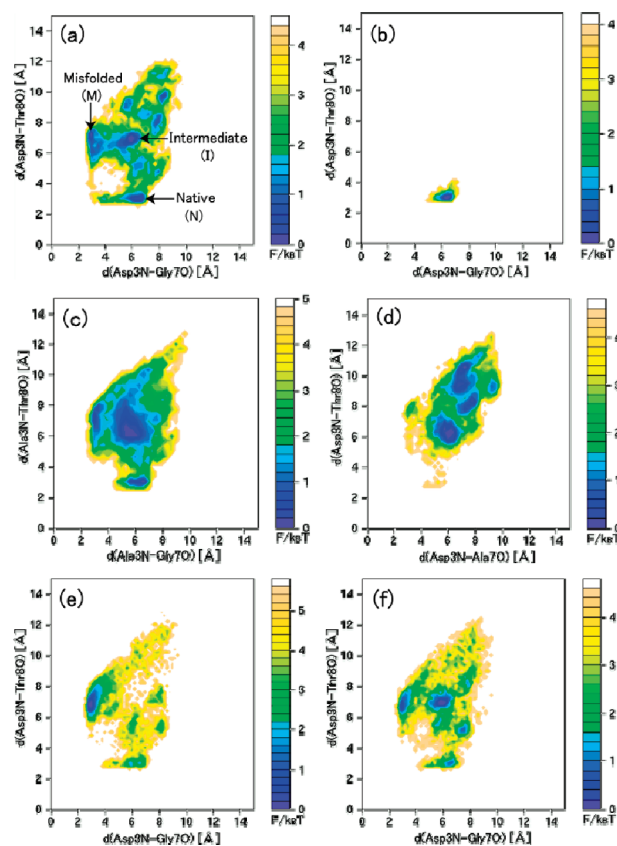

Figure 5. (a) The folding FEL of chignolin at 300 K projected onto the donor and acceptor distances in hydrogen bonds HB1 (ordinate) and HB2 (abscissa), under conditions that HB3 is formed. As above, but (b) a 100-ns all-atom CMD simulation, (c) mutant D3A, (d) G7A, (e) Y2A, and (f) W9A by the MSFEL (1 ns ×100 runs). A hydrogen bond is counted if the distance between two heavy atoms (N and O in this case) is <3.0 Å and the N−H···O angle is larger than 160°.

described above. The formation of the turn occurs in the first folding process from NC2 = 0 to NC2 = 1, followed by the formation of the hydrophobic core from NC1 = 0 to NC1 = 1. The basin of attraction around the native conformation (NC1 = 1 and NC2 = 1) is somewhat unstable due to fluctuations in the stacking of the side chains of the hydrophobic core. The results of this folding FEL analysis indicate the two-step folding process as shown in Figure 7. The turn region of chignolin first is formed to initiate the formations of the main-chain hydrogen bonds, and then the overall conformation is adjusted by the hydrophobic residue contacts between Tyr2 and Trp9. This process is similar to the folding events observed in 1.8 μs MD:[47] the main-chain hydrogen bonds form before the side chain of the hydrophobic residues contacts do. Although hydrophobic interaction between Tyr2 and Trp9 is important for the stabilizing the native conformation, it is not crucial for folding initiation consistent to the previous work.[47] This folding process is also similar to earlier reports of long-time-scale MD simulations of chignolin.[4]

Although the FEL analysis does not directly provide dynamic information, kinetics of the folding process can be roughly discussed using the free energy difference between the intermediate and native states. The free energy barrier between two states is estimated to be ∼2$k_B T$. Therefore, the transition from the intermediate to native states occurs relatively easily, consistent with the fast folding of this protein. The rate-limiting stage is the formation of the turn region
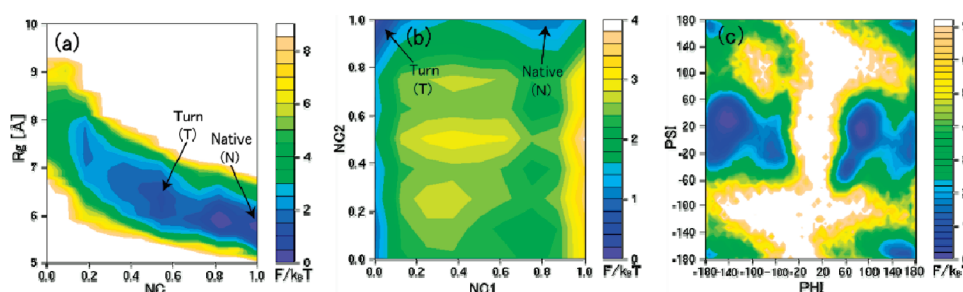
8810

dx.doi.org/10.1021/jp2008623 |*J. Phys. Chem. B* 2011, 115, 8806–8812

Figure 6. (a) The folding FEL projected onto the fraction of NC of each residue and the Rg. NC is counted when two C_α atoms are within 10 Å, and the Rg is calculated using non-hydrogen atoms. (b) The folding FEL projected onto the native contacts of the hydrophobic residues Tyr2—Trp9 (NC1), and Pro4—Gly7 located at the edge of the turn region (NC2). NC1 and NC2 are counted if two non-hydrogen atoms are located wihin 6.5 Å of each other. (c) The FEL projected onto the main-chain dihedral angles φ and φ of Gly7.
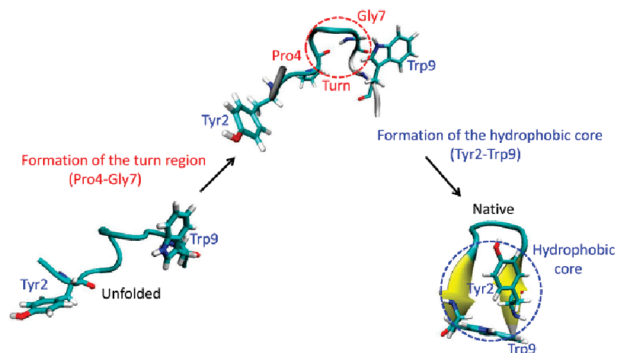

Figure 7. The presumed folding pathway of chignolin by the MSFEL.

from unfolded extend conformations whose energy barrier is $\sim 4k_BT$. Therefore, this process is considered to be a trigger of the folding.

3.6. Mutations of Trigger Residues. The folding FEL in Figure 5a indicates that the native state is characterized by the formation of HB1—HB3 and that the misfolded state is characterized by the formation of HB2—-HB3. The important hydrogen bond for triggering folding is HB3, between Asp3O and Gly7N. This hydrogen bond is necessary for folding to either the native or misfolded conformation.[46] To investigate the effect of mutations of the trigger on the folding process, the folding FEL of two mutants, D3A and G7A, were calculated and compared with the results of the wild-type: D3A is shown in Figure 5c, and G7A is shown in Figure 5d. Folding to the native or misfolded conformation is strongly affected by mutation G7A, whereas D3A shows only a slight effect. The position of the G7A global minimum is significantly shifted from the wild type. To examine the effect of G7A, we calculated the FEL of the wild type projected onto the dihedral angle space of the main-chain φ and φ angles of Gly7, as shown in Figure 6c. This FEL shows that Gly7 has several local energy minimum states, including a region prohibited to other (non-Gly) residues. This result indicates that the flexibility of Gly7 is a key in the folding pathway of chignolin and that G7A induces destabilization of the native conformation.

3.7. Mutations of the Hydrophobic Core. The relationship between stability and hydrophobic core formation in the folding process was investigated by mutating the hydrophobic residues Tyr2 and Trp9 to alanine, thus weakening the hydrophobic interactions. Figure 5e, f shows the folding FEL of Y2A and W9A, respectively. In both cases, the mutations destabilize the native conformation and stabilize the misfolded state, showing that the misfolded state is stabilized by weakened hydrophobic

interactions. As mentioned above, these hydrophobic residues are located on opposite sides of the β-hairpin (Figure 4b3), and these mutations do not strongly affect the stability of the misfolded conformation. The native state population of W9A is slightly higher than that of Y2A (Figure 5f). This is because the native conformation of chignolin is stabilized via an alternative hydrophobic interaction between Tyr2 and Pro4, as shown in Figure 4e. However, this new hydrophobic interaction is weak compared with that of the wild type, so this nativelike conformation is unstable. These results indicate that the folding process is significantly disrupted by breaking the hydrophobic interaction between Tyr2 and Trp9, leading to the formation of nonnative conformations.

■ CONCLUSIONS

In this study, the MSFEL was applied to the β-hairpin miniprotein, chignolin, solvated in explicit water, to investigate its folding mechanism. The MSFEL could sample several misfolded and intermediate conformations without being trapped in local minimum states, allowing calculation of the broad FEL and its comparison to a long (100 ns), full-atom CMD simulation. The folding FEL projected onto the fraction of native contacts with respect to the β-turn and hydrophobic core revealed a two-step mechanism: the folding process is initiated by the formation of the turn and is followed by contacts between hydrophobic residues Tyr2 and Trp9, resulting in folding into the native conformation. The FEL calculated for the mutation G7A showed that Gly7 is a key residue in stabilizing the native conformation because of its flexibility. Furthermore, the FEL calculated for the mutation Y2A in the hydrophobic core indicated that Tyr2 plays an important role in the folding and stabilizing of chignolin.

■ AUTHOR INFORMATION

Corresponding Author
*Phone: +81 3 5841-1472. Fax: +81 3 5841-2297. E-mail: kitao@iam.u-tokyo.ac.jp.

■ ACKNOWLEDGMENT

Research Center for Computational Scinece, Okazaki Reaserch Facilities, National Institite of Natural Science.

■ REFERENCES

(1) Fersht, A. R. *Curr. Opin. Struct. Biol.* **1997**, *7*, 3.

(2) Go, N. *Annu. Rev. Biophys. Bioeng.* **1983**, *12*, 183.

(3) Onuchic, J. N.; Wolynes, P. G. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70.

(4) Suenaga, A.; Narumi, T.; Futatsugi, N.; Yanai, R.; Ohno, Y.; Okimoto, N.; Taiji, M. *Chem—Asian J.* **2007**, *2*, 591.

(5) Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Duan, Y. *J. Mol. Biol.* **2003**, *327*, 711.

(6) Simmerling, C.; Strockbine, B.; Roitberg, A. E. *J. Am. Chem. Soc.* **2002**, *124*, 11258.

(7) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740.

(8) Duan, Y.; Wang, L.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9897.

(9) Freddolino, P. L.; Schulten, K. *Biophys. J.* **2009**, *97*, 2338.

(10) Tozzini, V. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144.

(11) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. *Biophys. J.* **2001**, *80*, 505.

(12) Bahar, I.; Atilgan, A. R.; Erman, B. *Fold. Des.* **1997**, *2*, 173.

(13) Tirion, M. M. *Phys. Rev. Lett.* **1996**, *77*, 1905.

(14) Honeycutt, J. D.; Thirumalai, D. *Biopolymers* **1992**, *32*, 695.

(15) Clementi, C.; Maritan, A.; Banavar, J. R. *Phys. Rev. Lett.* **1998**, *81*, 3287.

(16) Takada, S.; Luthey-Schulten, Z.; Wolynes, P. G. *J. Chem. Phys.* **1999**, *110*, 11616.

(17) Socci, N., D.; Onuchic, J. N.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 2031.

(18) Clementi, C.; Nymeyer, H.; Onuchic, J. N. *J. Mol. Biol.* **2000**, *298*, 937.

(19) Clementi, C.; Jennings, P. A.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5871.

(20) Plaxco, K. W.; Simons, K. T.; Baker, D. *J. Mol. Biol.* **1998**, *277*, 985.

(21) Fersht, A. R. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*; Freeman: New York, 1999.

(22) Koga, N.; Takada, S. *J. Mol. Biol.* **2000**, *313*, 171.

(23) Okazaki, K.; Koga, N.; Takada, S.; Onuchic, J. N.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11844.

(24) Okazaki, K.; Takada, S. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 11182.

(25) Harada, R.; Kitao, A. *Chem. Phys. Lett.* **2011**, *503*, 145.

(26) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195.

(27) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.

(28) Souaille, M.; Roux, B. *Comput. Phys. Commun.* **2001**, *135*, 40.

(29) Honda, S.; Yamasaki, K.; Sawada, Y.; Morii, H. *Structure* **2004**, *12*, 1507.

(30) Dyson, H. J.; Wright, P. E. *Curr. Opin. Struct. Biol.* **1993**, *3*, 60.

(31) Imperiali, B.; Ottesen, J. J. *J. Pept. Res.* **1999**, *54*, 177.

(32) Serrano, L. *Adv. Protein Chem.* **2000**, *53*, 49.

(33) Higo, J.; Umeyama, H. *Protein Eng.* **1997**, *10*, 373.

(34) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.

(35) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635.

(36) Gront, D.; Kmiecik, S.; Kolinski, A. *J. Comput. Chem.* **2007**, *28*, 1593.

(37) Canutescu, A. A.; Shelenkov, A. A.; Dunbrack, R. L. *Protein Sci.* **2003**, *12*, 2001.

(38) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(39) Case, D. A. *AMBER 9*; University of California: San Francisco, 2006.

(40) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712.

(41) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett.* **1974**, *28*, 578.

(42) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187.

(43) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

(44) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

(45) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.

(46) Satoh, D.; Shimizu, K.; Nakamura, S.; Terada, T. *FEBS Lett.* **2006**, *580*, 3422.

(47) Seibert, M. M.; Patriksson, A.; Hess, B.; van der Spoel, D. *J. Mol. Biol.* **2005**, *354*, 173.

8812

dx.doi.org/10.1021/jp2008623 |*J. Phys. Chem. B* 2011, 115, 8806–8812