

Different potential of mean force of two-state protein GB1 and downhill protein gpW revealed by molecular dynamics simulation

Xiaofeng Zhang(张晓峰), Zilong Guo(郭子龙), Ping Yu(余平), Qiushi Li(李秋实), Xin Zhou(周昕), Hu Chen(陈虎)

Citation: Chin. Phys. B . 2020, 29(7): 078701 . **doi:** 10.1088/1674-1056/ab8daf
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

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

Quantum intelligence on protein folding pathways

Wen-Wen Mao(毛雯雯), Li-Hua Lv(吕丽花), Yong-Yun Ji(季永运), You-Quan Li(李有泉)

Chin. Phys. B . 2020, 29(1): 018702 . **doi:** 10.1088/1674-1056/ab5fbe

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), Kaifu Gao(高恺夫)

Chin. Phys. B . 2018, 27(1): 018701 . **doi:** 10.1088/1674-1056/27/1/018701

Anisotropic formation mechanism and nanomechanics for the self-assembly process of cross-β peptides

Li Deng(邓礼), Yurong Zhao(赵玉荣), Peng Zhou(周鹏), Hai Xu(徐海), Yanting Wang(王延颋)

Chin. Phys. B . 2017, 26(12): 128701 . **doi:** 10.1088/1674-1056/26/12/128701

Combined effects of headgroup charge and tail unsaturation of lipids on lateral organization and diffusion of lipids in model biomembranes

Xiao-Jie Chen(陈晓洁), Qing Liang(梁清)

Chin. Phys. B . 2017, 26(4): 048701 . **doi:** 10.1088/1674-1056/26/4/048701

Proteins：From sequence to structure

Zheng Wei-Mou

Chin. Phys. B . 2014, 23(7): 078705 . **doi:** 10.1088/1674-1056/23/7/078705

SPECIAL TOPIC — Modeling and simulations for the structures and functions of proteins and nucleic acids

Different potential of mean force of two-state protein GB1 and downhill protein gpW revealed by molecular dynamics simulation*

Xiaofeng Zhang(张晓峰)[1], Zilong Guo(郭子龙)[1], Ping Yu(余平)[1], Qiushi Li(李秋实)[2], Xin Zhou(周昕)[2,†], and Hu Chen(陈虎)[1,‡]

[1] *Research Institute for Biomimetics and Soft Matter, Fujian Provincial Key Laboratory for Soft Functional Materials Research, Department of Physics, Xiamen University, Xiamen* 361005, *China*

[2] *School of Physical Sciences, University of Chinese Academy of Sciences, Beijing* 100190, *China*

(Received 30 March 2020; revised manuscript received 20 April 2020; accepted manuscript online 28 April 2020)

Two-state folding and down-hill folding are two kinds of protein folding dynamics for small single domain proteins. Here we apply molecular dynamics (MD) simulation to the two-state protein GB1 and down-hill folding protein gpW to reveal the relationship of their free energy landscape and folding/unfolding dynamics. Results from the steered MD simulations show that gpW is much less mechanical resistant than GB1, and the unfolding process of gpW has more variability than that of GB1 according to their force–extension curves. The potential of mean force (PMF) of GB1 and gpW obtained by the umbrella sampling simulations shows apparent difference: PMF of GB1 along the coordinate of extension exhibits a kink transition point where the slope of PMF drops suddenly, while PMF of gpW increases with extension smoothly, which are consistent with two-state folding dynamics of GB1 and downhill folding dynamics of gpW, respectively. Our results provide insight to understand the fundamental mechanism of different folding dynamics of two-state proteins and downhill folding proteins.

Keywords: protein folding, molecular dynamics simulation, umbrella sampling, potential of mean force

PACS: 87.14.E−, 87.10.Tf, 87.15.A−

DOI: 10.1088/1674-1056/ab8daf

1. Introduction

The folding and unfolding of proteins are fundamental biological processes in the cell. Fully understanding of the protein folding and unfolding dynamics at atomic resolution is of great importance for facilitating the prediction of protein structure from amino acid sequence,[1] the protein design,[2,3] and the study of molecular mechanism of neuron degenerative diseases which are due to protein misfolding and aggregation.[4]

Dynamics of the protein folding and unfolding processes can be studied by biochemical denaturation experiment[5] or single molecule manipulation techniques.[6] On the other hand, atomistic molecular dynamics (MD) simulation is a powerful theoretical approach to study the folding and unfolding dynamics of proteins.[7] At the end of last century, detailed simulation studies were first applied to characterize protein folding/unfolding pathways. The basic idea behind MD is to apply Newtons laws of motion to every atom of the system using empirical potential energy functions.[8] However, due to the limitation of computational power, the study of protein folding/unfolding by the atomistic MD simulation is usually feasible to only small proteins.

From physical point of view, multiple dimensional conformation space of protein can be described by a funnel-shape free energy landscape.[9,10] From both single molecule manipulation experiment and MD simulation, the free energy landscape or potential of mean force (PMF) of proteins need to be constructed. Usually a protein is either at native state or in unfolded state. There is an energy barrier (transition state) which separates the native state and unfolded state. This kind of proteins is called two-state proteins.[11] If the barrier is very small, at physiological condition the protein can fold to native state very fast almost barrierlessly. Such proteins are called downhill proteins.[12,13]

To build PMF from MD simulation, conformational space of protein needs to be sampled sufficiently. But normal MD simulation will spend long time in local traps of the free energy landscape. Using replica exchange method, extensive all-atom simulation has been performed to get the PMF of a downhill folding protein BBL as a function of four different reaction coordinates.[14] A folding free energy barrier of 1–2 $k_{\mathrm{B}}T$ was obtained from the PMF, and low folding cooperativity was confirmed for the BBL protein.

Typical two-state protein GB1 is the streptococcal B1 immunoglobulin-binding domain of protein G (Fig. 1(a)). A single domain of GB1 consists of 56 residues. Its structure has been obtained via NMR, as well as been characterized by

*Project supported by the National Natural Science Foundation of China (Grant Nos. 11874309, 11474237, and 11574310) and the 111 Project, China (Grant No. B16029).

†Corresponding author. E-mail: xzhou@ucas.ac.cn

‡Corresponding author. E-mail: chenhu@xmu.edu.cn

© 2020 Chinese Physical Society and IOP Publishing Ltd

MD simulations.[15–17] GB1 is an α/β-protein with the two terminal β-strands (strand 1 and strand 4) arranged in parallel. These strands are bonded by backbone hydrogen bonds to stabilize its native structure.[18] Because of its small size and highly symmetrical topology, GB1 has been an ideal model protein for protein folding studies. A vast number of previous researches have confirmed that GB1 folds in a standard two-state process in which only two distinct states are populated (the native and denatured states). Along the folding or unfolding pathway, the highest free energy point is at the transition state.[8,19,20] GB1 was found to have high mechanical stability in atomic force microscopy (AFM) experiment.[21,22]

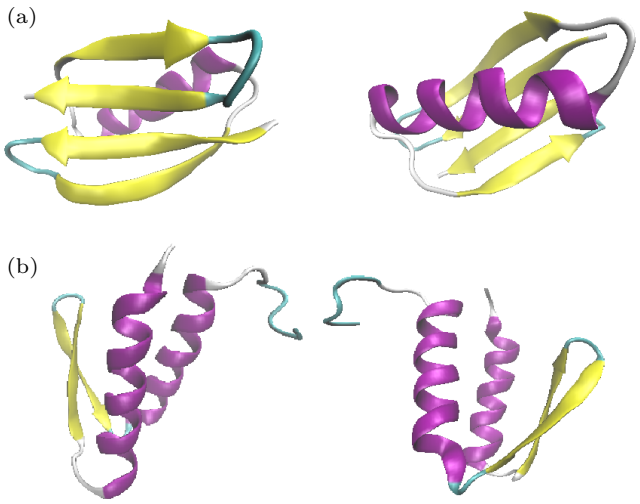


Fig. 1. The three-dimensional structures of proteins GB1 (a) and gpW (b) from two opposite perspectives.

Representative down-hill protein gpW is a 62-residue $\alpha+\beta$ protein that participates in virus head morphogenesis (Fig. 1(b)). It is a single gene product from bacteria λ phage virus.[23,24] Protein gpW has been proved experimentally to fold in a downhill regime with the maximal folding barrier of less than 1 $k_{\mathrm{B}}T$.[25]

In this paper, we compare the unfolding force-extension curves of GB1 and gpW by the steered molecular dynamics (SMD) simulation and find that GB1 is more mechanical resistant than gpW. PMF obtained from umbrella sampling reveals big difference between proteins GB1 and gpW, which gives physical foundation of their folding mechanism.

2. Simulation methods

2.1. Steered molecular dynamics simulation

The steered molecular dynamics (SMD) simulations mimic force spectroscopy measurement in single molecule manipulation experiment like AFM, but proceed on much shorter time scale.[26] In a SMD simulation, one terminal end of the protein is fixed, while the other end is coupled to one end of a dummy spring with elastic constant k. When the end of the dummy spring moves away from the protein with constant speed v, its position ξ is a linear function of time

$$\xi(t) = vt + x_0, \tag{1}$$

where x_0 is the initial value of the extension, i.e., the end-to-end distance of the protein, when it is aligned along the pulling direction.

During the pulling process, the force on the end of the protein, f, is calculated from the extension of the dummy spring,

$$f = k(\xi - x), \tag{2}$$

where x represents the extension of the protein along the force direction.[27] The response of the protein to force is investigated through the obtained force–extension curve during the stretching process (Fig. 2(a)).

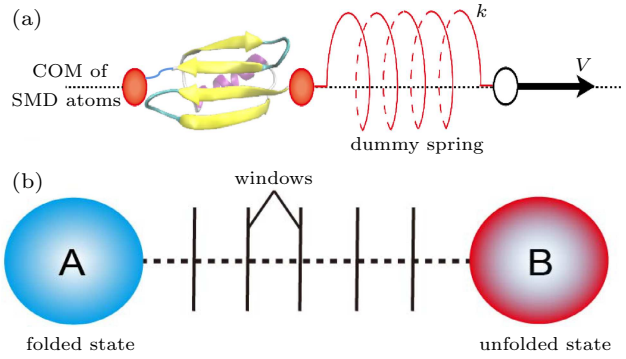


Fig. 2. Sketch of SMD simulation and the umbrella sampling method. (a) In SMD simulation, one terminus of the protein of interest is fixed, while the other terminus is coupled with a dummy spring with linear elastic constant k. The other end of the spring moves away from the protein with constant velocity v to stretch the protein. (b) In umbrella sampling, in conformational space projected to the extension as the reaction coordinate, the folded state and the unfolded state are located at each end of the coordinate. Between the folded state and the unfolded state, a spring is used to restrain conformations in a small window of conformational space. The system is well sampled in each window, and overlapping between neighboring windows helps to build PMF of the protein.

2.2. Umbrella sampling

One of the main challenges in computational biology is the calculation of free energy difference. Here we use umbrella sampling to construct PMF of proteins.[28] In umbrella sampling, a system is driven to conformations with high free energy by a bias potential along a reaction coordinate.[29] The conformations are covered by a series of overlapping windows (Fig. 2(b)), in each of which a bias potential, $w_i(x)$, serves to confine the conformational space to achieve a more efficient sampling.[30] In principle, the bias potentials can have any functional form, but harmonic potentials are generally chosen for their simplicity,

$$w_i(x) = \frac{k}{2}\left(\xi_i^{\mathrm{ref}} - x\right)^2, \tag{3}$$

where ξ_i^{ref} is the fixed position of the right end of the spring in the i-th window of the umbrella sampling simulations.

An accurate calculation of PMF requires an adequate number of windows (separate simulations) with enough overlap in conformational space.[28,31] Based on SMD simulation, a series of conformations were extracted from the SMD trajectory at regular interval of extension. Then each extracted conformation was used as the initial conformation to perform the umbrella sampling simulation in each window. When all simulations were done, the weighted histogram analysis method (WHAM), an extension of Ferrenberg and Swendsens multiple histogram technique, was applied to properly merge the data from all simulations to obtain the global PMF profile.[32,33]

Here we list detail parameters of SMD and umbrella simulations. Proteins GB1 and gpW with explicit waters were simulated in the NPT ensemble using GROMACS package 5.1.4, with the GROMOS96 53A6 all-atom force field at the temperature of 310 K. A time step of 2 fs was selected. Periodic boundary conditions were applied in all directions. Particle-mesh-Ewald summation was selected for the long-range Coulomb interactions. The cut-off distances for electrostatics and vdW interactions were set at 1.4 nm. The simulations started from an energy-minimized average NMR structure, deposited as entry 1PGA for protein GB1 and 2L6Q for protein gpW in the RCSB protein data bank. The structures of both proteins were solvated with SPC model water before simulations. The pulling velocity of the restraint point in the SMD simulation was 0.001 nm/ps, and the spring constant was 100 kJ/mol/nm^2. The umbrella simulations were performed in 95 extension windows for GB1 and 60 windows for gpW with spring constant of 500 kJ/mol/nm^2. Each simulation lasted for 15 ns, and the trajectory from 5 ns to 15 ns was used for the analysis while the first 5 ns simulations were applied for reaching equilibrium.

3. Results

3.1. Different mechanical resistances of GB1 and gpW

SMD simulation of GB1 and gpW began with an equilibrated folded structure and stopped when reaching an extension of $\sim$ 7.1 nm after $\sim$ 5 ns of simulation. GB1 and gpW show different capabilities to sustain the stretching force in SMD simulation with the same pulling velocity and spring constant.

Figure 3(a) shows the force–extension profiles of the pulling processes of protein GB1. Initially, the force grows linearly while the extension of the protein grows within a limited scale of $\sim$ 2.69 nm to $\sim$ 3.20 nm, which indicates that the protein maintains a mechanical stable structure due to the strong force-bearing hydrogen bonds between the first and last β strands. At the force peak with extension of $\sim$ 3.19 nm, the force drops suddenly and the extension increases to more than 7 nm due to broken of these hydrogen bonds. The corresponding conformation at the force peak is supposed to be the transition state of force-induced unfolding of GB1.

Figure 3(b) shows several typical SMD simulation processes. In all simulations, the structure of protein gpW extends readily under the maximum stretching force of $\sim$ 200 pN, which is much smaller than the maximum unfolding force of protein GB1 (above $\sim$ 600 pN). Though in all simulations, a force smaller than 200 pN is enough to destroy the native structure of gpW gradually, but the detailed force–extension curves in each simulation show little similarities, which indicates that the unfolding processes of protein gpW have large variability . In contrast to gpW, the force peaks of GB1 with the same simulation parameters are always located at the extension of 3.19 ± 0.02 nm with maximal force of 660 ± 60 pN.

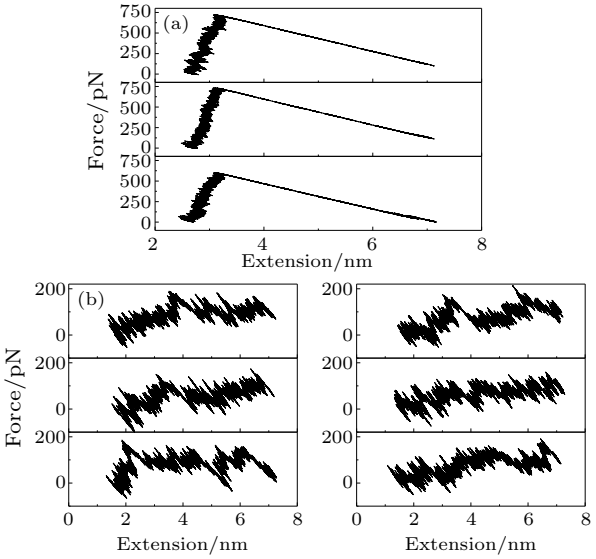


Fig. 3. Results of SMD simulation of GB1 and gpW proteins. (a) The force–extension curves of the unfolding process of protein GB1 show an force peak at the extension of $\sim$ 3.19 nm and forces all above 600 pN (710 pN, 714 pN, and 600 pN, respectively), which indicates the unfolding transition state. Three force–extension curves of the unfolding process of GB1 show very small difference in its unfolding process. (b) Six typical force–extension curves of the unfolding process of the protein gpW show multiple random force peaks. The peak forces are all below 200 pN, much smaller than that of GB1, which demonstrates that gpW is less mechanically resistant. In all SMD simulations, spring constant $k = 100$ kJ/mol/nm^2 and pulling velocity $v = 0.001$ nm/ps.

3.2. Different PMFs of GB1 and gpW

To obtain a more statistic and quantitative understanding of the unfolding process of GB1, the free energy profile of GB1 unfolding was calculated through a series of umbrella sampling simulations. Neighboring simulations have enough overlapping to build PMF over extension range of $\sim$ 2.4 nm to $\sim$ 4.1 nm for GB1 and $\sim$ 1.3 nm to $\sim$ 4.5 nm for gpW.

The PMF profiles of GB1 and gpW are shown in Fig. 4. For GB1, PMF is very sensitive to extension. When the extension is about 4.1 nm, PMF is about 34 k_BT. On the other hand, PMF of gpW increases with force much more slowly. PMF of gpW is only 18 k_BT at the extension of 4.5 nm.

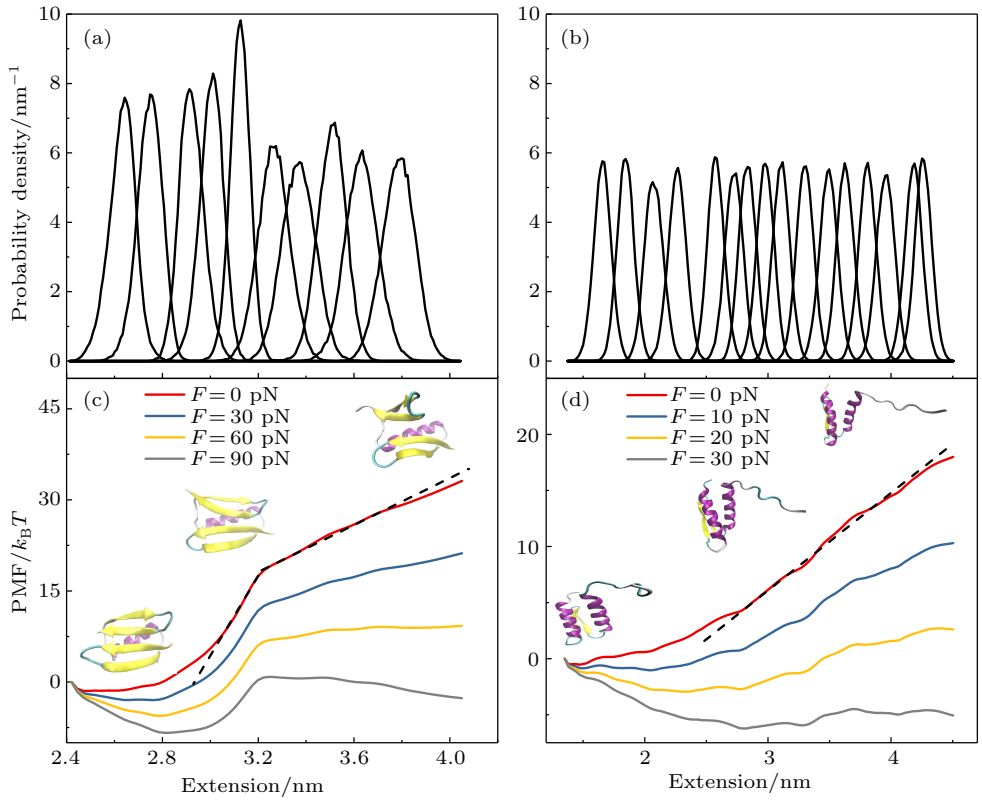


Fig. 4. The PMF profiles of proteins GB1 and gpW along the coordinate of extension obtained from umbrella sampling. Typical histograms of conformations along the reaction coordinate of GB1 and gpW are shown in panels (a) and (b). Just a few umbrella sampling windows are selected to show a sufficient sampling along the coordinate. Totally 95 umbrella sampling simulations were performed for GB1 and 60 for gpW. (c) PMF of GB1 protein at zero force (red curve) shows a kink at extension of ~ 3.18 nm, where the slope changes from ~ 292 pN to ~ 74 pN. At stretching forces of 30 pN, 60 pN, and 90 pN, PMF calculated from Eq. (4) is shown as blue, yellow, and gray curves, respectively. (d) PMF of gpW at zero force (red curve) shows a smooth curve which increases approximately linearly with extension. The slope of the curve is ~ 29 pN. Blue, yellow, and gray curves show PMF at stretching forces of 10 pN, 20 pN, and 30 pN, respectively. Insets of both figures show typical conformations of GB1 and gpW with certain extension from the simulation.

The PMF of gpW is almost a linear function of extension when the extension is longer than 2 nm. The slope changes smoothly along extension from 1.4 nm to 4.5 nm. This smooth PMF of gpW is consistent with its downhill folding dynamics obtained in biochemistry study.

Different from gpW, the PMF profile of GB1 has a transition point at 3.2 nm where the slope changes from 292 pN to 74 pN, i.e., the slope drops to about one quarter of its value at extension smaller than 3.2 nm. This transition point is very close to the force peak position in the SMD simulations.

When a constant stretching force is applied to the end of protein in single molecule manipulation experiment or MD simulations, the free energy profile will tilt. Therefore, force-dependent free energy profile $G_f(x)$ can be obtained from PMF (free energy profile at zero force, $G_0(x)$) according to the equation

$$G_f(x) = G_0(x) - f(x - x_0), \tag{4}$$

where x_0 represents the original extension of protein's native state, and x is the extension.

The $G_f(x)$ at different forces is also shown in Fig. 4. For GB1, the slope transition point becomes a peak when the force is 90 pN, which corresponds to the transition state of GB1 in its two-state folding/unfolding process.

As the PMF of gpW is almost a linear function of extension (Fig. 4(d)), its free energy profiles with stretching forces do not show any apparent barriers, which agrees with the downhill folding property of gpW. When the force is 20 pN, the PMF of gpW becomes almost a horizontal line. When the force is 30 pN or greater, the PMF decreases monotonically with extension. Three typical conformations with different extensions are shown for both proteins in the inset of Fig. 4.

4. Discussion and conclusion

Small single domain proteins usually fold through two-state transition. But some fast folding proteins fold to their native states downhill the free energy landscape without big free energy barrier. We use GB1 as a typical two-state protein, and gpW as a typical downhill folding protein to study their different mechanical stability and free energy landscape by MD simulation.

In summary, firstly we found that GB1 can sustain much larger force than gpW by SMD simulation, which might come from the shearing pulling geometry of GB1. To understand the

two-state property of GB1 and downhill folding mechanism of gpW, we focused our effort to build the free energy landscapes of GB1 and gpW from the PMF obtained by umbrella sampling MD simulation. PMF of GB1 increases with extension much faster than that of gpW, which is consistent with the different force resistance properties of GB1 and gpW obtained by SMD simulation. PMF of GB1 shows a kink point where the slope drops abruptly, corresponding to the conformation at the force peak during SMD simulation. In contrast to that of GB1, PMF of gpW increases smoothly with extension. Our simulation results are generally consistent with the two-state folding phenomena of protein GB1 and downhill folding behavior of gpW, and indicate that different protein folding dynamics is a natural consequence of their PMF.

In this paper, we only build PMFs of GB1 and gpW from their native states to partially extended conformations due to limited computational capability. Beyond the kink point of PMF of GB1, the free energy does not drop when there is no stretching force. Therefore, though the kink point is transition state of unfolding at high stretching forces, the transition state of unfolding at zero force might have longer extension. In the future if PMF from native state to fully unfolded peptide can be constructed, a whole picture of protein folding can be expected.

Free energy landscape can be also constructed from force-dependent folding and unfolding rates of proteins measured by single molecule stretching experiments.[34,35] Equilibrium folding and unfolding dynamics can be recorded by magnetic tweezers. At equilibrium critical force, the transition rates of downhill proteins are usually several orders of magnitude larger than those of two-state proteins, which is consistent with the free energy landscapes with low energy barrier for downhill folding proteins and high energy barrier for two-state proteins. Recently we found that the transition state of downhill protein gpW is located at longer extension than two-state protein GB1 from magnetic tweezers experiment, which agrees with the PMF of gpW from umbrella sampling MD simulation and indicates that the native state of gpW is very flexible. Single molecule experiments and MD simulations can work together to reveal the mechanism of protein folding.

References

[1] Honig B 1999 *J. Mol. Biol.* **293** 283
[2] Kuhlman B, Dantas G, Ireton G C, Varani G, Stoddard B L and Baker D 2003 *Science* **302** 1364
[3] Huang P S, Boyken S E and Baker D 2016 *Nature* **537** 320
[4] Soto C 2003 *Nat. Rev. Neurosci.* **4** 49
[5] McCallister E L, Alm E and Baker D 2000 *Nat. Struct. Mol. Biol.* **7** 669
[6] Chen H, Fu H, Zhu X, Cong P, Nakamura F and Yan J 2011 *Biophys. J.* **100** 517
[7] Beck D A C and Daggett V 2004 *Methods* **34** 112
[8] Fersht A R and Daggett V 2002 *Cell* **108** 573
[9] Leopold P E, Montal M and Onuchic J N 1992 *Proc. Natl. Acad. Sci. USA* **89** 8721
[10] Bryngelson J D, Onuchic J N, Socci N D and Wolynes P G 1995 *Proteins* **21** 167
[11] Jackson S E and Fersht A R 1991 *Biochemistry (Mosc.)* **30** 10428
[12] Garcia-Mira M M, Sadqi M, Fischer N, Sanchez-Ruiz J M and Munoz V 2002 *Science* **298** 2191
[13] Sadqi M, Fushman D and Munoz V 2006 *Nature* **442** 317
[14] Zhang J, Li W, Wang J, Qin M and Wang W 2008 *Proteins* **72** 1038
[15] Ding K, Louis J M and Gronenborn A M 2004 *J. Mol. Biol.* **335** 1299
[16] Schmidt H L, Sperling L J, Gao Y G, Wylie B J, Boettcher J M, Wilson S R and Rienstra C M 2007 *J. Phys. Chem. B* **111** 14362
[17] De Sancho D, Mittal J and Best R B 2013 *J. Chem. Theory Comput.* **9** 1743
[18] Cao Y and Li H 2007 *Nat. Mater.* **6** 109
[19] Jackson S E 1998 *Fold. Des.* **3** 81
[20] Barrick D 2009 *Phys. Biol.* **6** 015001
[21] Li H, Wang H C, Cao Y, Sharma D and Wang M 2008 *J. Mol. Biol.* **379** 871
[22] Puchner E M and Gaub H E 2009 *Curr. Opin. Struct. Biol.* **19** 605
[23] Murialdo H, Xing X, Tzamtzis D, Haddad A and Gold M 2003 *Biochem. Cell Biol.* **81** 307
[24] Sborgi L, Verma A, Munoz V and de Alba E 2011 *PLoS One* **6** e26409
[25] Fung A, Li P, Godoy-Ruiz R, Sanchez-Ruiz J M and Munoz V 2008 *J. Am. Chem. Soc.* **130** 7489
[26] Lu H and Schulten K 1999 *Proteins* **35** 453
[27] Lu H, Isralewitz B, Krammer A, Vogel V and Schulten K 1998 *Biophys. J.* **75** 662
[28] Kaestner J 2011 *Wiley Interdisciplinary Reviews-Computational Molecular Science* **1** 932
[29] Xu W, Li Y and Zhang Z 2012 *Chin. Phys. Lett.* **29** 068702
[30] Souaille M and Roux B 2001 *Comput. Phys. Commun.* **135** 40
[31] Torrie G M and Valleau J P 1977 *J. Comput. Phys.* **23** 187
[32] Kumar S, Bouzida D, Swendsen R H, Kollman P A and Rosenberg J M 1992 *J. Comput. Chem.* **13** 1011
[33] Kumar S, Rosenberg J M, Bouzida D, Swendsen R H and Kollman P A 1995 *J. Comput. Chem.* **16** 1339
[34] Chen H, Yuan G, Winardhi R S, Yao M, Popa I, Fernandez J M and Yan J 2015 *J. Am. Chem. Soc.* **137** 3540
[35] Yuan G, Le S, Yao M, Qian H, Zhou X, Yan J and Chen H 2017 *Angew. Chem.* **129** 5582