ELSEVIER

Biophysical Chemistry 115 (2005) 195–200

Biophysical Chemistry

http://www.elsevier.com/locate/biophyschem

Folding simulations of small proteins

Seung-Yeon Kim[a], Julian Lee[b], Jooyoung Lee[a,*]

[a]School of Computational Sciences, Korea Institute for Advanced Study, Seoul 130-722, South Korea
[b]Department of Bioinformatics and Life Sciences and CAMDRC, Soongsil University, Seoul 156-743, South Korea

Received 18 May 2004; received in revised form 1 November 2004; accepted 10 December 2004
Available online 6 January 2005

Abstract

Understanding how a protein folds is a long-standing challenge in modern science. We have used an optimized atomistic model (united-residue force field) to simulate folding of small proteins of various structures: HP-36 (α protein), protein A (β), 1fsd (α+β), and betanova (β). Extensive Monte Carlo *folding* simulations (ten independent runs with 10^9 Monte Carlo steps at a temperature) starting from non-native conformations are carried out for each protein. In all cases, proteins fold into their native-like conformations at appropriate temperatures, and glassy transitions occur at low temperatures. To investigate early folding trajectories, 200 independent runs with 10^6 Monte Carlo steps are also performed at a fixed temperature for a protein. There are a variety of possible pathways during non-equilibrium early processes (fast process, ~10^4 Monte Carlo steps). Finally, these pathways converge to the point unique for each protein. The convergence point of the early folding pathways can be determined only by direct folding simulations. The free energy surface, an equilibrium thermodynamic property, dictates the rest of the folding (slow process, ~10^8 Monte Carlo steps).

Keywords: Protein folding; Computer simulation; Folding pathways

1. Introduction

Protein folding is one of the most fundamental biophysical processes. It plays the most important role in controlling a wide range of cellular processes. The failure of a proper protein folding results in the malfunction of biological systems, leading to various diseases. Computer simulations have been extensively carried out to understand the mechanism of protein folding processes. The most microscopic level of the computational studies of protein folding uses atomic models of a protein as well as its environment. However, direct folding simulation using an atomistic model is a very difficult task and, moreover, it requires an astronomical amount of computational resources. For example, the 1-μs molecular dynamics simulation of the villin headpiece subdomain (HP-36) with an all-atom potential has produced only candidates for folding intermediates [1]. For this reason, direct folding simulations have been performed only for simplified models, such as the random energy model [2], lattice models [3,4], and off-lattice minimalist models [5,6]. An alternative indirect approach is to perform unfolding simulations [7,8] starting from the folded state of a protein. However, it is not obvious if the folding is the reverse of the unfolding [3,7,9].

In this study, we use an optimized united-residue (UNRES) potential [13], for which the global minimum-energy conformations for four small proteins are all native-like. The four proteins are HP-36 [1,10] (36 residues, a three-helix bundle), fragment B of staphylococcal protein A [7,11] (46 residues, a three-helix bundle), 1fsd [12] (28 residues, one β-hairpin, and one α-helix), and betanova [8] (20 residues, a three-stranded β-sheet). It should be noted that the UNRES potential includes all (native and non-native) interactions. We have carried out realistic *folding* simulations of the four proteins starting from non-native conformations to investigate the folding mechanism of the proteins.

* Corresponding author. Tel.: 8229583822; fax: 8229583820.
E-mail address: jlee@kias.re.kr (J. Lee).

0301-4622/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2004.12.040

2. UNRES force field

We study the folding processes of proteins using an optimized UNRES force field [13] where a protein is represented by a sequence of α-carbon (C^{α}) atoms linked by virtual bonds with attached united sidechains (SC) and united peptide groups located in the middle between the consecutive C^{α}'s. All the virtual bond lengths are fixed: the C^{α}–C^{α} distance is taken as 3.8 Å, and C^{α}–SC distances are given for each amino acid type. The energy of a protein is given by

$$E = \sum_{i} [U_{\mathrm{b}}(i) + U_{\mathrm{t}}(i) + U_{\mathrm{r}}(i)] + \sum_{i<j-1} U_{\mathrm{pp}}(i,j) + \sum_{i\neq j} U_{\mathrm{sp}}(i,j) + \sum_{i<j} [U_{\mathrm{ss}}(i,j) + U^{(4)}_{\mathrm{el-loc}}(i,j)] + U_{\mathrm{dis}}.$$

Here, the terms $U_{\mathrm{b}}(i)$, $U_{\mathrm{t}}(i)$ and $U_{\mathrm{r}}(i)$ denote the short-range interactions, corresponding to the energies of virtual angle bending, virtual dihedral angle torsions, and sidechain rotamers, respectively. The potential $U_{\mathrm{pp}}(i,j)$ accounts for the electrostatic interaction between the peptide groups of residues i and j, $U_{\mathrm{sp}}(i,j)$ corresponds to the excluded-volume interaction between the sidechain of residue i and the peptide group of residue j, and $U_{\mathrm{ss}}(i,j)$ represents the mean free energy of the hydrophobic (hydrophilic) interaction between the sidechains of residues i and j, which is expressed by Lennard–Jones potential. The four-body interaction term $U^{(4)}_{\mathrm{el-loc}}$ is from the cumulant expansion of the restricted free energy of the protein. U_{dis} denotes the energy term which forces two cysteine residues to form a disulfide bridge. The parameters of the UNRES force field are simultaneously optimized for HP-36, protein A, 1fsd, and betanova. The low-lying local-energy minima for these proteins are found by conformational space annealing [5]. The parameters are modified in such a way that the native-like conformations are energetically more favored than the others. After the parameter optimization, one set of the parameters is obtained for the proteins [13]. It should be noted that the approach proposed in this paper can be examined to other potential energy functions [13].

3. Monte Carlo dynamics

In the UNRES force field there are two backbone angles and two sidechain angles per residue (no sidechains for glycines). The values of these angles are perturbed one at a time, typically about 15°, and the backbone angles are chosen three times more frequently than the sidechain angles. The perturbed conformation is accepted according to the change in the potential energy, following the Metropolis rule. Since only small angle changes are allowed one at a time, the resulting Monte Carlo dynamics can be viewed as equivalent to the real dynamics [4]. During simulations the

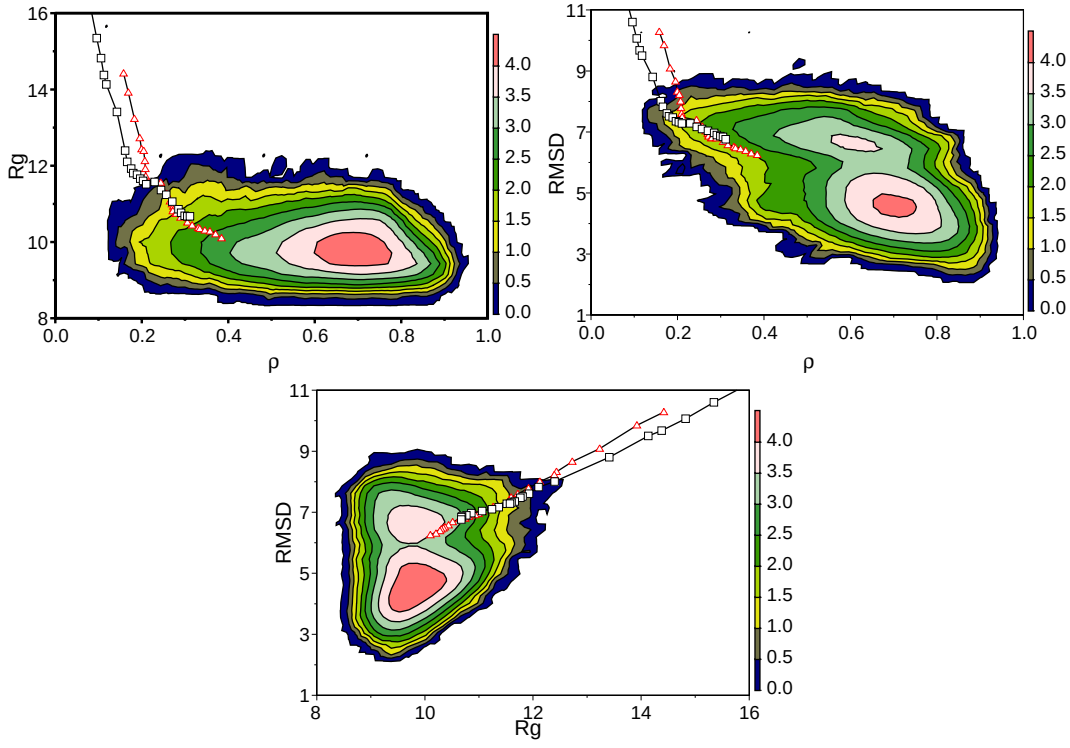

Fig. 1. The initial folding trajectories (from 200 independent short-time simulations) and the contour plots of the population distributions (from 10 independent long-time simulations) for HP-36 at T=70 as a function of (a) ρ and R_{g} (Å), (b) ρ and RMSD (Å), and (c) R_{g} and RMSD. The triangles represent the averages of 100 early folding trajectories starting from random conformations. The squares are from 100 early folding trajectories starting from fully extended conformations. The color scale of the contour plots indicates the exponent x of the population 10^{x+5}.

values of the root-mean-square deviation (RMSD) from the native structure and the radius of gyration (R_g) were calculated using C^α coordinates. The UNRES representation is simple, but it has C^α coordinates. Therefore, the RMSD between the experimental structure and the UNRES representation for a protein can be calculated only using C^α coordinates. The lowest RMSD values from folding simulations are 0.78 Å, 1.07 Å, 1.58 Å, and 2.07 Å for betanova, 1fsd, HP-36, and protein A, respectively. The fraction of the native contacts (ρ) [7,8,14] was also measured during simulations.

At a fixed temperature, at least ten independent simulations starting from various non-native states of a protein were performed up to 10^9 Monte Carlo steps (MCS), which we call *long-time* simulations. All together, hundreds of independent long-time simulations were conducted for each protein. To investigate the early folding trajectories, 200 independent simulations (10^6 MCS for each run) were also performed at a fixed temperature for a protein, which we call *short-time* simulations. In Figs. 1, 3, 4, and 5, we show various quantities averaged over early folding trajectories. We divide the initial 10^6 MCS into 28 intervals (first ten 10^3 MCS, subsequent nine 10^4 MCS, and the next nine 10^5 MCS), and averages are taken over the whole conformations for each interval. These averages are averaged again over 100 independent simulations starting from random conformations. The same procedure is repeated for 100 independent simulations starting from fully extended conformations.

4. Helical proteins

We observe that all proteins fold into their native-like conformations at appropriate temperatures [9,14]. Collapse occurs at a very early stage (~10^4 MCS) for all four proteins [9,14], but the details of each folding process appear to depend on the secondary structure contents. We have analyzed details of the folding behavior for each protein.

For HP-36 at T=60, the population distributions of various quantities such as RMSD for ten independent runs (10^9 MCS each) depend on initial conformations, showing its glassy behavior [9,14]. At higher temperatures ($T \geq 70$) this non-ergodic glassy behavior disappears. Fig. 1 shows initial folding trajectories (from 200 independent short-time simulations) and contour plots of population distributions (from 10 independent long-time simulations) at T=70. Regardless of its initial conformation (either random or fully extended), the initial folding trajectories in the figure converge to (ρ, R_g, RMSD)~(0.21, 11.5 Å, 7.4 Å) around at 10^4 MCS. The collapsed structures at the convergence point eventually fold into native-like structures (ρ, R_g, RMSD)~(0.7, 9.8 Å, 4.5 Å) after about 3×10^8 MCS. The most prominent characteristic of the contour plots for HP-36 is the two-peak structure in RMSD distributions as shown in Fig. 1b and c: (ρ, R_g, RMSD)~(0.6, 9.8 Å, 6.5 Å) and (ρ, R_g, RMSD)~(0.7, 9.8 Å, 4.5 Å). It should be noted that the values of R_g from the two peaks coincide. The conformations centered at the higher value of RMSD come from a variety of collapsed states. The conformations from the other peak are native-like. When we examine them, the helix I (residues 4–8) is stably formed, while the others are fluctuating. This stability of helix I is consistent with the recent experimental result [10]. Fig. 2 shows snapshots of a typical folding trajectory for HP-36 at T=80. The conformation in Fig. 2b is a collapsed state with no stable secondary structure formed and it corresponds to the convergence point of the initial folding trajectories in Fig. 1. The conformation in Fig. 2c is located close to the most dominant peaks of the contour plots in Fig. 1.

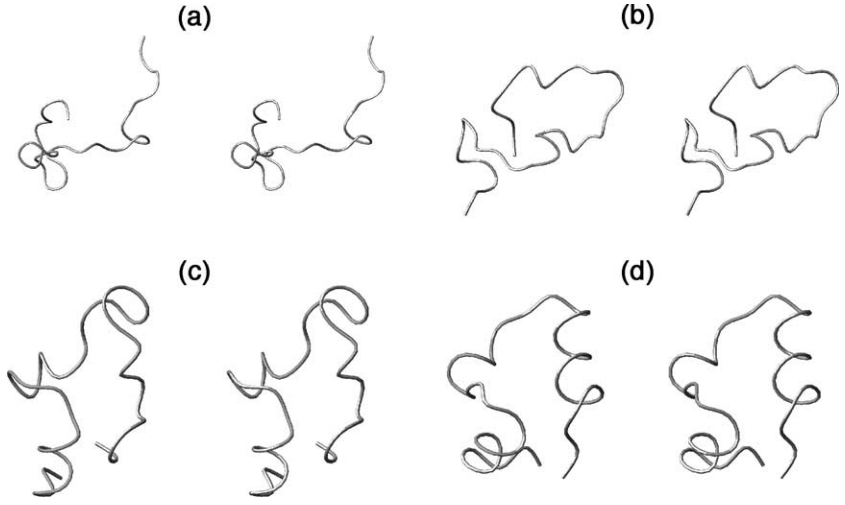

Fig. 2. Snapshots of a typical folding trajectory for HP-36 at T=80. In the figures the left side is the N-terminal part. (a) Initial random conformation: ρ=0.27, R_g=14:5 Å, and RMSD=11.1 Å. (b) Conformation at 3.1×10^5 MCS: ρ=0.20, R_g=11.6 Å, and RMSD=7.5 Å. This conformation corresponds to a collapsed state. (c) Conformation at 5×10^7 MCS: ρ=0.63, R_g=9.5 Å, and RMSD=3.6 Å. The helices I and II appear but the helix III is not yet formed. (d) Conformation at 1.602×10^8 MCS: ρ=0.85, R_g=9.3 Å, and RMSD=2.0 Å. The conformation is almost identical to the native one.

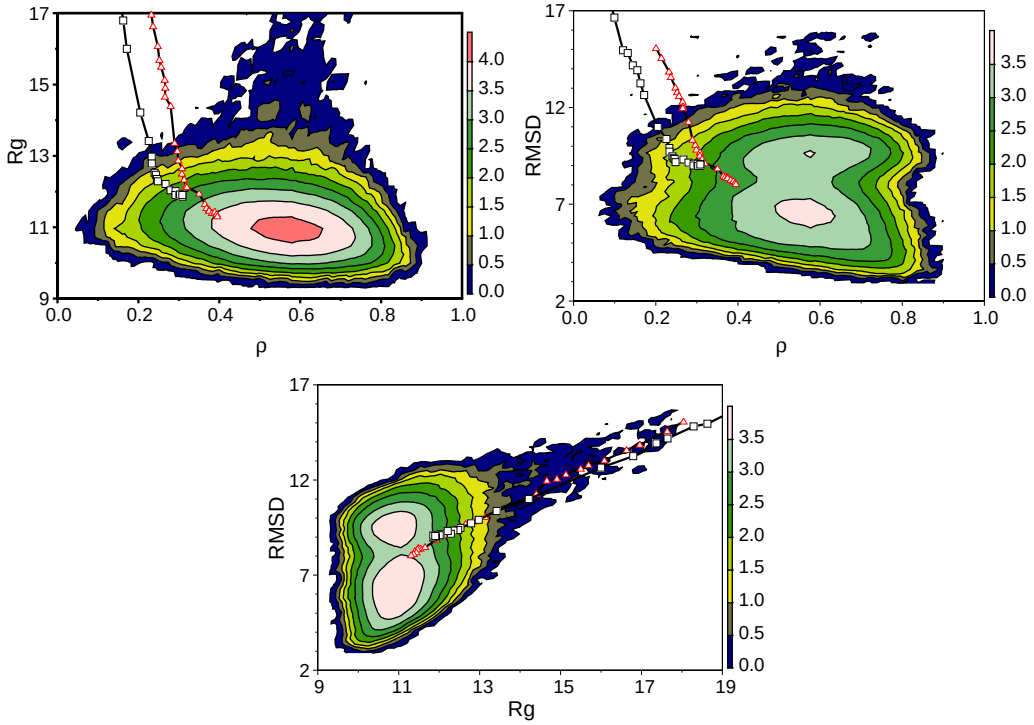

Fig. 3. The initial folding trajectories and the contour plots of the population distributions for protein A at T=80.

The overall folding characteristics of protein A are similar to those of HP-36. Fig. 3 shows the initial folding trajectories and the contour plots of the population for protein A at T=80. The initial folding trajectories converge to (14.4 Å, 11.3 Å) in the (R_g, RMSD) plane, and then they collapse to (ρ, R_g, RMSD)~(0.31, 12.1 Å, 9.2 Å) at about 9×10^4 MCS. These collapsed states fold into native-like structures in a fashion similar to the case of HP-36. The contour plots for protein A also show the two-peak structure in the RMSD distributions as shown in Fig. 3b and c: (ρ, R_g,

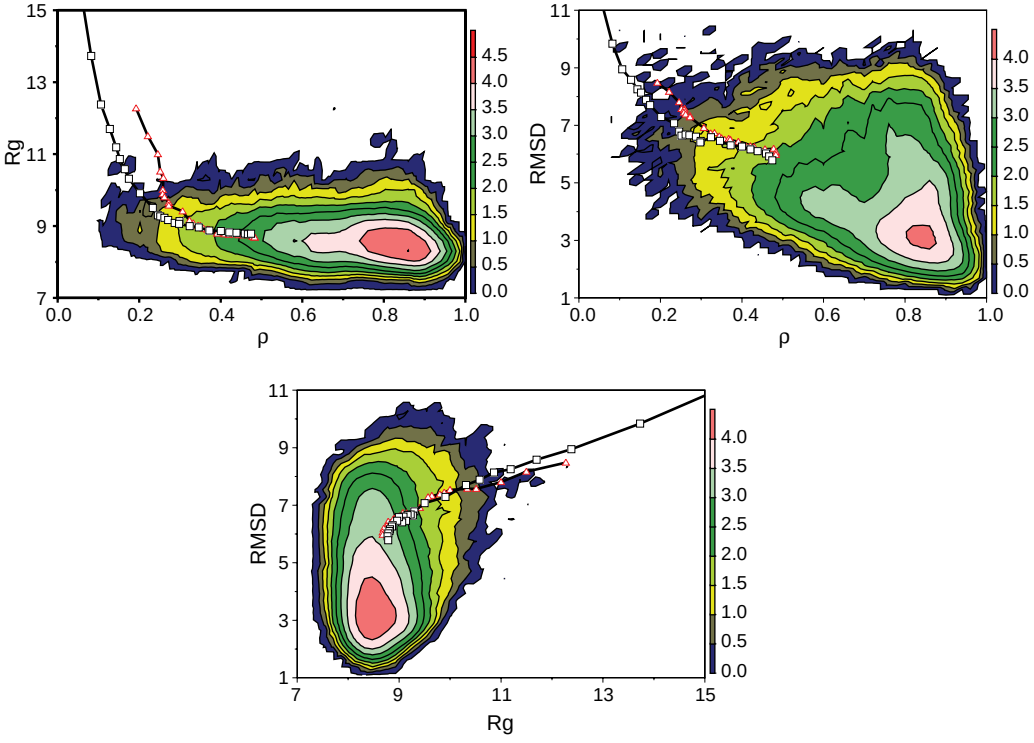

Fig. 4. The initial folding trajectories and the contour plots of the population distributions for 1fsd at T=70.

RMSD)~(0.57, 11 Å, 9.7 Å) and (0.6, 11 Å, 6 Å). The two peaks are separated only in their RMSD values. The region (ρ, R_g, RMSD)~(0.6, 11 Å, 6 Å) is reached after 2.4×10^8 MCS. When we examine the native-like conformations with 3 Å≤RMSD≤4 Å, the helix III (residues 42–55) is most stably formed [14]. This is in agreement with the recent experimental investigation [11].

5. Proteins with beta strands

For 1fsd the population distributions of various quantities for ten independent runs (10^9 MCS each) show glassy behavior for $T\leq50$. Again, the non-ergodic glassy behavior disappears at higher temperatures ($T\geq70$). Fig. 4 shows the initial folding trajectories and the contour plots of the population distributions at T=70. Again after the initial collapse to (R_g, RMSD)~(10.2 Å, 7.6 Å), the average trajectories converge to (ρ, R_g, RMSD)~(0.35, 9 Å, 6.5 Å) at about 6×10^4 MCS. The states at the convergence point usually have the β-hairpin formed, and they move into the most dominant region, (ρ, R_g, RMSD)~(0.85, 8.5 Å, 3 Å) after 6.5×10^7 MCS, whose structures are very similar to the native one [14]. For proteins with β strands, only one peak is observed in the RMSD distributions.

For betanova at low temperatures ($T\leq30$), the glassy behavior is observed. This glassy behavior disappears at higher temperatures ($T\geq40$). The initial folding trajectories and the population distributions at T=40 are shown in Fig. 5. The average pathways to the folded conformation initially converge to (ρ, R_g, RMSD)~(0.3, 9.5 Å, 6.7 Å) around 6×10^3 MCS, and then they move to the most dominant conformations (ρ, R_g, RMSD)~(0.95, 7.8 Å, 2.3 Å) after 5.3×10^7 MCS. The collapsed structures at the convergence point of the initial folding trajectories usually have one β-hairpin formed, and the conformations at the most populated region are almost identical to the native structure [14]. The populated states around (ρ, R_g)~(0.4~0.5; 10~12 Å) come from the thermal fluctuation that temporarily kicks the protein out of its native structure. This observation is contrasted to the recent conjecture from unfolding simulation augmented by free energy calculation [8] that these states are from initial folding trajectories. This kind of kinetic information can be captured only by direct folding simulations, and is difficult to be described by free energy calculations alone.

6. Discussion

We have performed folding simulations in a realistic setting, where all proteins fold into their native-like conformations at appropriate temperatures. In all cases, rapid collapse is followed by a subsequent folding process that takes place in a longer time scale. We also observe that glassy transitions occur at low temperatures. The folding mechanism observed in this study is as follows: there are two aspects of folding dynamics, (i) non-equilibrium kinetic

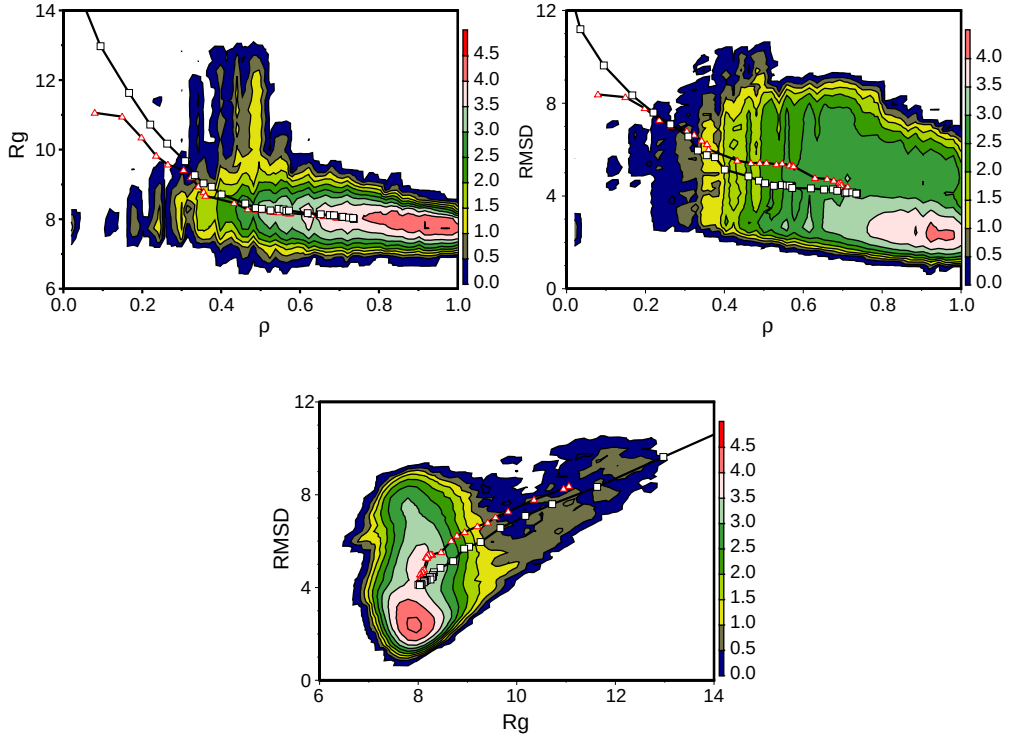

Fig. 5. The initial folding trajectories and the contour plots of the population distributions for betanova at T=40.

properties and (ii) equilibrium thermodynamic properties. The non-equilibrium kinetic properties are relevant to the early folding trajectories (fast process, ~10^4 MCS). There are a variety of possible pathways during these non-equilibrium kinetic processes. Finally, these pathways converge to the state that is unique for each protein. The convergence point of the early folding pathways can be determined only by direct folding simulations. The free energy surface, an equilibrium thermodynamic property, dictates the way states at the convergence point complete their folding (slow process, ~10^8 MCS).

Acknowledgments

This work was supported by grants No. R01-2003-000-11595-0 (Jooyoung Lee) and No. R01-2003-000-10199-0 (Julian Lee) from the Basic Research Program of the Korea Science and Engineering Foundation.

References

[1] Y. Duan, P.A. Kollman, Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution, Science 282 (1998) 740–744.
[2] J.D. Bryngelson, P.G. Wolynes, Spin glasses and the statistical mechanics of protein folding, Proc. Natl. Acad. Sci. 84 (1987) 7524–7528.
[3] A.R. Dinner, M. Karplus, Is protein unfolding the reverse of protein folding? A lattice simulation analysis, J. Mol. Biol. 292 (1999) 403–419.
[4] A. Kolinski, J. Skolnick, Lattice Models of Protein Folding, Dynamics, and Thermodynamics, Chapman and Hall, New York, 1996.
[5] S.-Y. Kim, S.J. Lee, J. Lee, Conformational space annealing and an off-lattice frustrated model protein, J. Chem. Phys. 119 (2003) 10274–10279.
[6] S.-Y. Kim, S.J. Lee, J. Lee, The energy landscape of a BLN protein with beta-hairpin shape, J. Korean Phys. Soc. 44 (2004) 589–593.
[7] J.-L. Shea, C.L. Brooks, From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding, Annu. Rev. Phys. Chem. 52 (2001) 499–535.
[8] B.D. Bursulaya, C.L. Brooks, Folding free energy surface of a three-stranded β-sheet protein, J. Am. Chem. Soc. 121 (1999) 9947–9951.
[9] S.-Y. Kim, J. Lee, J. Lee, Folding dynamics of small proteins with different types of structures, in: J. Klein-Seetharaman (Ed.), Biological Language Conference Proceedings, Carnegie Mellon University, Pittsburgh, 2003, pp. 218–228.
[10] Y. Tang, D.J. Rigotti, R. Fairman, D.P. Raleigh, Peptide models provide evidence for significant structure in the denatured state of a rapidly folding protein: the villin headpiece subdomain, Biochemistry 43 (2004) 3264–3272.
[11] Y. Bai, A. Karimi, H.J. Dyson, P.E. Wright, Absence of a stable intermediate on the folding pathway of protein A, Protein Sci. 6 (1997) 1449–1457.
[12] B.I. Dahiyat, S.L. Mayo, De novo protein design: fully automated sequence selection, Science 278 (1997) 82–87.
[13] J. Lee, S.-Y. Kim, J. Lee, Design of a protein potential energy landscape by parameter optimization, J. Phys. Chem., B 108 (2004) 4525–4534.
[14] S.-Y. Kim, J. Lee, J. Lee, Folding of small proteins using a single continuous potential, J. Chem. Phys. 120 (2004) 8271–8276.