

## **Switch from thermal to force-driven pathways of protein refolding**

Maksim Kouza, Pham Dang Lan, Alexander M. Gabovich, Andrzej Kolinski, and Mai Suan Li

Citation: [The Journal of Chemical Physics](#) **146**, 135101 (2017); doi: 10.1063/1.4979201

View online: <http://dx.doi.org/10.1063/1.4979201>

View Table of Contents: <http://aip.scitation.org/toc/jcp/146/13>

Published by the [American Institute of Physics](#)

---

---

**COMPLETELY  
REDESIGNED!**



**PHYSICS  
TODAY**

*Physics Today* Buyer's Guide  
Search with a purpose.

# Switch from thermal to force-driven pathways of protein refolding

Maksim Kouza,<sup>1,a)</sup> Pham Dang Lan,<sup>2</sup> Alexander M. Gabovich,<sup>3</sup> Andrzej Kolinski,<sup>1,b)</sup> and Mai Suan Li<sup>4,c)</sup>

<sup>1</sup>Faculty of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland

<sup>2</sup>Institute for Computational Science and Technology, Tan Chanh Hiep Ward, District 12, Ho Chi Minh City, Vietnam and Department of Theoretical Physics, Faculty of Physics and Engineering Physics, Ho Chi Minh University of Science, Vietnam

<sup>3</sup>Institute of Physics, National Academy of Sciences of Ukraine, 46, Nauka Ave., Kyiv 03680, Ukraine

<sup>4</sup>Institute of Physics, Polish Academy of Science, Al. Lotników 32/46, 02-668 Warsaw, Poland

(Received 18 October 2016; accepted 13 March 2017; published online 3 April 2017)

The impact of the quenched force on protein folding pathways and free energy landscape was studied in detail. Using the coarse-grain Go model, we have obtained the low, middle, and high force regimes for protein refolding under the quenched force. The folding pathways in the low force regime coincide with the thermal ones. A clear switch from thermal folding pathways to force-driven pathways in the middle force regime was observed. The distance between the denatured state and transition state  $x_f$  in the temperature-driven regime is smaller than in the force-driven one. The distance  $x_f$  obtained in the middle force regime is consistent with the available experimental data suggesting that atomic force microscopy experiments deal with the force-regime which is just above the thermal one. Published by AIP Publishing. [<http://dx.doi.org/10.1063/1.4979201>]

## I. INTRODUCTION

Despite substantial advance has been achieved in recent years,<sup>1–3</sup> deciphering free energy landscape (FEL) of biomolecules remains challenging in molecular biology. Single molecule force spectroscopy (SMFS) experiments which are able to distinguish the fluctuations of individual molecules from the ensemble average behavior are intensively used to probe FEL and molecular interactions. The atomic force microscopy (AFM) is one of the most suitable tools to study proteins<sup>4</sup> which are mechanically more stable than DNA and RNA. In AFM and other SMFS techniques, the external mechanical force  $f$  is used as an additional parameter to gain information on FEL. If  $f$  is sufficiently small then one can expect that the external force moves the FEL profile leaving the distance between the transition state (TS) and the native state (NS),  $x_u$ , and the distance between TS and the denatured state (DS),  $x_f$ , unchanged. Then the unfolding barrier is reduced by an amount of  $\Delta\Delta G_{TS-NS} = -fx_u$ , while the folding barrier is elevated by  $\Delta\Delta G_{TS-DS} = fx_f$  as the force obstacles refolding.<sup>5</sup> This situation is often referred to as the Bell approximation,<sup>5</sup> where the dependence of folding time  $\tau_f$  on  $f$  is given by the following equation:

$$\tau_f = \tau_f^0 \exp(fx_f/k_B T). \quad (1)$$

Here  $k_B$ ,  $T$ , and  $\tau_f^0$  are Boltzmann's constant, temperature, and the folding time in the absence of external force, respectively.

To go beyond the Bell approximation, i.e., to obtain the force dependence of unfolding/folding times for a wide range

of  $f$ , various approximations have been proposed.<sup>6–10</sup> The existence of parallel (un)folding pathways, the description of which is not possible within the Bell theory, has been tackled in the field experimentally.<sup>11</sup> Despite clear improvement of the Bell theory, the validity of these new theories remains restricted to some force regions. On the other hand, because the Bell formula is widely used to interpret experimental data, here we follow a primitive point of view that the Bell theory works for arbitrary forces but the value of  $x_f$  (and  $x_u$ ) depends on force regions.

The unfolding FEL of proteins, in particular,  $x_u$ , has been studied thoroughly,<sup>2,12–14</sup> whereas the refolding under the quenched force attracts much less attention. The distance  $x_f$  was experimentally measured for some proteins<sup>15–19</sup> (Table I) but its theoretical estimation was made only for immunoglobulin domain Ig27<sup>20</sup> and ubiquitin.<sup>21</sup> Thus, understanding of refolding FEL parameters of other experimentally studied proteins is of interest. More importantly, it remains unclear that in which force regime the distance between the denatured and native states has been measured in experiments and how refolding pathways depend on the quenched force.

The most detailed information on FEL may be gained from the all-atom simulations but this approach is restricted to rather short peptides and proteins<sup>22,23</sup> due to its high computational expenses. In the presence of external force, the problem becomes even more critical because the refolding times exponentially increase with  $f$  (Eq. (1)) and it is not feasible even on fastest computers.<sup>24</sup> Therefore, coarse-grained models are more suitable for this purpose. In this paper, we have chosen the structure-based Go model which has the additional advantage that the native state of proteins is *a priori* known and may be taken from the protein data bank (PDB).

<sup>a)</sup>Electronic mail: mkouza@chem.uw.edu.pl

<sup>b)</sup>Electronic mail: kolinski@chem.uw.edu.pl

<sup>c)</sup>Electronic mail: masli@ifpan.edu.pl

TABLE I. Summary of the distances from the unfolded to the transition state obtained from simulation in LF and MF regimes and from experimental data.  $x_f$  is measured in nm.  $T_{min}$ , obtained by simulation, is in  $\epsilon_H/k_B$ . References to experimental works are shown in the last column.

Protein	PDB code	Class	$T_{min}$ , Sim	$x_F$ , LF	$x_F$ , MF	$x_F$ , Expt.	Reference
Ubiquitin	1UBQ	$\alpha/\beta$	0.55	0.32	1.01	0.8	15
Src SH3 A7C/N59C	1SRL	$\beta$	0.53	0.88	4.86	$5.25 \pm 0.25$	17
ACBP 1-86	1NTI	$\alpha$	0.45	1.12	6.6	$6.7 \pm 0.10$	18
~70aa C-terminal cam	1CFC	$\alpha$	0.35	1.02	6.14	8.0	19
GB protein	1PGA	$\alpha/\beta$	0.53	0.23	1.9	2.1	36
Titin I27	1TIT	$\beta$	0.44	0.12	0.84	N/A	
Apomyoglobin N/C (molten globule)	1BZ6	$\alpha$				$14.9 \pm 1.5$	16

Using the Go model,<sup>25</sup> we have studied refolding of several proteins under quenched forces in a range from 0 to  $\sim 14$  pN. It was shown that there exist low, middle, and high force regimes with different  $x_f$ . At low forces, folding pathways were found to coincide with the thermal ones. The crossover from the low- to middle-force regimes occurs at  $f_{switch} \approx 3\text{-}7$  pN. The simulation values of  $x_f$  obtained in the middle force range agree with the experimental ones implying that experiments have been carried out in the second force regime which is just above the temperature-driven region.

The results of our simulations demonstrated that regardless of the structural type of protein, there is a switch from thermal to force-driven pathways and the crossover between different force regimes which corresponds to the emergence of an additional energy barrier, analogously to what we observed in our previous studies on mechanical unfolding.<sup>26,27</sup> These results could delve into certain deep and unanswered questions of unfolding and refolding studies. For example, one can use the force dependencies on time to estimate  $f_{switch}$ . We noted that a switch is observed at  $f_{switch}$  where the linear fitting based on Bell's theory ceases to work. Another consequence of this observation is that the thermal folding pathway can be probed by a force-clamp technique if the quenched force is smaller than  $f_{switch}$ . However, if the quenched force falls within the interval [ $f_{switch} - f_c$ ], which still allows a protein to refold, the pathways will be different from the thermal ones.

## II. MATERIAL AND METHODS

### A. Coarse-grained structure-based model

Folding of proteins takes microseconds to seconds. This time scale can be currently reached for small- and middle-sized proteins in the explicit solvent by the most powerful supercomputer.<sup>24</sup> On the other hand, the folding time grows with force exponentially that does not allow for estimating  $x_f$  by all-atom molecular dynamics (MD) simulations. Therefore, we will employ the coarse-grained Go-like model<sup>25</sup> where only  $C_\alpha$ -carbons are taken into account. The rationale of our choice is based on the fact that the protein folding is largely defined by the topology of native conformation.<sup>28</sup> Native structures for Go modeling were retrieved from protein data bank (PDB) using codes listed in Table I.

The interactions between residues are assumed to be Go-like, and the energy of such a model is as follows:<sup>25</sup>

$$\begin{aligned} E = & \sum_{bonds} K_r(r_i - r_{0i})^2 + \sum_{angles} K_\theta(\theta_i - \theta_{0i})^2 \\ & + \sum_{dihedral} \{K_\phi^{(1)}[1 - \cos(\phi_i - \phi_{0i})] \\ & + K_\phi^{(3)}[1 - \cos 3(\phi_i - \phi_{0i})]\} \\ & + \sum_{i>j-3}^{NC} \epsilon_H \left[ 5\left(\frac{r_{0ij}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{0ij}}{r_{ij}}\right)^{10} \right] \\ & + \sum_{i>j-3}^{NNC} \epsilon_H \left( \frac{C}{r_{ij}} \right)^{12} - |\vec{f} \cdot \vec{R}|. \end{aligned} \quad (2)$$

Here  $\Delta\phi_i = \phi_i - \phi_{0i}$ ,  $r_{i,i+1}$  is the distance between beads  $i$  and  $i + 1$ ,  $\theta_i$  is the bond angle between bonds  $(i - 1)$  and  $i$ , and  $r_{ij}$  is the distance between the  $i$ th and  $j$ th residues. Subscripts “0,” “NC,” and “NNC” refer to the native conformation, native contacts, and non-native contacts, respectively. Residues  $i$  and  $j$  are in native contact if  $r_{0ij}$  is less than a cutoff distance  $d_c$  taken to be  $d_c = 6.5$  Å, where  $r_{0ij}$  is the distance between the residues in the native conformation.

The first harmonic term in Eq. (2) accounts for chain connectivity, and the second term represents the bond angle potential. The potential for the dihedral angle degrees of freedom is given by the third term in Eq. (2). The interaction energy between residues that are separated by at least 3 beads is given by 10-12 Lennard-Jones potential. A soft sphere repulsive potential (the fourth term in Eq. (2)) disfavors the formation of non-native contacts. The last term is used to account the force applied to C and N termini along the end-to-end vector  $\vec{R}$  in the case of constant force simulations. We added an energy  $-|\vec{f} \cdot \vec{R}|$  to the total energy of the system, where  $\vec{R}$  is the end-to-end vector and  $\vec{f}$  is the force applied to both termini.

We choose  $K_r = 100\epsilon_H/\text{\AA}^2$ ,  $K_\theta = 20\epsilon_H/\text{rad}^2$ ,  $K_\phi^{(1)} = \epsilon_H$ , and  $K_\phi^{(3)} = 0.5\epsilon_H$ , where  $\epsilon_H$  is the characteristic hydrogen bond energy and  $C = 4$  Å. In our model, the force unit  $[f] = \epsilon_H/\text{\AA} = 68$  pN.<sup>13</sup> The dynamics of the polypeptide chain can be described by the Langevin equation which was numerically solved using the Verlet algorithm<sup>29</sup> with time step  $\Delta t = 0.005\tau_L$ . Here  $\tau_L = (ma^2/\epsilon_H)^{1/2} \approx 3$  ps,  $m$  is the typical

mass of amino acids, and  $a$  is the distance between neighboring beads.<sup>30</sup>

## B. Tools and measures used in the analysis

### 1. Folding time and $T_{\min}$

In order to study the crossover from the temperature-driven (low force) regime to the force-driven one, we have studied 5 proteins for which the experimental values of  $x_f$  are available (Table I). The folding time  $t_{\text{fold}}$  is a median value of folding times collected from individual MD runs. To obtain reliable results for a given value of force, a pool of 50–100 trajectories has been generated. All simulations were performed at temperature  $T_{\min}$ , at which the folding is fastest in the absence of force.

### 2. Folding pathway

In order to study the folding pathway sequence, we monitored the fraction of native contacts of the  $\beta$ -strands and  $\alpha$ -helices as a function of progress variable  $\delta = \frac{t}{t_{\text{fold}}}$ . The advantage of  $\delta$  is that it permits one to study folding events in the unique window  $0 < \delta \leq 1$  despite huge diversity of folding times obtained from different MD trajectories. To get averaged folding pathway, we averaged the fraction of native contacts over a unique window  $0 < \delta \leq 1$  and monitor the folding sequencing with the help of the progressive variable  $\delta$ . The refolding pathway is represented by an averaged refolding pathway which generally may not necessarily be one of the individual refolding pathways. Our previous study has shown that the averaged refolding pathway is in good agreement with the most probable refolding pathway.<sup>21</sup>

### 3. All-atom reconstruction

To reconstruct the atomic details from the alpha carbon trace, we followed a two-step procedure involving (1)

reconstruction of the backbone atoms by Modeller<sup>31</sup> and (2) adding of the remaining atoms of side-chains followed by their optimization to remove structural inaccuracies using the SCWRL 4.0 package<sup>32</sup> described in more detail in Ref. 33.

### 4. Free energy landscape

The two-dimensional free energy landscape surface, constructed along the number of native contacts ( $N_{\text{NC}}$ ) as one axis and the end-to-end distance  $R_{\text{NC}}$  as the other, was given by  $\Delta G(N_{\text{NC}}, R_{\text{NC}}) = -k_B T [\ln P(N_{\text{NC}}, R_{\text{NC}}) - \ln P_{\max}]$ , where  $P(N_{\text{NC}}, R_{\text{NC}})$  was the probability distribution obtained from a histogram of MD data.  $P_{\max}$  was subtracted to ensure that  $\Delta G = 0$  at the global minimum of free-energy.

### 5. Crossover point between different force regimes

We define (albeit roughly) the crossover point as the point at which correlation coefficients calculated individually for each regime reach their maximum level.

## III. RESULTS AND DISCUSSIONS

First, we have computed  $T_{\min}$  in the absence of external force. The results are shown in Table I. As expected,  $\alpha$ -proteins have lower  $T_{\min}$  than  $\beta$ -proteins being less stable. The refolding times were computed at  $T_{\min}$  for various values of the external force  $f$ .

### A. Existence of three force regimes

The pioneer experimental work on refolding under quenched force was done for ubiquitin.<sup>15</sup> Our previous simulation work on this protein was restricted to a narrow force region.<sup>21</sup> Here we extend our study to higher forces, and as evident from Fig. 1(a) one has three different slopes  $x_f = 0.32, 1.01$ , and  $3.24$  nm corresponding to low, middle, and high force (HF) regimes. As expected  $x_f$  elevates

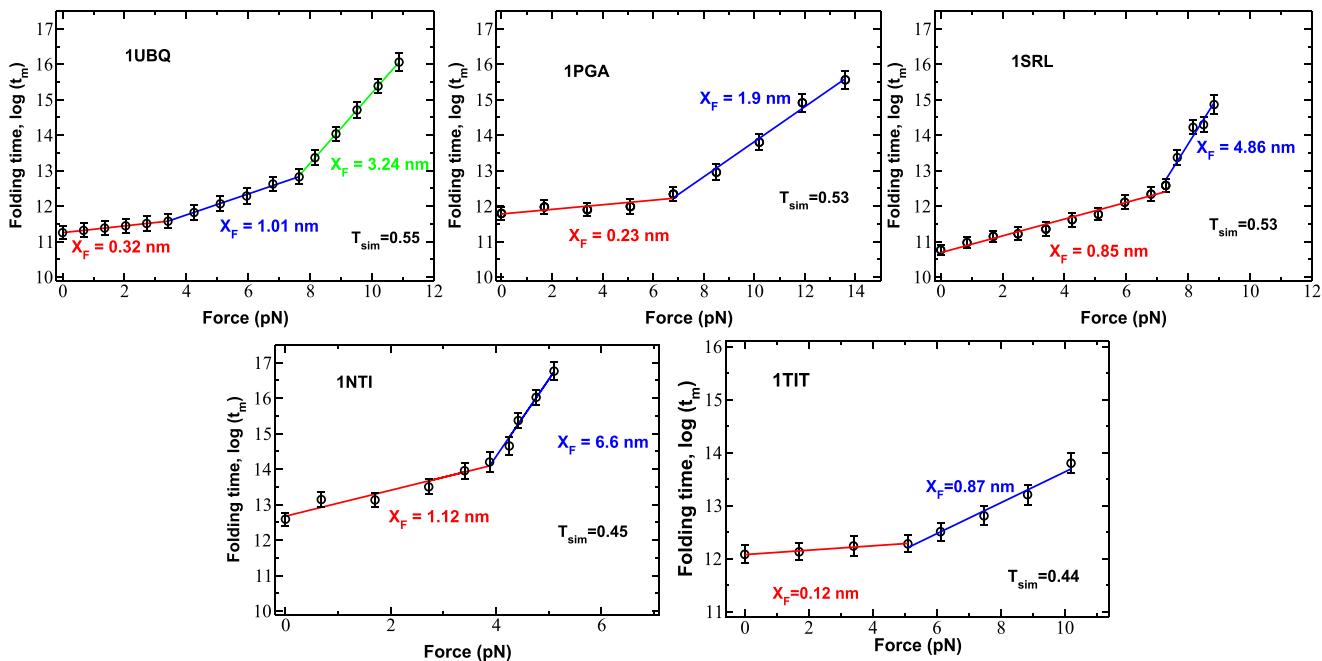


FIG. 1. Distance  $x_f$  in different force regimes. We show three regimes for ubiquitin and two regimes for the remaining proteins. Results were obtained at  $T_{\min}$  except the titin domain I27 for which the simulation was conducted at  $T_{\exp} = 0.44$  (or 294 K).

with  $f$  because the increase in folding times becomes steeper as force is increased. The value of  $x_f$  obtained in the middle force (MF) regime is close to the experimental result of  $\approx 0.8$  nm.<sup>15</sup> Using the same Go model, we have obtained  $x_f = 0.96$  nm<sup>13</sup> which is also consistent with the present estimate. But one should stress that in the prior theoretical<sup>13</sup> and experimental<sup>15</sup> works, the fit to the Bell theory (Eq. (1)) mixed the low force (LF) and partially MF regimes. The main feature of this paper is that we clearly distinguish different force scenarios.

One may think that 1UBQ data could have been divided into only two regimes, while 1SRL data could be divided into three regimes instead of two (Fig. 1). To clarify this issue, we divide UBQ data into two regimes at  $F = 7.85$  pN (Fig. S1 in the *supplementary material* (SM)) obtaining  $x_f = 0.71$  nm (correlation level  $R = 0.97$ ) for the low force regime. Here we adopt the criterion that a given force interval is not divided further into two regimes if the values of  $x_f$  obtained for two regimes do not differ from that of one regime more than 0.2 nm. Based on this criterion, one has to divide  $[0, 7.85$  pN] into two regions because  $x_f = 0.71$  nm differs from  $x_f = 0.32$  and  $1.01$  nm (Fig. S1) more than 0.2 nm, implying that 1UBQ has three force regimes.

To address the question whether 1SRL data could be divided into three regimes, we assume that the crossover at 3.5 pN exists (Fig. S2 in the *supplementary material*). For the  $[0, 7.25$  pN] interval, we obtain  $x_f = 0.85$  nm differing from  $x_f = 0.74$  and  $1.01$  nm of two sub-intervals less than 0.2 nm. Therefore, for 1SRL, we keep only two force regimes.

Because a clear discrimination between the LF and MF regimes is crucial for comparison with experiments (see below), we have implemented a simple but reasonable procedure to separate them. Namely, several choices for two force intervals have been probed, but the best choice should correspond to the highest fitting correlations for both regimes. For illustration, we consider protein 1PGA (Fig. S3 in the *supplementary material*). The correlation level  $R$  for the MF regime is nearly the same for three choices, but in the LF regime it visibly varies depending on how many data points have been picked up. Therefore, we choose the best case with the highest value  $R = 0.877$  as shown in Fig. 1.

The distance  $x_f$  was not experimentally measured for the titin domain I27 (PDB ID: 1TIT) but its critical force was obtained by the magnetic tweezer experiment of Chen *et al.*<sup>34</sup> Therefore we have also performed simulations for this protein at the experimental temperature  $T_{\text{exp}} = 294$  K<sup>34</sup> (Fig. 1). Because the folding temperature of I27 is 333 K<sup>35</sup> which

corresponds to  $T_F = 0.5\epsilon_H/k_B$  in our Go model,<sup>27</sup> we have  $T_{\text{exp}} = \frac{294}{333} \times 0.5 = 0.44$  in  $\epsilon_H/k_B$ .

## B. Experimentally measured $x_f$ in middle force regime

Because the HF regime is not interesting in terms of comparison with experiments and the estimations of folding times in this regime are too CPU consuming, we will focus on the LF and MF regions. Fig. 1 shows the results obtained for three proteins, 1PGA, 1SRL, and 1NTI, that are representative for  $\alpha/\beta$ -,  $\beta$ -, and  $\alpha$ -proteins. Their PDB structures are shown in Fig. 2. In the LF regime, we obtained  $x_f = 0.23, 0.88$ , and  $1.12$  nm for 1PGA, 1SRL, and 1NTI, respectively (Fig. 1). These values are far from the experimental estimates listed in Table I. This observation also holds for  $\alpha$ -protein 1CFC. The experimental data are also available for protein 1BZ6, but the theoretical estimation of  $x_f$  for this protein is beyond our computational facility because it has large  $x_f$  (Table I) corresponding to very slow folding.

Contrary to the LF case, the distances between TS and DS calculated in the MF regime are close to the experimental ones (Table I). For 1PGA, the agreement is nearly perfect as one has  $x_f = 1.9$  nm against 2.1 nm followed from the experiment of Cao and Li.<sup>36</sup> Overall the correlation between the experimental and simulation sets is very high with the correlation level  $R \approx 0.976$  (Fig. 3). Thus, in terms of  $x_f$  experiments probed the MF force regime.

Although  $x_f$  obtained in the LF region are far below experimental estimations they are highly correlated. Using the data from Table I, one can show that the correlation coefficient  $R \approx 0.90$ .

## C. Temperature dependence of $x_f$

Because the computation of refolding times under quenched force, in particular, at low temperatures, is very time consuming, we restricted our study to the region near  $T_{\min}$  and  $T_F$  and to protein 1PGA. Using the replica exchange method, we obtained the folding temperature  $T_F = 0.57$  (see Fig. S4 and simulation details in the *supplementary material*), which is a bit higher than  $T_{\min} = 0.53$  (Fig. S5). Because the folding temperature of 1PGA is 360.5 K,<sup>37</sup> the optimal temperature  $T_{\min} = \frac{0.53}{0.57} \times 360.5 = 335$  K. Thus  $T_{\min}$  is higher than  $T_{\text{exp}} = 277$  K<sup>36</sup> (or about 0.44 in  $\epsilon/k_B T$ ), where  $x_f$  was experimentally measured. To directly compare  $x_f$  with the experiment, we have performed additional simulations at  $T_{\text{exp}}$ . In the LF and MF regimes,  $x_f = 0.25$  and  $1.85$  nm (Fig. S6) which are close to 0.23 and 1.9 nm obtained at  $T_{\min}$  for 1PGA (Table I). The reason for similar results at  $T_{\min}$  and  $T_{\text{exp}}$  is

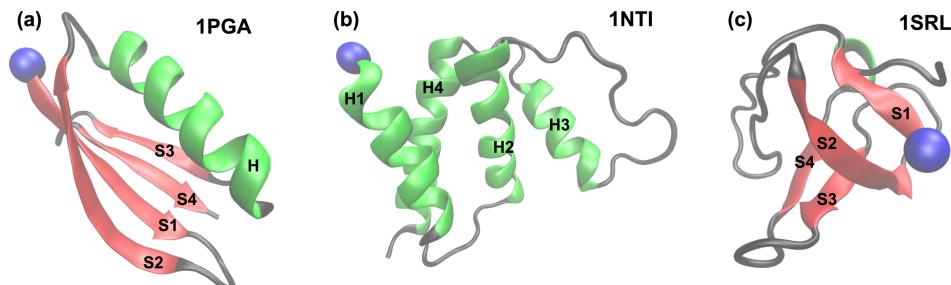


FIG. 2. Secondary structures of three representative proteins in the native state. Blue balls refer to the N-terminal.

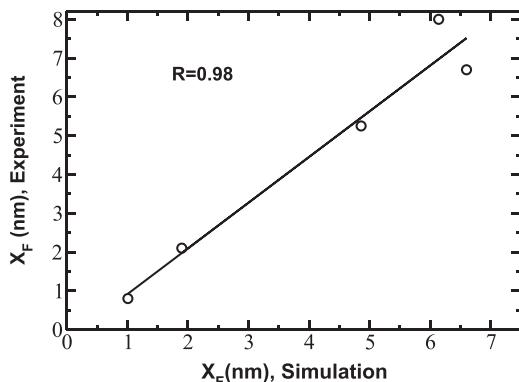


FIG. 3. Correlation between theoretical values of  $x_f$ , obtained in the MF regime, and experimental ones. The correlation level  $R = 0.98$ .

presumably two-fold. First, the experiment<sup>36</sup> was carried out not too far from  $T_{\min}$  which is located together with the folding temperature  $T_F$  at the wide bottom of the U-shape curve where the folding time is plotted as a function  $T$  (Fig. S5). Because in the absence of the external force the folding times at the bottom are nearly temperature-independent, one can expect that folding properties at relatively low  $f$  are not very different at  $T_{\min}$  and  $T_{\exp}$  leading to close values of  $x_f$  at these temperatures. Second, as  $T$  is decreased, the slope  $\gamma$  in Fig. S6 increases but  $x_f$  does not change much because  $x_f = k_B T \gamma$  according to Eq. (1). Thus, based on the results obtained for 1PGA, we anticipate that  $x_f$  is nearly the same at  $T_{\min}$  and  $T_{\exp}$  for other proteins.

For ubiquitin, using the same Go model (see Section II), one obtained  $T_F = 0.625 \epsilon_H/k_B$ <sup>13</sup> corresponding to the experimental  $T_F = 332.5$  K.<sup>38</sup> For this protein, we have performed simulations at  $T_{\min} = 0.55 \epsilon_H/k_B \approx 271$  K (Table I) which is very close to  $T_{\exp} = 277$  K where  $x_f$  was estimated.<sup>15</sup> Thus, our theoretical value of  $x_f$  at  $T_{\min}$  should not deviate much from that at  $T_{\exp}$ .

#### D. Switch from temperature-driven to force-driven folding pathways

The folding of B domain of protein G (GB) has been extensively investigated experimentally<sup>39,40</sup> and theoretically.<sup>41–43</sup> One notes that there is a controversy regarding the folding mechanism of GB. Protein engineering experiments have provided evidence that a well-formed C-terminal hairpin and partially folded helix are presented in the transition state ensemble.<sup>39</sup> This suggestion was also confirmed by simulations using a coarse-grained CABS model.<sup>42</sup> On the other hand, a number of alternative folding pathways involving the formation of helix-C-terminal hairpin or helix-N-terminal or helix-both hairpins, which can be additionally splitted into more pathways, have been reported.<sup>40,41,43</sup>

As individual folding pathways can be highly diverse and a consensus has not been reached on the specific folding pathways of GB, it is better to consider an ensemble of trajectories. For this reason, we focus not on the folding pathway of individual trajectories but also on the averaged one. Figs. 4(a)–4(c) show the dependence of the number of native contacts (NCs) of individual secondary structure elements on progress variable  $\delta$  for GB at  $f = 0$ , 3.4, and 10.2 pN. At  $f = 3.4$  pN, the acquisition of the native structure proceeded

via the pathway

$$(H, S2, S3) \rightarrow S1 \rightarrow S4. \quad (3)$$

We observed the same sequencing in the absence of the external force. Thus, below the borderline value  $f_{\text{switch}} \approx 7$  pN (Fig. 1), the protein folding is driven by thermal fluctuations. According to the low force scenario, the folding initiates from the formation of helix and N- and C-terminal hairpins (Figs. 4(a) and 4(b)). The helix is formed and dissolves intermediately, keeping an average number of native contacts about 80% through the simulation. Folding of C- and N-terminal hairpins indicates the complete formation of S2 and S3 strands, as those strands have native contacts only with adjacent S1 and S4 strands. The next folding event is the association of terminal hairpins into a  $\beta$ -sheet and simultaneous formation of contacts between the helix and  $\beta$ -sheet. Once the later contacts are established and S1 and S4 strands come together to form a  $\beta$ -sheet of four  $\beta$ -strands, the native-like conformation is formed.

At  $f = 10.2$  pN, i.e., above  $f_{\text{switch}}$ , as protein refolds, the sequence of the folding events obeys the following order:

$$(H, S2) \rightarrow S3 \rightarrow S1 \rightarrow S4. \quad (4)$$

According to the middle force regime, the first folding event is the formation of N-terminal hairpin (S1 with S2 strands) and helix (Fig. 4(c)). Interestingly, although a force is applied locally to both, N- and C-terminus, it seems to interfere with the C-terminal hairpin formation. The formation of the C-terminal hairpin (S3 with S4 strands) seems to be the rate limiting step in the folding in the middle force regime. Once C-terminal hairpin folded, the formation of  $\beta$ -sheet and helix- $\beta$ -sheet contacts completes the folding process.

As follows from Eqs. (3) and (4) and clearly seen from Fig. 5, refolding pathways in the force-driven regime are different from the thermal ones. This conclusion is also valid for protein 1SRL and protein 1NTI. In Sec. III E, we will analyze their folding pathways in more detail.

#### E. Is the switch from temperature-driven to force-driven folding pathways observed for other proteins?

An interesting question is whether the switch between temperature-driven and force-driven folding pathways observed for GB is also valid for other proteins? To check this, we chose conceptually very different proteins compared to GB: all- $\beta$  protein, the SH3 domain of Src (Src SH3, PDB ID: 1SRL), and the four- $\alpha$ -helix acyl-CoA binding protein (ACBP, PDB ID: 1NTI).

Src SH3 consists of four  $\beta$ -strands, S1, S2, S3, and S4, connected by loops between S1 and S2 as well as S3 and S4 (Fig. 2(c)). To obtain the sequence of folding events, we use the evolution of fraction of native contacts of  $\beta$ -strands. From Figs. 4(d)–4(f), we have the following refolding pathways for two force regimes:

$$S3 \rightarrow S2 \rightarrow S4 \rightarrow S1, \text{ low force pathway of Src SH3}, \quad (5a)$$

$$S3 \rightarrow S4 \rightarrow S2 \rightarrow S1, \text{ high force pathway of Src SH3}. \quad (5b)$$

We have obtained the same sequencing at  $f = 0.0$  and below  $f_{\text{switch}} \approx 7$  pN. According to the low force scenario

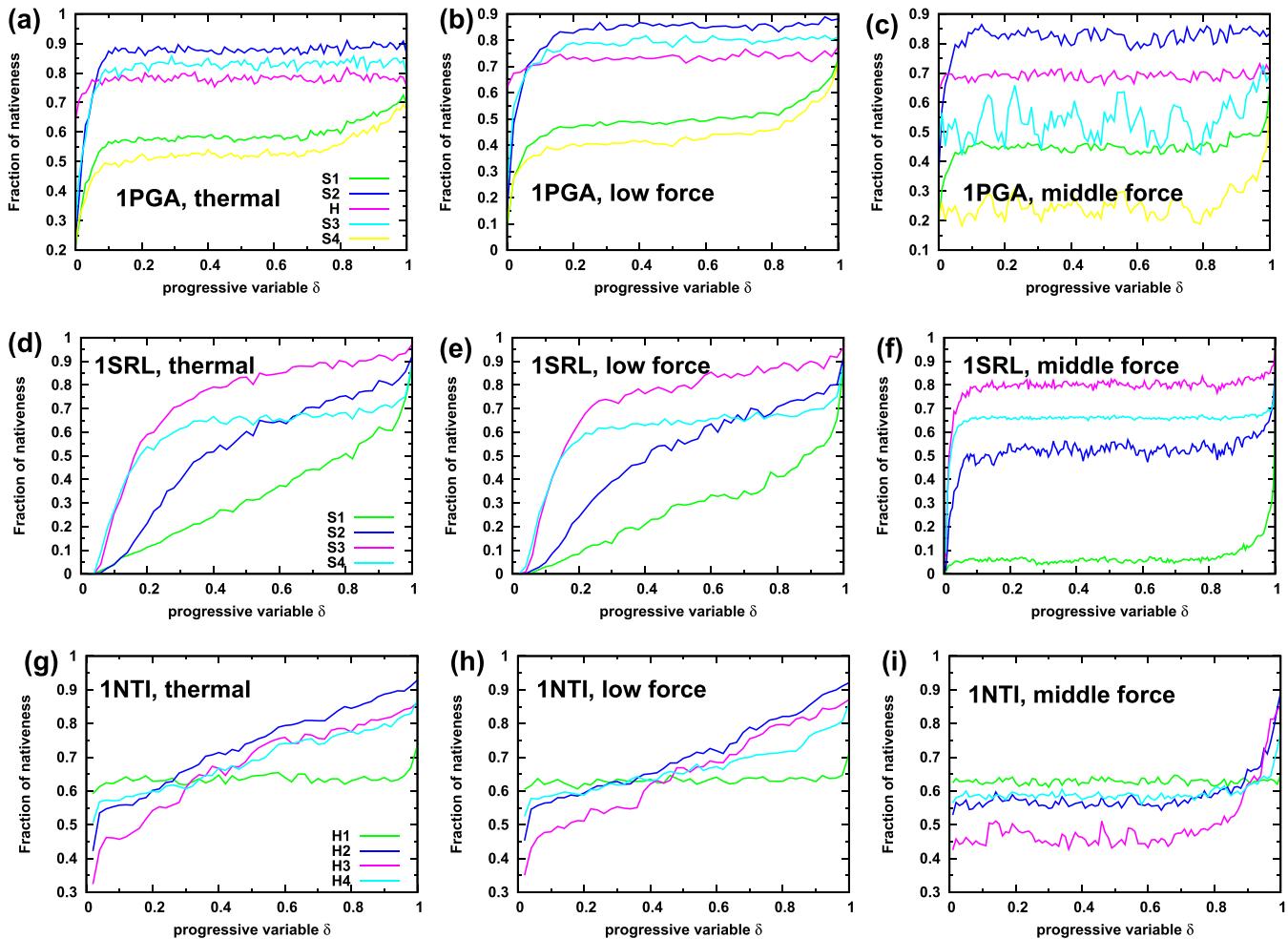


FIG. 4. Thermal and refolding pathways under force quench for three representative proteins. Temperature-driven and low force pathways are the same.

(Eq. (5a)), the folding originates from the formation of hairpin between the third and fourth  $\beta$ -strands (S3 and S4). Once formed, the hairpin seems to initiate the formation of contacts between the S2 strand with S3 and S1. After complete

formation of S3 and S2 strands, the S4 strand makes long range contacts with a loop between S1 and S2. The folding is finalized by the docking of C-terminal end of protein to the S1 strand.

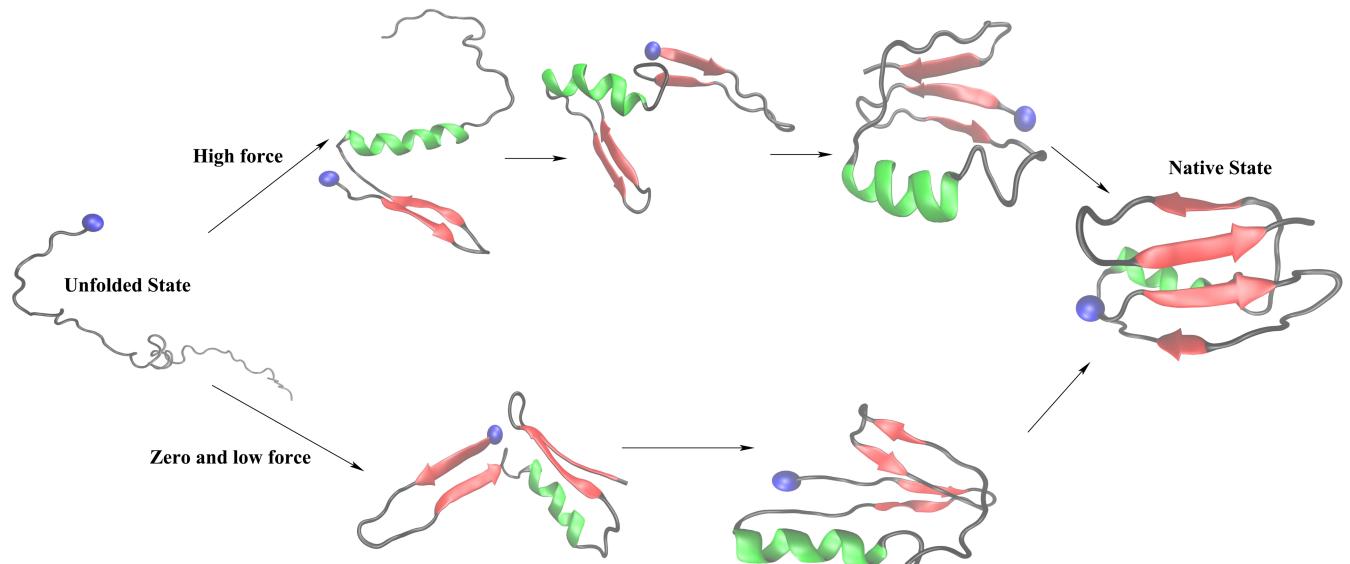
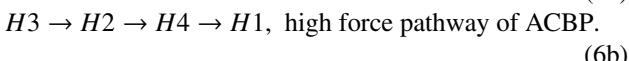
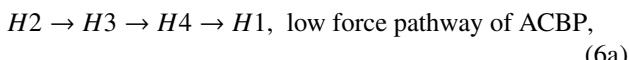


FIG. 5. Refolding pathways of the B domain of protein G with representative snapshots.

For the values of external force greater than  $f_{\text{switch}}$ , Fig. 4(f) gives the sequence of refolding events presented by Eq. (5b). Folding starts with the formation of hairpin between the third and fourth  $\beta$ -strands (S3 and S4) which induces the establishment of contacts between S2 with S3, but not S1 (as in the low force region). It seems that high force suppresses the interactions between S1 and S2 strands. As a consequence, S2 refolds faster than S1 and the refolding pathway becomes different compared to the low force regime. Once contacts between S1 and S2 are established, the C-terminal end docks to the S1 strand. A native-like structure is formed.

ACBP consists of four helices, H1, H2, H3, and H4, connected by loops between H1 and H2 as well as H3 and H4 (Fig. 2(b)). To obtain the sequence of folding events, we use the evolution of fraction of individual secondary structure elements. From Figs. 4(g)–4(i), we have the following refolding pathways for two force regimes:



Folding in all three cases (zero force, low, and middle force regimes) starts with the rapid formation of individual helices. Note that the sequence in Eqs. (6a) and (6b) was identified based on the fraction of individual helices which includes both intra- and inter-helical contacts. On average 30% of native contacts of the individual helices are intra-helix contacts. Figs. 4(g)–4(i) show that at the beginning of simulations, the total number of all-native contacts of individual helices reaches 30%–60% indicating the complete formation of local structures. This is followed by the formation of inter-helical contacts and contacts between helices and loops. At zero force and low force regimes (below

$f_{\text{switch}} \approx 3.4 \text{ pN}$ ), we observe the sequence of folding event described by Eq. (6a) which can be classified as follows. After folding of individual helices, H2-H3 and H2-H4 contacts are formed which is followed by the formation of contacts between H3 and loop connecting H2 and H3. The spatial arrangement between H1 and H4 ends the refolding process.

In the MF region, above  $f_{\text{switch}}$ , from Fig. 4(i) we obtain the different refolding sequence (Eq. (6b)). Folding pathways in the low and middle force regimes are different although they share the common feature that the folding starts with occurrence of individual helices and the formation of H1-H4 contacts completes the refolding process. The formation of contacts between H2 and H3 as well as between H3 and loop connecting H2 and H3 proceeds after folding of individual helices, but it precedes the formation of H2-H4 contacts.

## F. Relationship between critical and switch force

Because below switch force  $f_{\text{switch}}$  refolding pathways become identical to the thermal pathways, one can anticipate that  $f_{\text{switch}}$  coincides with the equilibrium critical force separating the folded and unfolded states,  $f_{\text{crit}}$ . Experimentally,  $f_{\text{crit}}$  can be approximately estimated, for instance, from the dependence of rupture force on pulling speeds.<sup>44</sup> Recently, using ultra stable magnetic tweezers and the exact criterion that at critical force the probabilities of folded and misfolded events are equal, Chen *et al.* have reported that  $f_{\text{crit}} \approx 5.4 \text{ pN}$  for the I27 domain<sup>34</sup> at  $T = 294 \text{ K}$ . This value is very close to our estimate  $f_{\text{switch}} \approx 5 \text{ pN}$  (Fig. 1) at the same temperature. Thus, it is reasonable to assume that  $f_{\text{switch}}$  coincides with  $f_{\text{crit}}$  not only for I27 but also for other proteins. In other words, the crossover from the temperature-driven regime to the force-driven regime occurs at the critical force  $f_{\text{crit}}$ .

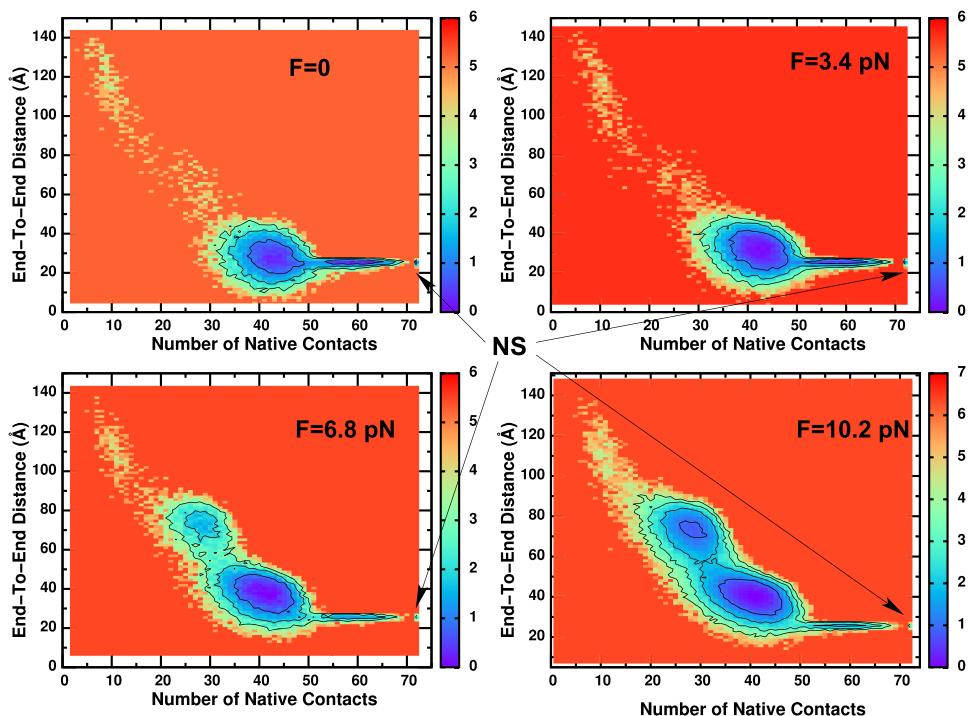


FIG. 6. Two-dimensional free energy landscapes as a function of number of native contacts and for 1PGA protein at  $T_{\min} = 0.53$ . The results are averaged over 100 trajectories. Surfaces are shown with contour lines indicating the relative  $k_BT$  slope of the surface.

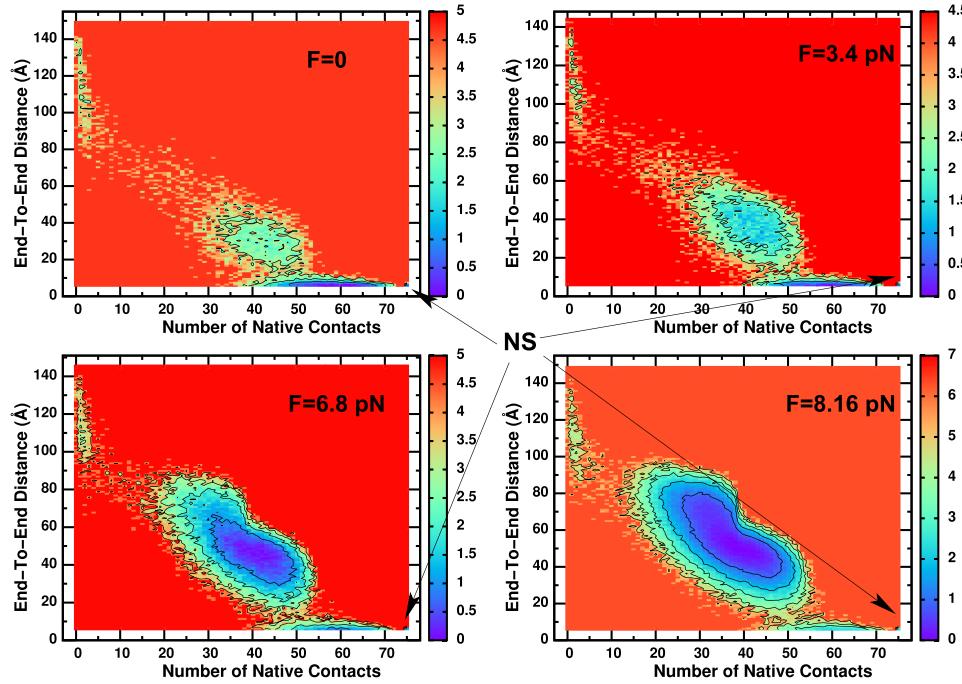


FIG. 7. Two-dimensional free energy landscapes as a function of number of native contacts and for 1SRL protein at  $T_{min} = 0.53$ . The results are averaged over 100 trajectories. Surfaces are shown with contour lines indicating the relative  $k_B T$  slope of the surface.

### G. Go model probes collapsed and native conformation ensembles

Fernandez and Li<sup>15</sup> have applied the force-clamp technique to probe refolding of ubiquitin under quench force  $f_q$ . In their experimental study, the protein is first extended to a well-defined state and then its subsequent refolding to a collapsed conformation ensemble under a low force was monitored as a function of protein extension over time.<sup>15</sup> Recently, follow-up experiments provided evidence that collapsed dynamics was probed in Ref. 15, but not final folding which occurs on a much delayed time scale.<sup>45,46</sup> Moreover, a number of

studies have shown that the final folding step occurs when a molten-globule like conformations matures into the native state.<sup>16,47</sup>

Then an interesting question emerges that whether the simulations, based on the Go model, probe only the collapse or both collapse and final folding conformations.

Because end-to-end distance  $R_{NC}$  might fail to describe the entire range of complexity of collapsed and final folding conformation ensembles due to the same end-to-end distance of folded and collapsed states, nonequilibrium free energy surfaces were constructed as a function of the number of native contacts  $N_{NC}$  and  $R_{NC}$ . As follows from

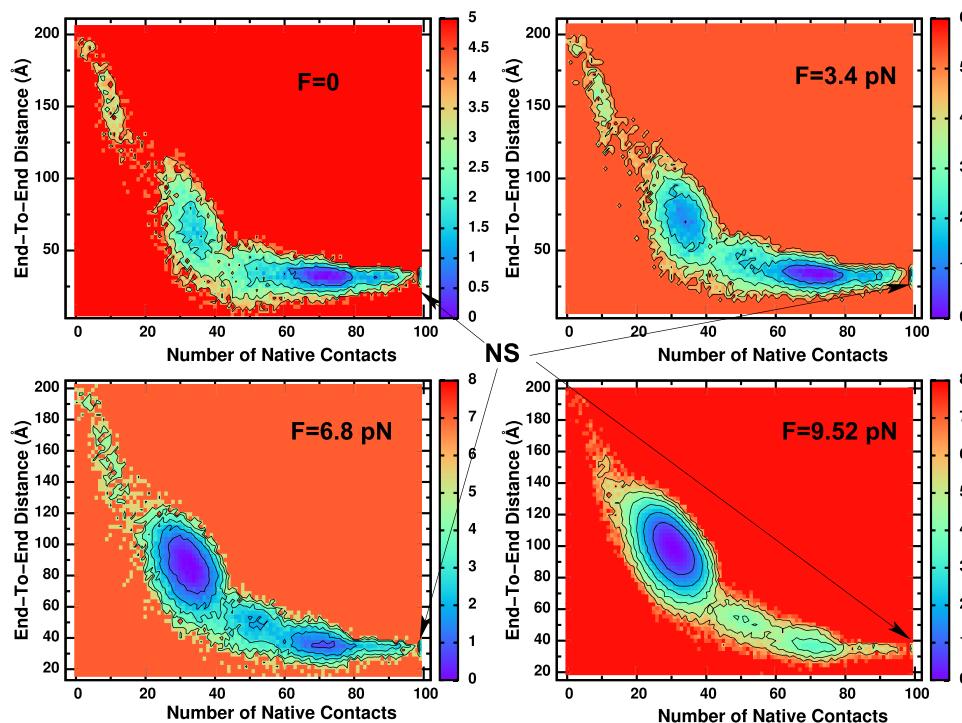


FIG. 8. Two-dimensional free energy landscapes as a function of number of native contacts and for 1UBQ protein at  $T_{min} = 0.55$ . The results are averaged over 100 trajectories. Surfaces are shown with contour lines indicating the relative  $k_B T$  slope of the surface.

Figs. 6–8, FEL of each protein has a broad energy minimum with the end-to-end distance of folded protein while the number of native contacts fluctuating down to 50% of  $N_{NC}$  of the folded protein. Although located in the unfolded domain, the compactness of the conformations in this minimum is very close to the native structure. The conformations achieved are distributed very broadly along  $R_{NC}$  and  $N_{NC}$  ranges. As the quenched force is increased, the position of the major local minimum moves towards the larger end-to-end distances (Figs. 6–8). In all studied proteins, the NS was sampled during simulations implying that the Go model probes not only collapsed and but also native conformation ensembles.

In order to better understand the impact of force on refolding kinetics, we monitor the time dependence of the end-to-end distance, gyration radius ( $R_g$ ), and native contacts in three force regimes for ubiquitin as a representative protein. In the LF regime with  $f = 1.34$  pN, folding pathways are diverse (Fig. 9). In (b) and (d), there is no appreciable plateau in the time dependence of both quantities implying that folding proceeded without intermediates. However, in (a), a long-lived plateau occurs between about 28 and 125 ns which represents a robust intermediate making folding slower compared to other trajectories. For the case (c), three relatively short-lived plateaus appeared on the route to NS. One can show that folding without intermediates was observed in about 50% trajectories. For all trajectories, refolding undergoes three major stages. The first transition, marked by (1) (Fig. 9), is from the fully extended state to unfolded or denatured ensemble (DE) conformations with the end-to-end distance and gyration radius larger than in NS. At this stage,  $R_g$  is reduced substantially. The second

transition initiates from DE conformations to molten globule or collapsed ensemble (CE) conformations where  $R_{NC}$  is almost the same as in NS but with less native contacts. The final acquisition of the native state occurs without a visible change in  $R_{NC}$  and  $R_g$ .

In the middle force regime ( $f = 6.8$  pN), the diversity in folding pathways becomes less compared to the LF regime with intermediates sampled en route to the native basin in all trajectories as shown in Fig. 10. Overall, folding without intermediates occurs only in 5% simulation runs.

Individual trajectories of ubiquitin at  $f = 9.5$  pN (HF regime) are presented in Fig. 11. Kinetics intermediates occurred in all trajectories. Quick transition from the fully extended state to unfolded state marked by (1) is followed by a long-lived intermediate state at which  $R_{NC}$  fluctuates a lot. This stage was also observed and termed as force-induced metastable intermediates for the RNA hairpin<sup>48</sup> and in more general context of biopolymers.<sup>49</sup> The next transition to the collapsed ensemble conformations for which  $R_{NC}$  and  $R_g$  are the same as in NS is marked by (2). Note that in the trajectories shown in Figs. 11(a) and 11(b), once the protein has reached the collapsed state ensemble, we observed shortly (200–300 ns) the final transition into the native state which is highlighted by (3). These collapsed structures represent the on-pathway intermediate state of the protein folding pathway and it was suggested that such collapsed conformations are necessary precursors of the folded state.<sup>46</sup> However, it is not the case for the trajectories shown in Figs. 11(c) and 11(d). After reaching the collapsed state ensemble the protein might undergo the transition back to the unfolded state. For those

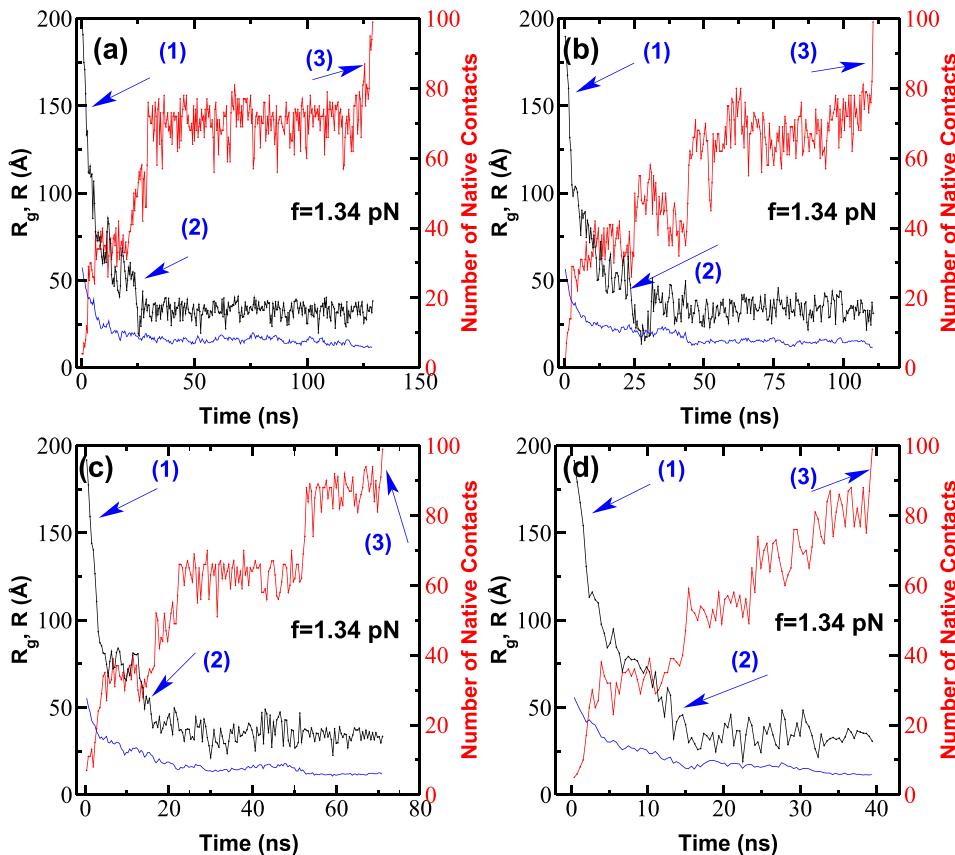


FIG. 9. Time dependence on end-to-end distance (black), gyration radius (blue), and number of native contacts (red) for individual trajectories of ubiquitin at  $f = 1.34$  pN (LF regime). The time dependence of end-to-end distance and number of native contacts are shown in black and red, respectively. (1) marks the transition from stretched to denatured ensemble (DE) conformations which has larger values of  $R_{NC}$  and gyration radius than in NS. (2) refers to the transition from DE to collapsed or molten globule ensemble conformations (CE) for which  $R_{NC}$  and  $R_g$  are the same as in NS. (3) marks the transition from CE conformations to NS.

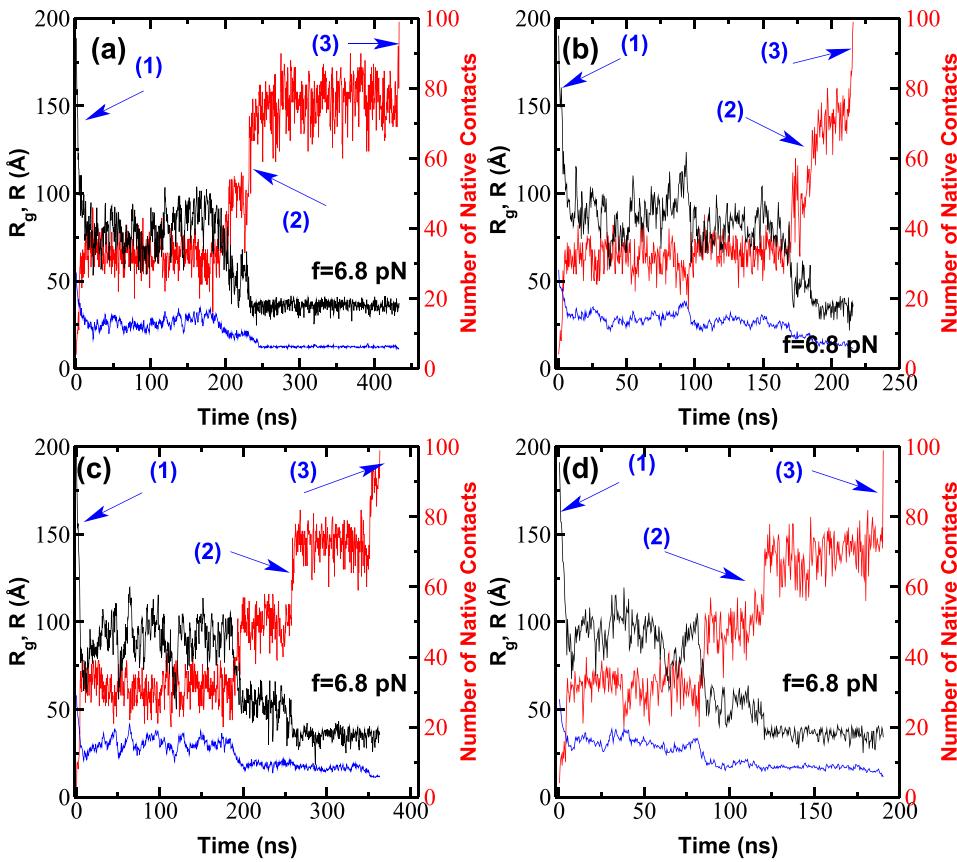


FIG. 10. The same as in Figure 9 but for  $F = 6.8$  pN (middle force regime).

two cases (Figs. 11(c) and 11(d)), the final folding implies going from one collapsed state ensemble configurations to the other through the ensemble of unfolded configurations. As a

consequence, the final folding step occurred on a much delayed time scale, e.g.,  $\sim 1000$  ns and  $\sim 2000$  ns for the trajectories shown in Figs. 11(c) and 11(d), respectively.

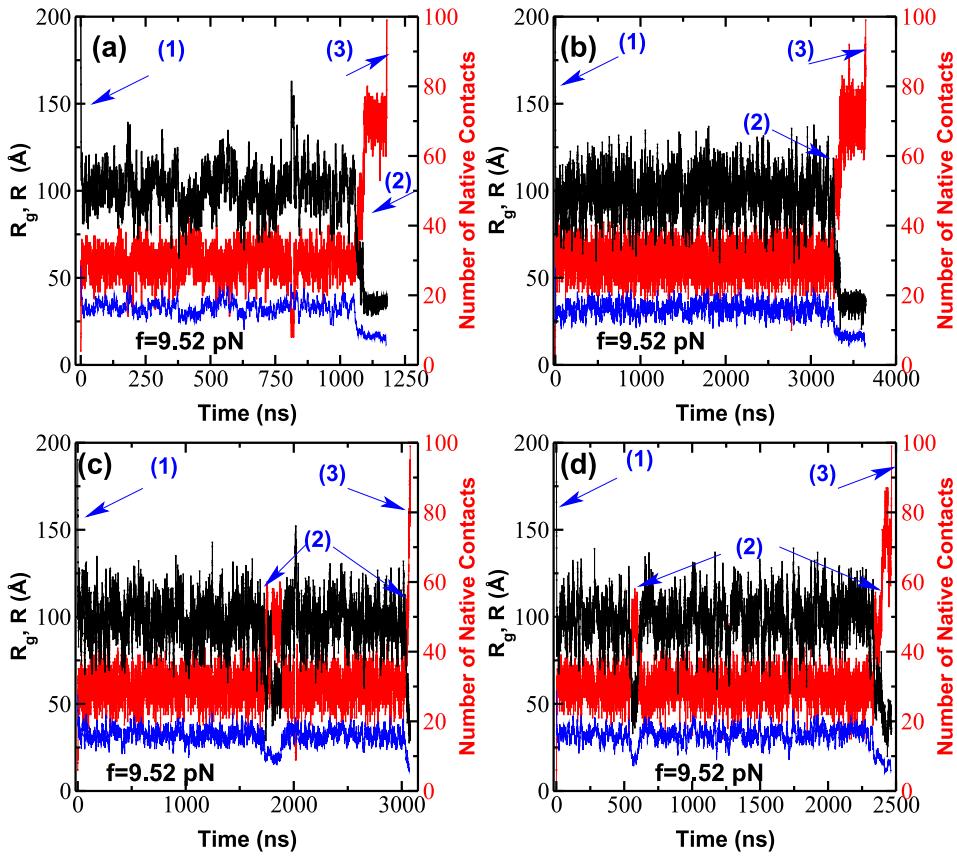


FIG. 11. The same as in Figure 9 but for  $f = 9.5$  pN (high force regime).

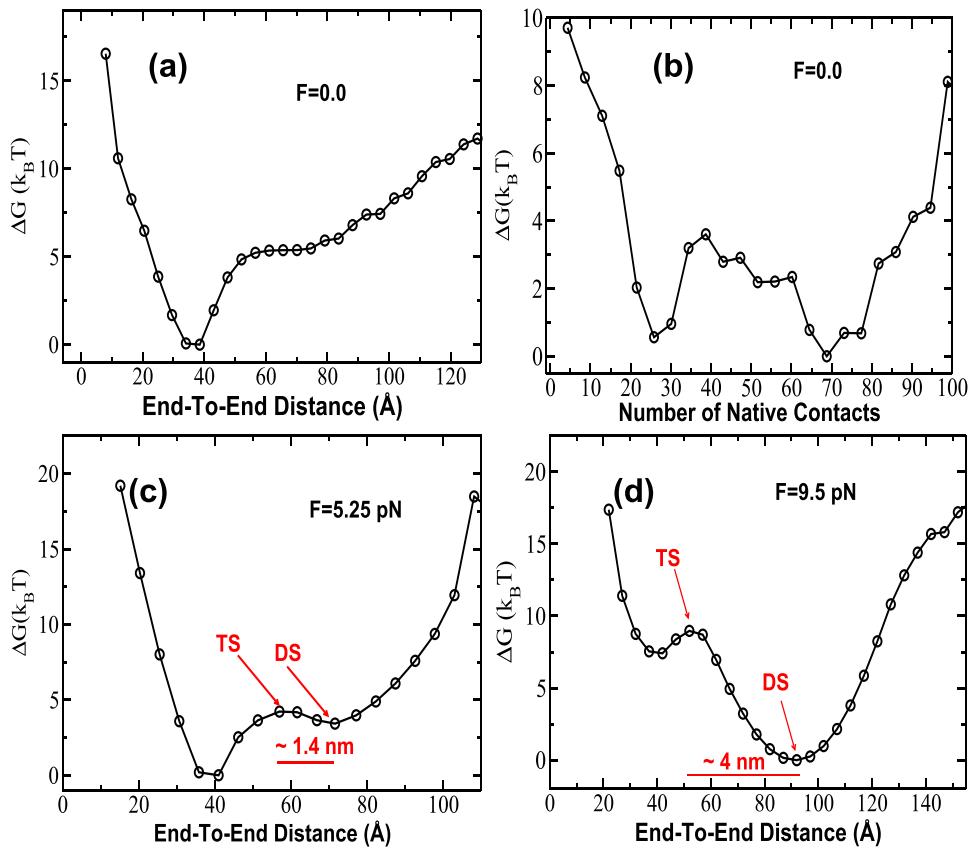


FIG. 12. Free energy landscapes of ubiquitin as a function of  $R_{nc}$  ((a), (c), and (d)) for different forces and as a function of the number of native contacts in the absence of force (b).

#### H. Extraction of $x_f$ from free energy landscape

The distance  $x_f$  may be extracted from the free energy landscape plotted as a function of the end-to-end distance  $R_{NC}$ .<sup>8</sup> For this purpose, one can project the two-dimensional free energy surfaces (Figs. 6–8) onto a single reaction coordinate  $R_{NC}$ . For ubiquitin, in the absence of force the FEL, plotted as a function of the end-to-end distance, has only one minimum (Fig. 12(a)) which corresponds to downhill folding. This implies that  $R_{NC}$  is not a good reaction coordinate because ubiquitin is known as a two-state folder (see Ref. 13 and the references therein). In the absence of force, the number of native contacts is a better reaction coordinate for describing folding as evident from Fig. 12(b) showing two minima.

Because for the LF interval, the FEL plotted as a function of  $R_{NC}$  does not display the transition state (Fig. 12(a)); we are not able to extract  $x_f$ . In the MF regime with  $f = 5.25$  pN, the distance between TS and DS  $x_f \approx 1.4$  nm (Fig. 12(c)) which is not far from the estimation obtained in the Bell theory (Table I). In the high force regime ( $f = 9.5$  pN), we obtained  $x_f \approx 4$  nm (Fig. 12(d)) which is higher than the Bell value 3.24 nm (Fig. 1). However, given a crude approximation of using a one-dimensional FEL, the agreement between two approaches may be considered as satisfactory.

Using the results presented in Figs. 6 and 7, we obtained FELs as a function of  $R_{nc}$  for 1PGA and 1SRL (results not shown). As in the ubiquitin case in the LF regimes, 1D FELs are downhill not allowing us for extracting  $x_f$ . However, we obtained  $x_f \approx 1.7$  and 4.2 nm in the MF regime for 1PGA and 1SRL, respectively. These results are in reasonable agreement

with those obtained from the dependence of refolding time on the external force (Table I). Thus one can expect that for force-driven regimes  $x_f$  may be extracted from one-dimensional FEL plotted as a function of  $R_{NC}$ .

It is well known that the free energy landscape of protein is complicated being multi-dimensional. So mapping it into a single reaction coordinate such as the end-to-end distance is a crude approximation. From this point of view, the fact that the  $x_f$  values extracted from the free-energy profiles agree quite well with those obtained from the kinetic simulations for a few force values does not mean that the free energy-based  $x_f$  varies in a step-wise manner with force.

#### IV. CONCLUSION

Having used the structure based models, we have shown that AFM experiments probe protein refolding below the HF region. At very low forces, refolding pathways coincide with the thermal ones. The crossover from LF to MF regimes accompanies the switch in folding pathways as well as in the elevation of  $x_f$ . This happens at the external quenched force of a few pN. Our results obtained for  $x_f$  in the MF regime are in good agreement with experiments. We have shown that in force-driven regimes, where the end-to-end distance is a good reaction coordinate  $x_f$  can be read out from one-dimensional FELs.

Using a simple coarse-grained model,<sup>50,51</sup> we have also studied the impact of external force on the refolding of homopolymer. Due to a flat (non-funnel shape) free energy landscape, homopolymers refold very slowly and our study

was restricted to short sequences. Similar to proteins, we have also observed the crossover from the temperature-driven to force-driven regime with different values of  $x_f$ .

A potential weakness of our study is the use of a coarse-grained structure-based model. The choice was motivated by the need to break computationally the fundamental barrier of experimental time scales unreachable by classical simulation tools. It is worth noting that assumptions underlying this model including neglect of the non-native interactions and explicit solvent can lead to the emergence of certain artifacts which indeed were observed in protein unfolding studies.<sup>27,52</sup> Nonetheless, we believe that the coarse-grained structure-based modeling is a promising and efficient alternative to mechanical refolding studies and allows us to capture some of the important mechanisms of mechanical refolding. The rationale of this approach is partially supported by the agreement between the simulation and experiment.

## SUPPLEMENTARY MATERIAL

See [supplementary material](#) for two and three force regimes for ubiquitin (Fig. S1) and 1SRL (Fig. S2); different fitting choices for LF and MF regimes for 1PGA (Fig. S3); procedure for the determination of folding temperature for 1PGA; temperature dependence of  $C_V(T)$  and  $df_N/dT$  (Fig. S4) and of folding times (Fig. S5) of 1PGA; and force dependence of refolding times of 1PGA at  $T_{\min}$  and  $T_{\exp}$  (Fig. S6).

## ACKNOWLEDGMENTS

This work was supported by Department of Science and Technology at Ho Chi Minh city, Vietnam; the Polish NCN Grant No. 2015/19/B/ST4/02721; the Foundation for Polish Science TEAM Project No. (TEAM/2011-7/6) co-financed by the European Regional Development Fund operated within the Innovative Economy Operational Program; Iuventus Plus Project No. IP2012 016872 from the Polish Ministry of Science and Higher Education; National Science Center Grant No. [MAESTRO 2014/14/A/ST6/00088]. M.K. acknowledges the Polish Ministry of Science and Higher Education for financial support through Mobilnosc Plus Program No. 1287/MOB/IV/2015/0.

- <sup>1</sup>J. N. Onuchic and P. G. Wolynes, *Curr. Opin. Struct. Biol.* **14**, 70 (2004).
- <sup>2</sup>S. Kumar and M. S. Li, *Phys. Rep.* **486**, 1 (2010).
- <sup>3</sup>L. Sun, J. Noel, J. Sulkowska, H. Levine, and J. Onuchic, *Biophys. J.* **107**, 2950 (2014).
- <sup>4</sup>M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, *Science* **276**, 1109 (1997).
- <sup>5</sup>G. I. Bell, *Science* **100**, 618 (1978).
- <sup>6</sup>D. Bartolo, I. Derenyi, and A. Ajdari, *Phys. Rev. E* **65**, 051910 (2002).
- <sup>7</sup>G. Hummer and A. Szabo, *Biophys. J.* **85**, 5 (2003).
- <sup>8</sup>M. Schlierf and M. Rief, *Biophys. J.* **90**, L33 (2006).

- <sup>9</sup>O. K. Dudko, G. Hummer, and A. Szabo, *Phys. Rev. Lett.* **96**, 108101 (2006).
- <sup>10</sup>R. W. Friddle, A. Noy, and J. J. De Yoreo, *Proc. Natl. Acad. Sci. U. S. A.* **109**, 13573 (2012).
- <sup>11</sup>Z. T. Yew, M. Schlierf, M. Rief, and E. Paci, *Phys. Rev. E* **81**, 031923 (2010).
- <sup>12</sup>M. S. Li, *Biophys. J.* **93**, 2644 (2007).
- <sup>13</sup>M. S. Li, M. Kouza, and C. K. Hu, *Biophys. J.* **92**, 547 (2007).
- <sup>14</sup>Y. Chen, S. E. Radford, and D. J. Brockwell, *Curr. Opin. Struct. Biol.* **30**, 89 (2015).
- <sup>15</sup>J. M. Fernandez and H. Li, *Science* **303**, 1674 (2004).
- <sup>16</sup>P. J. Elms, J. D. Chodera, C. Bustamante, and S. Marqusee, *Proc. Natl. Acad. Sci. U. S. A.* **109**, 3796 (2012).
- <sup>17</sup>B. Jagannathan, P. J. Elms, C. Bustamante, and S. Marqusee, *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17820 (2012).
- <sup>18</sup>P. O. Heidarsson *et al.*, *J. Am. Chem. Soc.* **134**, 17068 (2012).
- <sup>19</sup>J. P. Junker, F. Ziegler, and M. Rief, *Science* **323**, 633 (2009).
- <sup>20</sup>M. S. Li, C. K. Hu, D. K. Klimov, and D. Thirumalai, *Proc. Natl. Acad. Sci. U. S. A.* **103**, 93 (2006).
- <sup>21</sup>M. Kouza, C.-K. Hu, and M. S. Li, *J. Chem. Phys.* **128**, 045103 (2008).
- <sup>22</sup>S. Gnanakaran, H. Nyrmeyer, J. Portman, K. Y. Sanbonmatsu, and A. E. Garcia, *Curr. Opin. Struct. Biol.* **13**, 168 (2003).
- <sup>23</sup>P. H. Nguyen, G. Stock, E. Mittag, C. K. Hu, and M. S. Li, *Proteins: Struct., Funct., Bioinf.* **61**, 795 (2005).
- <sup>24</sup>D. E. Shaw *et al.*, *Science* **330**, 341 (2010).
- <sup>25</sup>C. Clementi, H. Nyrmeyer, and J. N. Onuchic, *J. Mol. Biol.* **298**, 937 (2000).
- <sup>26</sup>G. Arad-Haase *et al.*, *Biophys. J.* **99**, 238 (2010).
- <sup>27</sup>M. Kouza, C. K. Hu, M. S. Li, and A. Kolinski, *J. Chem. Phys.* **139**, 065103 (2013).
- <sup>28</sup>S. Jackson, *Folding Des.* **3**, R81 (1998).
- <sup>29</sup>W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, *J. Chem. Phys.* **76**, 637 (1982).
- <sup>30</sup>M. Kouza *et al.*, *Biophys. J.* **89**, 3353 (2005).
- <sup>31</sup>N. Eswar *et al.*, *Curr. Protoc. Bioinform.* **5**, 5.6 (2006).
- <sup>32</sup>G. S. Krivov, M. V. Shapovalov, and R. L. J. Dunbrack, *Proteins: Struct., Funct., Bioinf.* **77**, 778 (2009).
- <sup>33</sup>M. Blaszczyk *et al.*, *Methods* **93**, 72 (2016).
- <sup>34</sup>H. Chen *et al.*, *J. Am. Chem. Soc.* **137**, 3540 (2015).
- <sup>35</sup>A. S. Politou, M. Gautel, M. Pfuhl, S. Labeit, and A. Pastore, *Biochemistry* **33**, 4730 (1994).
- <sup>36</sup>Y. Cao and H. Li, *Nat. Mater.* **6**, 109 (2007).
- <sup>37</sup>P. Alexander, J. Obrian, and P. Bryan, *Biochemistry* **31**, 7243 (1992).
- <sup>38</sup>S. T. Thomas, V. V. Loladze, and G. I. Makhatadze, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10670 (2001).
- <sup>39</sup>E. McCallister, E. Alm, and D. Baker, *Nat. Struct. Biol.* **7**, 669 (2000).
- <sup>40</sup>L. Lapidus *et al.*, *Biophys. J.* **107**, 947 (2014).
- <sup>41</sup>J. Shimada and E. Shakhnovich, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 11175–11180 (2002).
- <sup>42</sup>S. Kmiecik and A. Kolinski, *Biophys. J.* **94**, 726–736 (2008).
- <sup>43</sup>M. Kouza and U. H. E. Hansmann, *J. Phys. Chem. B* **116**, 6645–6653 (2012).
- <sup>44</sup>E. Evans and K. Ritchie, *Biophys. J.* **72**, 1541 (1997).
- <sup>45</sup>R. Berkovich, S. Garcia-Manyes, M. Urbakh, J. Klafter, and J. M. Fernandez, *Biophys. J.* **98**, 2692 (2010).
- <sup>46</sup>S. Garcia-Manyes, L. Dougan, C. L. Badilla, J. Bruijc, and J. M. Fernandez, *Proc. Natl. Acad. Sci. U. S. A.* **106**, 10534 (2009).
- <sup>47</sup>L. E. Rosen, K. B. Connell, and S. Marqusee, *Proc. Natl. Acad. Sci. U. S. A.* **111**, 14746 (2014).
- <sup>48</sup>C. Hyeon and D. Thirumalai, *J. Am. Chem. Soc.* **130**, 1538 (2008).
- <sup>49</sup>C. Hyeon, G. Morrison, D. L. Pincus, and D. Thirumalai, *Proc. Natl. Acad. Sci. U. S. A.* **106**, 20288 (2009).
- <sup>50</sup>M. Cieplak, T. Hoang, and M. Robbins, *Phys. Rev. E* **70**, 011917 (2004).
- <sup>51</sup>W. Graessley, R. Hayward, and G. Grest, *Macromolecules* **32**, 3510 (1999).
- <sup>52</sup>M. Kouza, C. K. Hu, H. Zung, and M. Li, *J. Chem. Phys.* **131**, 215103 (2009).

# Supplementary Material

## Switch from thermal to force-driven pathways of protein refolding

Maksim Kouza<sup>\*</sup>

*Faculty of Chemistry, University of Warsaw,  
Pasteura 1 02-093- Warsaw, Poland*

Pham Dang Lan

*Institute for Computational Science and Technology,  
Tan Chanh Hiep Ward, District 12, Ho Chi Minh City, Vietnam and  
Department of Theoretical Physics, Faculty of Physics and Engineering Physics,  
Ho Chi Minh University of Science, Vietnam*

Alexander M Gabovich

*Institute of Physics, National Academy of Sciences of Ukraine,  
46, Nauka Ave., Kyiv 03680, Ukraine*

Andrzej Kolinski<sup>†</sup>

*Faculty of Chemistry, University of Warsaw,  
Pasteura 1 02-093 Warsaw, Poland*

Mai Suan Li<sup>‡</sup>

*Institute of Physics, Polish Academy of Science Al. Lotnikow 32/46 02-668 Warsaw, Poland*

(Dated: February 22, 2017)

---

<sup>\*</sup>Electronic address: [mkouza@chem.uw.edu.pl](mailto:mkouza@chem.uw.edu.pl)

<sup>†</sup>Electronic address: [kolinski@chem.uw.edu.pl](mailto:kolinski@chem.uw.edu.pl)

<sup>‡</sup>Electronic address: [masli@ifpan.edu.pl](mailto:masli@ifpan.edu.pl)

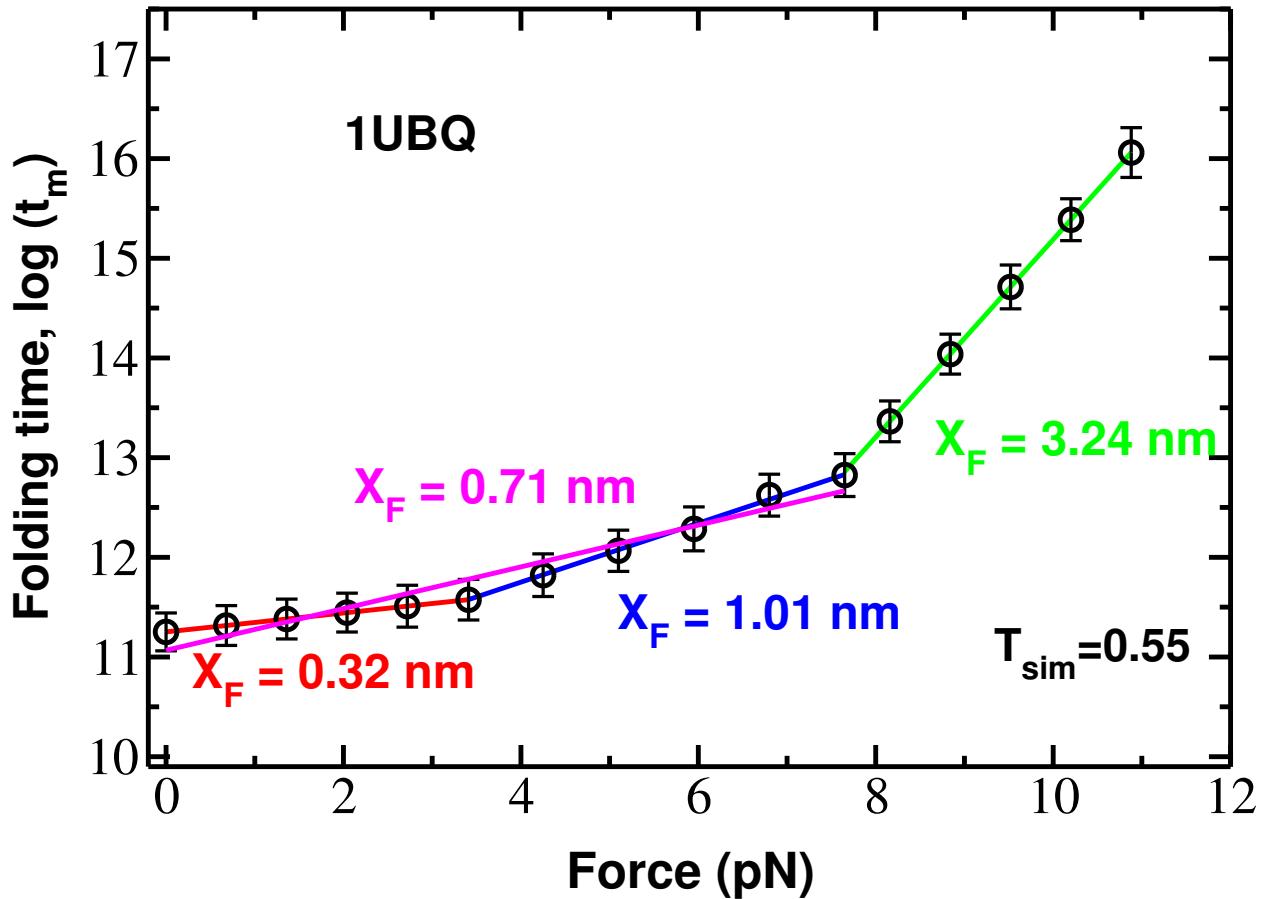


FIG. S1: Two and three force regimes for Ubiquitin. Three regimes are the same as in Fig. 1 in the main text. In the two-regime scenario data are separated at 7.85 pN. For the [0, 7.85 pN] interval  $x_f=0.71 \text{ nm}$  with the correlation level  $R=0.97$ . For sub-regions one has  $x_f=0.32$  and  $1.01 \text{ nm}$  with  $R = 1$  and  $0.99$ , respectively.

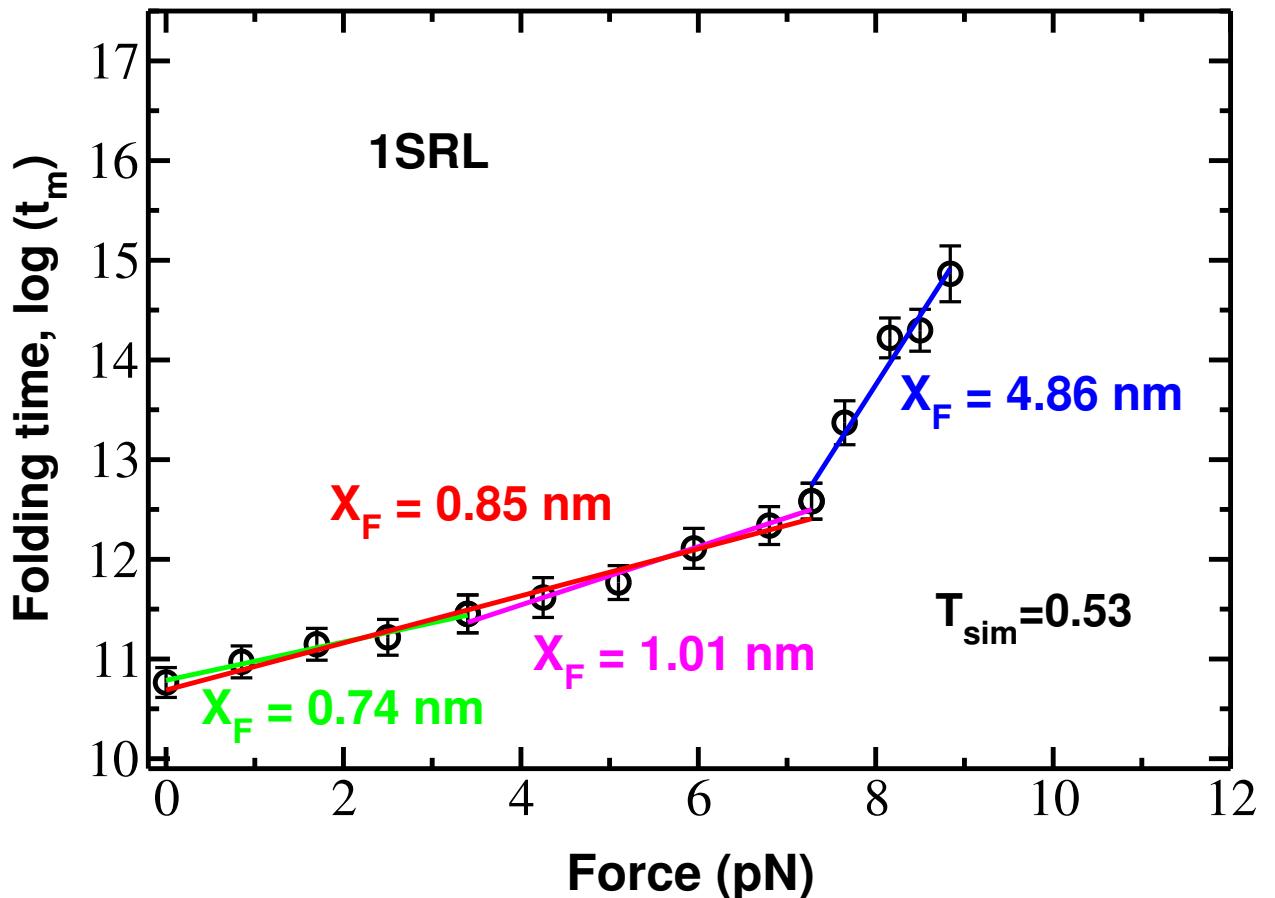


FIG. S2: Two and three force regimes for 1SRL. Two force regimes are the same as in Fig. 1 in the main text. In the three force regimes scenario we assume that the crossover at 3.5 pN exists. Then we obtain  $x_f=0.74$  and 1.01 nm for sub-regions with  $R=0.99$  and 0.98.

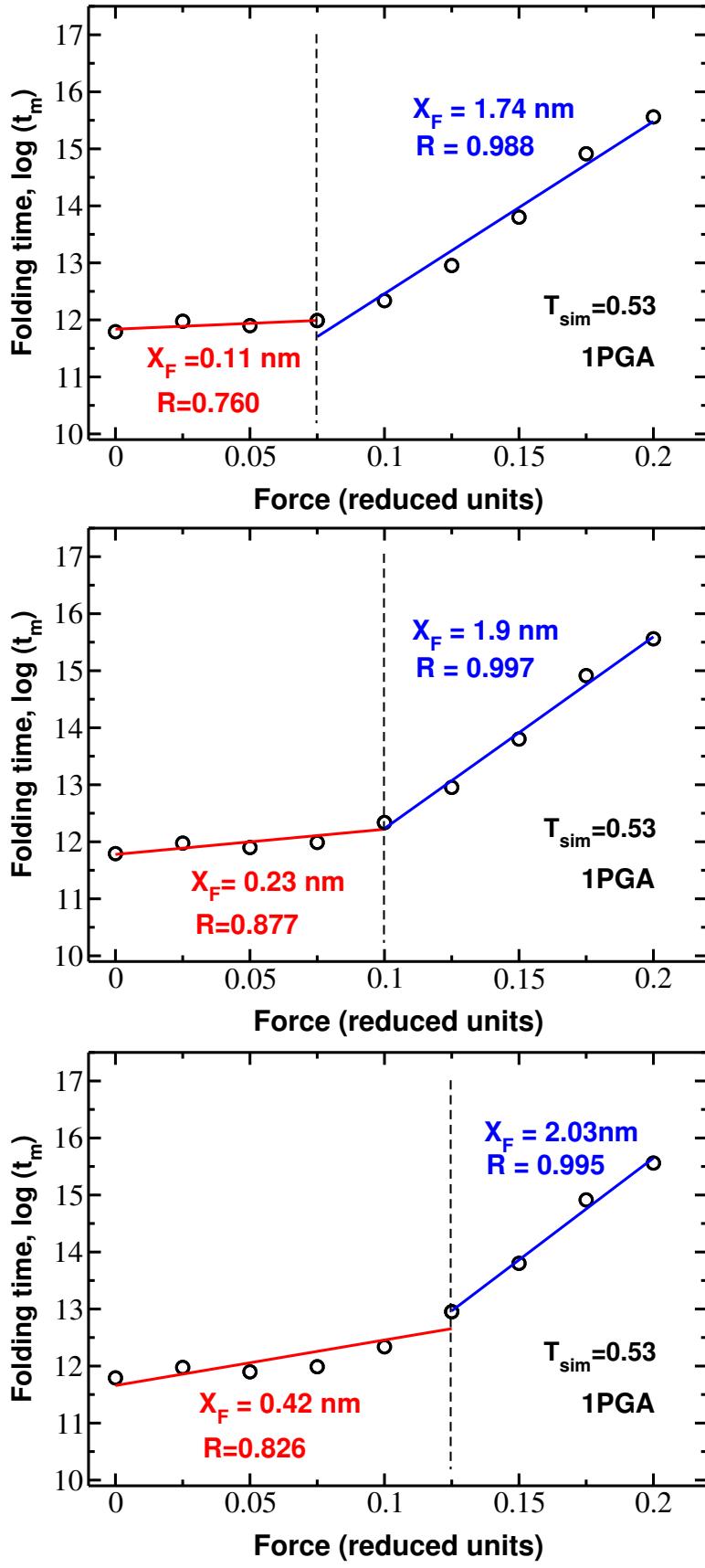


FIG. S3: Different choices for LF and MF regimes for 1PGA. The correlation level for the MF is nearly the same for three choices. In the LF regime  $R=0.760, 0.877$  and  $0.826$  for choices with 4, 5 and 6 points, respectively.

## Determination of $T_F$ for 1PGA

To quantify thermodynamics quantities of GB1 protein (1PGA) we performed an equilibrium simulation by a structure-based model, described in Refs. [1, 2] (see Material and Method). We apply standard replica-exchange method in temperature space to accelerate the equilibration and obtain the thermodynamics quantities. We chose 16 temperature replicas in the interval [0.32, 1.0] selected such that the averaged replica-exchange acceptance ratio was 24.2%. The system was equilibrated during first  $8 \times 10^5 \tau_L$ , after which histograms for the energy, native contacts, radius of gyration and end-to-end distances were collected for  $12 \times 10^5 \tau_L$ . All other parameters are the same as previously described in Refs. [1, 2]. Using the extended reweighting technique [3] and the data from the replica-exchange simulations we obtained the thermodynamic quantities of the GB1 protein. The folding temperature equals  $T_F = 0.57\epsilon_H/k_B$ , the temperature where we observed the maximum of  $C_v$  or  $df_N/dT$  (Fig. S4).

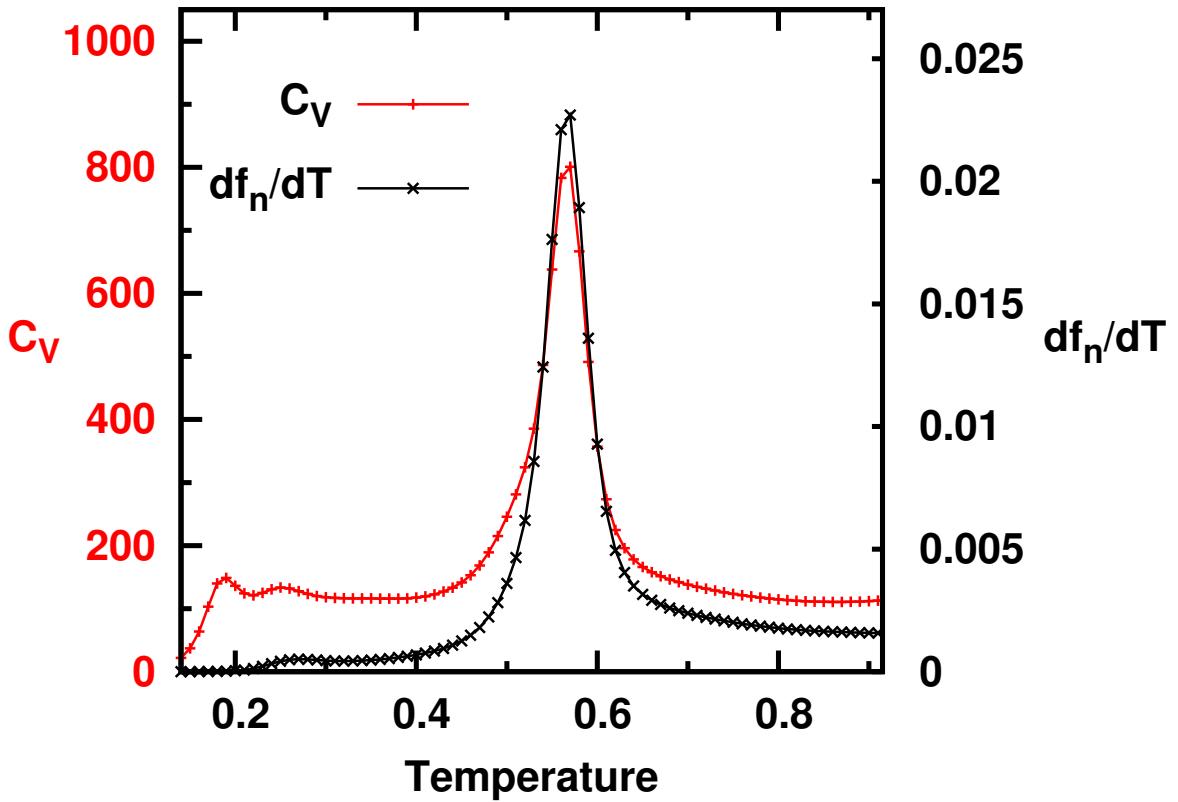


FIG. S4: The temperature dependence of the heat capacity  $C_V(T)$  is presented by red curve. The black curve refers to the temperature dependence of the structural susceptibility,  $df_N/dT$ . The scale for  $C_V$  is given on the left, while right scale corresponds to  $df_N/dT$ . The collapse temperature at which  $C_V$  has maximum,  $T_\theta$ , coincides with the folding temperature  $T_F$  at which strutural susceptibility is maximal.  $T_\theta = T_F = 0.57$ . The data obtained from structure-based simulations as described in text.

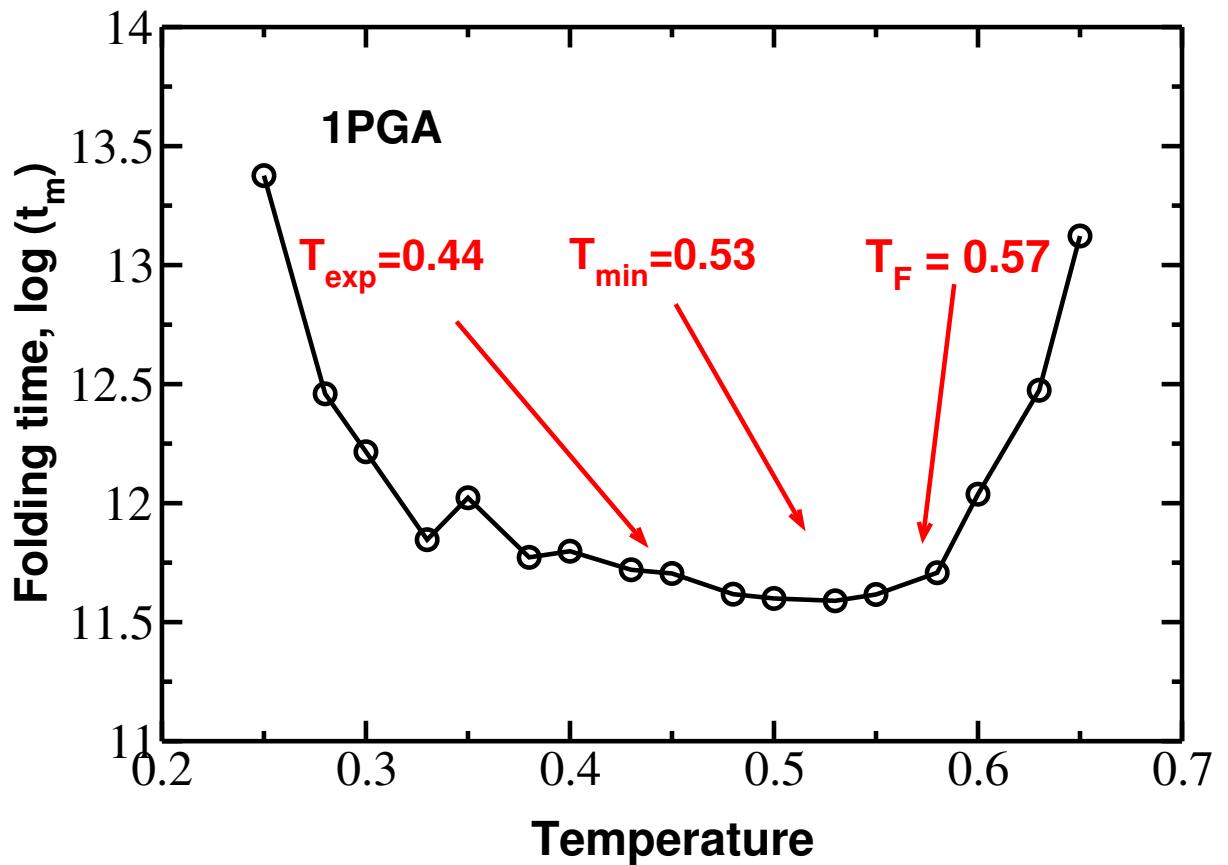


FIG. S5: The temperature dependence of refolding times of 1PGA. The result was obtained using 100 independent trrajectories and the Go models described in Material and Method in the main text. Positions of  $T_f$ ,  $T_{\text{min}}$  and  $T_{\text{exp}}$  are also shown.

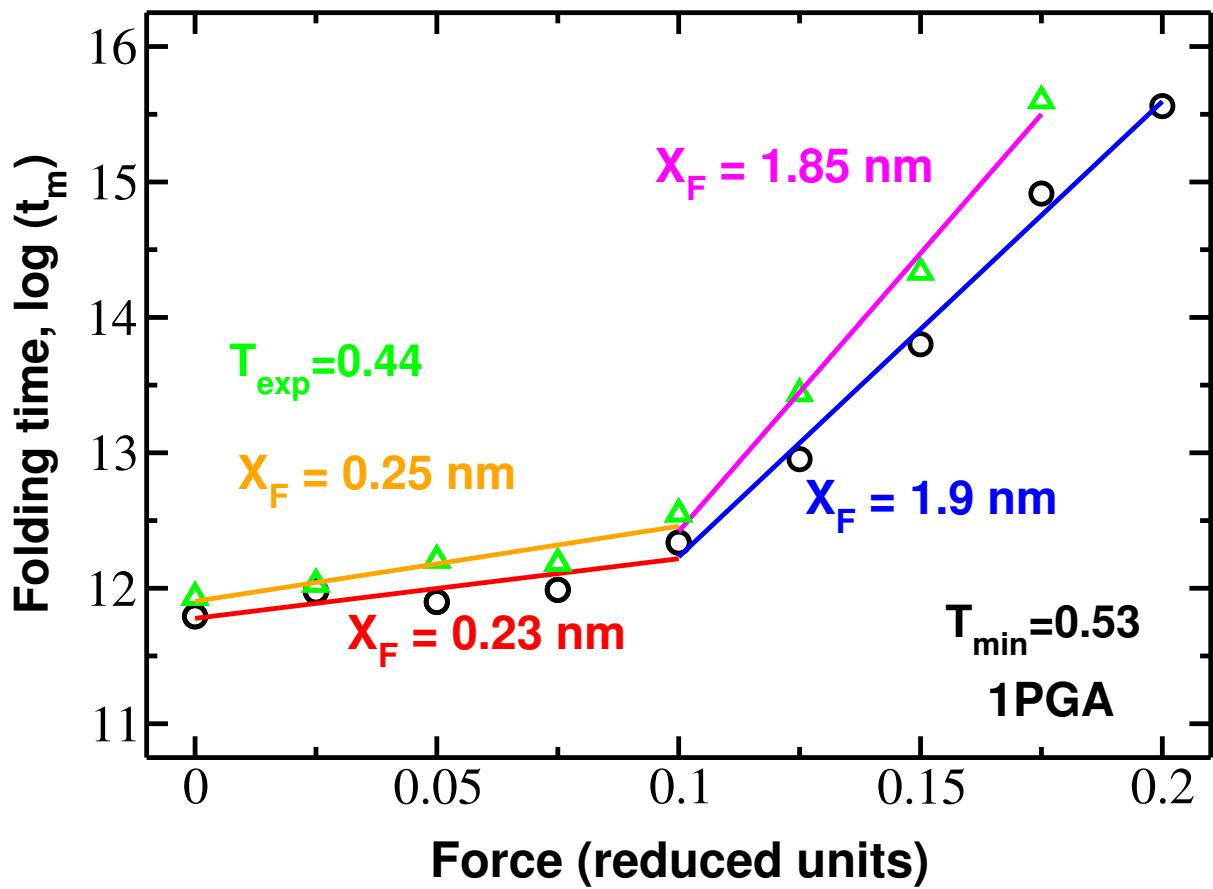


FIG. S6: The force dependence of refolding times of 1PGA at  $T_{\min}$  (black circles) and  $T_{\exp}$  (green triangles). The values of  $x_f$  in the LF and MF regimes are shown next to the fits.

- 
- [1] M. S. Li, M. Kouza, and C. K. Hu, *Biophys. J.* **91**, 547 (2007).
  - [2] M. Kouza, C.-K. Hu, and M. S. Li, *J. Chem. Phys.* **128**, 045103 (2008).
  - [3] A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **63**, 1195 (1989).