ISSN 1674-1056
http://cpb.iphy.ac.cn
http://iopscience.iop.org/cpb
中国物理 B
Chinese Physics
中国物理
Chinese Physics B

A network of conformational transitions in an unfolding process of HP-35 revealed by high-temperature MD simulationand a Markov state model

D, an Shao(邵丹丹), Kaifu Gao(高恺夫)
Citation: Chin. Phys. B . 2018, 27(1): 018701. **doi:** 10.1088/1674-1056/27/1/018701

Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Computational study of non-catalytic T-loop pocket on CDK proteins for drug development

Huiwen Wang(王慧雯), Kaili Wang(王凯丽), Zeyu Guan(管泽雨), Yiren Jian(简弋人), Ya Jia(贾亚), Fatah Kashanchi, Chen Zeng(曾辰), Yunjie Zhao(赵蕴杰)
Chin. Phys. B . 2017, 26(12): 128702. **doi:** 10.1088/1674-1056/26/12/128702

A damping boundary condition for atomistic-continuum coupling

Jie Zhang(张杰), Kiet Tieu, Guillaume Michal, Hongtao Zhu(朱洪涛), Liang Zhang(张亮), Lihong Su(苏利红), Guanyu Deng(邓关宇), Hui Wang(王辉)
Chin. Phys. B . 2017, 26(6): 068702. **doi:** 10.1088/1674-1056/26/6/068702

Computational design of proteins with novel structure and functions

Wei Yang(杨为) and Lu-Hua Lai(来鲁华)
Chin. Phys. B . 2016, 25(1): 018702. **doi:** 10.1088/1674-1056/25/1/018702

Coupling of two-dimensional atomistic and continuum models for dynamic crack

Ren Guo-Wu, Tang Tie-Gang
Chin. Phys. B . 2014, 23(11): 118704. **doi:** 10.1088/1674-1056/23/11/118704

SPECIAL REVIEW — Soft matter and biological physics

A network of conformational transitions in an unfolding process of HP-35 revealed by high-temperature MD simulation and a Markov state model*

Dandan Shao(邵丹丹) and Kaifu Gao(高恺夫)†

Institute of Biophysics and Department of Physics, Central China Normal University, Wuhan 430079, *China*

(Received 9 July 2017; revised manuscript received 1 October 2017; published online 22 December 2017)

An understanding of protein folding/unfolding processes has important implications for all biological processes, including protein degradation, protein translocation, aging, and diseases. All-atom molecular dynamics (MD) simulations are uniquely suitable for it because of their atomic level resolution and accuracy. However, limited by computational capabilities, nowadays even for small and fast-folding proteins, all-atom MD simulations of protein folding still presents a great challenge. An alternative way is to study unfolding process using MD simulations at high temperature. High temperature provides more energy to overcome energetic barriers to unfolding, and information obtained from studying unfolding can shed light on the mechanism of folding. In the present study, a 1000-ns MD simulation at high temperature (500 K) was performed to investigate the unfolding process of a small protein, chicken villin headpiece (HP-35). To infer the folding mechanism, a Markov state model was also built from our simulation, which maps out six macrostates during the folding/unfolding process as well as critical transitions between them, revealing the folding mechanism unambiguously.

Keywords: molecular dynamics simulation, Markov state model, folding/unfolding, HP-35

PACS: 87.14.E−, 87.15.ap, 87.15.Cc **DOI:** 10.1088/1674-1056/27/1/018701

1. Introduction

Despite decades of active research, protein folding remains one of the most important unsolved problems in molecular biology. An understanding of protein folding/unfolding has important implications for all biological processes, including protein degradation, protein translocation, aging, and diseases.[1,2] Recently, there has been substantial insight regarding the knowledge of native states as well as the stable intermediate states of several proteins.[3–6] However, less information is available about the path followed by a protein to fold to the native state and the characteristics of intermediates along the way. These intermediates are mostly short-lived and thus very difficult to trace.[7]

All-atom molecular dynamics (MD) simulations are uniquely suitable for the study of protein folding because of their atomic level resolution and accuracy. However, even for small and fast-folding proteins, all-atom MD simulations of protein folding still presents a great challenge, because the timescale of protein folding with all-atom MD simulations will exceed the computational capabilities of modern computers. The timescale of protein folding is on the order of 1–106 ms,[8] and the currently known fastest folding proteins fold over 0.7–1.0 μs, while only recently technological progress has extended all-atom MD simulations to ms timescales.[9–14]

An alternative way to overcome the challenge above is to study unfolding process using MD simulations at high temperature. High temperature provides more energy to overcome energetic barriers to unfolding. Moreover, there are numerous advantages to study unfolding rather than folding. Such simulations begin from a well-defined starting point, a crystal or NMR structure, which improves the odds of sampling experimentally relevant regions of conformational space, while simulations from an arbitrary extended structure present too many conformational possibilities, and the search problem becomes insurmountable. Furthermore, the system is less likely to become trapped in a local minimum during unfolding, which is common in attempts to simulate the folding reaction. Another advantage of studying unfolding is that the full reaction coordinate from the native to denatured states can be explored. Since the mechanism of folding can be probed from both directions, information obtained from studying unfolding can be shed light on the mechanism of folding.[1,15]

Markov state models (MSMs) are a powerful approach to map out conformational transitions as well as relevant conformational states. Rather than solely based on geometric criteria, MSMs are kinetic network models that can be used to identify metastable states and calculate their equilibrium thermodynamics and kinetics.[16–20] The models focus on metastable regions of phase space and partition conformational space into a number of metastable states such that intra-state transitions

*Project supported by the National Natural Science Foundation of China (Grant Nos. 11175068 and 11474117) and the Self-determined Research Funds of CCNU from the Colleges Basic Research and Operation of MOE, China (Grant No. 230-20205170054).

†Corresponding author. E-mail: gaokaifu@mail.ccnu.edu.cn

© 2018 Chinese Physical Society and IOP Publishing Ltd

are fast but inter-state transitions are slow. This separation of timescales ensures that an MSM is Markovian and allows an MSM built from short simulations to model long timescale events.[21] Totally, MSMs can provide a reduced view of the ensemble of spontaneous fluctuations that the molecule undergoes at equilibrium, as well as the population and transition rates between key conformational states,[22–24] which has been demonstrated by many recent studies of protein folding[25,26] and protein dynamics.[27]

35-residue villin headpiece subdomain (HP-35, PDB entry 1vii)[28] is one of the smallest proteins that can fold autonomously. It contains only natural amino acids without any disulfide bonds, oligomerization, or ligand binding for stabilization.[29] Although it is very small, HP-35 comprises three α-helices (Fig. 1) as well as one compact hydrophobic core, thereby bearing characteristics of large folded proteins.[30] The three α-helices are denoted as Helix 1 (Asp-4 to Lys-8), Helix 2 (Arg-15 to Phe-18), and Helix 3 (Leu-23 to Glu-32). The biological activity is believed to be centered on Helix 3.[31] With a folding time detected to be about 4 μs,[32] it is feasible to simulate the folding of HP-35 using MD simulations at current computer power. HP-35 has been used as model protein in many studies for investigating the pathways and landscapes of protein folding.[29,33–38]

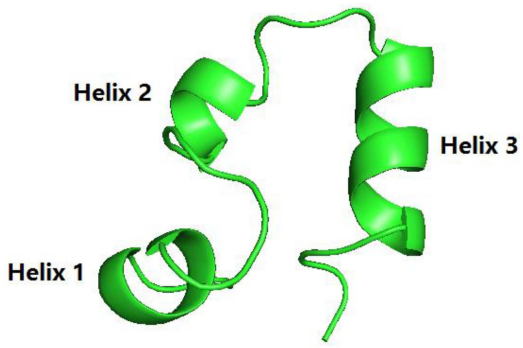


Fig. 1. (color online) The native structure of HP-35. The three helices are Helix 1 (Asp-4 to Lys-8), Helix 2 (Arg-15 to Phe-18), and Helix 3 (Leu-23 to Glu-32).

In the present study, a 1000-ns MD simulation at high temperature (500 K) was performed to investigate the unfolding process of HP-35. From its native state, fully extended conformations were obtained by our 500-K MD simulation. Then, based on a MSM, the conformational dynamics during the unfolding process is modeled as transitions between kinetically metastable states, giving a reduced kinetic network between multiple conformational states of the protein. We revealed the key intermediate states along the unfolding pathway as well as the time scale associated with this process. This study reveals intermediate states as well as transition pathways during the unfolding process of HP-35 and hence greatly enhances our understanding of the folding/unfolding mechanism of the protein.

2. Methods

2.1. MD simulation

The simulation was performed using the PMEMD simulation module for graphics processing units (pmemd.cuda) in the MD program package AMBER14[39–41] and the AMBER ff14SB force field.[42] The initial coordinates for the MD simulation were taken from the NMR structure of HP-35 (PDB entry 1vii (Fig. 1)). The protein molecule was placed in an octahedral TIP3P water box,[43] with the edge of the water box located at least 10 Å from the protein atoms. The system consisted of 596 protein atoms, 2 Cl- ions, and 28200 solvent atoms under periodic boundary conditions. To remove the bad contacts, the initial model was subjected to two rounds of energy minimization: first 1000 steps of steepest descent followed by 1000 steps of conjugate gradient minimization with the protein atoms constrained by a harmonic potential with a force constant of 100 kcal/mol, and then 2000 steps of steepest descent and 3000 steps of conjugate gradient minimization without any constraints. The system was heated at a constant volume from 0 to 500 K over 100 ps. The MD simulations were then run under isothermal–isobaric conditions (NPT) with the temperature maintained at the set temperatures by Langevin's thermostat,[44] with a collision frequency of 1.0 ps^{-1} and the pressure maintained at 1.0 atm using isotropic positional scaling with a pressure relaxation time of 2.0 ps. The particle-mesh Ewald method[45] was used to estimate the long-range electrostatic energies. The non-bonded cutoff was set to 10 Å. All bonds involving hydrogen atoms were fixed to their equilibrium values using the SHAKE algorithm[46] to allow a 2.0 fs time step. The snapshots were recorded at 1 ps intervals.

2.2. Secondary structure assignment

DSSP program[47] were used to measure the change in the secondary structure of the protein during its unfolding process. The secondary structure types were assigned based on H-bonding patterns: n-turn with an H- bond between the CO of residue i and the NH of residue $i+n$, where $n = 3, 4, 5$; bridge with H bonds between residues not near each other in sequence. These two types of pattern essentially exhaust all backbone-backbone H bonds. Repeating 4-turns define α-helices and repeating bridges define β-structure are in good agreement with intuitive assignments. All other occurrences of the basic patterns provide an interesting survey of 3_{10}-helices, π-helices, single turns, and single β-bridges.

2.3. Markov state modeling

The program MSMBuilder3[48] was employed to build the MSM of the conformational landscape of HP-35 in four steps. (i) Each frame of the MD trajectory was transformed into a vector composed by the dihedral angles phi and psi, by which different conformations can be discriminated. (ii) Using time-structure independent components analysis (tICA),[49]

the conformational space was reduced to six dimensions. On the basis of the tICA projections, the conformations were clustered into 1000 states (microstates) using the mini-batch k-medoids clustering algorithm. The conformations in the same state are geometrically so similar that one can reasonably assume that they can interconvert much faster within the state than between states. (iii) A transition matrix was constructed between these microstates at a proper lag time, in which the implied time scales are converged so that the microstate model can be demonstrated as Markovian. The free energy of each microstate was also generated as $F(x) = -\log \pi_{i(x)}/kT$, where $\pi_{i(x)}$ is the stationary probability of state x estimated from the first eigenvector of the MSM transition matrix.[50] (iv) A macrostate model was obtained by that kinetically related microstates were lumped into macrostates using the robust perron cluster analysis (PCCA+) algorithm. Moreover, to elucidate the transitions between these macrostates, a macrostate transition matrix was also constructed at a proper lag time in which the macrostate model is Markovian. From this matrix, the transition path theory (TPT)[51] was applied to extract the mean first passage time (MFPT) and net flux between each pair of macrostates as well as the highest-flux pathway. The free energies and MSM network were plotted using the program MSMExplorer.[52]

2.4. Principle component analysis (PCA)

Using the cpptraj module of AmberTools,[39] PCA was performed on the C_α atoms of HP-35. The coordinates of the trajectory were superposed by means of a least-squares fit with the first snapshot as a reference. Then the covariance matrix of the atomic fluctuations C[53,54] was calculated, and its eigenvalues λ_i and eigenvectors L_i were yielded by diagonalization of the matrix. These eigenvectors represent the principal modes of motion, and the eigenvalues represent the mean-square fluctuations along the eigenvector coordinates.

The percentage v_i of the total variance contained in a given eigenvector L_i is given by

$$v_i = \frac{\lambda_i}{\sum_{k=1}^{3N} \lambda_k}, \tag{1}$$

where v_i measures the contribution of a given principle component (PC) to the overall protein fluctuations. The cumulative contribution from a set of PCs is

$$V_n = \sum_{i=1}^{n} v_i, \quad V_{3n} = 1. \tag{2}$$

Typically, the first few PCs capture most of the intramolecular fluctuation of a protein.[55–58]

3. Results and discussion

3.1. RMSD and Rg during the unfolding simulation

Since our simulation was run at high temperature (500 K), the root-mean- square deviation (RMSD) increase very rapidly at the early stage of the MD simulation: from 0 to 8 Å in less than 2 ns (see Fig. 2(a)). However, such increasing only lasts for several ns. During the most time of the simulation, the values of the RMSD fluctuate, indicating the existence of intermediate states.

As shown in Fig. 2(b), the radius of gyration (R_g) also exhibits a rapid increase and then fluctuates during the most time of the simulation. As the radius is the largest around 600 ns, the protein represents fully extended conformations around this point. Therefore, our MD simulation at least includes an entire unfolding process.

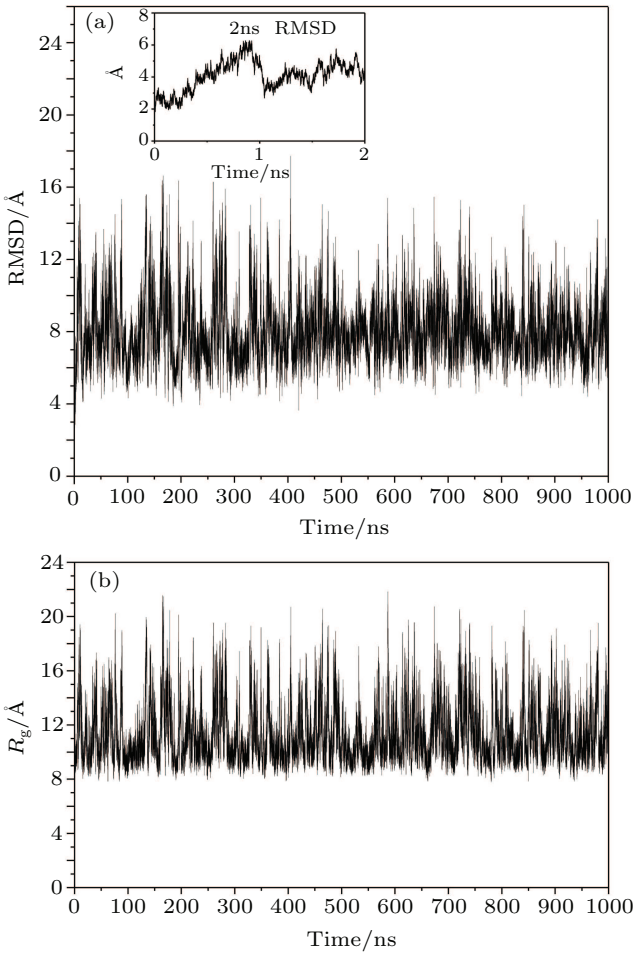


Fig. 2. (a) Time evolution of backbone RMSD during the MD simulation. (b) The Radius of gyration derived from the MD simulation.

3.2. The intactness of the hydrophobic core

The extent of folding/unfolding is best measured by the intactness of the hydrophobic core. Here, the intactness of the hydrophobic core is represented by three pairwise distances of the atoms that compose the hydrophobic core: the distance between the HZ atom of PHE7 and the HB2 atom of PHE18, the distance between the HE1 atom of PHE11 and the HB3 atom of LEU29, as well as the distance between the HZ atom of PHE18 and the HG3 atom of LYS30 (marked as 1, 2, and 3 in Fig. 3(a)). Three distances in the initial structure (the folded state) are respectively 1.67 Å, 2.82 Å, and 1.82 Å, indicating that in the folded state the hydrophobic residues are close to

each other, and the hydrophobic core are fully intact. However, in our unfolding simulation, these distances increases very rapidly to about 30 Å in only several ns and then greatly fluctuate in very large ranges (less than 3 Å to more than 40 Å, Figs. 3(b)–3(d)). In the most simulation time, the distances are larger than 15 Å, which means that the hydrophobic core fully collapse and the protein extends. Only time to time the distances decreases to less than 5 Å, but it collapses again very rapidly.

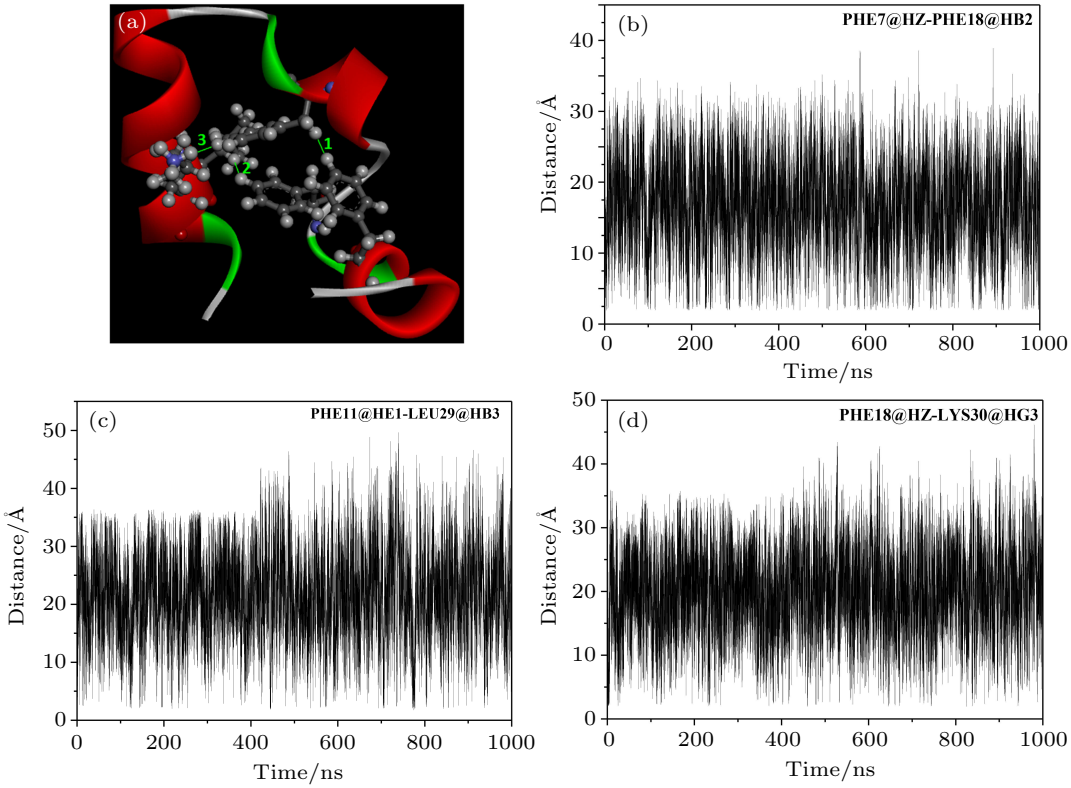


Fig. 3. (color online) (a) Three pairwise distances of the atoms that compose the hydrophobic core. The distance between the HZ atom of PHE7 and the HB2 atom of PHE18 is marked as 1, the distance between the HE1 atom of PHE11 and the HB3 atom of LEU29 is marked as 2, and the distance between the HZ atom of PHE18 and the HG3 atom of LYS30 is marked as 3. (b)–(d) The time evolution of these three distances.

3.3. Secondary structures of the protein

To elucidate the change in the secondary structure, all types of secondary structure elements were extracted from our simulation trajectory data, which is shown in Fig. 4.

As mentioned above, HP-35 consists only three helices with turn and bend structures between them. As shown in Fig. 4(b), the residues 1, 2, 35, and 36 at the N and C terminals are so flexible that no specific structures were represented during the unfolding process. In the three helices, Helix 3 lasts for the longest time and in contrast Helix 2 lasts much shorter, indicating that the unfolding of Helix 2 is much easier. Especially the residue 18 is the first residue to unfold. As Helix 3 is believed to play the biological function,[33] Helix 3 is the last one to unfold. When Helix 3 is fully unfolded, the protein is fully extended. During the unfolding process, the bend and the turn structures are relatively stable, but the three helices fluctuate dramatically. Some anti-parallel sheet conformations appears after 700 ns, but then they transit to other conformations very rapidly.

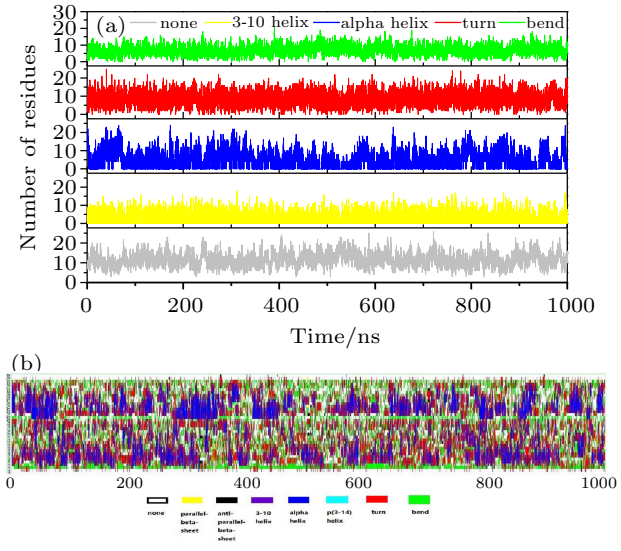


Fig. 4. (color online) The change of secondary structures of HP-35 during the unfolding process. (a) Time evolution of the number of residues with all types of secondary structures. The bend is marked in green, the turn in red, the α-helices in blue, and the 310 helices in yellow. (b) Time evolution of the types of secondary structures of each residue.

3.4. Markov state model of HP-35

To elucidate the conformational changes of HP-35, a kinetic network MSM was built from our simulation data, shown in Fig. 5. As mentioned in Section 2, the simulation data were first discretized into 1000 microstates. To verify that the microstate model is Markovian, the implied time scale was plotted as a function of the lag time, shown in Fig. S1. The implied time scales in the microstate model level are off at a lag time larger than 800 ps, implying that the microstate model is Markovian in this time range. So a lag time of 800 ps was applied to build the transition matrix in the microstate model (Fig. S1(a)). The free energy of each microstate was then estimated from the first eigenvector of the transition matrix and plotted in Figs. 5(a) and 5(b). Using the PCCA+ algorithm, six MSM macrostates were successfully extracted from the 1000 microstates. Once again, an 800 ps lag time was found to yield Markovian behavior for the macrostate model (Fig. S2(b)). So it was chosen to build the macrostate transition matrix.

As illustrated in Fig. 5(a), the partitioning of these six states is consistent with the energetic wells on the free energy map, which ascertains the robustness of the MSM analysis. The transition network between these six states is also unambiguously depicted by the MSM, which is shown in Fig. 5(c).

Our MSM provides novel insights into the conformational changes of HP-35 during the unfolding process. In these six metastable states, major conformational changes are detected in the three helix structures, which agrees well with the secondary structure analysis (Figs. 4(a) and 4(b)).

Figure 5(c) is the transition network between the six macrostates extracted from the MSM. In this network, the most important states are the states 3, 5, and 6. The states 3 and 5 correspond to the fully extended state. The states 6 includes the folded state and some partial folded states close to the folded state, which has the largest population. Here we infer the folding mechanism from the unfolding network. As shown in Fig. 5(c), from the fully extended (the states 3 and 5) to folded state (the states 6), there are two main pathways: the state 5 $\rightarrow$ the states 4 $\rightarrow$ the states 6 and the state 5 $\rightarrow$ the states 2 $\rightarrow$ the states 6. The pathway the state 5 $\rightarrow$ the states 4 $\rightarrow$ the states 6 has the highest probability flux. Along this pathway, the helices are first formed, and then these secondary structures get close to each other and collapse into the final tertiary structure, which fits well with the framework model[4] of protein folding. The other main pathway is the state 5 $\rightarrow$ the states 2 $\rightarrow$ the states 6 with the net probability flux only a little lower than the state 5 $\rightarrow$ the states 4 $\rightarrow$ the states 6. Although the net flux from the state 5 to the states 4 is about ten times as large as that from state 5 to the states 2, the total net flux of each pathway is determined by the slowest step, i.e., states 2 to states 6 and states 4 to states 6 here. Since the flux from states 2 to states 6 (8.82×10^{-4}) is close to that from states 4 to the states 6 (9.97×10^{-4}), the pathway state 5 $\rightarrow$ states 2 $\rightarrow$ states 6 is also very popular. Along this way, the hydrophobic core first collapses, so it can be described as the hydrophobic collapse model.[59,60] Besides

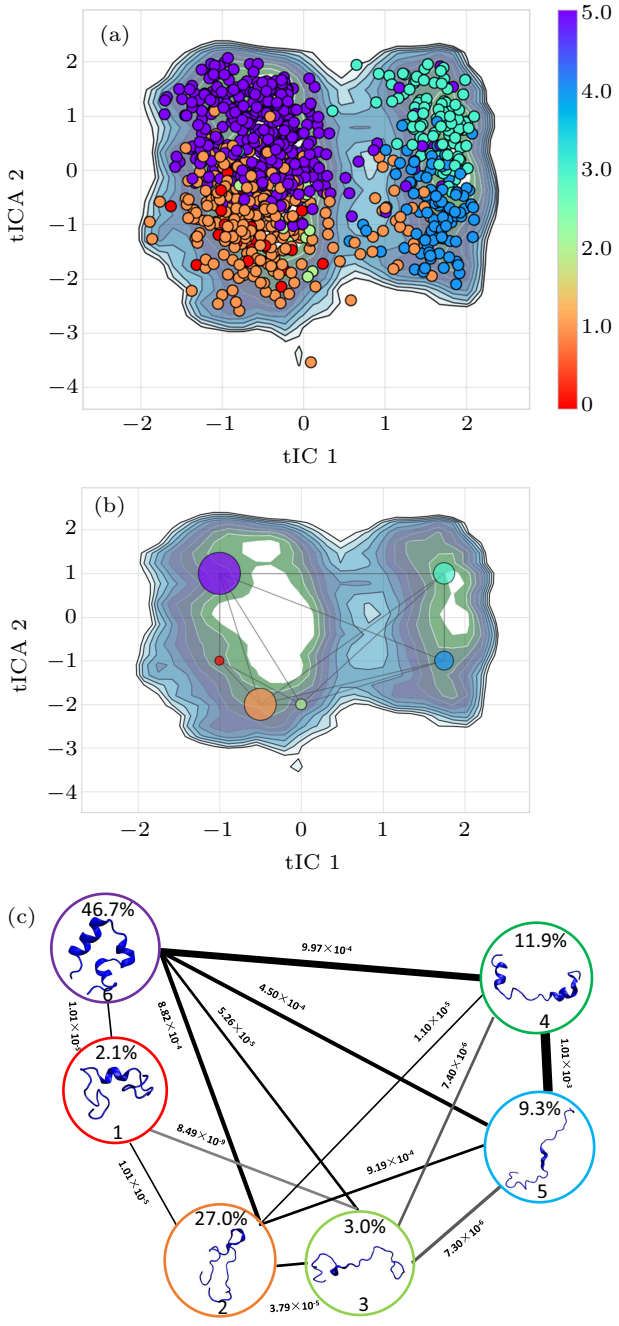


Fig. 5. (color online) MSM free energy map and high-flux pathway built from our unfolding simulation of HP-35. (a) The 1000-state microstate model, (b) the six state macrostate model extracted from the microstate model, and (c) the relative population of each macrostate and the net probability flux between each macrostate contributing to the transition from the fully extended state (the state 5) to folded state (the states 6). In panel (a), the circles represent the microstates and are colored according to their membership in the macrostate model. In panel (b), the circles represent the centers of the macrostates, with their sizes indicating the relative populations of the macrostates. In panels (b) and (c), the connecting lines between these circles mean the transitions between the macrostates, with the bold ones corresponding to the high-flux pathway from the fully extended state to the folded state. The free energy map is both shown in panels (a) and (b) as a contour plot.

them, with the net flux as half as the two main pathways, the pathway directly from state 5 to states 6 cannot be ignored.

3.5. PCA of the unfolding of HP-35

All C_α atoms were used to define the backbone conformation of HP-35 for PCA, 108 PCs from the 36 C_α atoms by the form of 108 eigenvectors and their associated eigenvalues. Consistent with the earlier researches, the first few PCs, especially the first three PCs, capture the most motion of the protein, whose eigenvalues are much larger than the rest (see Fig. 6). Hence, it is sufficient to analyze the first three PCs.

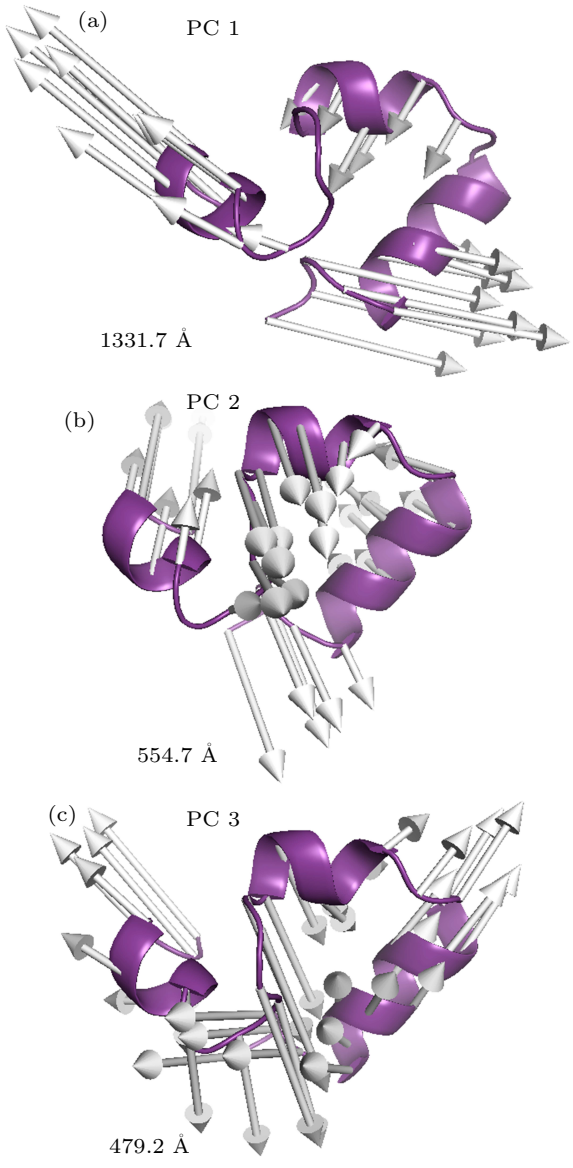


Fig. 6. (color online) Vector field representation and the corresponding eigenvalues of the first three PCs obtained from the MD simulation.

The vector field representation of the first three PCs obtained from the MD simulation as well as the corresponding eigenvalues is shown in Fig. 6. This figure clearly indicates that the dramatic motion which occurs during the unfolding process is that Helices 1 and 3 move far away from each other, so the protein extends, which is shown in PC 1 with much larger eigenvalues than other PCs. Well consistent with the transition network (Fig. 5(c)), the dominating mode of the unfolding of HP-35 is that protein extends and then the helices are unfolded. Another unfolding motions can be found in PC 2 and PC 3, especially PC 3, in which the residues in the same helix (Helices 1, 2, and 3) move to different directions, making the helices uncoil.

4. Conclusion

Limited by computational capabilities, nowadays even for small and fast-folding proteins, all-atom MD simulations of protein folding still presents a great challenge. However, there is an alternative way: studying unfolding process using MD simulations at high temperature. High temperature provides more energy to overcome energetic barriers to unfolding. More importantly, based on the previous investigation,[1] the mechanisms of protein folding can be reflected by unfolding process. However, although there is agreement between the unfolding simulations and experiments that investigate both folding and unfolding, it is acknowledged that sampling in MD is limited and that unfolding may not always be the same as folding. Some folding intermediates may disappear at high temperatures and some forbidden pathways may make contributions to the unfolding and refolding of the protein. For example, previous investigation shows that the higher- temperature simulations failed to capture the desolvation barrier involving the packing of the hydrophobic core that was observed at lower temperatures.

In the present study, a long time (1000 ns) MD simulation at high temperature (500 K) was performed to investigate the unfolding process of HP-35. To infer the folding mechanism, a MSM was also built from our simulation, which gives a reduced kinetic network between multiple conformational states of the protein. In our simulation, six macrostates coexist and transitions between these states constantly occur. Among these states, the highest-flux pathway of HP-35 folding is the state 5 $\rightarrow$ the states 4 $\rightarrow$ the states 6. Along this pathway, the helices are first formed, and then these secondary structures get close to each other and collapse into the final tertiary structure, which fits well with the framework model of protein folding. Such a detail picture is outside the limits of temporal and spatial resolution of current experimental techniques, which demonstrates the mightiness of MD simulations as well as the MSMs.

References

[1] Daggett V 2006 *Chem. Rev.* **106** 1898
[2] Toofanny R D and Daggett V 2012 *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2** 405
[3] Karplus M and Weaver D L 1976 *Nature* **260** 404

[4] Kim P S and Baldwin R L 1982 *Annu. Rev. Biochem.* **51** 459
[5] Weissman J S and Kim P S 1991 *Science* **253** 1386
[6] Radford S E, Dobson C M and Evans P A 1992 *Nature* **358** 302
[7] Jackson S E and Fersht A R 1991 *Biochemistry* **30** 10436
[8] Bowman G R, Voelz V A and Pande V S 2011 *Curr. Opin. Struct. Biol.* **21** 4
[9] Piana S, Lindorff-Larsen K and Shaw D E 2012 *Proc. Natl. Acad. Sci. USA* **109** 17845
[10] Piana S, Klepeis J L and Shaw D E 2014 *Curr. Opin. Struct. Biol.* **24** 98
[11] Banushkina P V and Krivov S V 2013 *J. Chem. Theory Comput.* **9** 5257
[12] Jain A and Stock G 2014 *J. Phys. Chem. B* **118** 7750
[13] Mori T and Saito S 2016 *J. Phys. Chem. B* **120** 11683
[14] Ma B G 2016 *Chin. Sci. Bull.* **61** 2670
[15] He E B, Guo Z Y and Mao Y L 2009 *Acta Biophys. Sin.* **25** 396
[16] Noé F and Fischer S 2008 *Curr. Opin. Struct. Biol.* **18** 154
[17] Chodera J D, Singhal N, Pande V S, Dill K A and Swope W C 2007 *J. Chem. Phys.* **126** 155101
[18] Buchete N V and Hummer G 2008 *J. Phys. Chem. B* **112** 6057
[19] Bowman G R, Huang X and Pande V S 2009 *Methods* **49** 197
[20] Bowman G R, Beauchamp K A, Boxer G and Pande V S 2009 *J. Chem. Phys.* **131** 124101
[21] Muff S and Caflisch A 2009 *J. Chem. Phys.* **130** 125104
[22] Prinz J H, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera J D, Schtte C and Noé F 2011 *J. Chem. Phys.* **134** 174105
[23] Noé F, Horenko I, Schütte C and Smith J C 2007 *J. Chem. Phys.* **126** 155102
[24] Pande V S, Beauchamp K and Bowman G R 2010 *Methods* **52** 99
[25] Beauchamp K A, McGibbon R, Lin Y S and Pande V S 2012 *Proc. Natl. Acad. Sci. USA* **109** 17807
[26] Lane T J, Shukla D, Beauchamp K A and Pande V S 2013 *Curr. Opin. Struct. Biol.* **23** 58
[27] Shukla D, Hernández C X, Weber J K and Pande V S 2015 *Acc. Chem. Res.* **48** 414
[28] McKnight C J, Matsudaira P T and Kim P S 1997 *Nat. Struct. Biol.* **4** 180
[29] Duan Y and Kollman P A 1998 *Science* **282** 740
[30] Ghosh R, Roy S and Bagchi B 2013 *J. Phys. Chem. B* **117** 15625
[31] Doering D S and Matsudaira P 1996 *Biochemistry* **35** 12677
[32] Kubelka J, Eaton W A and Hofrichter J 2003 *J. Mol. Biol.* **329** 625
[33] Duan Y, Wang L and Kollman P A 1998 *Proc. Natl. Acad. Sci. USA* **95** 9897
[34] Jang S, Kim E, Shin S and Pak Y 2003 *J. Am. Chem. Soc.* **125** 14841
[35] Lei H, Wu C, Liu H and Duan Y 2007 *Proc. Natl. Acad. Sci. USA* **104** 4925
[36] Koulgi S, Sonavane U and Joshi R 2010 *J. Mol. Graph. Model.* **29** 481
[37] Lei H and Duan Y 2007 *J. Mol. Biol.* **370** 196
[38] Lu Y, Zhou X and Ou-Yang Z 2017 *Chin. Phys. B* **26** 50202
[39] Case D A, Berryman J T, Betz R M, *et al.* 2015 *Amber*
[40] Götz A W, Williamson M J, Xu D, Poole D, Le Grand S and Walker R C 2012 *J. Chem. Theory Comput.* **8** 1542
[41] Salomon-Ferrer R, Götz A W, Poole D, Le Grand S and Walker R C 2013 *J. Chem. Theory Comput.* **9** 3878
[42] Maier J A, Martinez C, Kasavajhala K, Wickstrom L, Hauser K E and Simmerling C 2015 *J. Chem. Theory Comput.* **11** 3696
[43] Pomelli C S, Tomasi J and Barone V 2001 *Theor. Chem. Acc.* **105** 446
[44] Loncharich R J, Brooks B R and Pastor R W 1992 *Biopolymers* **32** 523
[45] Darden T, York D and Pedersen L 1993 *J. Chem. Phys.* **98** 10089
[46] Ryckaert J P, Ciccotti G and Berendsen H J C 1977 *J. Comput. Phys.* **23** 327
[47] Kabsch W and Sander C 1983 *Biopolymers* **22** 2577
[48] Beauchamp K A, Bowman G R, Lane T J, Maibaum L, Haque I S and Pande V S 2011 *J. Chem. Theory Comput.* **7** 3412
[49] Schwantes C R and Pande V S 2015 *J. Chem. Theory Comput.* **11** 600
[50] Sadiq S K, Noé F and De Fabritiis G 2012 *Proc. Natl. Acad. Sci. USA* **109** 20449
[51] Noé F, Schütte C, Vanden-Eijnden E, Reich L and Weikl T R 2009 *Proc. Natl. Acad. Sci. USA* **106** 19011
[52] Cronkite-Ratcliff B and Pande V 2013 *Bioinformatics* **29** 950
[53] Mesentean S, Koppole S, Smith J C and Fischer S 2007 *J. Mol. Biol.* **367** 591
[54] Skjaerven L, Martinez A and Reuter N 2011 *Proteins* **79** 232
[55] García A 1992 *Phys. Rev. Lett.* **68** 2696
[56] Mesentean S, Fischer S and Smith J C 2006 *Proteins* **64** 210
[57] Lou H and Cukier R I 2006 *J. Phys. Chem. B* **110** 24121
[58] Tournier A L and Smith J C 2003 *Phys. Rev. Lett.* **91** 208106
[59] Dill K A 1985 *Biochemistry* **24** 1501
[60] Dill K A 1990 *Biochemistry* **29** 7133