PCCP

Accepted Manuscript

This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the **Information for Authors**.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard **Terms & Conditions** and the **Ethical guidelines** still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ARTICLE TYPE

Direct Sampling of Multiple Single-Molecular Rupture Dominant Pathways Involving Multistep Transition

Huijun Jiang, Huai Ding, and Zhonghuai Hou*

We report a novel single-molecular rupture mechanism revealed by direct sampling of the dominant pathway using a self-optimized path sampling method. Multiple dominant pathways involving multistep transitions are identified. The rupture may take place via a direct unfolding from the native state to the unfolding state, or through a two-step pathway bypassing a distinct intermediate metastable state (IMS). This scenario facilitates us to propose a three-state kinetic model, which can produce a nonlinear dependence of the rupture time on pulling force similar to ones reported in literature. In particular, molecule conformations in the IMS maintain an elongation of the tail at one terminal, by which external pulling will enhance the relative stability of IMS. Consequently, even though the overall transition rate of the multistep pathway is relative small, the molecule still has to be ruptured via it rather than the direct pathway. Thus, our work demonstrates an IMS trapping effect induced rupture mechanism involving an abnormal switching from a fast dominant pathway to a slow one.

1 Introduction

Facilitated by the development of atomic force microscope[1] and laser tweezer technique[2,3], single-molecule micromanipulation has led to increasing new insights into an expanding variety of fundamental biological problems. One of its important applications is the mechanical rupture of a single molecule under pulling forces, which reveals the mechanical properties and dynamics of individual molecules or their complexes[2–8]. In experiments, pullings are applied to the end of the molecule, and rupture times or rupture forces are recorded for constant force or constant speed pulling, respectively. Generally, for a constant force F, the rupture time shows exponential-decay dependence on F as indicated by the so-called Bell's formula[9–12]. Of particular interest, however, a rather counterintuitive effect was also found in pulling experiments[6–8] that the rupture time shows an abnormal increase for low forces, followed by a maximum and then decrease at higher forces. So far, there are several catch-bond models[13] have been proposed to understand such a novel "rollover" behavior, which can be mainly divided into two types. For the first one, the nonmonotonic dependence is considered to be arisen from a force distorted dominant pathway[14–17]. A minimal model proposed in Ref. 15 has revealed the existence of a spectrum of unusual scenarios for the force-dependent lifetime. On the other hand, the second one attributes the "rollover" to the switching of the dominant unfolding pathway from the one which is fast for s-

mall pullings and not sensitive to the force, to a force-sensitive one which turns to be fast for large forces[7,8,14,18]. Therefore, it is desired to investigate directly the rupture pathways to enable discrimination between the two possible mechanisms, which, however, still remains a big challenge for theoretical study due to that the rupture process is a very rare and complicated event especially for small forces.

In this paper, we address such a problem by direct sampling of rupture pathways with the help of a self-optimized path sampling method we proposed. Clear evidences reveal that the rupture of the considered protein involves multiple pathways with multistep transitions. The protein may unfold via a one-step pathway (path I) connecting directly the native state and the unfolding state, or through a multistep pathway (path II) with an evident intermediate metastable state (IMS). With increasing force, the protein get more trapped into the IMS such that the protein mainly unfold through path II, although the overall rate of which is slower than path I. A multi-pathway kinetic model consisting of the native state **A**, the unfolding state **B** and the IMS is proposed, which allows us to analyze explicitly how the average rupture time depends on the pulling force. With the sampled kinetic parameters associated with the protein, similar *rollover* feature is successfully reproduced as that in literature. Remarkably, our work also unravels a novel mechanism for the non-monotonic dependence of rupture time on pullings. In contrary to the reported multipathway mechanism where the rupture process is always via the fastest pathway[7,8,14,18], we find an abnormal switching from a fast dominant pathway to a slow one, which is induced by force enhanced IMS-trapping effect.

Department of Chemical Physics & Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China, Hefei, Anhui 230026, China; E-mail: hzhlj@ustc.edu.cn

2 Model and Method

2.1 Go-type model

Dynamic simulations are performed with a coarse-grained model of proteins in which each residue is represented by a particle with the same mass centered on the α carbons. The beads are connected via rigid bonds along the protein backbone which is realized in simulations by using the SHAKE[19] algorithm. Besides, the interactions between beads are described by a Go-like potential[20], including harmonic terms for each bond angle, dihedral terms reflecting the conformational preferences of the backbone, attractive interactions between residues that are in native-contact, and repulsive interactions between other pairs of residues. It has been shown that such Go-models can reproduce the folding process of small proteins significantly better than their all-atom counterparts[21–23]. The molecule we study in the present work is protein L (1HZ6 on pdb.org[24]) consisting of β sheets separated by an α helix and a small tail at the N-terminal, for which the "rollover" of rupture time was reported[25]. The Go-like potential for this protein is calculated following the method proposed by Karanicolas and Brooks[20,26]. A fundamental energy scale is determined, such that rescaling of the potentials described above would result in a folding transition temperature T_f^*. In simulations, the temperature is fixed as $T = 6T_f^*/7$.

2.2 The self-optimized forward flux sampling method

Although pulling can accelerate the rupture process, the unfolding event, especially for small forces, is still so rare such that it is very hard to provide statistically reliable dominant rupture pathway by brute-force simulations, and one usually turns to path sampling methods. For nonequilibrium processes such as mechanical unfolding investigated here, forward flux sampling (FFS) has been a widely-used choice[27,28]. Generally, FFS assumes an order parameter λ to distinguish the initial state **A** from the final state **B**. A series of interfaces $\{\lambda_i, i = 0, ..., N\}$ in between are used to calculate the transition rate and to sample the transition path ensemble step by step. FFS generates partial trajectories between adjacent interfaces which are integrated forward in time only, and calculates the transition rate via $k_{AB} = \phi_0 \prod_{i=0}^{N-1} P(\lambda_{i+1}|\lambda_i)$ from **A** to **B**, where ϕ_0 is the effective forward flux leaving **A** and reaching the first interface, $P(\lambda_{i+1}|\lambda_i)$ is the conditional probability that a trajectory coming from **A** crosses interface i for the first time and then reaches the interface $i+1$ before returning to **A**. FFS has been widely used in a variety of systems, e.g., the flipping of genetic toggle switch[27], nucleation process[29], polymer translocation[30], protein conformational changes[31], to list just a few.

However, difficulties arise when we try to apply FFS to the present system, where complex pathways may exist. The order parameter is normally chosen as the end-end distance of the protein with one end fixed and the other subjected to the force. According to standard FFS procedures, the interfaces $\{\lambda_i\}$ are set with equal λ-distance between adjacent ones. Nevertheless, this may lead to considerable computational inefficiency because the potential difference between adjacent interfaces are usually not equal, such that a particularly sharp increase of potential would hinder the sampling process. In addition, FFS may fail if some unknown IMSs hide between the initial and final states. To overcome these two difficulties, here we propose a self-optimized approach which self-adaptively locates the interfaces and searches for IMSs *on the fly* during the sampling process[32]. The basic idea is simply that an optimized set of interfaces λ_i should maintain equal conditional transition probabilities $P(\lambda_{i+1}|\lambda_i)$ between adjacent interfaces, rather than equal order-parameter distance. This is done by firstly running local trajectories starting from the current interface λ_i to get the conditional probability distribution $P_c(\lambda > \lambda_i|\lambda_i)$, and then determining the next interface λ_{i+1} by equalling $P_c(\lambda_{i+1}|\lambda_i)$ to a given value p_0. With these optimized interfaces, FFS can be run in a much more efficient way on the fly. In addition, the approach facilitates us to find the IMSs conveniently, by monitoring some special long trajectories that neither end at the initial state nor reach the next interface, the number of which will increase sharply from zero if such IMS is encountered.

A detailed procedure is described as follows:

(a) Run a long enough trajectory with time length T_1 starting from **A** till the conditional probability $P_c(\lambda|\mathbf{A})$ converges. Interface λ_0 is located at where the conditional probability is of a fixed value $P_c(\lambda|\mathbf{A}) = p_0$. Store configurations that cross λ_0 and count its number n_0, and calculate ϕ_0 by n_0/T_1.

(b) Run trajectories with fixed time length T_2 starting from random-chosen configurations at interface λ_i till the conditional probability $P_c(\lambda_{i+1}|\lambda_i)$ for $\lambda \geq \lambda_i$ converges. Interface λ_{i+1} is located at where the conditional probability $P_c(\lambda_{i+1}|\lambda_i)$ takes value p_0. Store configurations crossing λ_{i+1}.

(c) Calculate the exact conditional probability by $P(\lambda_{i+1}|\lambda_i) = n_{i,1}/M_i$, where $n_{i,1}$ is the number of successful trajectories which cross interface λ_{i+1} before returning to **A**, and M_i is the total number of trajectories. Notice that, trajectories which reach neither interface λ_{i+1} nor the initial state **A** during T_2 should be further run till ends at one of the two states. Count $n_{i,3}$ the number of trajectories which end neither at λ_{i+1} nor at **A** for a sufficient long time.

(d) If $n_{i,3}$ is substantially larger than zero, run a long enough trajectory starting from a configuration picked up randomly from these trajectories till the local phase space density $\rho(\lambda)$ converges. Locate the IMS **A'** at where $\rho(\lambda = \lambda_{IMS})$ reaches the maximal value. Replace **A** by the IMS **A'**, repeat steps (a) to (d).

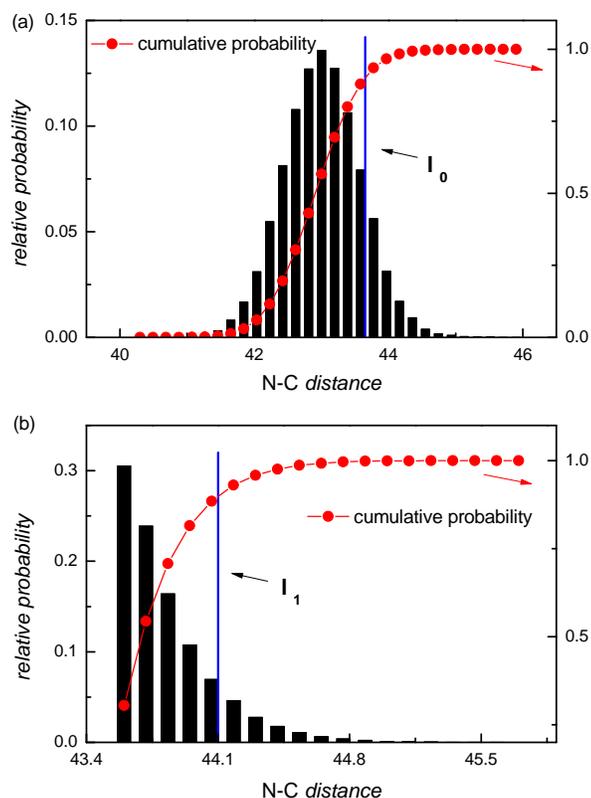(e) Repeat step (b) to (d) till the final state is reached.

Fig. 1 (Color online) Sampled conditional local phase space distribution and cumulative distribution (both in relative probability) starting from (a) the native state **A** and (b) a known interface (here interface λ_0 as an example). The new interface is located at where the cumulative distribution takes the value 0.9.

In practice, the conditional probability $P_c(\lambda_{i+1}|\lambda_i)$ can be replaced by conditional local phase space distribution $\rho(\lambda|\lambda_i)$ of forward trajectories staring from interface λ_i, which further improves the computational efficiency. The difference between $\rho(\lambda|\lambda_i)$ and $P_c(\lambda_{i+1}|\lambda_i)$ lies in that trajectories that cross a given λ for multiple times contribute many times to the former but only once to the latter. Surely $\rho(\lambda|\lambda_i)$ only approximates $P_c(\lambda_{i+1}|\lambda_i)$, nevertheless, we found it has already worked very well. In addition, one may also use cumulative distribution $\rho_c(\lambda|\lambda_i) = \int_{\lambda_i}^{\lambda} \rho(\lambda|\lambda_i)d\lambda$ instead of $\rho(\lambda|\lambda_i)$ itself since the cumulative distribution converges more rapidly. We have validated the approach in a two-state model system and a two-dimensional lattice gas Ising model[32]. Our approach is shown to be much more efficient than the conventional FFS method without losing accuracy, and it can also well reproduce the two-step nucleation scenario of the Ising model with easy identification of the hidding IMS.

2.3 Details of the sampling process

In simulations, the sampling is done by using the single order parameter $\lambda = x$, the end-to-end distance.

Determining the interface location–For a moderate strength of the pulling force (e.g., $F = 1.0$ in arbitrary unit), to sample the rupture of the protein, we start from the native configuration and run a long trajectory to determine the location of the first interface λ_0. The conditional local distribution $\rho(\lambda|\mathbf{A})$ (in relative probability) and cumulative distribution $\rho_c(\lambda|\mathbf{A})$ are plotted in Fig.1(a). $\rho(\lambda|\mathbf{A})$ shows a well Gaussian-like distribution. The first interface λ_0 determined by the self-optimized approach is located at $\lambda_0 \approx 43.6$ where $\rho_c(\lambda|\mathbf{A}) = 0.9$. Then, configurations on λ_0 are sampled by using the obtained long trajectory and the location of first interface.

Based on a known interface λ_i, the next interface λ_{i+1} can be determined as follows. We run sufficient many forward trajectories (whose length $T_2 = 10^3 dt$ with time step $dt = 0.01$) starting from configurations on λ_i to get the conditional local distribution $\rho(\lambda|\lambda_i)$ and cumulative distribution $\rho(\lambda|\lambda_{i+1})$. As an example, Fig.1(b) shows the corresponding distributions with the known interface λ_0. Similarly, the next interface λ_1 is located at $\lambda_1 = 44.1$ where $\rho(\lambda|\lambda_{i+1}) = 0.9$.

Searching for unknown IMS–To determine the existence of possible IMS, we need to monitor special trajectories that end neither at **A** nor at **B**. In simulations, a trajectory is considered to be a special trajectory for a sufficient long time $T_s = 10^5 dt$ which is larger than typical relaxation time to the native state **A** (which is of magnitude $10^4 dt$). As shown in Fig.2, the number of special trajectories increases sharply around $\lambda \approx 49$ demonstrating an IMS nearby. To further validate the existence of an IMS, we obtain the local phase space distribution starting from a configuration picked up random-
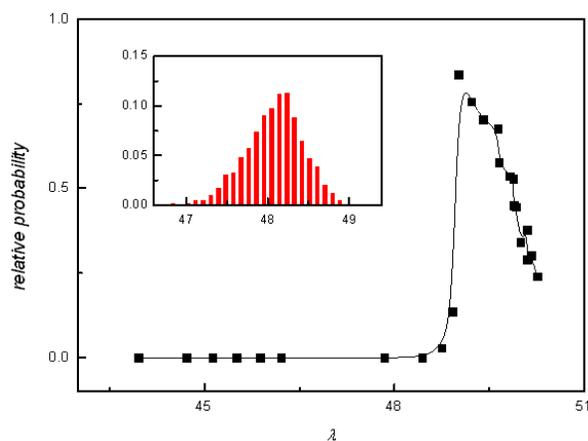
Fig. 2 (Color online) Ratio of special trajectories that end neither at the initial state **A** nor at the next interface to the total number of trajectories. Inset shows the local phase space distribution starting from a configuration picked up randomly from these special trajectories.

ly from these special trajectories. The result is plotted in the inset of Fig.2, in which a clear peak can be observed at about $\lambda \approx 48.1$.

Sampling of the transition rate–For small or large enough forces, the rupture takes almost all the direct or multistep pathway, respectively. Transition rates for these parameter regions are sampled straightforward by the self-optimized sampling method, and are averaged over 20 repeating simulations for each of a given force. However, for moderate pulling forces where both of the rupture pathways are taken at the same time, direct sampling of the precise transition rate is impossible. The main reason is due to the fact that the two pathways are not well distinguished for the earlier rupture stage. On the one hand, to sample correctly the transition rate, configurations recorded on these interfaces should represent correctly the relative probability that the protein unfolds via each of the pathways (if not, the deviation will be amplified exponentially and the obtained transition rate will diverge far from its real value), on the other hand, to obtain the relative probability, the transition rate should be correctly sampled at first. Thus, the sampled results can only provide qualitative properties of the complex rupture process (such as, two rupture pathways with multistep transition) but not quantitative information about it (e.g., the transition rate).
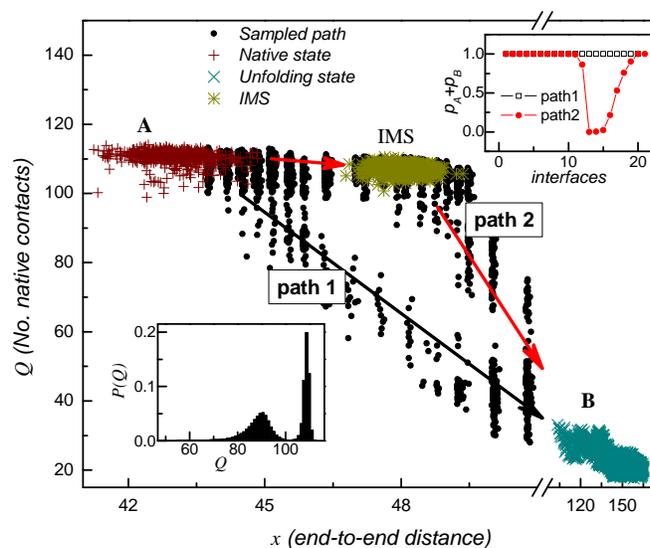

Fig. 3 (Color online) Two dominant pathways of protein unfolding: A direct unfolding pathway (path 1, indicated by the black arrow) from the native state (**A**) to the unfolding state (**B**), and a two-step unfolding pathway (path 2, indicated by the red arrows) with an intermediate metastable state (IMS). The top-right inset shows the summation of committor functions $p_A(\lambda_i)$ and $p_B(\lambda_i)$ (see the text for their definition) as a function of interface λ_i for path 1 (black open square) and path 2 (red filled cycle), indicating a one-step unfolding pathway and a two-step unfolding pathway involving three stable states, respectively. The bottom-left inset presents a bimodal distribution of Q for configurations generated from the interface about $x = 46$.

3 Results

3.1 Multiple dominant rupture pathways involving multistep transition

By applying our self-optimized approach, we have successfully obtained the dominant mechanical unfolding pathways for protein L. The sampling is done by using the single order parameter x, the end-to-end distance, but the configurations along the dominant pathways are plotted in the $Q - x$ plane, where Q is the number of remaining native contacts during the rupture process. Fig.3 presents the results for a moderate strength of pulling force, where two distinguished rupture pathways are clearly shown in the $Q - x$ plane. For path 1 indicated by the black arrow, the protein unfolds directly from the native state **A** to the unfolding state **B**. This is a one-step transition process, for which any long-enough trajectory either ends at **A** or reaches **B**. Accordingly, the summation of the committor functions $p_A(\lambda_i)$ and $p_B(\lambda_i)$ would be 1 for any interface λ_i, as plotted in the top-right inset of Fig.3. Here $p_A(\lambda_i)$ (or $p_B(\lambda_i)$) is defined as the probability that a trajectory initiated from configurations at λ_i reaches **A** before **B**(or **B** before **A**)[28], respectively. For path 2 indicated by the red arrows, on the other hand, the whole rupture process shows a clear feature of two-step transition involving an apparent IMS: The protein first unfolds from **A** to IMS and then to **B**. The emergence of the IMS leads to very long trajectories trapped around, which is verified by the abrupt decrease of $p_A(\lambda_i) + p_B(\lambda_i)$ for interfaces near the IMS as also plotted in the top-right inset of Fig.3. Moreover, the existence of two different pathways is also evident from the bimodal distribution of Q for configurations at a fixed interface, as shown for $x = 46$ in the bottom-left inset of Fig.3. In short, Fig.3 validates clearly a complex protein unfolding mechanism via multiple dominant pathways involving multistep transitions.

While Fig.3 presents a typical picture of multiple-pathway unfolding, which pathway dominates the process is dependent on the strength of the force F. Here are some important observations from our simulation. If F is very small, the protein unfolds only through path I and no IMS exists. With increasing F, the probability of path II increases rapidly and more and more trajectories starting from the native state **A** will enter the IMS. If the force is not too large, these trajectories along path II will be *trapped* for a long time in the IMS, followed by a sudden transition to the final state **B**. If the force is large enough, only the indirect unfolding pathway II can be observed. Note that the transition from **A** to IMS is always much easier than that from IMS to **B**, i.e., the latter is the rate-limiting step of path II.

Based on above picture, it is feasible for us to build up a simple kinetic model involving the transitions among the three
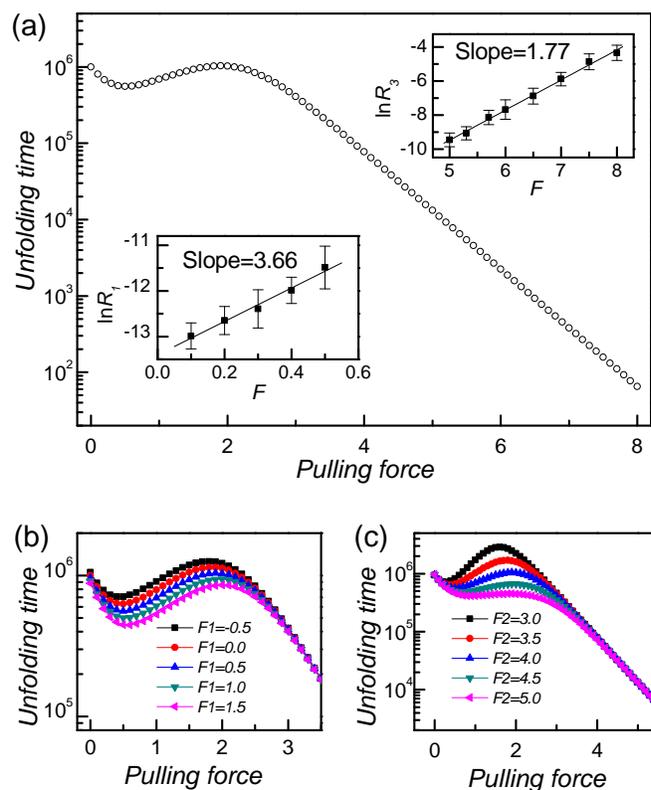

Fig. 4 (Color online) (a) Unfolding time τ as a function of external force F, and robustness of the roll-over behavior for the switch parameters (b) F_1 and (c) F_2. τ is calculated by the sampled parameters associated with the protein we used. The bottom-left and top-right insets in (a) present the dependencies of unfolding rates R_1 and R_2 on external pulling F sampled for small and large forces, respectively, both of which can be well fitted by Bell's formula $R_i = m_i e^{n_i F}$ with $m_1 = 1.52 \times 10^{-6}$, $n_1 = 3.66$ and $m_2 = 1.07 \times 10^{-8}$, $n_2 = 1.77$. The other switching parameter is $F_2 = F_1 + 3.5$ in (b) and $F_1 = 0.5$ in (c).

states **A**, IMS and **B** as follows,

$$A \xrightarrow{R_1} B \tag{1a}$$

$$A \underset{R_{-2}}{\overset{R_2}{\rightleftharpoons}} IMS \tag{1b}$$

$$IMS \xrightarrow{R_3} B. \tag{1c}$$

The direct unfolding pathway I is given by Eq.(1a), while the indirect two-step pathway II is contained in Eq.(1b) and(1c). Note that the rupture process is completed once the final state **B** is reached, such that Eq.(1a) and (1c) are irreversible. But it is possible that the molecule could change from the IMS back to the initial state **A**, hence Eq.(1b) is reversible (Surely in the presence of pulling force, the rate R_{-2} is small compared to R_2). Generally, the ensemble-averaged rupture time τ may be written as $\tau = w_1\tau_1 + (1 - w_1)\tau_2$ where w_1 is the relative probability that the protein unfolds via the direct pathway I, τ_1 or τ_2 are the rupture time through path I or path II, respectively. Based on this kinetic model, we may estimate w_1 by $R_1/(R_1 + R_2)$, taking into account that $R_1(R_2)$ is proportional to the number of trajectories out from **A** per unit time that follows path I (path II) respectively. The rupture time τ_1 is simply given by $1/R_1$, while for path II we have $\tau_2 \simeq 1/R_3$ because $R_3 \ll R_2$ as discussed in the last paragraph. Hence we have

$$\tau \simeq w_1 \frac{1}{R_1} + (1 - w_1)\frac{1}{R_3} = \frac{R_2 + R_3}{R_1 + R_2}\frac{1}{R_3} \tag{2}$$

To investigate the dependence of τ on the force F, we now need to know how the rates R_i $(i = 1, 2, 3)$ depend on F. As reported in many experimental and theoretical studies, R_i as a function of F generally can be described by the well-known Bell's formula[33],

$$R_i(F) = m_i e^{n_i F} \tag{3}$$

where m_i and n_i $(i = 1, 2, 3)$ are positive constants. As stated above, the protein unfolds via path I or II for small or large force respectively, which facilitates us to get the parameters for R_1 and R_3. These are shown in the insets of Fig.4(a), where good fittings to Bell's formula are demonstrated, with $m_1 = 1.52 \times 10^{-6}$, $n_1 = 3.66$ and $m_3 = 1.07 \times 10^{-8}$, $n_3 = 1.77$ (Note that the rate R_i and force F are shown in arbitrary units here, and its scaling with real unit is not important for the present study). Nevertheless, the parameters for R_2 is not easy to get, because a trajectory may jump between two different paths when they coexist and it is hard to distinguish clearly which path it belongs to. To proceed, we again use the fact that *almost all* paths are along path I (path II) for F less (larger) than some critical value $F_1(F_2)$, respectively. Without lose of generality, for instance, we may set this *almost all* value to be 99%. For the threshold force F_1, we may choose it as the value where the Bell's formula starts to break, which can

be estimated as $F_1 \simeq 0.5$ according to the left-bottom inset in Fig.4(a). Similarly, an acceptable estimation of F_2 can be $F_2 \simeq 4.0$ or less. For these parameters, our *almost all* criteria means that $w_1(F_1) = R_1(F_1)/[R_1(F_1) + R_2(F_1)] = 0.99$ and $w_1(F_2) = 0.01$, which gives that $R_2/R_1 = 3.22 \times 10^{-3}e^{2.63F}$ corresponding to $m_2 = 4.89 \times 10^{-9}$ and $n_2 = 6.29$.

It is now instructive to investigate how the rupture process depends on the force based on these parameters. If F is small, R_2 is much smaller than R_1 and the protein mainly unfolds via path I as expected. With increasing F, R_2 increases rather rapidly and path II emerges and begins to dominate. Once the protein selects path II, however, it will be trapped into the IMS because both R_3 and R_{-2} are generally small. Note that, for any force strength, R_3 is always much smaller than R_1, such that τ_2 given approximately by $1/R_3$ is always much larger than $\tau_1 = 1/R_1$. Therefore, the force-enhanced-trapping of protein into the IMS along path II will lead to increase of the total rupture time τ with F in the small to moderate force range. Of course, if F is too large, only path II can be observed and τ will decrease with F again. Consequently, the dependence of τ on F is plotted in Fig.4(a), where a clear-cut maximum appears in the moderate force range. The initial decrease of τ with F in the small force range is consistent with the fact that only path I can be observed there and R_1 increases with F according to the Bell's formula. In a word, the proposed rupture mechanism associated with IMS and multiple pathways here provides a rather novel and reasonable interpretation of the rollover feature of rupture time observed in single-molecule experiments.

Since parameters F_1 and F_2 are somewhat arbitrarily chosen, it is necessary to investigate the robustness of the rollover rupture behavior to these parameters. In Fig.4(b) and (c), we show the dependence of rupture time on the pulling force for various F_1 and F_2. Obviously, the rollover of the rupture time exists for a wide range of parameter values. F_1 affects qualitatively little the dependence but only quantitatively the value of parameters for path switching, while smaller F_2 results in a more sharp rollover of the rupture time. Interestingly, broadening of the switching region results in a series of rupture times which roll over more and more smoothly, which is quite similar to the phenomenon observed in the pulling simulations of ubiquitin with a change of the pulling direction[14].

3.2 A trapping effect of IMS induced abnormal dominant pathway switching

To further illustrate the underlying physics of above mechanism, we have plotted typical configurations of the native state **A**, the IMS and the unfolded state **B** in Fig.5(a). Clearly, the IMS is quite similar in configuration to the native state **A** except that the tail at the N-terminal is straightened, which results in an increment of the end-to-end distance $\Delta x \approx 5nm$. Simi-
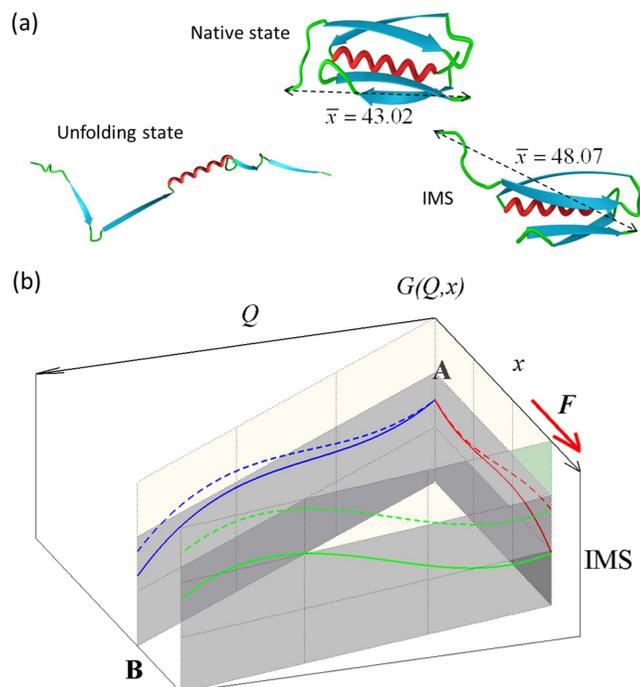
Fig. 5 (Color online) (a) Typical configurations of the native state, IMS and the unfolding state. The configuration of IMS is quite similar to the one of the native state except that the tail at the N-terminal is straightened. (b) A schematic diagram for the rupture mechanism of an IMS induced abnormal switching of dominant unfolding pathways from the thermodynamically favored one to a mechanically preferred one. The dash(solid) lines denote conceptual free energy contours for small(large) external forces.

lar extension before a full unfolding event has also been observed experimentally in the elongation of single titin proteins using the atomic force microscope, which was considered to be an important 'unfolding intermediate' unrecognized before[5]. Since the applied pulling force is always along the end-to-end direction, the elongation of the tail results in a change of pulling direction relative to the main body (β sheets) of the protein (as indicated by arrows in Fig.5(a)), which could affect strongly the dependence of unfolding rate on the external force[14]. Note that such a type of IMS cannot be stable if no external force is present. What's more, as a result of the special configuration of IMS, external pulling will enhance the relative stability of IMS. As illustrated in Fig.5(b), the key of the revealed novel rupture mechanism is that, IMS attracts more and more transition trajectories flowing out of the native state as external force increases. For large forces, these attracted trajectories have to overcome a high barrier back to the native state to take the direct unfolding pathway. Consequently, even though the overall transition rate of the multistep pathway is relative small, the protein still has to unfolding via it rather than the direct pathway forbidden by the IMS.

4 Conclusion

In summary, direct sampling of multiple single-molecular rupture pathways involving multistep transition has been realized by the self-optimized sampling method we proposed. A previously unrecognized intermediate stable state (IMS) was identified. Remarkably, we revealed a force-enhanced trajectory trapping effect of the IMS, which provides a novel mechanism for the "rollover" of the rupture time and also deep insights for the protein mechanical unfolding. We believe that direct sampling of complex dominant pathways can inspire intensive research interests and open a wide range of perspectives on the study of single-molecular pulling dynamics as well as other similar problems via unknown IMS and complex transition pathways.

5 Acknowledgments

References

1 G. Binnig, C. F. Quate and C. Gerber, *Phys. Rev. Letter*, 1986, **56**, 930.
2 L. Tskhovrebova, J. Trinick, J. A. Sleep and R. M. Simmons, *Nature*, 1997, **387**, 308.
3 M. S. Z. Kellermayer, S. B. Smith, H. L. Granzier and C. Bustamante, *Science*, 1997, **276**, 1112.
4 M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez and H. E. Gaub, *Science*, 1997, **276**, 1109.

5 P. E. Marszalek, H. Lu, H. Li, M. Carrion-Vazquez, A. F. Oberhauser, K. Schulten and J. M. Fernandez, *Nature*, 1999, **402**, 100.

6 B. T. Marshall, M. Long, J. W. Piper, T. Yago, R. P. McEver and C. Zhu, *Nature*, 2003, **423**, 190.

7 E. Evans, A. Leung, V. Heinrich and C. Zhu, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 11281.

8 V. Barsegov and D. Thirumalai, *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 1835.

9 G. Hummer and A. Szabo, *Biophys. J.*, 2003, **85**, 5.

10 O. K. Dudko, A. E. Filippov, J. Klafter and M. Urbakh, *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 11378.

11 O. K. Dudko, G. Hummer and A. Szabo, *Phys. Rev. Lett.*, 2006, **96**, 108101.

12 L. Freund, *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 8818.

13 S. Rakshit and S. Sivasankar, *Phys. Chem. Chem. Phys.*, 2014, **16**, 2211.

14 R. B. Best, E. Paci, G. Hummer and O. K. Dudko, *J. Phys.Chem. B*, 2008, **112**, 5968.

15 Y. Suzuki and O. K. Dudko, *Phys. Rev. Lett.*, 2010, **104**, 048101.

16 Y. Suzuki and O. K. Dudko, *J. Chem. Phys.*, 2011, **134**, 065102.

17 S. S. Konda, J. N. Brantley, B. T. Varghese, K. M. Wiggins, C. W. Bielawski and D. E. Makarov, *J. Am. Chem. Soc.*, 2013, **135**, 12722.

18 Y. V. Pereverzev, O. V. Prezhdo, M. Forero, E. V. Sokurenko and W. E. Thomas, *Biophys. J.*, 2005, **89**, 1446.

19 J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327.

20 J. Karanicolas and C. L. BrooksIII, *Protein Sci.*, 2002, **11**, 2351.

21 H. S. Chan and K. A. Dill, *Proteins*, 1998, **30**, 2.

22 H. Nymeyer, A. E. Garcia and J. N. Onuchic, *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 5921.

23 N. Koga and S. Takada, *J. Mol. Biol.*, 2001, **313**, 171.

24 J. W. O'Neill, D. E. Kim, D. Baker and K. Y. Zhang, *Acta. Crystallogr. Sect. D: Biol. Crystallogr.*, 2001, **57**, 480.

25 D. K. West, P. D. Olmsted and E. Paci, *J. Chem. Phys.*, 2006, **124**, 154909.

26 J. Karanicolas and C. L. BrooksIII, *J. Mol. Biol.*, 2003, **334**, 309.

27 R. J. Allen, P. B. Warren and P. R. ten Wolde, *Phys. Rev. Lett.*, 2005, **94**, 018104.

28 R. J. Allen, C. Valeriani and P. R. ten Wolde, *J. Phys.: Condens. Matter*, 2009, **21**, 463102.

29 T. Li, D. Donadio and G. Galli, *Nature Communications*, 2013, **4**, 1887.

30 R. J. Allen, D. Frenkel and P. R. ten Wolde, *J. Chem. Phys.*, 2006, **124**, 024102.

31 C. Velez-Vega, E. E. Borrero and F. A. Escobedo, *J. Chem. Phys.*, 2009, **125**, 164904.

32 H. Jiang, M. Pu and Z. Hou, *Sci. Chin. Chem.*, 2014, **57**, 165.

33 G. I. Bell, *Science*, 1978, **200**, 618.