PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Understanding the roles of amino acid residues in tertiary structure formation of chignolin by using molecular dynamics simulation

Tohru Terada,[1]* Daisuke Satoh,[1,2] Tsutomu Mikawa,[3,4] Yutaka Ito,[5] and Kentaro Shimizu[1]

[1] Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan
[2] Intelligent Modeling Laboratory, The University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-8656, Japan
[3] International Graduate School of Arts and Sciences, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan
[4] RIKEN Harima Institute, 1-1-1 Kouto, Sayo-cho, Sayo-gun, Hyogo 679-5148, Japan
[5] Department of Chemistry, Tokyo Metropolitan University, 1-1 Minami-Osawa, Hachioji, Tokyo 192-0397, Japan

ABSTRACT

Chignolin is a 10-residue peptide (GYDPETGTWG) that forms a stable β-hairpin structure in water. However, its design template, GPM12 (GYDDATKTFG), does not have a specific structure. To clarify which amino acids give it the ability to form the β-hairpin structure, we calculated the folding free-energy landscapes of chignolin, GPM12, and their chimeric peptides using multicanonical molecular dynamics (MD) simulation. Cluster analysis of the conformational ensembles revealed that the native structure of chignolin was the lowest in terms of free energy while shallow local minima were widely distributed in the free energy landscape of GPM12, in agreement with experimental observations. Among the chimeric peptides, GPM12(D4P/K7G) stably formed the same β-hairpin structure as that of chignolin in the MD simulation. This was confirmed by nuclear magnetic resonance (NMR) spectroscopy. A comparison of the free-energy landscapes showed that the conformational distribution of the Asp3-Pro4 sequence was inherently biased in a way that is advantageous both to forming hydrogen bonds with another β-strand and to initiating loop structure. In addition, Gly7 helps stabilize the loop structure by having a left-handed α-helical conformation. Such a conformation is necessary to complete the loop structure, although it is not preferred by other amino acids. Our results suggest that the consistency between the short-range interactions that determine the local geometries and the long-range interactions that determine the global structure is important for stable tertiary structure formation.

Proteins 2008; 73:621–631.
© 2008 Wiley-Liss, Inc.

Key words: protein folding; molecular dynamics simulation; multicanonical method; chignolin; β-hairpin.

INTRODUCTION

One of the most important goals of studying protein folding is to clarify the atomistic mechanism by which an amino acid sequence determines its native structure. Among the methods used in the study of protein folding, reproducing the folding process using molecular dynamics (MD) has been considered the most straightforward approach.[1,2] However, this approach has suffered from a large gap between the timescale of actual folding and that of MD simulation. In general, the timescale of folding is longer than 1 ms, whereas MD simulations are currently limited to less than 1 μs. To overcome this problem, researchers have searched for rapidly folding proteins and have found several small proteins that fold within microseconds.[2–10] The use of these kinds of proteins is therefore expected to promote understanding of the protein-folding mechanism, enabling direct comparison between the results from simulations and experiments. To date, folding simulations starting from extended structures have been successful for Trp-cage (20 residues),[11–16] BBA5 (23 residues),[6,17,18] Trp zippers (12–16 residues),[8] and villin headpiece (35 residues).[19]

Chignolin (GYDPETGTWG), an artificially designed 10-residue peptide, is one such protein. This peptide was designed to have the same β-hairpin structure as residues 45–52 of the B1 domain of protein G and was shown to have a structure quite similar to the template by nuclear magnetic resonance (NMR).[20] It meets the requirements for a protein: it folds into a unique structure and has a cooperative thermal transition between its unfolded and folded states.[20] Therefore, it is

Grant sponsor: Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT); Grant numbers: 16014204, 16041206.
*Correspondence to: Tohru Terada, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan.
E-mail: tterada@iu.a.u-tokyo.ac.jp.
Received 8 November 2007; Revised 21 March 2008; Accepted 25 March 2008
Published online 12 May 2008 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.22100

currently the smallest protein.[20] Independent MD simulations have successfully reproduced its native structure.[21–24] Although the folding time constant has not yet been determined experimentally, it was estimated to be 1.0 μs from the MD trajectories.[25] Interestingly, the peptide having the template sequence (GYDDATKTFG), termed GPM12, does not have a specific structure.[20] Except for terminal Gly residues, the sequence of GPM12 is originally part of a peptide known as G-peptide, which corresponds to residues 41–56 of the B1 domain of protein G (GEWTYDDATKTFTVTE; underlined residues are shared by GPM12). G-peptide was shown to have a β-hairpin structure,[26] and its folding has been studied extensively by MD simulations.[27–30] In the simulations, the peptide has six or seven backbone hydrogen bonds between the β strands.[27–30] Since the removal of three residues from each terminal decreases the number of possible backbone hydrogen bonds by three, the inability of GPM12 to form the β-hairpin structure is probably explained by the loss of these intramolecular interactions. Although the sequence of chignolin was designed on the basis of statistical analysis of Protein Data Bank (PDB) data,[20] the atomistic mechanism by which the amino acids stabilize the β-hairpin structure is still unclear.

The difference in the folding behavior of the two peptides (chignolin and GPM12) can be ascribed to the difference in the amino acids at positions 4 (Pro/Asp), 5 (Glu/Ala), 7 (Gly/Lys), and 9 (Trp/Phe). To clarify the roles of these residues in the tertiary structure formation, we compared the folding free-energy landscapes of chignolin, GPM12, and their chimeric peptides. As in a previous study,[22] we used multicanonical MD to calculate the free-energy landscapes. This method can sample conformational spaces of biomolecules having rough energy landscapes more efficiently than the conventional method and can produce more accurate free-energy landscapes.[31–33] From a comparison of the free-energy landscapes, we determined a set of mutations required for GPM12 to form the β-hairpin structure and investigated the mechanism by which an amino acid sequence determines its tertiary structure.

MATERIALS AND METHODS

MD simulations with implicit solvent model

The MD simulations were performed for chignolin, GPM12, three chimeric peptides, [GPM12(D4P), GPM12(K7G), and GPM12(D4P/K7G)], and an Asp-Pro fragment. The initial structure used in each simulation was generated in an extended conformation by using the LEaP module of the Amber 6.0 software.[34] The N- and C-termini of the chignolin, GPM12, and their chimeric peptides were left uncapped, bearing a positive and negative charge, as in the experiment.[20] Those of the Asp-Pro fragment were capped with acetyl and *N*-methyl groups. After initial equilibration runs at a constant-temperature of 300 K, the multicanonical potential functions were determined so that the multicanonical MD simulations produced flat energy distributions covering energy regions corresponding to 290–700 K by using a method described elsewhere.[35] Production runs of the multicanonical simulations were carried out for 300 ns for the chignolin, GPM12, and their chimeric peptides and for 100 ns for the Asp-Pro fragment. During the simulations, the coordinates were recorded every ps. We used a modified version of parm99 as the force-field parameters[11,36] and the generalized Born/surface area (GB/SA) model to calculate the solvation free energy.[37] The parameters for the GB model were taken from Tsui and Case.[38] The dielectric constant of the solute interior was 1.0, and that of the solvent was 78.5. Salts were not included in the system. The value of the surface tension used in the SA model was 20.9 J mol^{-1} $Å^{-2}$. Non-bonded interactions were calculated without cut-off operations. Hydrogen-containing groups including XH_n groups (X = C, N, and O, and n = 1, 2, and 3) and aromatic rings were treated as rigid bodies to enable the use of a longer time step (2 fs) for integration with a small error.[35,39] The simulations were carried out using a modified version of the SANDER module of the Amber 6.0 software.[34,35]

MD simulations with explicit solvent model

MD simulations with the explicit solvent model were also performed for chignolin and GPM12. For the chignolin, three structures were selected from the largest cluster, whereas for the GPM12, one structure was selected from each of the three largest clusters. Each structure was placed in a water box containing about 1200 water molecules. Sodium ions were added to neutralize the system. The modified version of the parm99 force-field parameters[11,36] was used for the solute, and the TIP3P model[40] was used for the solvent. After energy minimization, the system was equilibrated with a 700-ps MD simulation, during which harmonic restraints to the initial structure were imposed on the non-hydrogen atoms and the force constant was gradually decreased. Production runs were performed for 30 ns for each system. In both the equilibrating and production runs, the temperature and pressure were kept close to 300 K and 10^5 Pa, respectively, by using the weak-coupling algorithm.[41] The SHAKE algorithm was used to constrain the bond lengths involving hydrogen atoms and to enable the use of a long time step (2 fs).[42] The electrostatic interactions were calculated with the particle mesh Ewald method.[43] The van der Waals interactions were calculated with a cut-off distance of 9 Å. The original version of the SANDER module of the Amber 6.0 software was used.[34]

For the Asp-Pro fragment, a multicanonical MD simulation was also performed in explicit solvent. The

fragment structure generated in an extended conformation was placed in a water box containing 1304 water molecules. One sodium ion was added to neutralize the system. The system was first equilibrated with a 100-ps constant-temperature, constant-pressure MD simulation at 300 K and 10^5 Pa. Then, the multicanonical MD simulation was performed for 100 ns with a multicanonical potential function, which was determined so as to produce a flat energy distribution covering the energy range corresponding to 290–700 K. Note that the volume was kept constant during the multicanonical MD simulation. The multicanonical simulation was performed using the same method as that used in the simulations with the GB/SA model, except that the electrostatic interactions were calculated with the particle mesh Ewald method.

Cluster analysis

The conformational ensembles obtained from the multicanonical MD simulations were converted into canonical ones at 300 K by calculating weight factors for the snapshot structures using a reweighting formula.[32,44] The structures were then classified into clusters on the basis of their structural similarities in two stages. In the first stage, the structures having heavy-atom RMSDs of less than 1.0 Å from the cluster center were grouped into a cluster. The atoms of the terminal residues (Gly1 and Gly10) were excluded from the calculation. The grouping started with a tentative cluster center using one of the snapshot structures and was repeated until there was convergence with the center structure (calculated as the average of the member structures). We refer to the clusters generated during this stage as "subclusters." In the second stage, an agglomerative hierarchical clustering[45] was performed on the subclusters using the pair-wise C^{α} RMSDs between the centers of the subclusters as the distance measure. The distance between clusters was defined as the average of the pair-wise RMSDs between their subclusters weighted by the sum of the weight factors of the structures in the subclusters. The dendrogram was truncated at a distance of 0.5 Å to produce "clusters." In the PDB survey, the structures were clustered similar to the structure clustering in the second stage. Each PDB structure was weighted equally, and the dendrogram was truncated at a distance of 1.0 Å.

NMR measurements

The GPM12(D4P/K7G) peptide was purchased from the Toray Research Center. The purity of the sample was checked by using HPLC and was more than 95%. It was dissolved at 2 mM in 90% ^1H_2O/10% ^2H_2O NMR buffer (pH 5.5) containing 20 mM of sodium phosphate. The 2D NOESY, ROESY, and total correlation spectroscopy (TOCSY) spectra were measured at 277 K using a Bruker DRX600 spectrometer (600.03 MHz for ^1H). The mixing times for NOESY and ROESY were 200 ms, and that of TOCSY was 50 ms. From the analysis of the NOESY and ROESY spectra, 119 distance constraints were obtained, and the structures were calculated with these constraints by using the Crystallography and NMR System.[46] Finally, 23 structures having no distance violation greater than 0.3 Å were obtained from 200 calculations. Their quality was evaluated using the PROCHECK-NMR program.[47] The atomic coordinates, experimental restraints, and resonance assignments have been deposited in PDB under accession code 2E4E.

RESULTS

Chignolin

Prior to calculating the folding free-energy landscapes of GPM12 and the chimeric peptides, we recalculated that of chignolin using a longer (300 ns) multicanonical MD simulation than in the previous study (180 ns).[22] Since a refined multicanonical potential function was used in the simulation, the flatness of the energy distribution was improved. As a result, the frequency of folding from unfolded states [root-mean-square deviations (RMSDs) from the average of 18 NMR structures calculated for the C^{α} atoms of residues 2–9 being greater than 4.0 Å] into near-native states (the RMSDs being less than 1.0 Å) increased from 0.65 to 0.87 ns^{-1} [Fig. 1(a)]. Because free energy is a statistical quantity, its accuracy depends on the number of independent samples. In the case of conformational sampling using MD simulation, the number is roughly proportional to the number of conformational transitions. Therefore, the free-energy landscapes obtained here are more accurate than those obtained in the previous study.[22] Figure 1(b) confirms that the populations of the largest and the second largest clusters converged to the final values (32.7 and 16.4% of the whole ensemble) within the simulation time. Figure 2 shows superposition of the MD structures for the largest cluster on a representative NMR structure, which is the closest to the average NMR structure. As in the previous study,[22] these structures agreed well with the average NMR structure with C^{α} RMSD values of 0.46 Å $\pm$ 0.14 Å. The hydrogen bonds important for β-hairpin structure formation (Asp3N-Thr8O, Asp3O-Gly7N, and Asp3O^{δ}-Glu5N) were correctly formed in these structures [Fig. 2(b)].

Since the native structure was the lowest in terms of free energy under the conditions of the MD simulation, the current method performed well in the calculation of the folding free-energy landscape of chignolin. We used the generalized Born/surface area (GB/SA) model to approximately calculate the solvation free energy. With this model, the protein does not experience the solvent's viscosity, and the conformational transitions of the protein occur much more frequently than in the explicit solvent.[2,49] This characteristic is advantageous for efficient

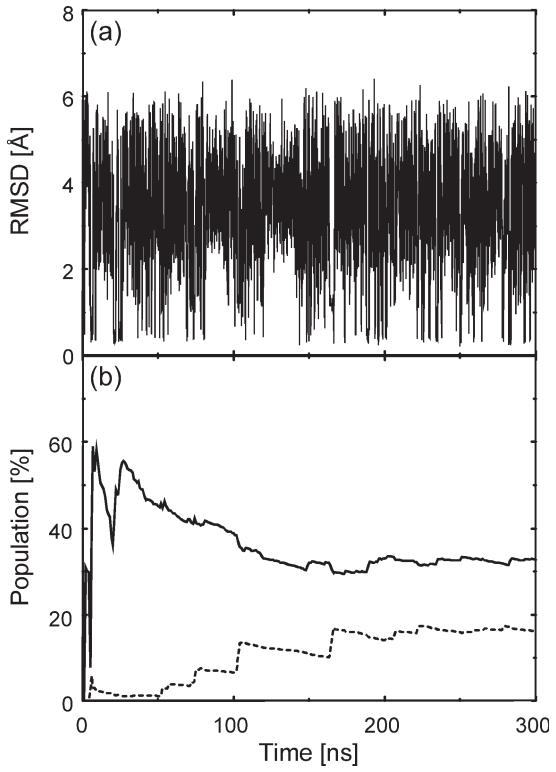

Figure 1

(**a**) Time evolution of C^{α} RMSD calculated for residues 2–9 from average NMR structure of chignolin during multicanonical MD simulation of chignolin. (**b**) Plots of populations of largest (solid line) and second largest (dashed line) clusters as function of simulation time. Value at time *T* is population in the ensemble from time 0 to time *T*.

conformational sampling. However, the use of the GB/SA model inevitably causes an error in the calculation of the solvation free energy.[49–51] To check whether the error of the GB/SA model affects the free-energy landscape, we performed constant-temperature MD simulations in the explicit solvent, starting from three different structures in the largest cluster. The characteristic hydrogen bonds (Asp3O-Gly7N and Asp3N-Thr8O) were stable in two runs. The average distances of the hydrogen bonds were 2.89 Å ± 0.14 Å and 2.96 Å ± 0.18 Å. In the third run, they were stably maintained for 28.5 ns with average distances of 3.06 Å ± 0.37 Å and 3.02 Å ± 0.32 Å, although the Asp3N-Thr8O hydrogen bond was broken after that. Since the structures were maintained for sufficiently long time in the explicit solvent, the GB/SA model does not bias the free-energy landscape of chignolin and thus provides a reliable free-energy landscape as does the explicit solvent model.

GPM12

We next performed a multicanonical MD simulation on GPM12. The conformational clusters calculated from the MD trajectory were ranked on the basis of their populations. The structures for the top three clusters (14.8, 7.4, and 3.9% of the whole ensemble) are shown in Figure 3. Obviously, the structures differ from that of chignolin and also from that of the corresponding region in the B1 domain of protein G. On the whole, they have a bent α-helical conformation; however, the structures from the two largest clusters were significantly different, with C^{α} RMSD values of 2.33 Å ± 0.13 Å. Figure 4 shows plots of the accumulative population against the subcluster rank. Since each subcluster occupies a separate space in the conformational space with roughly equal volume, the plots indicate that the conformation of GPM12 is more diversified than that of chignolin. In other words, shallow free-energy minima are broadly

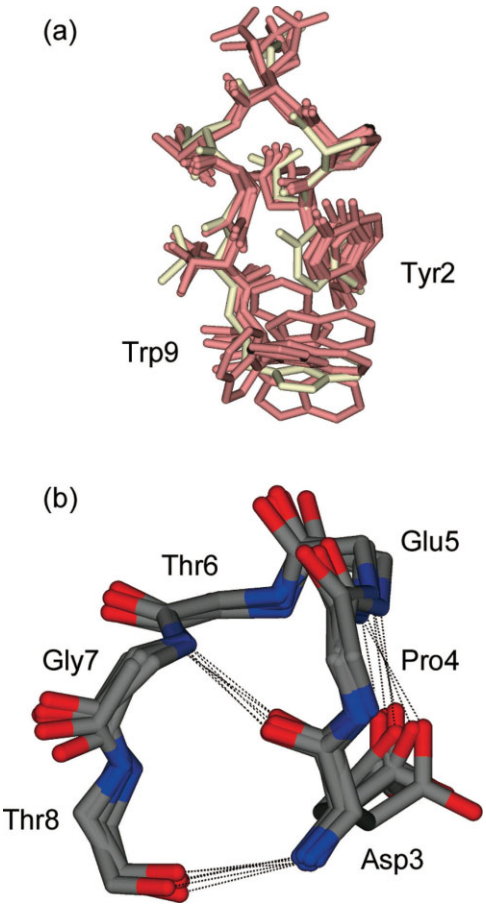

Figure 2

(**a**) Superposition of chignolin MD structures from cluster having largest population (pink) on representative NMR structure (ivory). Each MD structure represents center of subcluster with probability of existence larger than 1%. Only non-hydrogen atoms of residues 2–9 are shown. (**b**) Close-up views of residues 3–8 of MD structures from cluster having largest population. Backbone non-hydrogen atoms and side-chain non-hydrogen atoms of Asp3 are shown with stick model. Carbon, nitrogen, and oxygen atoms are colored gray, blue, and red. Hydrogen bonds are indicated with dotted lines. All structure figures were produced with Molscript.[48]

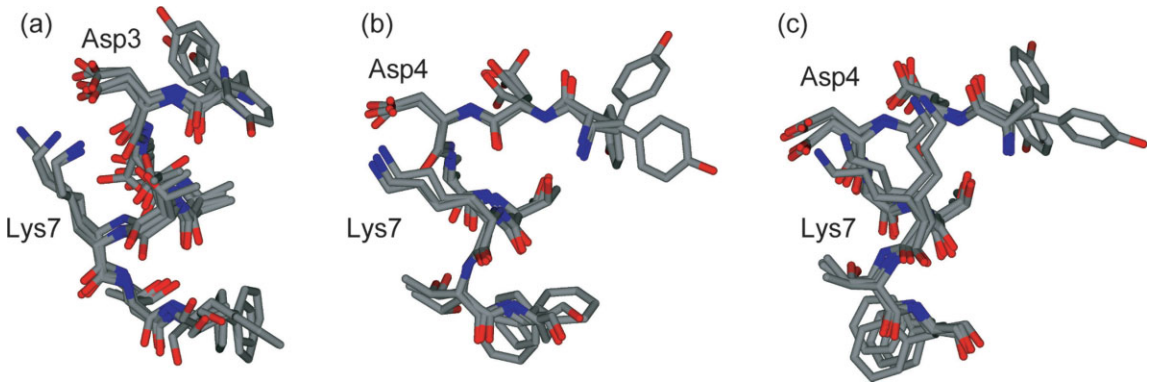

Figure 3

GPM12 MD structures from clusters having largest (**a**), second largest (**b**), and third largest (**c**) populations. Only non-hydrogen atoms of residues 2–9 are shown (with stick model). Carbon, nitrogen, and oxygen atoms are colored gray, blue, and red.

distributed in the conformational space of GPM12. This observation is consistent with the experimental finding that GPM12 does not have a specific structure.[20]

In most of the structures from the top three clusters, salt bridges formed between Asp3 or Asp4 and Lys7 that do not exist in chignolin (see Fig. 3). Since the backbone structure varies with the salt-bridging mode, the salt bridge might be responsible for the diversity of the GPM12 conformations. However, the GB/SA model tends to overestimate the strength of the salt bridge formed between corresponding residues in G-peptide.[50,51] Therefore, we performed constant-temperature MD simulations starting from the structures from the top three clusters. Although the salt bridge was frequently broken, it formed again after a short period (see Fig. 5). The quick and large change in the conformation is probably consistent with the free-energy landscape calculated with the GB/SA model, in which shallow free-energy minima are broadly distributed. These observations indicate that the salt bridge is formed at least weakly, resulting in the conformational diversity of GPM12. However, a more important cause is probably a lack of intramolecular interactions. In G-peptide, the β-hairpin structure is stabilized by a hydrophobic cluster of four residues and by

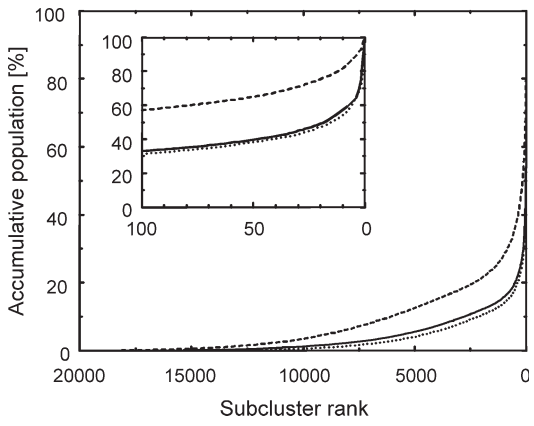

Figure 4

Accumulative population of subclusters of chignolin (solid line), GPM12 (dashed line), and GPM12(D4P/K7G) (dotted line). Subclusters are ranked in accordance with their populations. Region corresponding to 100th subcluster or higher is shown in inset.

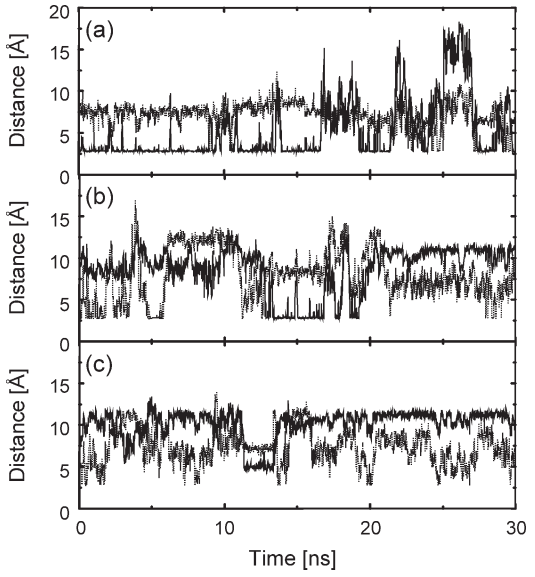

Figure 5

Time evolution of distance between Asp3O^{δ} and Lys7N^{ζ} (solid line) and between Asp4O^{δ} and Lys7N^{ζ} (dashed line) during MD simulations with explicit water model starting from structures selected from three largest clusters of GPM12 (**a**)–(**c**). Distances were calculated for two possible pairs, that is, $O^{\delta 1}$-N^{ζ} and $O^{\delta 2}$-N^{ζ}, and smaller values were plotted.

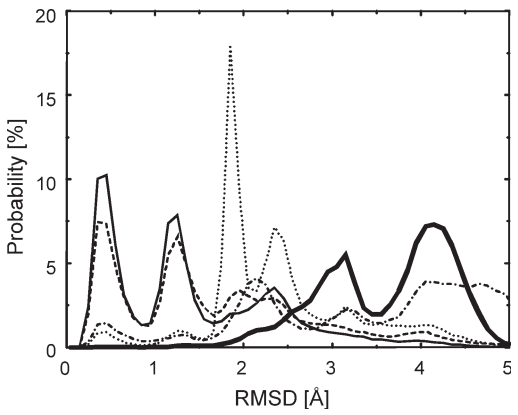

Figure 6

Probability distribution as function of C^{α} RMSD calculated for residues 2–9 from average NMR structure of chignolin in MD ensembles of chignolin (solid thin line), GPM12 (solid thick line), GPM12(D4P) (dash-dotted line), GPM12(K7G) (dotted line), and GPM12(D4P/K7G) (dashed line).

six or seven backbone hydrogen bonds,[27–30,50,51] whereas in GPM12, the hydrophobic cluster and the hydrogen bonds are incomplete due to the deletion of the six residues.

Chimeric peptides

To address the question of why chignolin can form the β-hairpin structure stably despite having the same sequence length as GPM12, we analyzed the free-energy landscapes of the chimeric peptides. Among the residues different between chignolin and GPM12, the residues at positions 4 (Pro/Asp) and 7 (Gly/Lys) are especially different in their physicochemical properties. We therefore examined two single mutants, GPM12(D4P) and GPM12(K7G). Here, GPM12(X*n*Y) stands for a mutant of GPM12 in which amino acid X at position *n* is replaced with amino acid Y. Figure 6 shows plots of the probability distribution as a function of the C^{α} RMSD from the average NMR structure of chignolin in the canonical ensembles obtained from the MD simulations. The highest peak for chignolin was at around 0.4 Å, indicating that it has the native structure, whereas the probability of GPM12 having a β-hairpin structure is almost zero. The two single mutants had small but evident peaks at around 0.4 Å, indicating that the residues introduced into GPM12 helped stabilize the β-hairpin structure.

However, the populations of the β-hairpin structure were much smaller than that in the chignolin ensemble, suggesting that single mutations are insufficient. Therefore, we next examined a double mutant, GPM12(D4P/K7G). In Figure 6, the plot for this mutant is overlaid on the plots for the other peptides. Since this mutant had its highest peak at around 0.4 Å, it can efficiently form the β-hairpin structure, like chignolin can. The MD structures from the most populated cluster (25.2% of the whole ensemble) fit well on the NMR structure of chignolin (C^{α} RMSD values of 0.47 Å ± 0.15 Å). These structures had the same hydrogen bonds as chignolin. In addition, the structures of the cluster with the second largest population (10.4% of the whole ensemble) had a structure quite similar to that of the chignolin cluster with the second largest population. These results suggest that, in the MD simulation, the double mutation (D4P/K7G) converted the folding behavior of GPM12 into one like that of chignolin.

NMR structure of GPM12(D4P/K7G)

To confirm the prediction that GPM12(D4P/K7G) has the same β-hairpin structure as chignolin, we determined its tertiary structure by NMR spectroscopy. Statistics for the determined structure are summarized in Table I. The NMR spectra were measured under the same condition as that of Honda *et al.*[20] They showed that chignolin is monomeric under the condition.[20] For a comparison, we measured the NMR spectra of chignolin and found that the line shapes of the NMR spectra of GPM12(D4P/K7G) were similar to those of chignolin. Therefore, it is probable that GPM12(D4P/K7G) is also monomeric. As shown in Figure 7, the long-range NOEs characteristic of the β-hairpin structure were clearly observed, and the structure was well determined. The structure was quite close to that of chignolin (pair-wise C^{α} RMSDs of 0.67 Å ± 0.22 Å), as predicted by the MD simulations. The MD structures of GPM12(D4P/K7G) for the cluster with the largest population had C^{α} RMSD values of 0.66 Å ± 0.20 Å from the average NMR structure and satisfied 88.7% of the NMR restraints on average within the violation criterion (0.5 Å for distance restraints). From these

Table I

Statistics for GPM12(D4P/K7G) Structure Determined by NMR Spectroscopy

Number of constraints	
Intra-residue NOEs	57
Sequential NOEs ($\|i-j\|=1$)	36
Medium-range NOEs ($\|i-j\|=2,3,4$)	18
Long-range NOEs ($\|i-j\|>4$)	8
Total	119
RMSD from constraints (Å)	0.063 ± 0.001
RMSDs from ideal geometry	
Bonds (Å)	0.0071 ± 0.0001
Angles (degrees)	0.58 ± 0.01
Impropers (degrees)	0.27 ± 0.02
RMSDs about mean coordinate positions (residues 2–9)	
Backbone (N, C^{α}, and C) atoms (Å)	0.11 ± 0.04
Nonhydrogen atoms (Å)	0.50 ± 0.06
Ramachandran map analysis by PROCHECK-NMR (%)	
Most favored region	65.2 ± 4.7
Additional allowed region	32.6 ± 6.0
Generously allowed region	2.2 ± 5.6
Disallowed region	0.0 ± 0.0

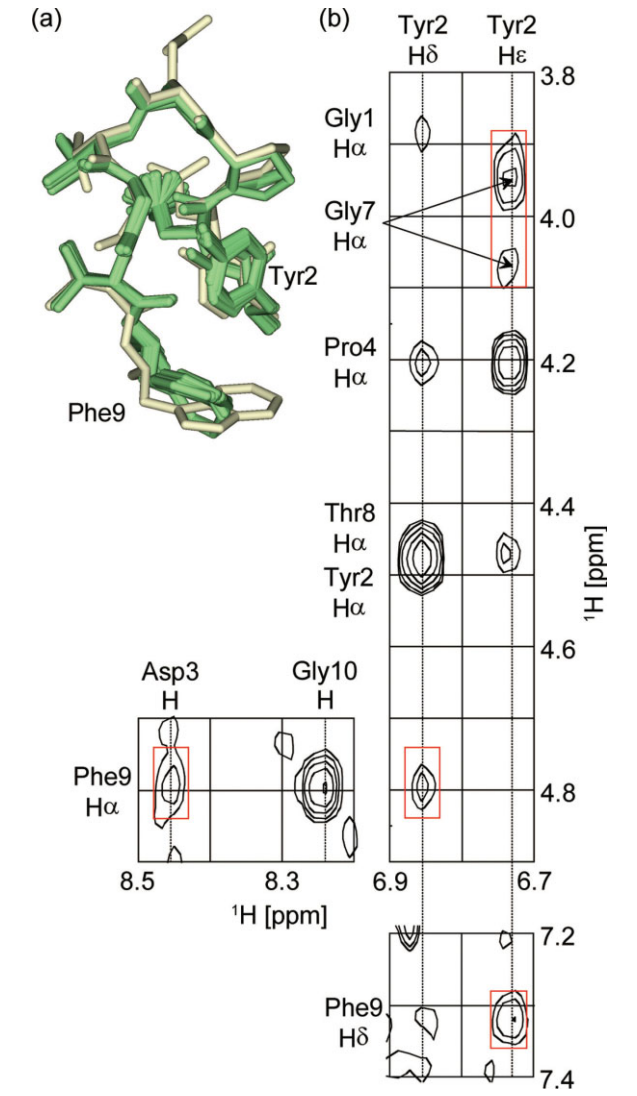

Figure 7

(**a**) Superposition of NMR structures of GPM12(D4P/K7G) (green) on representative NMR structure of chignolin (ivory). (**b**) Parts of NOESY and ROESY spectra of GPM12(D4P/K7G). Crosspeaks of long-range NOEs characteristic of β-hairpin structure are marked with red boxes.

results, we concluded that D4P/K7G is the minimum mutation required for GPM12 to form the β-hairpin structure.

DISCUSSION

Role of Asp3-Pro4 sequence

To understand why the D4P/K7G mutation enables GPM12 to form the β-hairpin structure, we compared the distributions of the backbone dihedral angles of chignolin with those of GPM12. We found that the distribution maps of the Asp3 ψ − Pro4 ϕ ($\psi_3 - \phi_4$) dihedral angles were clearly different between chignolin and GPM12. In the chignolin plot [Fig. 8(a)], there is one minor (5.3%) and one major (94.7%) peak centered on about (ψ_3, ϕ_4) = (−60°, −60°) and (120°, −60°). In the GPM12 plot [Fig. 8(b)], there is a broader distribution with three peaks at around (ψ_3, ϕ_4) = (−40°, −60°) (55.3%), (160°, −60°) (26.8%), and (−60°, −150°) (15.8%). It is obvious that the narrower distribution of ϕ_4 for chignolin was due to the ring structure of Pro4, in which possible conformations are highly restricted. There are two possible reasons for the biased distribution of ψ_3 for chignolin. One is that the distribution is inherent in the Asp-Pro sequence. Another is that the distribution results from the interaction between Asp3-Pro4 and the rest of the peptide. To distinguish these two possibilities, we performed 100 ns multicanonical MD simulations on an Asp-Pro fragment, of which N- and C-termini were capped with acetyl and *N*-methyl groups, respectively. We found that the distribution map of Asp ψ − Pro ϕ of the fragment was almost the same as that of the corresponding residues of chignolin. The populations of the major peak were 97.3%. Cluster analysis revealed that 50.6% of the conformational ensemble of the Asp-Pro fragment was grouped in a cluster having a structure quite similar to that of the corresponding region of chignolin [Fig. 8(c,d)]. Similar results (97.8% for the major peak population and 54.3% for the cluster population) were obtained from a simulation with the explicit water model. These results indicate that the biased distribution of the Asp ψ dihedral angle is inherent in the Asp-Pro sequence. As can be seen from [Fig. 8(e)], structures having Asp ψ dihedral angles of around 0° and −120° are prohibited due to strong collisions of the $C^{\delta}H_2$ group of Pro with the amide and $C^{\beta}H_2$ groups of Asp, respectively. In addition, Asp ψ dihedral angles of around −180°, −60°, and 60° are unfavorable due to gauche-type interactions between the bulky tertiary amide group of Pro and the amide and $C^{\beta}H_2$ groups of Asp. In contrast, the favorable conformation having an Asp ψ dihedral angle of around 120° is further stabilized by a hydrogen bond between the O^{δ} of Asp and the amide hydrogen of the *N*-methyl group, which corresponds to that of Glu5 in chignolin [see Fig. 8(c)]. This interaction limits the range of the ψ dihedral angle of Pro to about −60° to 60°.

The bias of the Asp ψ dihedral angle toward 120° is advantageous for β-sheet formation because the N—H and C=O bonds of Asp can point in the same direction. According to a classification of β-hairpin conformations by Sibanda *et al.*,[52] the structure of chignolin can be classified as a four-residue, doubly hydrogen-bonded (4:4) loop. The four residues comprising the loop typically take an α_R-α_R-γ_R-α_L conformation, where α_R and α_L are right- and left-handed α-helical conformations, and γ_R is a conformation with backbone ϕ and ψ dihedral angles of about −90° and 0°. In chignolin, the four

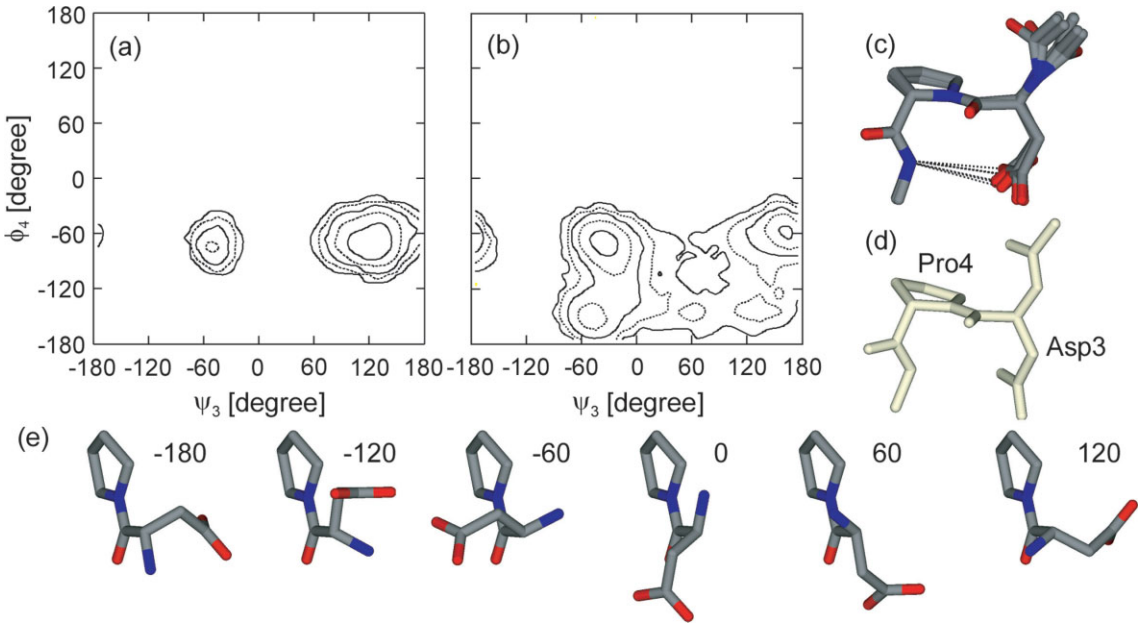

Figure 8

Probability distributions as function of ψ_3 and ϕ_4 dihedral angles of chignolin (**a**) and GPM12 (**b**) using same contour levels for both plots. Level of n-th contour is given by $4^{(n-1)} \times 10^{-4}$. MD structures of capped Asp-Pro fragment from cluster having largest population (**c**) and corresponding region (Asp3-Pro4) of representative NMR structure of chignolin (ivory) (**d**). (**e**) Conformations of Asp-Pro fragment for different ψ dihedral angles. Numbers indicate ψ dihedral angles in degrees. Pro carbonyl group is omitted for simplicity.

residues correspond to Pro4-Glu5-Thr6-Gly7. Together with the restriction on the Pro4 ϕ dihedral angle to about -60°, the bias in the distribution of the Pro4 ψ dihedral angle caused by the hydrogen bond between the O^δ of Asp3 and the backbone amide hydrogen of Glu5 increases the probability that Pro4 takes an α_R conformation, which is favorable to loop structure initiation.

Role of Gly7

Gly7 corresponds to the last residue of the loop and thus must have an α_L conformation. GPM12 has a lysine residue at this position, and the α_L conformation is less favorable than the α_R conformation due to the gauche-type interaction between the carbonyl group of the preceding residue and the $C^\beta H_2$ group of the lysine residue. In contrast, glycine can take α_R and α_L conformations equally and stably since it lacks a side chain and therefore has a rotationally symmetric Ramachandran plot. Thus, enabling α_L conformation is the role of Gly7 in β-hairpin structure formation.

Roles of other residues

Our results indicate that the differences in positions 5 (Glu/Ala) and 9 (Trp/Phe) are not critical to β-hairpin structure formation. The circular dichroism spectrum of a triple mutant, GPM12(D4P/K7G/F9W) [i.e., chignolin(E5A)], was almost the same as that of chignolin (data not shown), which supports the conclusion that the Glu/Ala difference is not important for β-hairpin structure formation. Since Trp9 makes tight contact with Tyr2 in the NMR structure, it has been suggested that the interaction between Tyr2 and Trp9 might promote stability of the β-hairpin structure.[20] It should be noted here that our results do not deny the possibility that the interaction between the aromatic rings contributes to the stabilization of the β-hairpin structure.

Survey of PDB

To examine whether the Asp-Pro-X-X-Gly sequence (X stands for any amino acid except proline) can form β-hairpin structures in other proteins, we searched the PDB for this sequence. We analyzed a non-redundant set of PDB data generated by the PISCES server[53] with a percentage identity cut-off of 90% and a resolution cut-off of 2.5 Å. From 10,813 chains in the set, we found 457 fragments having the Asp-Pro-X-X-Gly sequence. The coordinates of Asp-Pro-X-X-Gly plus one N- and two C-terminal residues were extracted from the PDB, and the residues were numbered from 2 to 9 in accordance with the residue number of chignolin. The structures thus obtained were classified into clusters on the basis of their structural similarities measured as pair-wise C^α RMSDs.

Table II
List of Major Clusters Generated by Classifying PDB Structures Having Asp-Pro-X-X-Gly Sequence in Accordance with Structural Similarities

Cluster ID	Number of structures	Ratio (%)	C^{α} RMSD from chignolin (Å)
1	77	16.8	0.62 ± 0.17
2	21	4.6	3.76 ± 0.16
3	17	3.7	1.31 ± 0.19
4	14	3.1	4.75 ± 0.09
5	11	2.4	1.88 ± 0.14

Table II lists the clusters with more than 10 structures. The structures in the largest cluster were quite similar to that of chignolin with C^{α} RMSD values of 0.62 Å ± 0.17 Å. Figure 9(a) shows the complete structure of a cluster member, adrenodoxin reductase (PDB ID: 1CJC).[54] The backbone structure of residues 343–349 (FDPLKGVV) of the protein agreed well with that of chignolin with C^{α} RMSD values of 0.30 Å [Fig. 9(b)]. As can be seen from Figure 9(a), the region of the sequence is in a long loop and was exposed to the solvent. It is therefore plausible that this region autonomously adopts a β-hairpin structure.

Correspondence with protein-folding theories

Finally, we interpret our results, connecting them with existing protein-folding theories. The Asp-Pro sequence plays two roles: it arranges the amide and carbonyl groups of Asp for hydrogen bonding with another β strand, and it initiates loop structure formation. These groups enable the formation of "long-range" hydrogen bonds, Asp3N-Thr8O and Asp3O-Gly7N, as a result of "short-range" interactions within the Asp-Pro sequence. Such cooperation between the "short-range" and "long-range" interactions is important for tertiary structure formation of protein, as proposed by Gō as the consistency principle[55] and by Bryngelson and Wolynes as the principle of minimal frustration.[56] The replacement of Lys7 in GPM12 with glycine resolves the conflict between the free-energy gain due to β-hairpin structure formation and the loss due to taking the α_L conformation unfavorable for lysine, which can be viewed as an increase in the consistency or a decrease in the frustration. In normal-sized proteins, a small inconsistency or frustration might be compensated for by favorable interactions in other parts of the protein. In mini-proteins, where the number of possible intramolecular interactions is limited, consistency between the "short-range" and "long-range" interactions is especially important and is still the key to understanding the mechanism of tertiary structure formation.

CONCLUSIONS

In the present study, we compared the folding free-energy landscapes of chignolin, GPM12 and their chimeric peptide to clarify the reason why chignolin can form a stable β-hairpin structure in water, in spite that its design template, GPM12, does not have a specific structure. We used the multicanonical MD method to

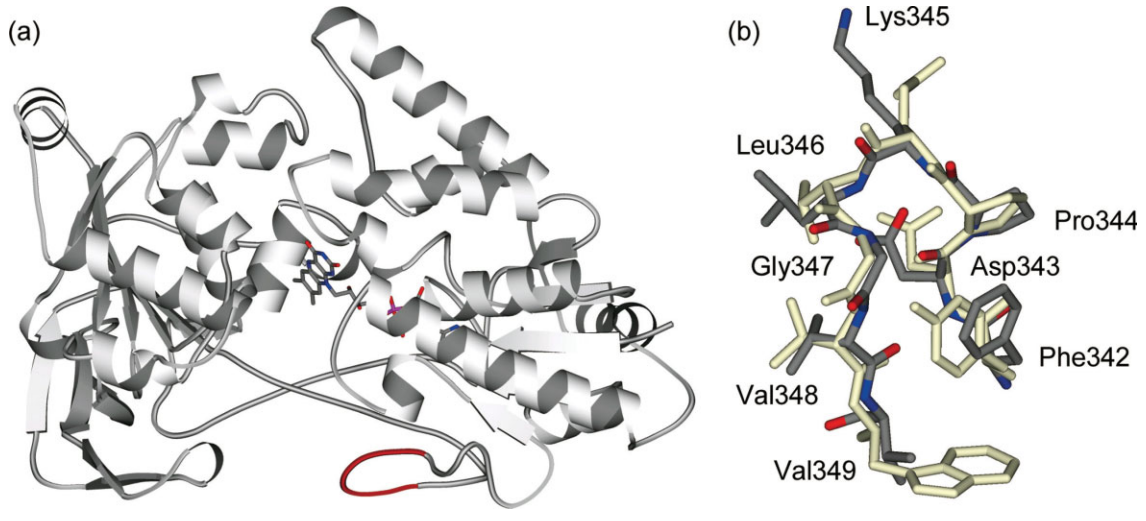

Figure 9
(**a**) Schematic representation of backbone structure of adrenodoxin reductase (PDB ID: 1CJC). Residues 342–349 are colored red. Flavin adenine dinucleotide is shown with stick model. (**b**) Superposition of structure of residues 342–349 of adrenodoxin reductase on representative NMR structure of chignolin (ivory). Carbon, nitrogen, and oxygen atoms of adrenodoxin reductase are colored gray, blue, and red.

obtain accurate free-energy landscapes, and those of chignolin and GPM12 were consistent with the experimental observations. Therefore, we next calculated folding free-energy landscapes for several chimeric peptides. Among them, GPM12(D4P/K7G) stably formed the same β-hairpin structure as that of chignolin in the MD simulation, which was verified experimentally by NMR spectroscopy. The comparison of the free-energy landscapes suggested that the Asp3-Pro4 sequence and Gly7 of chignolin play important roles in the β-hairpin formation: The Asp3-Pro4 sequence arranges the amide and carbonyl groups of Asp for hydrogen bonding with another β strand and initiates loop structure formation. Gly7 completes the loop structure by having a left-handed α-helical conformation, which is not preferred by other amino acids. Our results suggest that the consistency between the short-range interactions that determine the local geometries and the long-range interactions that determine the global structure is important for stable tertiary structure formation.

ACKNOWLEDGMENTS

The authors thank Prof. Akinori Kidera of Yokohama City University for the fruitful discussions we had with him. The authors thank Dr. Ryuichiro Ishitani for the useful input he provided when he was at the University of Tokyo.

REFERENCES

1. Fersht AR, Daggett V. Protein folding and unfolding at atomic resolution. Cell 2002;108:573–582.
2. Snow CD, Sorin EJ, Rhee YM, Pande VS. How well can simulation predict protein folding kinetics and thermodynamics? Annu Rev Biophys Biomol Struct 2005;34:43–69.
3. Neidigh JW, Fesinmeyer RM, Andersen NH. Designing a 20-residue protein. Nat Struct Biol 2002;9:425–430.
4. Qiu L, Pabit SA, Roitberg AE, Hagen SJ. Smaller and faster: the 20-residue Trp-cage protein folds in 4 μs. J Am Chem Soc 2002; 124:12952–12953.
5. Struthers M, Ottesen JJ, Imperiali B. Design and NMR analyses of compact, independently folded BBA motifs. Fold Des 1998;3:95–103.
6. Snow CD, Nguyen H, Pande VS, Gruebele M. Absolute comparison of simulated and experimental protein-folding dynamics. Nature 2002;420:102–106.
7. Cochran AG, Skelton NJ, Starovasnik MA. Tryptophan zippers: stable, monomeric β-hairpins. Proc Natl Acad Sci USA 2001;98:5578–5583.
8. 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.
9. McKnight CJ, Doering DS, Matsudaira PT, Kim PS. A thermostable 35-residue subdomain within villin headpiece. J Mol Biol 1996;260:126–134.
10. Kubelka J, Eaton WA, Hofrichter J. Experimental tests of villin subdomain folding simulations. J Mol Biol 2003;329:625–630.
11. Simmerling C, Strockbine B, Roitberg AE. All-atom structure prediction and folding simulations of a stable protein. J Am Chem Soc 2002;124:11258–11259.
12. Snow CD, Zagrovic B, Pande VS. The Trp cage: folding kinetics and unfolded state topology via molecular dynamics simulations. J Am Chem Soc 2002;124:14548–14549.
13. Chowdhury S, Lee MC, Xiong G, Duan Y. *Ab initio* folding simulation of the Trp-cage mini-protein approaches NMR resolution. J Mol Biol 2003;327:711–717.
14. Pitera JW, Swope W. Understanding folding and design: replica-exchange simulations of "Trp-cage" miniproteins. Proc Natl Acad Sci USA 2003;100:7587–7592.
15. Zhou R. Trp-cage: folding free energy landscape in explicit water. Proc Natl Acad Sci USA 2003;100:13280–13285.
16. Ota M, Ikeguchi M, Kidera A. Phylogeny of protein-folding trajectories reveals a unique pathway to native structure. Proc Natl Acad Sci USA 2004;101:17658–17663.
17. Rhee YM, Pande VS. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. Biophys J 2003; 84:775–786.
18. Rhee YM, Sorin EJ, Jayachandran G, Lindahl E, Pande VS. Simulations of the role of water in the protein-folding mechanism. Proc Natl Acad Sci USA 2004;101:6456–6461.
19. Lei H, Wu C, Liu H, Duan Y. Folding free-energy landscape of villin headpiece subdomain from molecular dynamics simulations. Proc Natl Acad Sci USA 2007;104:4925–4930.
20. Honda S, Yamasaki K, Sawada Y, Morii H. 10 residue folded peptide designed by segment statistics. Structure 2004;12:1507–1518.
21. Seibert MM, Patriksson A, Hess B, van der Spoel D. Reproducible polypeptide folding and structure prediction using molecular dynamics simulations. J Mol Biol 2005;354:173–183.
22. Satoh D, Shimizu K, Nakamura S, Terada T. Folding free-energy landscape of a 10-residue mini-protein, chignolin. FEBS Lett 2006;580:3422–3426.
23. Kannan S, Zacharias M. Enhanced sampling of peptide and protein conformations using replica exchange simulations with a peptide backbone biasing-potential. Proteins 2007;66:697–706.
24. Suenaga A, Narumi T, Futatsugi N, Yanai R, Ohno Y, Okimoto N, Taiji M. Folding dynamics of 10-residue β-hairpin peptide chignolin. Chem Asian J 2007;2:591–598.
25. van der Spoel D, Seibert MM. Protein folding kinetics and thermodynamics from atomistic simulations. Phys Rev Lett 2006;96: 238102.
26. Blanco FJ, Rivas G, Serrano L. A short linear peptide that folds into a native stable β-hairpin in aqueous solution. Nat Struct Biol 1994;1:584–590.
27. García AE, Sanbonmatsu KY. Exploring the energy landscape of a β hairpin in explicit solvent. Proteins 2001;42:345–354.
28. Zhou R, Berne BJ, Germain R. The free energy landscape for β hairpin folding in explicit water. Proc Natl Acad Sci USA 2001; 98:14931–14936.
29. Nguyen PH, Stock G, Mittag E, Hu C, Li MS. Free energy landscape and folding mechanism of a β-hairpin in explicit water: a replica exchange molecular dynamics study. Proteins 2005;61:795–808.
30. Yoda T, Sugita Y, Okamoto Y. Cooperative folding mechanism of a β-hairpin peptide studied by a multicanonical replica-exchange molecular dynamics simulation. Proteins 2007;66:846–859.
31. Hansmann UHE, Okamoto Y, Eisenmenger F. Molecular dynamics. Langevin and hybrid Monte Carlo simulations in a multicanonical ensemble. Chem Phys Lett 1996;259:321–330.
32. Nakajima N, Nakamura H, Kidera A. Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides. J Phys Chem B 1997;101:817–824.
33. Mitsutake A, Sugita Y, Okamoto Y. Generalized-ensemble algorithms for molecular simulations of biopolymers. Biopolymers (Pept Sci) 2001;60:96–123.
34. Case DA, Pearlman DA, Caldwell JW, Cheatham TE, III, Ross WS, Simmerling CL, Darden TA, Merz KM, Stanton RV, Cheng AL, Vincent JJ, Crowley M, Tsui V, Radmer RJ, Duan Y, Pitera J, Massova I,

Seibel GL, Singh UC, Weiner PK, Kollman PA. AMBER 6, San Francisco: University of California; 1999.
35. Terada T, Matsuo Y, Kidera A. A method for evaluating multicanonical potential function without iterative refinement: application to conformational sampling of a globular protein in water. J Chem Phys 2003;118:4306–4311.
36. Wang J, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J Comput Chem 2000;21:1049–1074.
37. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 1990;112:6127–6129.
38. Tsui V, Case DA. Theory and applications of the generalized Born solvation model in macromolecular simulations. Biopolymers (Nucleic Acid Sci) 2001;56:275–291.
39. Terada T, Kidera A. Generalized form of the conserved quantity in constant-temperature molecular dynamics. J Chem Phys 2002;116: 33–41.
40. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
41. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
42. Ryckaert J, Ciccotti G, Berendsen HJC. Numerical integration of the Cartesian equation of motion of a system with constraints: molecular dynamics of *n*-alkanes. J Comput Phys 1977;23:327–341.
43. Darden T, York D, Pedersen L. Particle mesh Ewald: an $N\cdot\log(N)$ method for Ewald sums in large systems. J Chem Phys 1993;98: 10089–10092.
44. Ferrenberg AM, Swendsen RH. New Monte Carlo technique for studying phase transitions. Phys Rev Lett 1988;61:2635–2638.
45. Späth H. Cluster analysis algorithms for data reduction and classification of objects. England: Ellis Horwood Limited; 1980.
46. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D 1998;54:905–921.
47. Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 1996;8:477–486.
48. Kraulis PJ. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 1991; 24:946–950.
49. Ishizuka T, Terada T, Nakamura S, Shimizu K. Improvement of accuracy of free-energy landscapes of peptides calculated with generalized Born model by using numerical solutions of Poisson's equation. Chem Phys Lett 2004;393:546–551.
50. Zhou R, Berne BJ. Can a continuum solvent model reproduce the free energy landscape of a β-hairpin folding in water? Proc Natl Acad Sci USA 2002;99:12777–12782.
51. Zhou R. Free energy landscape of protein folding in water: explicit vs. implicit solvent. Proteins 2003;53:148–161.
52. Sibanda BL, Blundell TL, Thornton JM. Conformation of β-hairpins in protein structures. A systematic classification with applications to modelling by homology, electron density fitting and protein engineering. J Mol Biol 1989;206:759–777.
53. Wang G, Dunbrack RL, Jr. PISCES: recent improvements to a PDB sequence culling server. Nucleic Acids Res 2005;33:W94–W98.
54. Ziegler GA, Vonrhein C, Hanukoglu I, Schulz GE. The structure of adrenodoxin reductase of mitochondrial *P*450 systems: electron transfer for steroid biosynthesis. J Mol Biol 1999;289:981–990.
55. Gō N. The consistency principle in protein structure and pathways of folding. Adv Biophys 1984;18:149–164.
56. Bryngelson JD, Wolynes PG. Spin glasses and the statistical mechanics of protein folding. Proc Natl Acad Sci USA 1987;84: 7524–7528.