Extreme Mechanics Letters 8 (2016) 251–256

Contents lists available at ScienceDirect

Extreme Mechanics Letters

journal homepage: www.elsevier.com/locate/eml

Force-dependent mechanical unfolding pathways of GFP

Penghui Cao [a,b], Weiwei Tao [a], Harold S. Park [a,*]

[a] Department of Mechanical Engineering, Boston University, Boston, MA 02215, United States

[b] Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge MA 02139, United States

ARTICLE INFO

Article history:
Received 2 October 2015
Received in revised form 2 December 2015
Accepted 4 December 2015
Available online 8 December 2015

Keywords:
GFP
Force clamping
Intermediate configuration

ABSTRACT

We characterize the force-dependent unfolding pathways and intermediate configurations of the green fluorescence protein (GFP) using novel atomistic simulations based on potential energy surface exploration. By using this approach, we are able to unfold GFP to significantly longer end-to-end distances, i.e. 40 nm, as compared to that seen in previous atomistic simulation studies. We find that there are four intermediate states between 5 and 40 nm end-to-end distance, where the intermediate configurations and unfolding pathways are strongly force-dependent. We additionally calculate the force-dependent lifetime of the 14 nm $\alpha\beta_1$ intermediate, and demonstrate that it obeys Bell's formula.

1. Introduction

Over the past decade, researchers have successfully exploited experimental advances in nanomechanical testing to mechanically unfold and refold individual proteins [1,2]. One of the proteins that has been studied most frequently is the green fluorescence protein (GFP), which because it shows green fluorescence when subject to light at a specific wavelength, has been proposed to be used as an optical strain sensor [3].

Because of its potential as a mechanically active biological sensor, the mechanical unfolding of GFP has been studied using both experiments [4–6] and computer simulations [7–10], where a focus of the simulations has been to identify its various intermediate configurations. The simulations have used various forms of atomistic modeling and approximation, including steered molecular dynamics (SMD) [10], an Ising-like model [9], and an elastic network model [8]. However, due to the known time scale limitation in SMD, the longest simulation time has been on the order of 100 ns, with the largest end to end distance (ETED) of 14 nm [10]. Thus, while a complete characterization of the intermediate configurations and unfolding pathways of GFP is necessary to understand its capacity as a strain sensor, this information, particularly with regard to its unfolding pathways for ETEDs larger than 14 nm, are still not clear.

In this work, we employ the self-learning metabasin escape (SLME) method [11–13], which couples potential energy surface exploration with constant tensile forces ranging from 100 to 300 pN to study the mechanical unfolding of GFP. We note that many previous works in utilizing potential energy surface exploration to study the unfolding dynamics and mechanisms of proteins have been performed. These works include the metadynamics approach of Marinelli et al. [14], which has recently been used to study the dynamic behavior of protein molecules [15–17], and the discrete path sampling (DPS) method of Wales [18,19], who used PES exploration to determine rate constants for different transition states, and the SLME method for ubiquitin [20] and prion [21].

Our key findings are associated with the unfolding of GFP well-beyond that seen in previous simulation studies, where we reach ETEDs of 40 nm. In doing so, we find that there are three intermediate states between 14 and 40 nm ETED, where the intermediate configurations and

* Corresponding author.
E-mail addresses: pcao@mit.edu (P. Cao), parkhs@bu.edu (H.S. Park).

http://dx.doi.org/10.1016/j.eml.2015.12.001
2352-4316/

unfolding pathways are strongly force-dependent. We additionally calculate the force-dependent lifetime of the 14 nm $\alpha\beta_1$ intermediate, and demonstrate that it obeys Bell's formula.

2. Methods

It is well-known that the SMD simulations for force-induced unfolding of proteins exhibit major drawbacks as compared to the corresponding force clamp experiments [22–25]. In particular, the time scales accessible by SMD are on the order of hundreds of nanoseconds, while the unfolding time of GFP is on the order of microseconds to milliseconds [4]. In order to study the mechanical unfolding of GFP at experimentally-relevant (100–300 pN) forces and longer time scales, we utilize the SLME method of Cao et al. [12,13], which couples potential energy surface (PES) sampling with tensile deformation, and which was recently used to bring new insights into the mechanical unfolding pathways of the proteins ubiquitin [20] and prion [21], as well as bringing new insights into the characteristics of shear transformation zones in amorphous solids [26,27].

The SLME method is an improved version of the original autonomous basin climbing (ABC) method of Kushima et al. [11]. This improvement was necessary due to the substantial increase in computational expense that occurs as PES exploration continues due to the need to store the penalty functions that are used to push the system out of local energy minima such that the system does not re-explore any energy basins. The ABC method has recently been used to probe extremely slow dynamical processes like diffusion in amorphous silicon [28], nanocrystal creep [29], void nucleation and growth [30] and dislocation–defect interactions at slow strain rates [31].

In the SLME approach, quartic penalty functions are utilized to push the system out of potential energy wells in which it can become stuck due to the relatively low forces (i.e. 100s of pN for proteins) that are constantly applied. Upon application of a sufficient number of penalty functions, the system escapes over the lowest energy barrier to a neighboring potential energy well, where penalty functions are again applied if the applied force is not sufficient to lower the energy barrier to enable the system to escape. Thus, the penalty functions can be physically interpreted as thermal activation that assists the mechanical force in enabling the system to escape from a local energy minimum. This procedure is repeated until GFP is unfolded. In going through this procedure, the system is able to find and pass through the relevant intermediate configurations.

The unfolding time is estimated using transition state theory [32,33] via

$$\tau = \sum_{i=1}^{N} \left(\nu \exp^{-Q_i/k_B T}\right)^{-1}, \tag{1}$$

where Q_i is the energy barrier separating energy minima $i-1$ and i, N is total number of energy minima explored on the unfolding path, ν is a frequency prefactor, T is the temperature and k_B is the Boltzmann constant. We used the value $\nu = 1e^9\ s^{-1}$ as was utilized in the experimental GFP studies of Dietz and Rief [4].

We employed the AMBER99sb potential field which utilizes an implicit solvent model for water, while we used the protein with the Protein Data Bank (PDB) structure of 1fgl [34] for the native configuration of GFP. The native GFP structure was first equilibrated at 300 K, and then energy minimization was performed to generate the corresponding local energy minimum. At both the N and C-termini, a constant pulling force ranging from 100 to 300 pN was subsequently applied using the SLME method [13]. All simulations were performed using the open source GROMACS simulation package [35,36].

To construct the free energy profile of GFP along an unfolding path, we performed umbrella sampling while accounting for the clamping force. We sampled every 1 Å of end-to-end distance using a 1000 kJ mol^{-1} $Å^{-2}$ force constant for a 1 ns equilibration period. The potential of mean force (PMF) was then extracted by the Weighted Histogram Analysis Method (WHAM) [37].

3. Results

We conducted 50 independent SLME simulations at clamping forces ranging from 100 to 300 pN, with 10 simulations performed at each force level. The unfolding pathways for the different forces are shown in Fig. 1 where the plateau regions indicate intermediate states. These intermediate states are separated by energy barriers which must be crossed for additional unfolding to occur. For such small forces, SMD is not able to simulate large barrier crossing due to the well-known time scale limitation. However, by adding penalty functions, the SLME simulations are able to access a much longer time scale, and as shown in Fig. 1(c), is able to unfold GFP to a distance of 40 nm, which far exceeds the 14 nm unfolding that was previously observed using SMD [10]. We note that while this 40 nm unfolding distance is larger than seen in previous SMD simulations, it is still shorter than the 76 nm distance that is needed to completely unfold GFP [4]. In agreement with previous SMD simulations [10], at 100 pN the corked state was observed at an ETED of 5 nm while the ripple state was observed at an ETED of 14 nm. Larger unfolding distances for the 100 pN clamping force were not achieved due to computational limitations.

In all constant force unfolding simulations for all force levels, the N terminal α-helix unravels first from the barrel, which agrees with earlier constant velocity coarse grain model simulations [5,9]. However, the constant velocity simulations exhibit a bifurcation in the unfolding pathways after the unfolding of the α-helix. Specifically, the $\alpha \rightarrow \beta_1$ pathway is observed in 72% of the simulations, while the $\alpha \rightarrow \beta_{11}$ pathway is observed in 28% of the simulations [5,9]. Those values for each pathway can be changed by cross-link mutations, or by changing the clamping force direction.

Fig. 1 also demonstrates that the unfolding pathways depend on the magnitude of the clamping force. At 300 pN, 90% of the unfolding trajectories take two unfolding steps to unfold to the ETED of 40 nm, while in contrast 70% of the simulations follow 3-step unfolding at the smaller

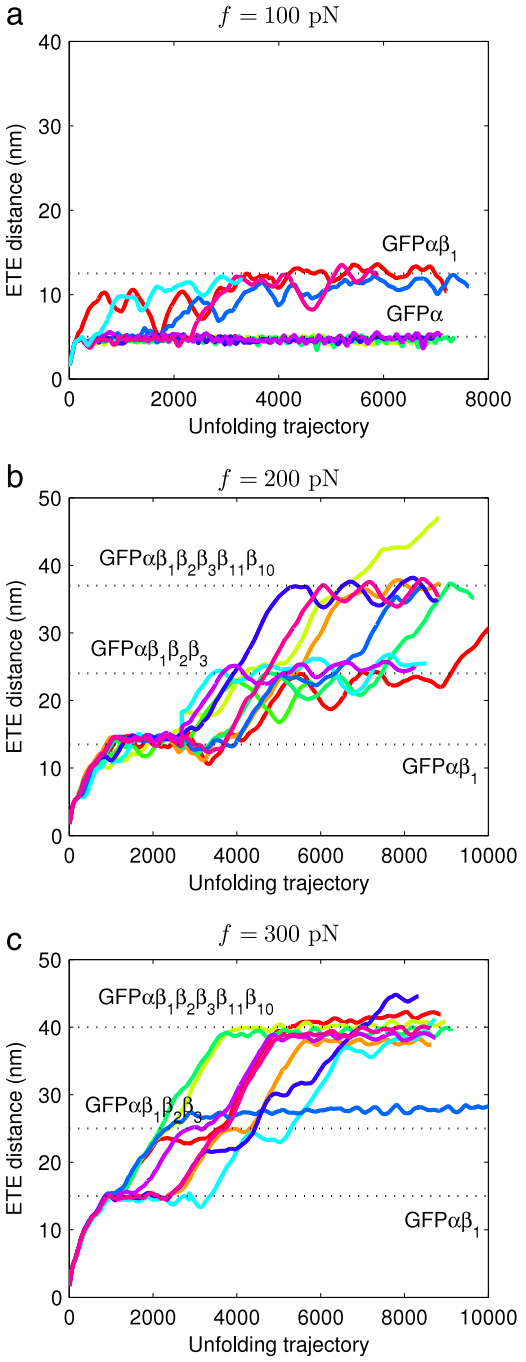

Fig. 1. The mechanical unfolding of GFP obtained using the SLME method as shown through the end-to-end distance vs. local minima explored for unfolding at 100, 200 and 300 pN. Unfolding trajectory refers to the total number of local energy minima found for the SLME simulation.

clamping force of 200 pN. For the 100 pN simulations, we were unable to, due to computational limitations, unfold GFP past an ETED of 14 nm.

We show in Fig. 2 two representative unfolding pathways for the 200 pN clamping force. The pathway analyzed in Fig. 2(a)–(c) passes through one intermediate at 14 nm, while the pathway analyzed in Fig. 2(d)–(f) passes through two intermediates, one at 14 and the other at 23 nm. Fig. 2(b) and (e) shows the heights of the energy barriers that are crossed over during the unfolding as a function of ETED. As can be seen, the pathway shown in Fig. 2(d) encounters many more large energy barriers starting around ETED in Fig. 2(e) as compared to the pathway shown in Fig. 2(b), which explains why it passes through an intermediate around ETED of 23 nm. The second salient point is the unfolding time as a function of ETED in Fig. 2(c) and (f). In particular, the unfolding distance from 0 to 38 nm ETED is on a time scale of tens of μs, which is a time scale that is difficult to reach using classical MD without substantial parallel computing resources. In contrast, all of the computations performed in this work were done on a standard Linux desktop computer, which demonstrates the potential of the SLME method for studying the unfolding of proteins at experimentally-relevant times scales.

4 intermediate states, α, $\alpha\beta_1$, $\alpha\beta_1\beta_2\beta_3$ and $\alpha\beta_1\beta_2\beta_3\beta_{11}\beta_{10}$ were observed in the SLME simulations. The α intermediate state is formed after the initial rupture of the N-terminal α-helix. This intermediate is observed only in the 100 pN pathways, and is not observed for larger clamping forces because the value of the clamping force required to unravel β_1 is around 100 pN [7,4]. The unraveling of β_1 followed by α forms the intermediate $\alpha\beta_1$, or the ripped state [10] as shown in Fig. 3(a), with a corresponding ETED for the $\alpha\beta_1$ intermediate state in the range of 12–14 nm. The unraveling of β_1 leaves $\beta_2\beta_3$ unshielded, and thus they begin unraveling and the length of the next intermediate, corresponding to $\alpha\beta_1\beta_2\beta_3$ in Fig. 3(b), occurs at an ETED of around 25 nm. After passing through this intermediate, $\beta_{11}\beta_{10}$ unfolds resulting in a new intermediate state of $\alpha\beta_1\beta_2\beta_3\beta_{11}\beta_{10}$ at an ETED of about 39 nm as shown in Fig. 3(c). These intermediates states correspond to the minima on the free energy profile, as we now discuss.

In Fig. 4, we report the PMF profile for a clamping force of 200 pN applied during the umbrella sampling simulation. Three distinct local minima can be observed for ETEDs of about 13, 23 and 36 nm. As previously discussed, these local minima correspond to the mechanical intermediate configurations of $\alpha\beta_1$, $\alpha\beta_1\beta_2\beta_3$ and $\alpha\beta_1\beta_2\beta_3\beta_{11}\beta_{10}$ shown in Fig. 3. Furthermore, the fact that these intermediate configurations correspond to local energy minima validates them as true intermediate states during the mechanical unfolding process. To compare with existing experimental data, we note that in the 2004 work of Dietz et al. [4], the energy barrier for transition from GFP$\Delta\alpha$ to GFP$\Delta\alpha\Delta\beta$ was found to be about 23 k_BT, or about 57 kJ/mol. Fig. 4(left) shows that the barrier for GFP$\Delta\alpha\Delta\beta$ is about 47 kJ/mol, which gives reasonably good agreement.

We now analyze the force-dependent dynamics of one particular intermediate configuration, the $\alpha\beta_1$ intermediate. This intermediate state was chosen as it is observed at all clamping force levels we used for the SLME simulations. We analyze the force-dependence of the lifetimes of this intermediate state using the Arrhenius equation, where the separation free energy barrier can be reduced by the application of mechanical force to a molecule, which exponentially increases the transition rate. The rupture of molecular bonds under mechanical force was first studied

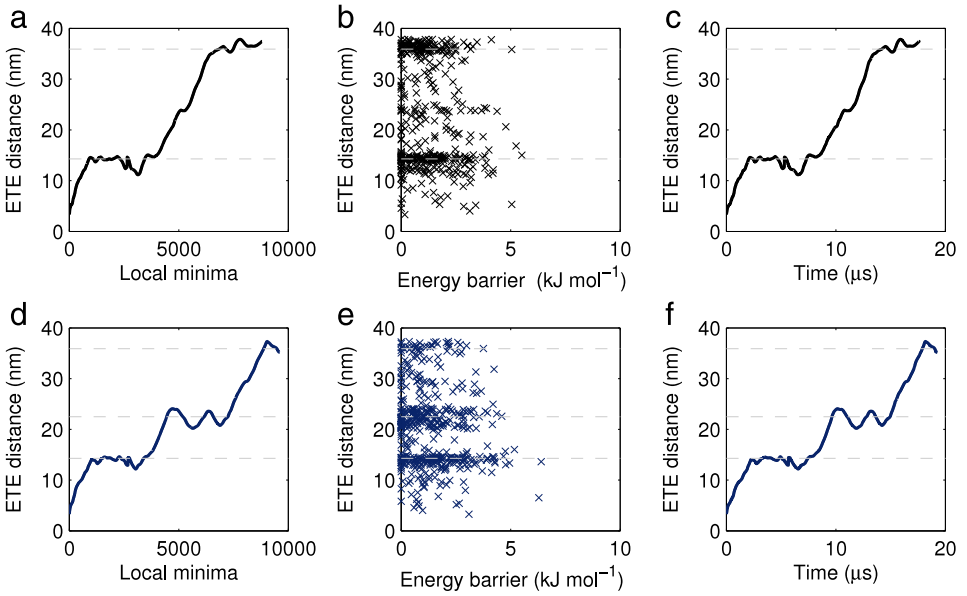

Fig. 2. (a)(d) Two representative unfolding pathways at 200 pN. (b)(e) Distribution of energy barriers along each unfolding pathway. (c)(f) Unfolding time as a function of ETED.

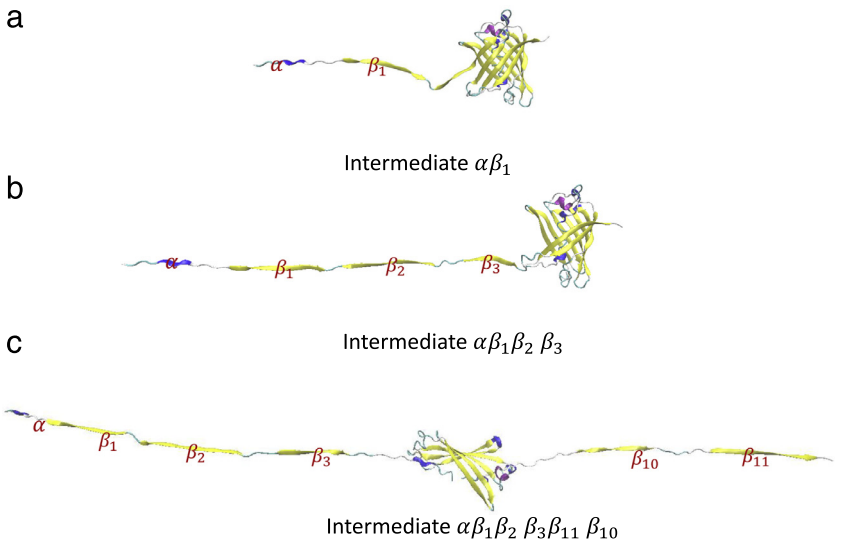

Fig. 3. Intermediate states observed during the mechanical unfolding GFP at 200 pN. (a) $\alpha\beta_1$, (b) $\alpha\beta_1\beta_2\beta_3$, (c) $\alpha\beta_1\beta_2\beta_3\beta_{11}\beta_{10}$.

by Bell [38], who predicted that the life-time decreases exponentially with the stretching force which is defined as $t(F) = t_0 \exp(-F\Delta x/k_BT)$, where t_0 is the lifetime without any clamping force and Δx denotes the potential width. To study the force-dependence of the lifetimes of the $\alpha\beta_1$ intermediate state, we computed and analyzed the lifetimes for clamping forces ranging from 150 to 300 pN. The lifetime is estimated using Eq. (1) with $\nu = 1e^9$ s^{-1} [4]. In our SLME simulations as shown in Fig. 5, the average lifetime of the $\alpha\beta_1$ intermediate decreases with increasing force. By fitting the curve using Bell's formula, we obtain $t_0 = 379$ μs and $\Delta x/k_BT = 0.02$ pN^{-1}.

Finally, we comment on the average lifetimes of the three intermediate configurations that were observed in the 300 pN unfolding simulations. In doing so, we find that the averaged lifetimes of $\alpha\beta_1$, $\alpha\beta_1\beta_2\beta_3$ and $\alpha\beta_1\beta_2\beta_3\beta_{11}\beta_{10}$ are 1.5, 1.8 and 6.89 μs, respectively. The intermediate states $\alpha\beta_1$ and $\alpha\beta_1\beta_2\beta_3$ have shorter lifetimes because the rupture of the hydrogen bonds is via tearing, with sequential breaking of bonds. However, for $\alpha\beta_1\beta_2\beta_3\beta_{11}\beta_{10}$, the pulling direction is perpendicular to the hydrogen bonds, resulting in a shearing mode of failure which is comparatively more difficult. These explanations are consistent with the explanation in the difference in lifetimes that can be obtained from the free energy profiles in Fig. 4. There, it is clear that $\alpha\beta_1$ and $\alpha\beta_1\beta_2\beta_3$ have smaller unfolding barriers and thus shorter unfolding time.

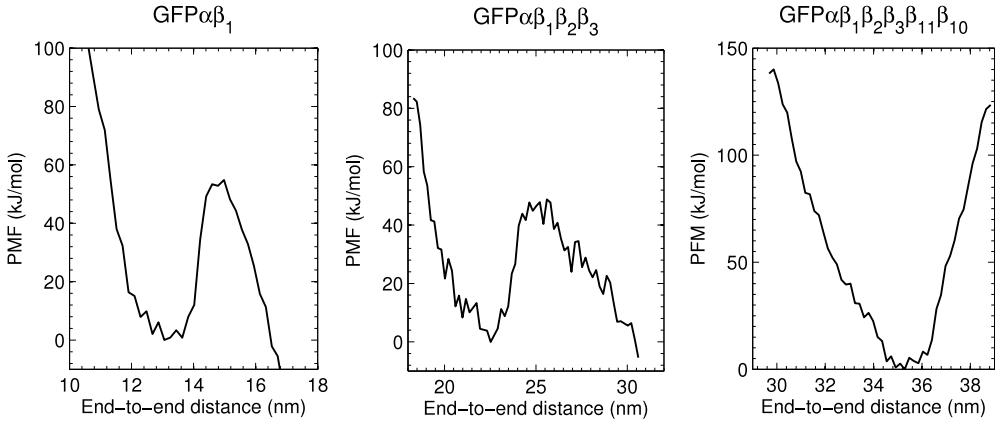

Fig. 4. Free energy profile as a function of end-to-end distance for intermediate states found during the 200 pN unfolding of GFP.

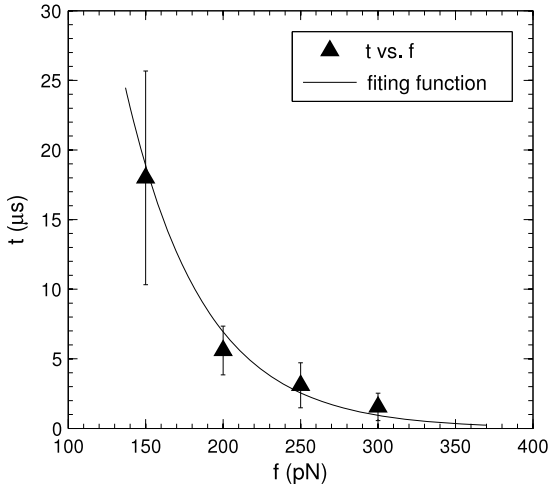

Fig. 5. Force dependent lifetime of intermediate state $\alpha\beta_1$. Fitting function is $t = 379 * \exp(-0.02 * F)$. Standard deviation for each force value shown in error bars.

4. Conclusions

In conclusion, we have used the SLME simulation methodology to examine the force-induced unfolding of the protein GFP. Our key finding is the discovery of four total, and two new intermediate states that exist between 5 and 40 nm end-to-end distance, where the intermediate configurations, unfolding pathways, and lifetimes of the intermediate states is strongly force-dependent. The ability to unfold GFP to substantially longer end-to-end distances as compared to previous atomistic simulations is due to the time-scale bridging nature enabled by the SLME methodology, and demonstrates the potential of this methodology for protein folding and unfolding problems.

Acknowledgment

HSP, WWT and PC acknowledge the support of NSF CMMI-1234183.

References

[1] K.C. Neuman, A. Nagy, Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy, Nat. Methods 5 (2008) 491–505.
[2] C. Bustamante, Y.R. Chemla, N.R. Forde, D. Izhaky, Mechanical processes in biochemistry, Annu. Rev. Biochem. 73 (2004) 705–748.
[3] M. Zimmer, Green fluorescent protein (GFP): applications, structure, and related photophysical behavior, Chem. Rev. 102 (2002) 759–782.
[4] H. Dietz, M. Rief, Exploring the energy landscape of GFP by single-molecule mechanical experiments, Proc. Natl. Acad. Sci. USA 101 (2004) 16192–16197.
[5] M. Mickler, R.I. Dima, H. Dietz, C. Hyeon, D. Thirumalai, M. Rief, Revealing the bifurcation in the unfolding pathways of GFP by using single-molecule experiments and simulations, Proc. Natl. Acad. Sci. 104 (2007) 20268–20273.
[6] R. Perez-Jimenez, S. Garcia-Manyes, S.R.K. Ainavarapu, J.M. Fernandez, Mechanical unfolding pathways of the enhanced yellow fluorescent protein revealed by single molecule force spectroscopy, J. Biol. Chem. 281 (2006) 40010–40014.
[7] C. Hyeon, R.I. Dima, D. Thirumalai, Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins, Structure 14 (2006) 1633–1645.
[8] E. Eyal, I. Bahar, Toward a molecular understanding of the anisotropic response of proteins to external forces: insights from elastic network models, Biophys. J. 94 (2008) 3424–3435.
[9] M. Caraglio, A. Imparato, A. Pelizzola, Direction-dependent mechanical unfolding and green fluorescent protein as a force sensor, Phys. Rev. E 84 (2011) 021918.
[10] J. Saeger, V.P. Hytönen, E. Klotzsch, V. Vogel, GFP's mechanical intermediate states, PLoS One 7 (2012) e46962.
[11] A. Kushima, X. Lin, J. Li, J. Eapen, J.C. Mauro, X. Qian, P. Diep, S. Yip, Computing the viscosity of supercooled liquids, J. Chem. Phys. 130 (2009) 224504.
[12] P. Cao, M. Li, R.J. Heugle, H.S. Park, X. Lin, A self-learning metabasin escape algorithm and the metabasin correlation length of supercooled liquids, Phys. Rev. E 86 (2012) 016710.
[13] P. Cao, H.S. Park, X. Lin, Strain-rate and temperature-driven transition in the shear transformation zone for two-dimensional amorphous solids, Phys. Rev. E 88 (2013) 042404.
[14] F. Marinelli, F. Pietrucci, A. Laio, S. Piana, A kinetic model of trp-cage folding from multiple biased molecular dynamics simulations, PLoS Comput. Biol. 5 (2009) e1000452.
[15] E. Formoso, V. Limongelli, M. Parrinello, Energetics and structural characterization of the large-scale functional motion of the adenylate kinas, Sci. Rep. 5 (2015) 8425.
[16] A. Berteotti, A. Barducci, M. Parrinello, Effect of urea on the β-hairpin conformational ensemble and protein denaturation mechanism, J. Am. Chem. Soc. 133 (2011) 17200–17206.
[17] A. Laio, M. Parrinello, Escaping free-energy minima, Proc. Natl. Acad. Sci. 99 (2002) 12562–12566.
[18] D.J. Wales, Discrete path sampling, Mol. Phys. 100 (2002) 3285–3305.
[19] D.J. Wales, Some further applications of discrete path sampling to cluster isomerization, Mol. Phys. 102 (2004) 891–908.

[20] P. Cao, G. Yoon, W. Tao, K. Eom, H.S. Park, The role of binding site on the mechanical unfolding mechanism of ubiquitin, Sci. Rep. 5 (2015) 8757.
[21] W. Tao, G. Yoon, P. Cao, K. Eom, H.S. Park, β-sheet-like formation during the mechanical unfolding of prion protein, J. Chem. Phys. 143 (2015) 125101.
[22] R. Elber, Long-timescale simulation methods, Curr. Opin. Struct. Biol. 15 (2005) 151–156.
[23] M. Carrion-Vazquez, H. Li, H. Lu, P.E. Marszalek, A.F. Oberhauser, J.M. Fernandez, The mechanical stability of ubiquitin is linkage dependent, Nat. Struct. Mol. Biol. 10 (2003) 738–743.
[24] J. Li, J.M. Fernandez, B.J. Berne, Water's role in the force-induced unfolding of ubiquitin, Proc. Natl. Acad. Sci. 107 (2010) 19284–19289.
[25] G. Stirnemann, D. Giganti, J.M. Fernandez, B.J. Berne, Elasticity, structure, and relaxation of extended proteins under force, Proc. Natl. Acad. Sci. 110 (2013) 3847–3852.
[26] P. Cao, X. Lin, H.S. Park, Surface shear transformation zones in amorphous solids, Phys. Rev. E 90 (2014) 012311.
[27] P. Cao, X. Lin, H.S. Park, Strain-rate and temperature-dependence of yield stress of amorphous solids via self-learning metabasin escape algorithm, J. Mech. Phys. Solids 68 (2014) 239–250.
[28] X. Yan, A. Gouissem, P. Sharma, Atomistic insights into li-ion diffusion in amorphous silicon, Mech. Mater. (2015).
[29] T.T. Lau, A. Kushima, S. Yip, Atomistic simulation of creep in a nanocrystal, Phys. Rev. Lett. 104 (2010) 175501.
[30] Y. Fan, A. Kushima, S. Yip, B. Yildiz, Mechanism of void nucleation and growth in bcc Fe: atomistic simulations at experimental time scales, Phys. Rev. Lett. 106 (2011) 125501.
[31] Y. Fan, Y.N. Osetskiy, S. Yip, B. Yildiz, Mapping strain rate dependence of dislocation-defect interactions by atomistic simulations, Proc. Natl. Acad. Sci. 110 (2013) 17756–17761.
[32] I. Popa, J.M. Fernández, S. Garcia-Manyes, Direct quantification of the attempt frequency determining the mechanical unfolding of ubiquitin protein, J. Biol. Chem. 286 (2011) 31072–31079.
[33] P. Hanggi, P. Talkner, M. Borkovec, Reaction-rate theory: fifty years after kramers, Rev. Modern Phys. 62 (1990) 251–342.
[34] F. Yang, L.G. Moss, G.N. Phillips Jr., The molecular structure of green fluorescent protein, Nat. Biotechnol. 14 (1996).
[35] Gromacs, 2014. http://www.gromacs.org/.
[36] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation, J. Chem. Theory Comput. 4 (2008) 435–447.
[37] J.S. Hub, B.L. de Groot, D. van der Spoel, g_wham—a free weighted histogram analysis implementation including robust error and autocorrelation estimates, J. Chem. Theory Comput. 6 (2010) 3713–3720.
[38] G.I. Bell, Models for the specific adhesion of cells to cells, Science 200 (1978) 618–627.