Ab Initio Folding of Helix Bundle Proteins Using Molecular Dynamics Simulations

4 AUTHORS, INCLUDING:

Eunae Kim
Chosun University
40 PUBLICATIONS **409** CITATIONS

SEE PROFILE

Seokmin Shin
Seoul National University
108 PUBLICATIONS **1,673** CITATIONS

SEE PROFILE

Youngshang Pak
Pusan National University
33 PUBLICATIONS **628** CITATIONS

SEE PROFILE

Ab Initio Folding of Helix Bundle Proteins Using Molecular Dynamics Simulations

Soonmin Jang,[†] Eunae Kim,[‡] Seokmin Shin,*,[†] and Youngshang Pak*,[‡]

Contribution from the School of Chemistry, Seoul National University, Seoul 151-747, Korea, and Department of Chemistry, Pusan National University, Pusan 609-735, Korea

Received February 17, 2003; E-mail: sshin@snu.ac.kr; ypak@pusan.ac.kr

Abstract: We have demonstrated that ab initio fast folding simulations at 400 K using a GB implicit solvent model with an all-atom based force field can describe the spontaneous formation of nativelike structures for the 36-residue villin headpiece and the 46-residue fragment B of Staphylococcal protein A. An implicit solvent model combined with high-temperature MD makes it possible to perform direct folding simulations of small- to medium-sized proteins by reducing the computational requirements tremendously. In the early stage of folding of the villin headpiece and protein A, initial hydrophobic collapse and rapid formation of helices were found to play important roles. For protein A, the third helix forms first in the early stage of folding and exhibits higher stability. The free energy profiles calculated from the folding simulations suggested that both of the helix-bundle proteins show a two-state thermodynamic behavior and protein A exhibits rather broad native basins.

1. Introduction

There has been significant progress toward the understanding of protein folding problems both experimentally and theoretically.[1] While characteristics of folded (native) structures and folding mechanisms may be deduced from experimental methods, Molecular Dynamics (MD) simulation studies can provide direct information on detailed dynamic folding events at molecular levels. Theoretical predictions of protein folding based only on sequences of proteins and their chemical/physical nature, called ab initio folding or de novo folding, have been one of the central issues in structural biology for several decades.[2] However, even with today's modern computers, tackling protein folding by computational methods still remains a challenging problem due to the complicated nature of potential energy surfaces involved in protein folding. Recently, several promising simulation strategies have been proposed to overcome these computational difficulties.[3] In particular, with the new computational schemes, such as distributed computing[4] and replica exchange methods,[5] folding predictions of small peptides, of either the implicit or explicit solvation model, becomes practically possible. In other approaches, unfolding studies such as high-temperature induced unfolding simulations with conventional MD methods have been performed and give valuable information on the folding events[6] despite concerns about possible discrepancies between the folding and unfolding mechanisms. For ab initio folding simulations, it may become feasible to fold small peptides directly from their unfolded states to nativelike structures by normal MD if one uses an implicit solvation model. Recently, it was shown that slow folding dynamics could easily be accelerated by a moderate increase of temperature, enabling one to observe fast reversible folding/unfolding events within several tens of nanoseconds.[7]

In general, MD simulations with implicit solvation models are computationally much less demanding than those with explicit solvent models. Thus, for the long time simulations usually required by protein folding studies, use of implicit solvation models may be a reasonable choice in practice. Among the various implicit solvation models developed so far, the generalized Born (GB) solvation model provides a good representation of solvation effects and appears to be one of the most promising models in terms of its efficiency and accuracy.[8] In a previous study, we showed that the normal MD combined with the GB solvation model at elevated temperatures successfully observed "nativelike" structures of various peptides of β-structures within several nanoseconds of simulation time.[9]

[†] Seoul National University.

[‡] Pusan National University.

(1) (a) Nolting, B. *Protein folding kinetics: Biophysical Methods*; Springer: Berlin, 1999. (b) Fersht, A, R. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*; Freeman: New York, 1999.

(2) (a) Berne, B. J.; Straub, J. E. *Curr. Opin. Struct. Biol.* **1997**, 7, 181. (b) Shea, J. E.; Brooks, C. L., III. *Annu. Rev. Chem.* **2001**, 52, 499. (c) Baker, D.; Sali, A. *Science* **2001**, 294, 93.

(3) Onuchic, J. N.; Nymeyer, H.; Garcia, A.; Chahine, J.; Socci, N. *Adv. Protein Chem.* **2000**, 53.

(4) Shirts, M. R.; Pande, V. *Science*, **2001**, 290, 1903.

(5) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. *Science* **1995**, 267, 5204.

(6) (a) Karplus, M.; Sali, A. *Curr. Opin. Struct. Biol.* **1995**, 5, 58. (b) Wang, L.; Duan, Y.; Shortle, R.; Imperiali, B.; Kollman, P. A. *Protein Sci.* **1999**, 8, 1292. (c) Finkelstein, A. V. *Protein Eng.* **1997**, 10, 843.

(7) (a) Cavalli, A,; Ferrara, P.; Catflish, A. *Proteins* **2002**, 47, 305. (b) Simmering, C.; Strockbine, B.; Roitberg, A. E. *J. Am. Chem. Soc.* **2002**, 124, 11258.

(8) (a) Constanciel, R.; Contreras, R. *Theor. Chim. Acta* **1984**, 65, 1. (b) Qui, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, 101, 3005. (c) Still, W. C.; Tempczyk, A.; Hawley R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, 112, 6127. (d) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, 246, 122. (e) Edinger, S. R.; Cortis, C.; Shenkin, P. S.; Friesner, R. A. *J. Phys. Chem. B* **1997**, 101, 1190. (f) Dominy, B. H.; Brooks, C. L., III. *J. Phys. Chem. B* **1999**, 103, 3765. (g) Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. *Theor. Chem. Acc.* **1999**, 101, 426. (h) Bursulaya, B. D.; Brooks, C. L., III. *J. Am. Chem. Soc.* **1999**, 121, 9947.

Direct MD simulations at moderately elevated temperatures with implicit solvation models are expected to provide useful means of describing the folding dynamics of peptides and small proteins.

In this contribution, we extend our MD simulation study to the folding dynamics of more complex systems: 36-residue villin headpiece (HP-36; PDB code 1VII) and 46-residue fragment B of Staphylococcal protein A (BpA; PDB code 1BDC). These proteins are known to be fast folders[10] (around several microseconds) and one can expect that direct folding simulations of them are computationally feasible by high-temperature MD in combination with a GB solvation model. There have been several simulation studies on HP-36 including high-temperature unfolding simulations. Duan and Kollman[11] performed a fully atomistic MD simulation with explicit water, revealing detailed information about the initial folding stage of HP-36. More recently, introducing an all atom-based Langevin dynamics (LD) scheme with the GB model (with a surface area correction), Shen and Freed[12a] produced a 200-ns folding trajectory in good agreement with the MD results of Duan and Kollman. In their simulation, HP-36 was found to be almost completely folded into the nativelike structure (rmsd value from the NMR structure less than 4 Å) within 60 ns starting from an unfolded state. Zagrovic et al.[12b] used distributed computing techniques and a supercluster of thousands of computer processors to study folding of the villin headpiece in atomistic detail. They simulated a total of nearly 300 μs of folding time and obtained an ensemble of folded structures. Srinivas and Bagchi[12c] used a minimalist model HP-36 to research qualitative folding features. Recent simulations by Fernandez et al. suggested that three-body correlations are important in the α-helix bundle formation considering low local helix propensities of the villin headpiece.[12d]

The three-helix bundle protein A is one of the simplest helix bundle proteins along with the villin headpiece, and there have been many unfolding studies ranging from simplified-model[13] to all atom model[14] simulations. Brooks and co-workers[15] carried out extensive MD simulation studies to elucidate folding pathways of this protein by constructing its free energy landscape. Zhou and Karplus[16] investigated detailed kinetic folding pathways of this protein, using a discrete molecular dynamics method[17] with a simplified coarse-grained force field based on the Go-type potential. Subsequently, Linhananta and Zhou[18] employed all atom version of the Go-type potential to take into account side-chain packing effects and indicated that two distinctive folding pathways may exist in this protein. Using the GB solvation model, Ghosh et al.[19] investigated the folding pathway of BpA using the novel stochastic differential equation method with all-atom potential.[20]

We carried out several independent MD simulations of these two proteins using an all-atom AMBER force field with the GB model. The simulation temperature was set to 400 K. Performing folding simulations at low (physiological) temperature usually takes very long time since the system can be trapped in local minima along the rugged energy landscape. The main purpose of using high temperature is to accelerate the folding process by keeping the system from such trappings. At moderately high temperatures, the free energy barriers get smoother, and there might exist faster folding pathways leading to the native structures. If the temperature is too high, one may not observe folding events to stable native structures. With the temperatures in the appropriate range, it is expected to observe reversible folding and search for most of the relevant local minimum structures involved in the final tune-up process of protein folding. The native structure, as a global minimum, can be predicted from the analysis of such structures. The appropriate temperature range for performing efficient folding simulations may depend on specific systems. To observe refolding processes, the simulation temperatures should be lower than the melting temperatures of the proteins. The melting point of HP-36 was observed to be around 350 K.[10c,g] It may be difficult to estimate the exact melting points for the proteins studied in the present simulations, considering the uncertainties in the interaction parameters and the approximations inherent in the use of the implicit solvation model. One can argue that the actual simulation temperatures are below the melting points of the protein models used in the present study, since most of the trajectories led to successful predictions of the native structures via fast reversible folding. It needs to be mentioned that the time scales of our simulations may be different from the actual physical time scales since the GB implicit solvation model lacks the friction existing in a realistic water model. We have not included friction terms in the equation of motions for MD simulations. Therefore, the results of the present simulations may not be suitable for quantitative estimation of the folding kinetics. It should also be noted that some of the folding simulations were started from the fully stretched conformation, which may not represent experimentally accessible unfolded states.[21]

2. Model and Simulation Methods

We have performed NVT molecular dynamics simulations by an AMBER7 package[22] using an all-atom AMBER force field (parm99) with the GB implicit water model by Tsui and Case.[23] The surface access area correction was made using linear combinations of pairwise overlap (LCPO) by Still et al.[24] The simulation time step was set to 1.5 fs, and each simulation was performed up to 1×10^7 time steps

(9) Jang, S.; Shin, S.; Pak, Y. *J. Am. Chem. Soc.* **2002**, 124, 4976.
(10) (a) Brescher, A.; Weber, K. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, 76, 2321. (b) Bottomley, S. P.; Popplewell, A. G. Scawen, M.; Wan, T.; Sutton, B. J.; Gore, M. G. *Protein Eng.* **1994**, 7, 1463. (c) McKnight, C. J.; Doering, D. S.; Matsudaira, P. T.; Kim, P. S. *J. Mol. Biol.* **1996**, 260, 126. (d) McKnight, C. J.; Matsudaira, P. T.; Kim, P. S. *Nat. Struct. Biol.* **1997**, 4, 180. (e) Bai, Y.; Karimi, A.; Dyson, H. J.; Wright, P. E.. *Protein Sci.* **1997**, 6, 1449. (f) Myers, J. K.; Oas, T. G. *Nat. Struct. Biol.* **2001,** 8, 552. (g) Wang, M.; Tang, Y.; Sato, S.; Vugmeyster, L.; MacKnight. C. J.; Raleigh, D. *J. Am. Chem. Soc.* **2003**, 125, 6032.
(11) (a) Duan, Y.; Kollman, P. A. *Science* **1998**, 282, 740. (b) Duan, Y.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, 95, 9897.
(12) (a) Shen, M.; Freed, K. F. *Proteins.* **2002**, 49 439. (b) Zagrovic, B.; Snow, C. D.; Shirts, M.; Pande, V. S. *J. Mol. Biol.* **2002**, 323, 927. (c) Srinivas, G.; Bagchi, B. *J. Chem. Phys.* **2002**, 116, 8579. (d) Fernandez, A.; Shen, M.; Colubri, A.; Sosnick, T. R.; Berry, S.; Freed, K. F. *Biochemistry,* **2003**, 42, 664.
(13) (a) Berriz, G. F.; Shakhnovich, E. I. *J. Mol. Biol.* **2001**, 310, 673. (b) Zhou, Y.; Karplus, M. *J. Mol. Biol.* **1999**, 293, 917. (c) Kolinski, A.; Galazka, W.; Skolnick, J. *J. Chem. Phys.* **1998**, 108, 2608. (d) Shea, J.-E.; Onuchic, J. N.; Brooks, C. L., III. *J. Chem. Phys.* **2000**, 113, 7663. (e) Favrin, G.; Irbäck, A.; Wallin, S. *Proteins* **2002**, 42, 99. (f) Kussell, E.; Shimada, J.; Shakhnovich, E. I. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99, 5343
(14) Alonso, D. O. V.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, 97, 133.
(15) (a) Guo, Z.; Brooks, C. L., III; Boczko, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **1997** 94, 10161. (b) Boczko, E. M.; Brooks, C. L., III. *Science* **1995**, 21, 393.
(16) Zhou, Y.; Karplus, M. *Nature* **1999**, 401, 400.

(17) Smith, A. V.; Hall, C. K. *J. Chem. Phys.* **2000** 113, 9331.
(18) (a) Linhananta, A.; Zhou, Y. *J. Chem. Phys.* **2002**, 117, 8983. (b) Linhananta, A.; Zhou, H.; Zhou, Y. *Protein Sci.* **2002**, 11, 1695.
(19) Ghosh, A.; Elber, R.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99, 10394.
(20) (a) Elber, R.; Shalloway, D. *J. Chem. Phys.* **2000**, 112, 5539. (b) Elber, R.; Ghosh, A.; Cardenas, A. *Acc. Chem. Res.* **2002**, 35, 396.
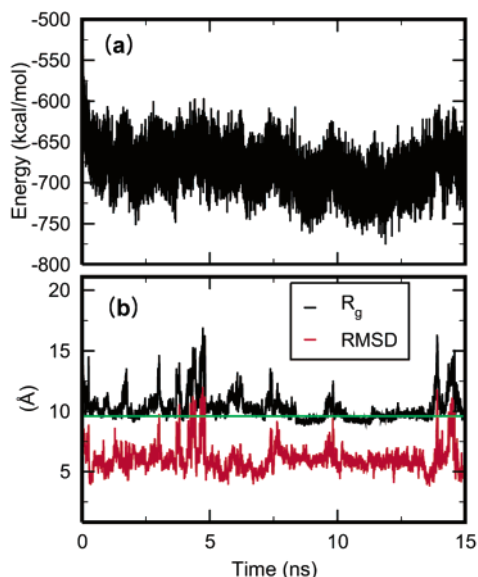(21) Shortle, D.; Ackernab, M. A. *Science* **2001** 293, 487.

Figure 1. Time course of (a) potential energy, (b) radius of gyration (black line), and main-chain rmsd (red line) for a representative simulation trajectory of the villin headpiece (HP-36). The green line represents the radius of gyration of the native state.

(15.0 ns). The bond distances between heavy and hydrogen atoms were fixed with the SHAKE algorithm. We used the Berendsen thermostat[25] for temperature control with coupling parameter 1.0 ps, and the simulation temperature was set to 400 K. In our simulations, the initial conformation was prepared as a fully stretched chain of a given amino acid sequence. Local energy minimization was performed with the initial conformation before starting the MD simulations. Several independent MD simulations were performed, and the conformations were saved at every 250 steps (0.375 ps) for analysis. The N terminal and C terminal of the protein were patched with acetyl groups and amine groups, respectively.

3. Results: Villin Headpiece (HP-36)

To characterize the "folded" structure of HP-36 we have performed 15 ns MD simulation starting from the native structure at 300 K. The average rmsd was found to be about 4.3 Å with fluctuations ranging from 3.2 to 5.0 Å. We have performed nine independent MD simulations of 15 ns at 400 K for HP-36 starting from the extended structures. All the trajectories exhibited structural fluctuations during the simulation leading to nativelike structures. The nativelike folded structure (rmsd less than 4.0 Å) begins to appear around at 4.0 ns depending, on the trajectories. All the trajectories show reversible folding events during simulation times. The potential energy, radius of gyration (R_g), and the rmsd as a function of time for a representative trajectory are plotted in Figure 1. The R_g of the native state (green line) is 9.4 Å. The results for R_g showed that the overall shape of the HP-36 undergoes an expansion−contraction process during simulations. Compared with the folding simulations at the physiological temperature (300 K),[11] the time scales of such large structural changes are much shorter for the present
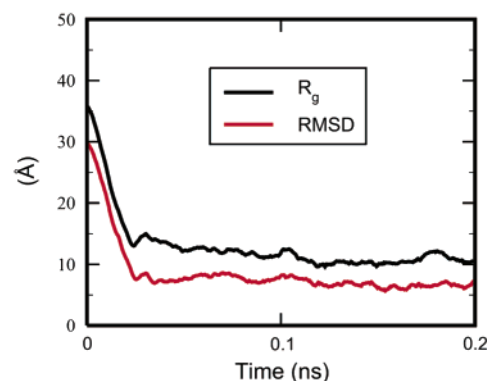
(22) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Wang, J.; Ross, W. S.; Simmering, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J. Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I. Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER7*; University of California: San Francisco, 2002.
(23) Tsui, V.; Case, D. A. *Biopolymers*, **2001**, 56, 275.
(24) Weiser, J.; Shenkin, P. S.; Still, W. C. *J. Comput. Chem.* **1999**, 20, 217.
(25) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, 81, 3684.

Figure 2. Radius of gyration (R_g: black line) and rmsd (red line) at the early stage of folding simulation for the villin headpiece (HP-36), illustrating fast collapse of the extended conformation.
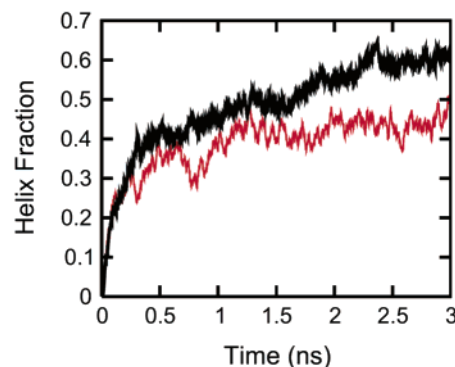

Figure 3. Average fraction of helical conformation for the two helix bundles (black line: 1st helix; red line: 2nd helix) of the villin headpiece (HP-36) as a function of time.

simulations at higher temperature (400 K). It has been noticed that the R_g of the simulation trajectory is frequently lower than that of the native state, which is possibly due to the absence of explicit water molecules in our simulations. The time-development of rmsd is closely correlated with R_g. It has been shown by many authors that the folding sequence of the villin headpiece involves an initial rapid collapse followed by slow structural adjustment.[11,12] Figure 2 shows R_g and rmsd at the early stage of the simulation, and it can be seen there is rapid decrease of rmsd and R_g. The initial extended conformation is quickly converted to a structure with R_g < 15 Å. Duan et al. found that the fully unfolded forms of the same protein after 1-ns unfolding simulations at 1000 K retained a turn feature with values of R_g less than 15 Å.[11b]

We have simulated 29 more trajectories up to 3.0 ns and calculated average fractions of helix contents of helix bundles using a total of 38 independent trajectories. Although the native structure appears to have three helices, a detailed analysis by DSSP[26] shows two helix bundles. Figure 3 displays the fractional native helical contents for the two helix bundles during the folding simulations. The helices started to form from a very early stage of the simulation and reached up to a 60% level after 2 ns. There appears to be no appreciable difference in the time scales of helix formation for the two helix bundles, while the relative stabilities show some differences. After the helix formation is completed at around 2 ns, the helical content remains almost the same for the rest of the simulation, indicating

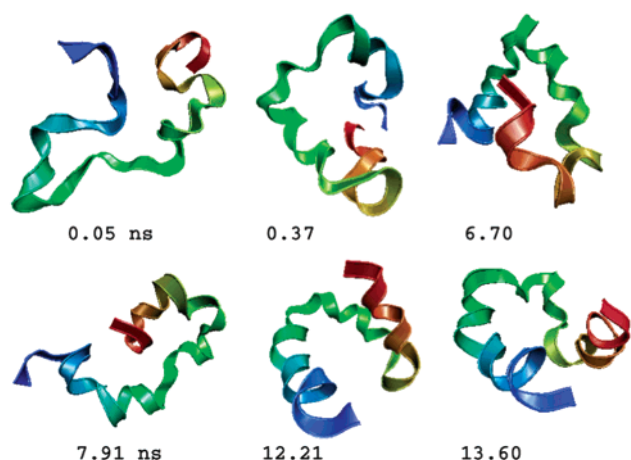(26) Kabsch, W.; Sander, C. *Biopolymers* **1983**, 22, 2577.

Figure 4. Snapshots of representative folding trajectories for the villin headpiece (HP-36). The backbones of the protein are shown in the ribbon representation with the C terminus on the left side. The numbers represent simulation times in ns.
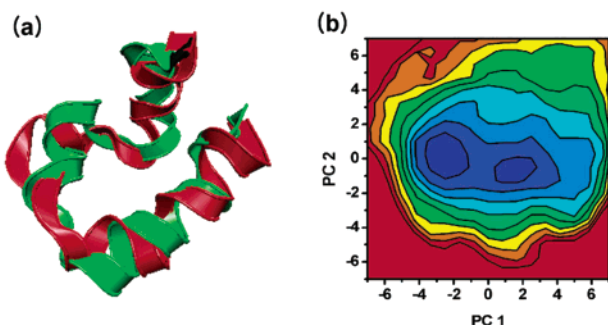

Figure 5. (a) Superimposed image of simulated (green) and native (red) structures of the villin headpiece (HP-36). The rmsd is 3.92 Å. (b) Free energy landscape calculated from the principal component analysis for the folding simulations of HP-36.

that once the helices are formed only reorientations of the helix bundles are occurring. During the folding simulations, we observed several nativelike conformations, i.e., those with rmsd less than 4.0 Å. Figure 4 illustrates sequences of conformations in a folding trajectory. At 6.70 ns, we obtained the lowest rmsd value (3.78 Å), and after unfolding to a random coil (7.91 ns) another nativelike conformation appeared at 12.21 ns (rmsd 3.9 Å). The superimposed image of the simulation structure with the native structure is shown in Figure 5a. The rmsd calculation we have used is based on a protein backbone match. To analyze the nativelike features of conformations, we also calculated the side-chain native contacts as defined by Brooks et al.[15a] Structures with low rmsd (less than 4.1 Å) were found to retain about 85% of the native contacts. The minimized energies of several low rmsd conformations were found to be lower (few kcal/mol) than the minimized energy of the native structure.

We have performed principal component (PC) analysis to examine the conformational sampling during the folding simulations. The PC analysis is based on the diagonalization of the covariance matrix built from positional fluctuations of the backbone atoms in MD trajectories. We added seven more folding simulations of 15 ns at 400 K, starting from arbitrary unfolded conformations. The PC analysis was performed using the total 80 000 frames from the 16 trajectories of 15-ns simulations. The eigenvalues of the first two PC components (PC 1 and PC 2) comprise about 60% of the total fluctuations.
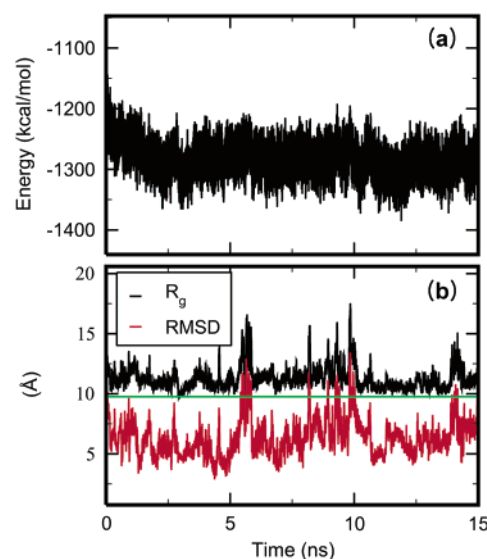
Figure 6. Time course of (a) potential energy, (b) radius of gyration (black line), and main-chain rmsd (red line) for a representative simulation trajectory of the BpA. The green line represents the radius of gyration of the native state.

We have used these two components to evaluate the "pseudo" free energy landscape. The free energy is calculated by $-RT \ln(P/P_0)$ where R is the gas constant, T is the temperature, P is the probability of finding the conformation at the given PC values (PC 1 and PC 2), and P_0 is the normalized probability. Figure 5b displays contours of the free energy landscape projected onto the first two principle component basis vectors. The convergence of PC analysis was checked by comparing the results from calculations using 9 and 16 trajectories, respectively. The free energy surface consists of two basins, indicating that HP-36 shows two-state thermodynamic behavior. PC analysis for the equilibrium simulation of the native structure at 300 K shows a similar free energy landscape with two basins. Although the exact positions are somewhat different, the minimum free energy regions of the native structure at 300 K are within the major basins of free energy landscape for folding simulations at 400 K. It can be concluded that folding simulations at 400 K cover wider regions of free energy landscape, including the native state conformations.

4. Results: Protein A (BpA)

The folding simulations of BpA show results similar to those of HP-36. All nine trajectories starting from the extended structures exhibit reversible folding and unfolding events during the simulation time (15 ns), and the initial stage of folding exhibits a rapid collapse of the extended structure. The time evolutions of potential energy, R_g, and rmsd for a representative trajectory are displayed in Figure 6. The behavior of R_g and rmsd illustrates the expansion−contraction processes involved in partial unfolding/refolding sequences. Somewhat different from HP-36, R_g of the native state forms the lower boundary, with conformations of R_g smaller than that of the native state rarely found. Very small rmsd (<3.0 Å) appear several times during the folding simulations with the lowest rmsd of 2.9 Å. Equilibrium MD simulation of the native structure at 300 K shows rmsd fluctuations ranging from 3 to 6 Å with average value of 4.2 Å. It is found that structures with low rmsd (less than 4.1 Å) contain about 87% of the native side-chain contacts.
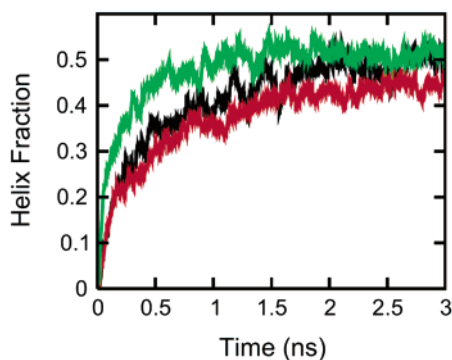
Figure 7. Average fraction of helical conformation for Helix 1 (black line), Helix 2 (red line), and Helix 3 (green line) of BpA as a function of time.
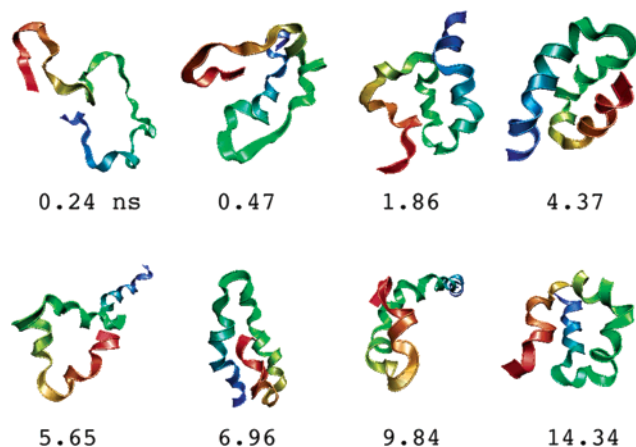

Figure 8. Snapshots of representative folding trajectories for BpA. The backbones of the protein are shown in the ribbon representation with the C terminus on the left side. Helix 1, Helix 2, and Helix 3 are in red, green, and blue, respectively. The numbers represent simulation times in ns.

As for HP-36, the minimized energies of several low rmsd conformations of BpA were found to be lower (few kcal/mol) than the minimized energy of the native structure.

There are several studies on the relative stabilities of the three helices of BpA. Boczko and Brooks[15b] concluded from the folding free energy study that the third helix is the least stable thermodynamically and folds last. In contrast, Ghosh et al.[19] showed that the third helix is the most stable and the formation of the third helix possibly precedes that of the other two helices, consistent with the experimental observations.[10b,e] The simulations by Alonzo and Daggett[14] and Linhananta and Zhou[18] showed similar results. Figure 7 displays the time evolution of the fractional native helical contents for the three helices averaged over 38 trajectories. It has been noticed that Helix 3 is clearly formed first in the early stage of folding and the overall average helix fraction is in the order of Helix 3 > Helix 1 > Helix 2, which is consistent with the results of the previous study by Ghosh et al.[19] As in the case of HP-36, the formation of the helix-bundle is completed around 2 ns, and then reorientations of the helices lead to the stable nativelike structure. Figure 8 illustrates sequences of conformations in a folding trajectory. The folding pathway illustrated in our simulations is similar to the "dominant pathway" proposed by Linhananta and Zhou,[18a] where the folding of BpA is initiated by rapid helical formation followed by the formation of an intermediate Helix 2–Helix 3 domain before the final docking of Helix 1. We also observed intermediate structures of "domain swapped"
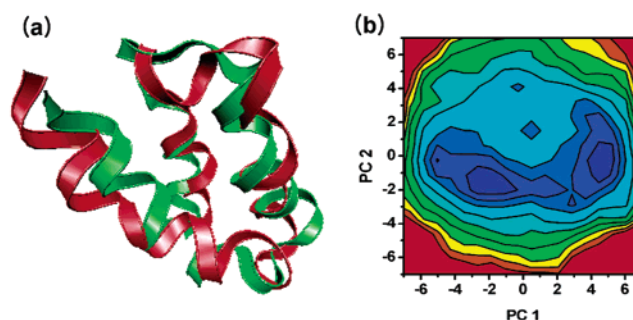

Figure 9. (a) Superimposed image of simulated (green) and native (red) structures of BpA. The rmsd is 3.92 A. (b) Free energy landscape calculated from the principal component analysis for the folding simulations of BpA.

forms,[16,18a] where the two helices are packed together and the other helix is extended from the domain.

The superposition of the low-rmsd conformation obtained from the folding simulation with the native structure is shown in Figure 9a. The free energy landscape of BpA resulting from the principal component (PC) analysis of the folding simulations is shown in Figure 9b. As in the case of HP-36, the free energy surface of BpA contains two major basins, illustrating a two-state thermodynamic behavior. The free energy landscape of BpA shows rather broad native basins, suggesting flexible structures of the protein.[15a] PC analysis for the equilibrium simulation of the native structure at 300 K shows that the major basins of free energy landscape at 400 K includes the minimum free energy regions of the native structure.

5. Concluding Remarks

We have demonstrated the feasibility of ab initio simulations of fast folding helix-bundle proteins. Spontaneous formation of nativelike structures for the 36-residue villin headpiece and the 46-residue fragment B of Staphylococcal protein A was observed during folding simulations at 400 K using a GB implicit solvent model with an all-atom-based force field. Under such mild unfolding conditions, the MD trajectories showed rapid folding, starting from the extended conformation, followed by partial unfolding and refolding processes within tens of ns simulations. The implicit solvent model combined with high-temperature MD makes it possible to perform direct folding simulations of medium-sized proteins by reducing the computational requirement tremendously.

Folding simulations for the villin headpiece have provided results consistent with those from previous simulation studies. The folding process of BpA described by our simulations mostly agrees with the overall picture developed by recent studies. In the early stage of folding for the helix-bundle protein, initial hydrophobic collapse and rapid formation of helices played important roles. For BpA, Helix 3 forms first in the early stage of folding and exhibits the higher stability compared with the stabilities of the other helices. The free energy profiles calculated form the folding simulations suggested that both of the helical proteins show a two-state thermodynamic behavior and BpA exhibits rather broad native basins. Comparisons with native-state simulations show that the free energy landscapes at 400 K contain the minimum free energy regions of the proteins at 300 K.

Several different models have been proposed for the detailed folding mechanism of BpA. Earlier kinetic refolding experiments

showed that BpA folds extremely rapidly by a two-state mechanism without formation of stable, partly folded intermediate.[10e] Free energy profiles calculated by Guo et al. suggested that folding of BpA involved on-pathway intermediates formed by initial formation of transient helical structure and subsequent collapse.[15] The diffusion−collision model has been successful in explaining many features of protein folding kinetics for helical proteins.[27,28] Zhou and Karplus[16] showed that, depending on the difference between the strength of native and nonnative contacts, the folding kinetics of a model helical protein is changed from a diffusion−collision mechanism to one that involves simultaneous collapse and secondary structure formation with the presence of on-pathway intermediates. The results of NMR experiments for BpA were found to be consistent with the diffusion−collision model, with the preorganization of elements of secondary structure in the unfolded protein.[10e] The study by Berriz and Shakhnovich emphasized the importance of the organization of the hinge between the helices, suggesting that the simple diffusion−collision mechanism is insufficient to explain the folding kinetics of BpA.[13a] Linhananta et al. showed that a locally interacting loop could promote folding or serve as the hinge region for domain swapping.[18b] The recent study by Ghosh et al. suggested that early events of BpA folding involve the formation of small nuclei with native hydrogen bonds mostly in the third helix and subsequent collapse into a molten-globule state.[19] Their picture is different from a pure hydrophobic collapse mechanism or a model assuming early formation of substantial secondary structures. In the case of forming on-pathway intermediates, the final tune-up processes toward the native state involve the concomitant formation of secondary and tertiary structures.[15,19] The results of the present folding simulations seemed to be consistent with a diffusion−collision mechanism for BpA, where significant portion of helices are formed in the early stage of folding after hydrophobic collapse. Free energy landscape from the principal component

analysis suggested that folding pathways of BpA are rather mobile, involving apparently two different thermodynamic states. As discussed previously, our simulations have limitations in the quantitative analysis of exact folding kinetics. One cannot rule out other possibilities in the detailed folding mechanism of BpA on the basis of the present simulations.

As shown in the case of villin headpiece, the initial extended conformation is quickly converted to a structure having the characteristics of unfolded conformations described in the previous studies. It may be argued that the folding scenarios suggested by simulations at 400 K could be different from the folding pathway at the physiological temperature. Similar criticism has been made concerning thermal denaturation simulations at high temperatures. Systematic unfolding simulation studies on many proteins with different architectures have showed mostly reliable agreement between the simulations and experiments, thus demonstrating that the overall pathway of unfolding is independent of temperature and the process is merely accelerated by increasing temperature.[29] We can conclude that our simulations at moderately elevated temperatures can provide plausible folding pathways, although other folding pathways are certainly possible. One needs to understand how the results of high-temperature folding simulations are related with detailed folding kinetics at physiological temperatures, which will be the subject of future studies.

(27) Islam, S. A.; Karplus, M.; Weaver, D. L. *J. Mol. Biol.* **2002**, 318, 199.

(28) Mayor, U.; Guydosh, N. R.; Johnson, C. M.; Grossmann, J. G.; Sato, S.; Jas, G. S.; Freund, S. M. V.; Alonso, D. O. V.; Daggett, V.; Fersht, A. R. *Nature* **2003**, 421, 863.
(29) Daggett, V. *Acc. Chem. Res.* **2002** 35, 422.