A structure-based model fails to probe the mechanical unfolding pathways of the titin I27 domain

Maksim Kouza, Chin-Kun Hu, Mai Suan Li, and Andrzej Kolinski

Re-register for Table of Content Alerts

Create a profile.

Sign up today!

A structure-based model fails to probe the mechanical unfolding pathways of the titin I27 domain

Maksim Kouza,[1,a)] Chin-Kun Hu,[2,b)] Mai Suan Li,[3,c)] and Andrzej Kolinski[1,d)]

[1]*Faculty of Chemistry, University of Warsaw, Pasteura 1 02-093 Warsaw, Poland*
[2]*Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan*
[3]*Institute of Physics, Polish Academy of Science, Al. Lotnikow 32/46 02-668 Warsaw, Poland*

We discuss the use of a structure based Cα-Go model and Langevin dynamics to study in detail the mechanical properties and unfolding pathway of the titin I27 domain. We show that a simple Go-model does detect correctly the origin of the mechanical stability of this domain. The unfolding free energy landscape parameters x_u and $\Delta G^{\ddagger}$, extracted from dependencies of unfolding forces on pulling speeds, are found to agree reasonably well with experiments. We predict that above $v = 10^4$ nm/s the additional force-induced intermediate state is populated at an end-to-end extension of about 75 Å. The force-induced switch in the unfolding pathway occurs at the critical pulling speed $v_{crit} \approx 10^6$–10^7 nm/s. We argue that this critical pulling speed is an upper limit of the interval where Bell's theory works. However, our results suggest that the Go-model fails to reproduce the experimentally observed mechanical unfolding pathway properly, yielding an incomplete picture of the free energy landscape. Surprisingly, the experimentally observed intermediate state with the A strand detached is not populated in Go-model simulations over a wide range of pulling speeds. The discrepancy between simulation and experiment is clearly seen from the early stage of the unfolding process which shows the limitation of the Go model in reproducing unfolding pathways and deciphering the complete picture of the free energy landscape. © *2013 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4817773]

I. INTRODUCTION

Over the last 15 years the mechanical unfolding of the titin I27 domain has been a subject of intense experimental and theoretical studies.[1–6] This domain consists of 8 β-strands (Fig. 1) that fold into two layers of β-sheets through backbone hydrogen bonds (HBs) and side-chain interactions and it has high resistance to external force. One of the most remarkable findings is that the force-extension profile of I27 displays a "hump," a previously overlooked pre-peak preceding the main peak.[4] It was interpreted by steered molecular dynamics (SMD) all-atom simulation in explicit solvent as a signature of an unfolding intermediate in which all hydrogen bonds between the A and B strands were broken.[7,8] Experimentally, this suggestion was confirmed by studies where a Lys to Pro point mutation at residue 6 disrupted hydrogen bonds connecting strand A with B and resulting peaks on the force-extension profile occurred without humps.[4] Two mutants, the I27 domain with a detached A strand and a destabilized A strand by a Val to Ala mutation at residue 4 have been used to verify whether the unfolding behavior would be affected. Both of them did not show the difference between the wild type and mutants of the I27 domain in unfolding forces or in the dependencies of unfolding forces on pulling speed,[5] suggesting the mechanical unfolding of titin as a two-step process.[4,5] The first step is a transition from the native (N) to intermediate state (IS), which corresponds to breaking of HBs between β-strands A and B and unfolding of A at the end-to-end extension ≈ 7 Å. A force of about 100 pN is necessary to cross the first transition state (TS1) and form a stable intermediate structure with the A strand detached.[4] The second stage, a transition from IS to the denatured state (DS), was initially believed to be solely associated with the cooperative rupture of six HBs between strands A′ and G. Subsequently, it was shown that together with hydrogen bonding interactions the side-chain packing in the A′G region plays an important role in the unfolding process and contributes to the mechanical stability of titin.[5,9] A force of about 200 pN is required to cross the second transition state (TS2) and unfold the protein completely. Despite the important role of water molecules,[8] implicit solvent simulations also revealed the detachment of the A strand as the first step in the unfolding process.[5,10] Thus, until now detailed all-atom simulations using the CHARMM force field either with the TIP3P water model[3,7] or a continuum representation of solvent[5,10] as well as experiments[4] showed that structure A unfolds first. However, deciphering the unfolding free energy landscape (FEL) of long proteins by all-atom simulations with explicit solvent is still computationally inaccessible. The time scale discrepancy (and the discrepancy in stretching forces required to induce unfolding) between AFM experiments and simulations can be reduced using Go models.[11,12] They have been successful in describing the folding and unfolding of a number of proteins,[13–21] but mechanical unfolding studies typically

a)Electronic mail: mkouza@chem.uw.edu.pl
b)Electronic mail: huck@phys.sinica.edu.tw
c)Electronic mail: masli@ifpan.edu.pl
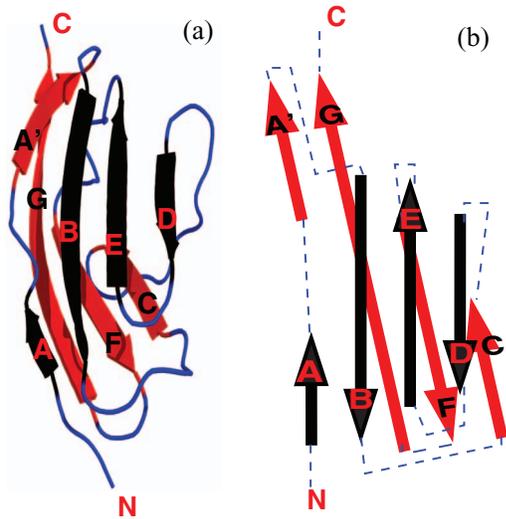d)Electronic mail: kolinski@chem.uw.edu.pl

FIG. 1. Native state conformation of the I27 domain of titin (PDB ID: 1TIT). There are 8 β-strands: A (4–7), A$'$ (11–15), B (18–25), C (32–36), D (47–52), E (55–61), F(69–75), and G (78–88). (a) PDB structure in cartoon representation. (b) Schematic view of the same structure. Red and black β-strands belong to different β-sheets. N- and C-terminal residues are marked N and C, respectively.

agree with experimental data in mechanostability properties, such as the unfolding force (F_{max}) and FEL parameters including the distance between the native and transition states x_u and the unfolding barrier $\Delta G^{\ddagger}$. Little success has been obtained for mechanical unfolding pathways as pathways of titin, Ubiquitin, and DDFLN4 predicted by Go-model[17,22,23] did not agree with experimental findings. In the case of Ubiquitin, the intermediate state was overlooked due to the lack of non-native interactions. The incorporation of these interactions helped to detect it correctly.[24] An analogous result was obtained for protein DDFLN4, in which non-native interactions were shown to lead to an intermediate state[25] previously undetected by Go-model simulations.[23]

For titin, it is unclear why the intermediate state cannot be captured by Go-model, as an inaccurate pathway was obtained not only by Go-model[17] but also by an all-atom simulation in implicit solvent,[26] in which non-native interactions were taken into account. Complete breaking of A$'$G contacts was shown to occur before the rupture of contacts between strands A and B. Consequently, it failed to capture the experimentally observed intermediate state as the crossing over the first transition state should be associated with a loss of native interactions or breaking of HBs between the A and B strands. On the other hand, it has been recently shown that extreme conditions change the unfolding pathway of other β-strand proteins (FnIII and DDFLN4).[23,27] Therefore, one of the possible reasons for the difference in titin unfolding pathways is that pulling speed, $v = 17 \times 10^6$ nm/ps, used in the simulations[17] is a few orders of magnitude higher than that used in AFM experiments.[28]

An interesting question arises whether the Go-model could correctly describe the unfolding pathways of the best-studied titin I27 domain at pulling speeds close to the experimental ones. In the present paper, we address this question using the Go model version developed in Ref. 12. We found

that the second peak that occurs at an end-to-end extension of ≈ 70 Å disappears at pulling speeds lower than $v = 10^4$ nm/s. The force-induced switch in the unfolding pathway was shown to occur at the critical pulling speed $v_{crit} \approx 10^6$–10^7 nm/s. We propose that this critical pulling speed constitutes an upper limit of the interval in which Bell's theory works. It is shown that unfolding pathways depend on pulling speeds but the Go model fails to describe the pathway observed in the experiments even at pulling speeds comparable to those used in experiments. To summarize, contrary to the common belief that structure-based models reproduce properly the key features of unfolding process, the results obtained by Go-model simulations should be taken with a grain of salt.

II. MATERIALS AND METHODS

A. Off-lattice Go model and Langevin dynamics

We use coarse-grained continuum representation for the I27 domain in which only the positions of C_α-carbons are retained. The interactions between residues are assumed to be Go-like and the energy of such a model is as follows:[12]

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

Here, $\Delta\phi_i = \phi_i - \phi_{0i}$, $R_{ij} = r_{0ij}/r_{ij}$; $r_{i,i+1}$ is the distance between beads i and $i + 1$, θ_i is the bond angle between bonds $(i - 1)$ and i, and ϕ_i is the dihedral angle around the ith bond and r_{ij} is the distance between the ith and jth 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 cut-off distance d_c taken to be $d_c = 6$ Å, where r_{0ij} is the distance between the residues in the native conformation. With this choice of d_c the native conformation from the PDB we have 86 native contacts in total.

The first harmonic term in Eq. (1) 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. (1). The interaction energy between residues that are separated by at least 3 beads is given by 10-12 Lennard-Jones potential. A soft sphere (last term in Eq. (1)) repulsive potential disfavors the formation of non-native contacts. We choose $K_r = 100\epsilon_H/\text{Å}^2$, $K_\theta = 20\epsilon_H/\text{rad}^2$, $K_\phi^{(1)} = \epsilon_H$, and $K_\phi^{(3)} = 0.5\epsilon_H$, where ϵ_H is the characteristic hydrogen bond energy and $C = 4$ Å. Since $T_F = 0.5\epsilon_H$ (see below) and $T_F = 333$ K,[29] we have $\epsilon_H = 1.37$ kcal/mol. Then the force unit $\epsilon_H/\text{Å} = 95$ pN. The dynamics of the system is obtained by integrating the

TABLE I. List of pulling speeds used in simulations. The upper limit for pulling speed used in the AFM experiment is 10^4 nm/s (taken from Ref. 28).

Pulling velocity (nm/s)	F_{max1}	F_{max2}	Number of trajectories
$v_1 = 2.5 \times 10^4$	210.5 ± 7.3	36.6 ± 8.5	50
$v_2 = 5.76 \times 10^4$	215.4 ± 13.3	44.2 ± 12.3	50
$v_3 = 1.29 \times 10^5$	221.0 ± 18.3	50.7 ± 11.4	50
$v_4 = 2.88 \times 10^5$	224.3 ± 24.1	57.8 ± 16.2	50
$v_5 = 6.44 \times 10^5$	232.4 ± 27.7	64.1 ± 17.1	50
$v_6 = 1.44 \times 10^6$	253.7 ± 29.3	80.7 ± 23.5	50
$v_7 = 3.22 \times 10^6$	264.4 ± 29.7	98.8 ± 29.5	50
$v_8 = 7.2 \times 10^6$	288.1 ± 33.2	119.9 ± 31.6	50
$v_9 = 1.6 \times 10^7$	352.1 ± 48.3	178.9 ± 44.2	50
$v_{10} = 3.6 \times 10^7$	422.6 ± 57.2	283.2 ± 34.2	50
$v_{11} = 8.05 \times 10^7$	549.8 ± 52.4	448.0 ± 52.3	50

following Langevin equation:[30, 31]

$$m \frac{d^2\vec{r}}{dt^2} = -\zeta \frac{d\vec{r}}{dt} + \vec{F}_c + \vec{\Gamma}, \qquad (2)$$

where m is the mass of a bead, ζ is the friction coefficient, $\vec{F}_c = -dE/d\vec{r}$. The random force $\vec{\Gamma}$ is a white noise, i.e., $\langle \Gamma(t)\Gamma(t') \rangle = 2\zeta k_B T \delta(t - t')$. It should be noted that the folding thermodynamics does not depend on the environment viscosity (or on ζ) but the folding kinetics depends on it. Most of our simulations (if not stated otherwise) were performed at the friction $\zeta = 2\frac{m}{\tau_L}$, where the folding is fast. Here, $\tau_L = (ma^2/\epsilon_H)^{1/2} \approx 3$ ps. The equations of motion were integrated using the velocity form of the Verlet algorithm[32] with the time step $\Delta t = 0.005\tau_L$. In order to check robustness of our predictions for unfolding pathways, limited computations were carried out for the friction $\zeta = 50\frac{m}{\tau_L}$ which is believed to correspond to the viscosity of water.[32] In this overdamped limit, we use the Euler method for integration and the time step $\Delta t = 0.1\tau_L$. Three types of Langevin dynamics simulations were carried out. (i) In the absence of force. (ii) A constant force was applied to both termini. In the latter case, one has to add to the energy (1) the term $-(\vec{f}\vec{r})$ where $\vec{r}$ is the end-to-end vector. (iii) In the constant velocity force simulation, we fix the N-terminal and pull the C-terminal by force, $f = K_r(vt - r)$, where r is the displacement of the pulled atom from its original position and the spring constant of cantilever, K_r, is set to be the same as the spring constant of the Go model. The pulling direction was chosen along the vector connecting N- and C-terminal atoms. List of pulling speeds used in simulations is shown in Table I.

B. Tools and measures used in the analysis

The temperature-force phase diagram and the thermodynamic quantities were obtained by the multiple histogram method[33] extended to the case when the external force is applied to the termini.[34] The reweighting is carried not only for temperature but also for force. We collected data for five values of T at $f = 0$ and for five values of f at a fixed value of T. The duration of MD runs for collection of histograms was chosen to be the same for all trajectories. In order to obtain
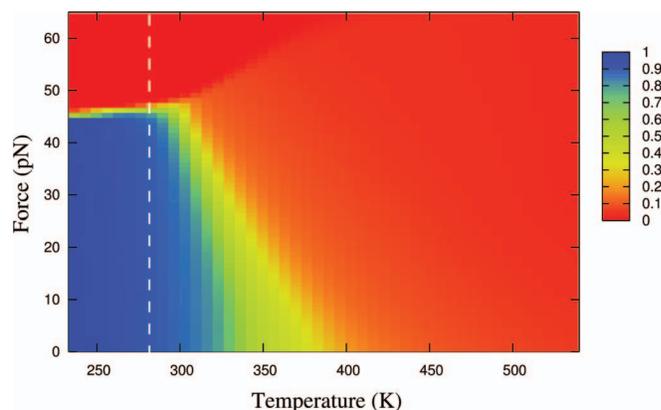

FIG. 2. The $f - T$ phase diagram obtained by the extended histogram method. The results were averaged over 30 trajectories. The vertical dashed line marks $T = 0.42\epsilon_H/k_B = 280$ K at which most of our calculations have been performed.

sufficient sampling, 30 independent trajectories were generated at each value of temperature and force.

We studied the unfolding pathways by monitoring the fraction of native contacts of each β-strand and their pairs as a function of end-to-end distance which is believed to be a good reaction coordinate. All-atom representations were obtained by reconstruction of the backbone and side chain atoms by BBQ method[35] and SCWRL 4.0 package[36] described in more detail in Ref. 37.

III. RESULTS

A. Temperature-force phase diagram

The $f - T$ phase diagram, obtained from the extended histogram method (see Sec. II) is shown in Fig. 2. The folding-unfolding transition is defined by the yellow region which is sharp in the low temperature region but becomes less cooperative (the fuzzy transition region is wider) as T increases. The weak reentrancy occurs at low temperatures, where a slight decrease in the critical force with T is observed. This seemingly strange phenomenon occurs as a result of competition between the energy gain and the entropy loss on stretching. A similar cold unzipping transition was also observed in a number of models for heteropolymers[38] and for proteins.[22] In the absence of force, the folding temperature, T_F, at which df_N/dT is maximum (results not shown), is equal to $T_F = 0.5\epsilon_H/k_B$. Equating this value with the experimental value $T_F = 333$ K,[29] we can extract the energy scale ϵ_H which is given in Sec. II. At $T = 0.42\epsilon_H/k_B = 280$ K, in which our simulations were carried out, the equilibrium critical force $f_{eq} = 0.42\epsilon_H/\text{Å} \approx 40$ pN. This value is higher than the experimental estimate $f_{eq} \approx 18$ pN.[2] Given the simplicity of the Go model we use, this agreement with the experiments is considered reasonable.

B. Force-extension profile: The second peak disappears at low pulling speeds

The force-extension profiles of I27 shown in Fig. 3 have two peaks over a wide range of pulling speeds. The main peak occurs at $\Delta R \approx 8$ Å for all speeds studied, while the
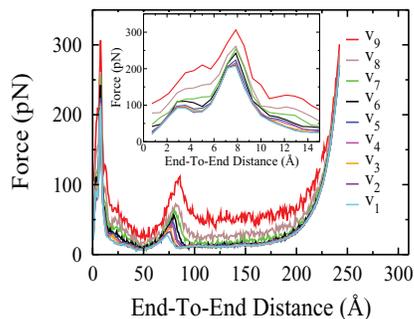
FIG. 3. Averaged force-extension profiles of the I27 titin domain. The results obtained by Go-model simulations performed at different pulling speeds are shown on the right. The inset shows an enlargement of the starting region. Results have been averaged over 50 trajectories. Values of pulling rates are given in Table I.

position of the second lower peak depends on loading rates. As follows from Figs. 3 and 4, the first peak is preceded by the hump which was also observed in AFM experiments and interpreted as a signature of the intermediate state with the A strand detached from the protein.[4] At first glance, the existence of the hump and the first peak agrees well with the AFM data, however, there is a substantial difference in the nature of the humps observed in the experiments and in our Go model. Since the hump seen in the experiments is caused by the detachment of the whole A strand, the peak associated with it may be considered as the first transition state (TS1) (Fig. S3 in the supplementary material[63]). The hump observed in Go simulations, as shown below, occurs due to breaking of only two out of the seven native contacts between A and B and its smeared peak at an extension of 4–5 Å cannot be interpreted as the transition state. Therefore, TS1 in the Go model is the main peak at ~ 8 Å in Fig. 3 (see also Fig. S3 in the supplementary material[63]), while the second peak at ~ 75 Å corresponds to the second TS2. The experimental TS2 is the main peak located at ~ 10 Å (Fig. S3 in the supplementary material[63]).

The second peak around 75–85 Å revealed by the Go-model simulations indicates that an additional mechanical intermediate becomes populated. Thus, the Go-model fails to predict the experimentally observed intermediate state with the A strand detached. Instead, it predicts the existence of the peak around 75–85 Å far from the native conformation.
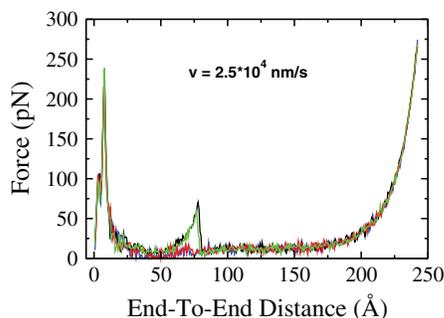

FIG. 4. Individual force-extension profiles of the I27 titin domain. Four individual trajectories at 2.5×10^4 nm/s obtained by Go-model simulations are presented. There is no second peak in two trajectories.

This result was also observed in earlier Go-model studies[17,39] and confirmed by all-atom simulation in explicit solvent (see Fig. S1 in the supplementary material[63]). It should be noted that the height of the second peak strongly decreases with the decreasing pulling speed (Fig. 3). Its average value is as low as 36 pN at the lowest velocity $v_1 = 2.5 \times 10^4$ nm/s. More importantly, at this speed 30% of trajectories do not reveal the second peak while at $v_2 = 5.76 \times 10^4$ nm/s this value becomes 18%. For illustration, we show four typical force-extension curves in Fig. 4, where trajectories shown in green and black proceed with clear second peaks, while those in red and green do not. The increasing probability of such a pathway suggests that the second peak might vanish in the experiment (see also below). As shown below, the Go model does not correctly describe the unfolding pathways even for trajectories that proceed with the second peak.

C. Protein unfolding pathway dependence on pulling speed

To monitor unfolding sequencing, we plot the fraction of native contacts formed by β-strands with the rest of the protein and native contacts formed by pairs of β-strands as a function of end-to-end distance, ΔR. Figs. 5 and 6 show the ΔR dependence of native contacts of all β-strands and their pairs for different pulling speeds.

1. High pulling speed regime, $v \lesssim 10^6$ nm/s

In this regime, where the pulling speed v is larger than 10^6 nm/s (Fig. 5(a)), unfolding starts from the C-terminus (Eq. (3a)). G and F strands are detached first followed by the simultaneous unfolding of strands A, A′, B, and C. Finally, unfolding of the most stable strands E and D occurs. Fig. 5(d) gives the following sequence for interstrand contacts for the high velocity regime (Eq. (3b)). Interstrand contacts begin to break down from AG and AG′ followed by the serial breaking of FG, CF, BE, and AB contacts. Breaking the DE contacts completes the unfolding process. The typical unfolding pathway at $v = 3.22 \times 10^6$ nm/s is shown in Figs. 8(e)–8(h). Clearly, that first peak in the force-extension profile at $\Delta R \approx 8$ Å corresponds to breaking the AG and A′G contacts [Figs. 5(d)–5(f)]. Once they are ruptured (Fig. 7(c)), the protein passes from the transition state into the intermediate one. A typical intermediate structure with the C-terminus unfolded is shown in Fig. 8

$$G \rightarrow F \rightarrow (A, A', B, C) \rightarrow E \rightarrow D, \qquad (3a)$$

$$AG \rightarrow A'G \rightarrow FG \rightarrow CF \rightarrow BE \rightarrow AB \rightarrow DE. \qquad (3b)$$

2. Low velocity regime, $v \lesssim 10^6$ nm/s

The unfolding sequence below $v = 2.88 \times 10^5$ nm/s is different. The G and F strands remain partially structured before the complete unfolding of the A, A′, B, and C strands (Fig. 5(c)). Only after the complete unfolding of the A, A′, B, and C strands, do F and G strands lose their secondary structures (Eq. (4a)). We observed a switch in the unfolding
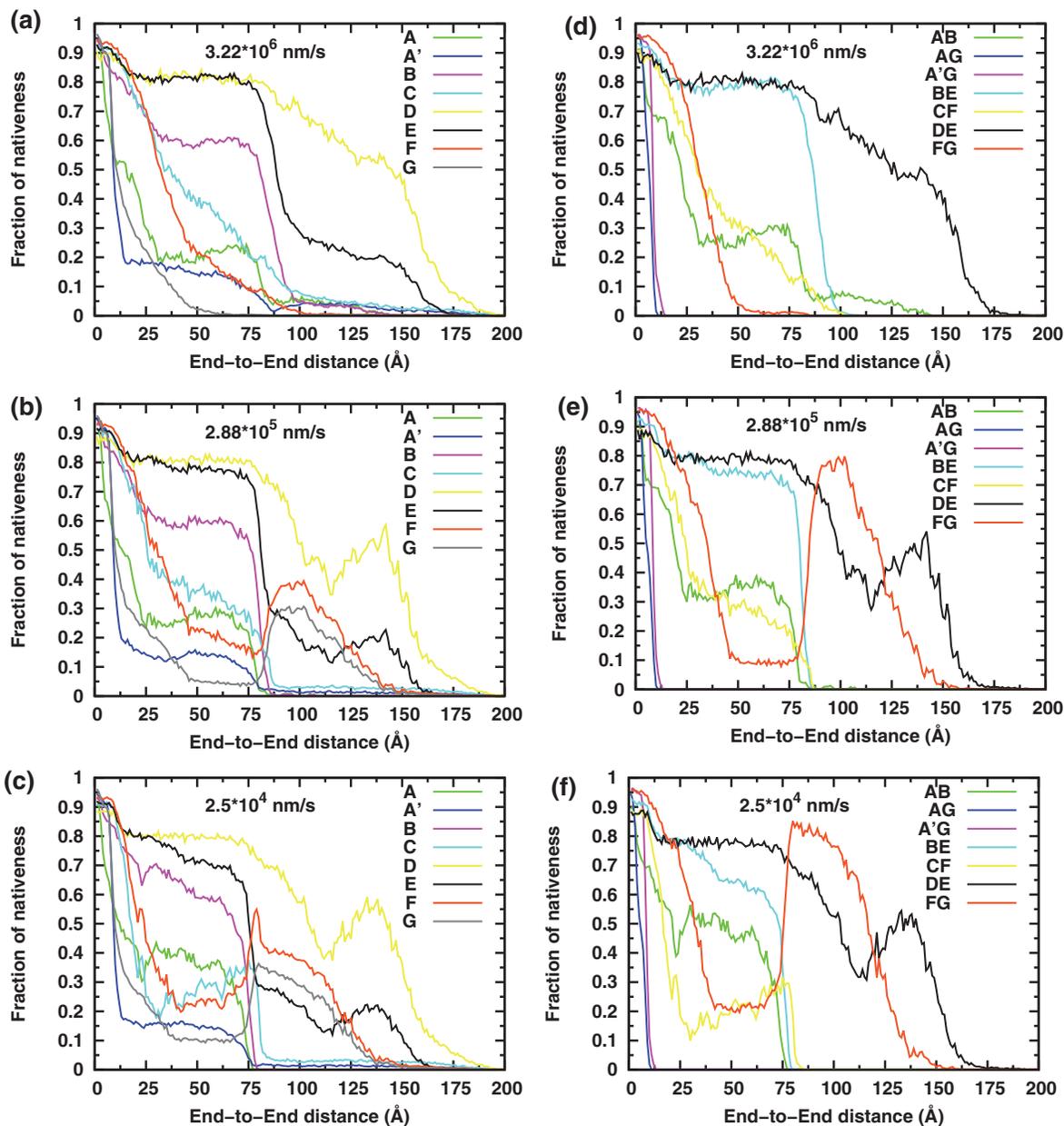
FIG. 5. (a)–(f) End-to-end distance dependence of averaged fractions of native contacts. Native contacts are formed by 8 β-strands marked in Fig. 1 at different loading rates. Clearly, the unfolding at high forces starts from the C-terminus detaching G-strand first. In contrast, at low forces the A and A′ strands are unfolded first, but it should be noted that the extension at which complete detachment of the A strand takes place is rather large, 75 Å.

mechanism – below $v = 2.88 \times 10^5$ nm/s the unfolding starts form the N-terminus, otherwise from the C-terminus. As follows from the dependence of intrastrand contacts on ΔR, there is a reverse order of events compared to the high velocity pulling regime, besides the initial and final stages of the unfolding process. Namely, the serial breaking AB, BE, CF, FG contacts proceeds after unraveling the AG and A′G contacts, but it precedes the final unwrapping of the DE contacts (Eq. (4b))

$$(A, A') \to B \to C \to G \to F \to E \to D, \qquad (4a)$$

$$AG \to A'G \to AB \to BE \to CF \to FG \to DE, \quad (4b)$$

$$(AG, AB) \to A'G \to ?. \qquad (4c)$$

According to the Go-model results, the first resistance point is the contact between the A and A′ with the G strand. Once it is ruptured, the force drops drastically. The first peak in the force-extension profile is robust for all pulling speeds studied. The nature of second resistance point (peak) is due to the core formed by either A, A′, B, C strands at low pulling speeds or by A, A′, B, C, F strands at high pulling speeds. As contacts begin to break down, the further unraveling proceeds smoothly without significant resistance.

At first glance, our Go-model data agree with the experiment where at low loading rates strand A is detached from the protein first. According to the AFM experiment, AB HBs provide the first line of defense against the external force. At about 100 pN, the A strand detaches out of the G and B strands leading to a ∼7–8 Å extension. The second line of defense
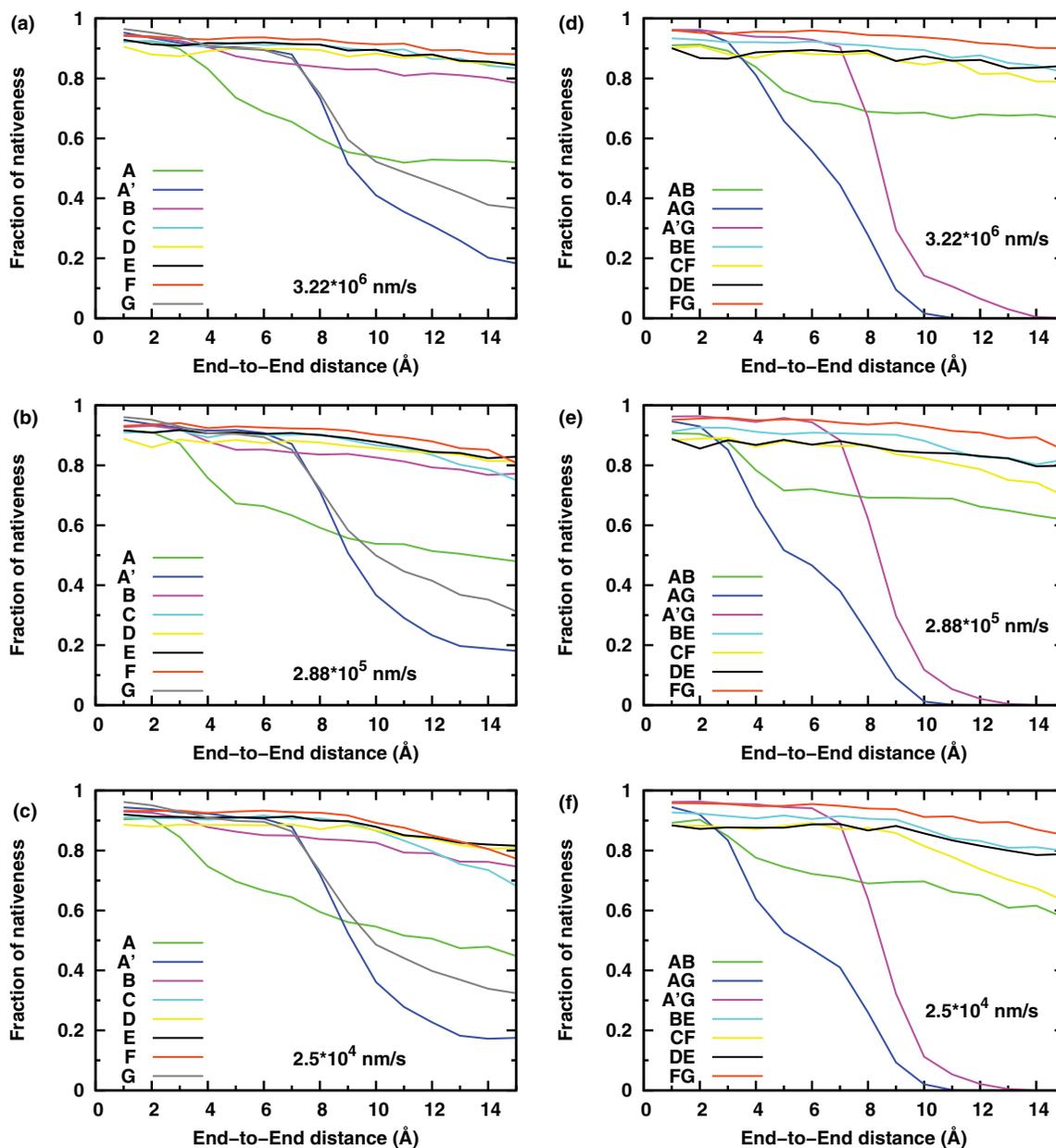
FIG. 6. (a)–(f) End-to-end distance dependence of averaged fractions of native contacts. Same as Fig. 5 but for the end-to-end distance up to 15 Å. Within 15 Å the detachment of the A strand out of the protein core is not observed for any speed studied.

lies in the A′G region. Once A′G HBs are broken the protein no longer resists to the force. However, as seen in Fig. 5, the complete detachment of the A strand takes place at a 75 Å extension almost simultaneously with the A′ and B strands. Interestingly, similar to the AFM experiment, the Go-model shows the hump in the rising phase of the first peak (Fig. 3). However, if we examine the molecular origin of the hump, agreement between the Go-model simulation and the experiment is not observed. Namely, instead of the full detachment of the A strand (or breaking all HBs between the A and B strands), we observe breaking 100% AG contacts on average (there are two native contacts between A and G) and only up to 30% AB contacts (2 out of 7 native contacts between A and B). Thus, unlike the experimentally confirmed unfolding pathway (Eq. (4c)), we do not observe the detachment of the A strand at small extensions within 10 Å over the all

pulling speeds studied. If we look at the evolution of pairs (Figs. 6(d)–6(f)), contacts between the A and B strands are always broken after those between A′ and G. Within a 10 Å extension more than 70% AB contacts remain formed. Typical conformations at an extension of 4 and 9 Å are shown in Figs. 7(b) and 7(c), respectively. Thus, the Go-model fails to reproduce the experimental observation that the A strand unfolds first regardless the applied pulling force. In 100% trajectories, the system is directed into an alternative pathway, where the breaking of A′G and AG contacts takes place first, while experiments showed that in the intermediate state the A strand should be detached from the protein completely (or both AG and AB contacts are broken). It is not clear whether the rupture events at a larger extension are correct, as even at a small extension the system is directed into the wrong pathway. Moreover, molecular interactions underlying the mechanical
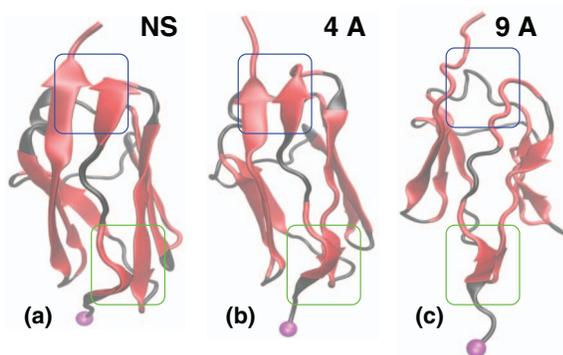
FIG. 7. Typical unfolding pathway of titin from Go-model simulation. Green and blue squares mark AB and A′G regions, respectively. The N-terminal residue is shown in magenta. (a) NS conformation. (b) Conformation at 4 Å extension, contacts between AG are missing, both A′G and AB remained formed. (c) Conformation at 9 Å extension (after the main peak), contacts between G with A and A′ are broken, those between the A and B strands are preserved. It should be emphasized that 100% trajectories at the beginning of unfolding (within 10 Å) proceed via the same pathway presented here regardless of the applied pulling rate.

resistance of the protein might be altered. Fortunately, it is not the case for titin, where the mechanical resistance of the protein lies in the A′G region, while AB contacts do not contribute to mechanostability.[5,9] However, one has to keep in mind, that, in general, the unfolding pathway probed by Go-models can differ from the pathway studied in AFM experiments and the molecular basis of mechanical stability might be affected.

It is worth mentioning that all-atom simulations in explicit solvent, show correctly not only the hump but also the molecular structure behind it indicating the presence of the experimentally observed intermediate state. All hydrogen bonds between the A and B strands are broken after the protein passes the first transition state. Snapshots are shown in Fig. S2 in the supplementary material.[63]

D. FEL parameters: Bell approximation and beyond

In experiments, one usually uses the Bell formula[40]

$$\tau_U = \tau_U^0 \exp(-x_u f / k_B T) \qquad (5)$$

to extract x_u for two-state proteins from the force dependence of unfolding times τ_U. Equation (5) is valid if the location of the transition state does not move under external force. Assuming that the force increases linearly with pulling speed v and the x_u does not depend on the external force, Evans and Ritchie[41] have shown that the distribution of unfolding force $P(f)$ obeys the following equation:

$$P(f) = \frac{k_u(f)}{v} \exp\left\{ \frac{k_B T}{x_u v}[k_u(0) - k_u(f)] \right\}, \qquad (6)$$

where $k_u(f) = \tau_U^{-1}$ is given by Eq. (5). Then, the most probable unbinding force or the maximum of force distribution f_{max}, obtained from the condition $dP(f)/df|_{f=f_{max}} = 0$, is

$$f_{max} = \frac{k_B T}{x_u} ln \frac{k v x_u}{k_u(0) k_B T}. \qquad (7)$$
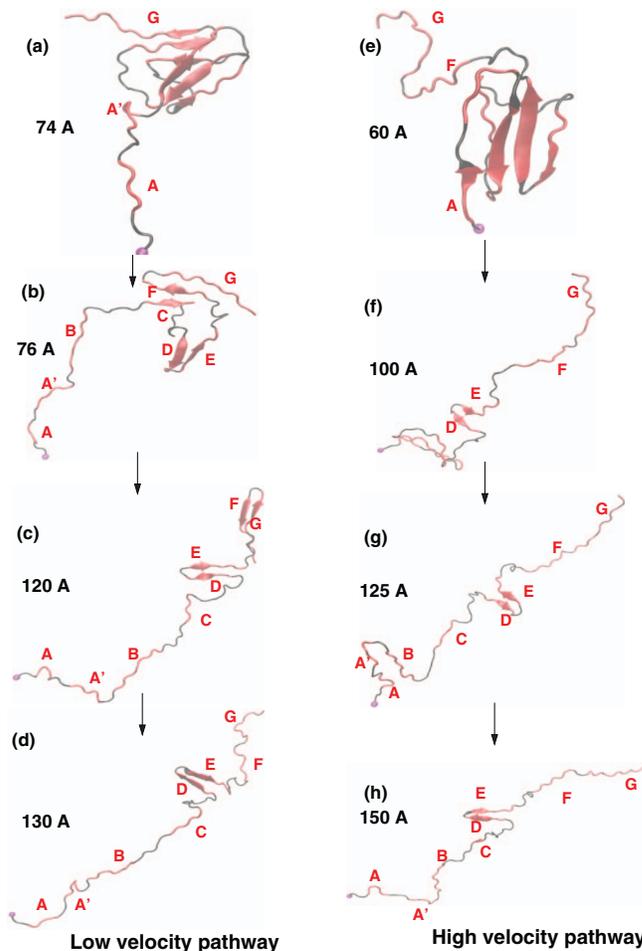

FIG. 8. Typical unfolding pathways of titin. (a)–(d) High pulling speed and (e)–(h) low pulling speed regions. The N-terminal residue is shown in magenta.

Schlierf and Rief[42] have shown that if location of transition state is sensitive to the applied force, one has to go beyond the Bell approximation. Dudko *et al.*[43] have proposed the following force dependence for the unfolding time:

$$\tau_u = \tau_u^0 \left(1 - \frac{\nu x_u}{\Delta G^\ddagger} \right)^{1-1/\nu}$$
$$\times \exp\left\{ -\frac{\Delta G^\ddagger}{k_B T}[1 - (1 - \nu x_u f / \Delta G^\ddagger)^{1/\nu}] \right\}. \qquad (8)$$

Here, $\Delta G^\ddagger$ is the unfolding barrier, and $\nu = 1/2$ and $2/3$ for the cusp[44] and the linear-cubic free energy surface,[45] respectively. $\nu = 1$ leads to the phenomenological Bell theory (Eq. (5)). Note that if $\nu \neq 1$, both x_u and ΔG can be determined.

Fig. 9 shows the most probable unbinding force as a function of pulling speed. As evident from the plot, there exists a critical speed $v_c \approx 10^6$–10^7 nm/s, separating the low and high pulling speed regimes. In the low pulling speed regime ($v \lesssim 4 \times 10^6$ nm/s), a linear fit works pretty well. Using a linear fit $y = 109.6 + 9.59 \ln(x)$ for F_{max1} and Eq. (7), we obtain the distance between N and TS1, $x_{u(NS \to TS1)} = 3.76$ Å which is comparable to the experimental values of 2.5–3.0 Å.[2,28] A nonlinear fit works for a much wider interval (Fig. 9). Using Eq. (8) with $\nu = 1/2$, we obtain $x_{u(NS \to TS1)} = 6.68$ Å and
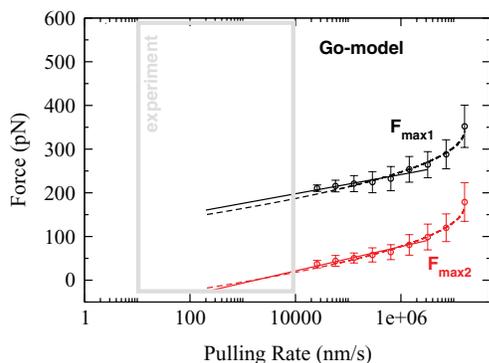
FIG. 9. Force dependence on pulling speeds at $T = 285$ K. Force at a given value of pulling speed is computed as an average of maximum forces over 50 trajectories. Grey boundaries of the polygon illustrate the interval of pulling rates used in the AFM experiment. Black circles correspond to data for F_{max1}. Solid and dashed black curves represent linear and nonlinear fits for F_{max1}. Similarly, red color is used for F_{max2}. Straight lines are fits to the Bell-Evans-Ritchie equation (Eq. (7)), $y = 109.6 + 9.59\ln(x)$ and $y = -90.332 + 12.108\ln(x)$ for F_{max1} and F_{max2}, respectively. Using a linear fit we found $x_{u(N \to TS1)} = 3.76$ Å and $x_{u(IS \to DS)} = 2.95$ Å for F_{max1} and F_{max2}, respectively. From a nonlinear fit and Eq. (8), we got $x_u = 6.68$ Å and $\Delta G^\ddagger = 32.48 k_B T$ for F_{max1} and $x_u = 3.88$ Å and $\Delta G^\ddagger = 9.22 k_B T$ for F_{max2}. Extrapolation to the experimental pulling speed, $v = 200$ nm/s, gives a negative value of F_{max2} regardless of the fit used. Extrapolated values of F_{max1} to 200 nm/s are 160 and 152 pN using linear and nonlinear fits, respectively.

$\Delta G_1^\ddagger = 32.48 k_B T$ for F_{max1}. Similar results for $x_{u(NS \to TS1)}$ and $\Delta G_1^\ddagger$ were found using a fit with $v = 2/3$ which works for a slightly narrower interval (results not shown). The distance to the transition state, $x_{u(NS \to TS1)}$, based on the nonlinear theory is close to the experimental value $x_u = 5.9$ Å reported by Williams *et al.*[46] Extrapolating results to the pulling speed $v = 200$ nm/s used in the experiments, we get $F_{max1} \approx 160$ pN (Fig. 9), which agrees quite well with the $F_{max} \approx 200$ pN obtained by stretching a polyprotein of identical I27 domains in the AFM experiment.[4]

In the case of F_{max2}, the linear fit gives $x_{u(IS \to TS2)} = 2.95$ Å, while from the nonlinear theory we obtain $x_{u(IS \to TS2)} = 3.88$ Å and $\Delta G_2^\ddagger = 9.22 k_B T$. This result is consistent with the previously reported values obtained by constant force Go-model simulations.[39]

As stated above, the number of trajectories without the second peak increases as pulling speed decreases. Using the results presented in Fig. 9, we can roughly estimate the v below which the second maximum disappears. Using $y = -90.332 + 12.108\ln(x)$, we estimated the second peak to be zero at a pulling speed of $v = 1738$ nm/s. It is worth noting that the extrapolated value of F_{max2} at the upper limit of AFM pulling speed, $v = 10^4$ nm/s, is only 21 pN which cannot be detected against the fluctuation background which could be as high as 30 pN (Fig. 3 in Ref. 2). Thus, it is clear why the second peak is not observed under experimental conditions. In addition, the Go model results suggest that in the high pulling speed regime, $v > \sim 10^4$ nm/s , unfolding becomes three-state. The applied force not only lowers the energy barrier, but also leads to an additional transition state not found at low pulling speeds.

E. Switch in unfolding pathways and applicability of Bell's theory

It is noted that, the pulling speed $v \approx 10^6$ nm/s, at which we observed the switch in the unfolding pathway, falls within the interval $\sim 10^6$–10^7 nm/s. On the other hand, above this loading rate the linear fitting based on Bell's theory ceases to work. Thus, the switch in the unfolding pathway might be interpreted as the sign that Bell's theory is applicable no more. This phenomenon was also observed for protein DDFLN4[23] but the reason behind it was not discussed. Here, we address this problem in more detail. At weak forces (low pulling speeds), the protein experiences the action of the external force uniformly along the chain and the secondary structure that has the weakest interaction with the rest would unfold first and so all. However, for the finite propagation speed of a perturbation caused by the force, the situation changes if a strong force is applied. In this case, it is not necessary that the weakest part unfolds first if it is located far from the point where the force is applied. The secondary structure at the terminus pulled by the force may be detached earlier.[23] Thus, the switch in mechanically unfolding pathways is associated with the crossover from the weak force to the strong force regime. On the other hand, Bell's theory, which is valid at low forces, is expected to fail in the strong force regime. Therefore, to the best of our knowledge for the first time, we observe a relationship between the switch in the unfolding pathway in Go models and the range of applicability of Bell's theory. This observation is general and should work for other models. It would be interesting to demonstrate this for other proteins using Go and more precise modeling.

It should be stressed that in our simulations the switch in unfolding pathways occurs at pulling speeds of 10^6–10^7 nm/s. In this regime, the deviation from Bell behavior may be captured by the 1D theory.[42,43] Combing the experimental data with results obtained from high speed pulling simulations, Schulten and Gao[47] suggested that Bell's theory is violated at $v \sim 10^8$ nm/s but they did not explicitly show the change in pathways. There exists another lower-limit of applicability of Bell's theory at very small forces close to 0.[20,48] Here, the change in unfolding pathways is observed at $v \sim 20$ nm/s. The dependence of the unfolding force on v cannot be fitted by either Bell's or 1D theory. The coexistence of force-induced and zero-force (thermal or denaturant) unfolding pathways[21] at small forces leads to non-zero unfolding forces which deviate significantly from what is predicted by these theories.[20] To correctly describe this phenomenon, one has to go beyond 1D models using the multidimensional energy landscape or taking into account alternative unfolding pathways.[20,48–51]

IV. CONCLUSIONS

The key result in this paper is that the mechanical unfolding pathway of the I27 domain probed by a structure-based $C\alpha$-model is not consistent with experimental observations even at low pulling speeds. Similarly to other β-strand proteins (DDFLN4 and FNIII),[23,27] the unfolding pathway depends on pulling speed: at large v the C-terminus unfolds first,

while at speeds close to the experimental ones the N-terminus unfolds before the C-terminus.

Comparing our simulations with the previous reports, we find that a similar deviation from experimentally detected unfolding pathways was observed previously not only in the Go-model[17,39] but also in the all-atom implicit solvent simulations.[26] Both models neglect the IS probing the artificial unfolding pathway. The slowest pulling speed, $v = 2.5 \times 10^4$ nm/s, used in our Go-model simulation corresponds to the upper limit of loading rates used in AFM experiments, $v = 10^4$ nm/s[28] and is nearly two orders of magnitude slower than previously reported.[17] It turns out that probing the wrong pathway by the Go-model is not related to the fast pulling, but rather indicates more fundamental problems. The failure of the Go-model to detect the first transition state (or the intermediate state), which should be associated with a loss of native interactions comes as a surprise. This finding is valuable as the unfolding by an external force is believed to be solely governed by the native topology of proteins. Thus, one has to maintain a healthy skepticism about systematic studies based on Go-models.[52,53] Although Go-models have been proved to provide reasonable estimates for mechanostability properties,[22,54–59] there is no guarantee that they can solve unfolding pathways. Benchmarking of simulation results on the molecular basis of protein mechanostability by experiment is necessary to make sure that properties measured in an experiment are properly reproduced by simulation.

The inclusion of more realistic interactions and explicit solvent influence into the model would help to obtain proper pathways.[6] However, at present, deciphering the unfolding FEL of long proteins by all-atom simulations with explicit water is still computationally prohibitive. Developing a model which is still computationally feasible and at the same time takes into account the contribution of side-chain and non-native interactions would be of great help. One of the possibilities is to use combination of structure-based models with more realistic interactions described by statistics-based potentials.[60,61] We are now using such a combination of the Go-model with CABS software[62] to test the effectiveness of this idea in ongoing simulations.

ACKNOWLEDGMENTS

[1] L. Tskhovrebova, K. Trinick, J. A. Sleep, and M. Simons, Nature (London) **387**, 308 (1997).

[2] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, Science **276**, 1109 (1997).

[3] H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten, Biophys. J. **75**, 662 (1998).

[4] P. E. Marszalek, H. Lu, H. Li, M. Carrion-Vazquez, A. F. Oberhauser, K. Schulten, and J. M. Fernandez, Nature (London) **402**, 100 (1999).

[5] S. Fowler, R. R. Best, J. L. Toca-Herra, T. Rutherford, A. Steward, E. Paci, M. Karplus, and J. Clarke, J. Mol. Biol. **322**, 841 (2002).

[6] E. H. Lee, J. Hsin, M. Sotomayor, G. Comellas, and K. Schulten, Structure **17**, 1295 (2009).

[7] H. Lu, V. Vogel, and K. Schulten, Chem. Phys. **247**, 141 (1999).

[8] H. Lu and K. Schulten, Biophys. J. **79**, 51 (2000).

[9] R. Best, S. Fowler, J. Herrera, A. Steward, E. Paci, and J. Clarke, J. Mol. Biol. **330**, 867 (2003).

[10] E. Paci and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **97**, 6521 (2000).

[11] N. Go, Annu. Rev. Biophys. Bioeng. **12**, 183 (1983).

[12] C. Clementi, H. Nymeyer, and J. N. Onuchic, J. Mol. Biol. **298**, 937 (2000).

[13] R. D. Hills and I. C. L. Brooks, Int. J. Mol. Sci. **10**, 889 (2009).

[14] M. Kouza and U. Hansmann, J. Chem. Phys. **134**, 044124 (2011).

[15] D. West, P. Olmsted, and E. Paci, J. Chem. Phys. **124**, 154909 (2006).

[16] M. Kouza, C. K. Hu, and M. S. Li, J. Chem. Phys. **128**, 045103 (2008).

[17] M. Cieplak, T. X. Hoang, and M. Robbins, Proteins: Struct., Funct., Bioinf. **49**, 114 (2002).

[18] S. Sharma, F. Ding, H. Nie, D. Watson, A. Unnithan, J. Lopp, D. Pozefsky, and N. V. Dokholyan, Bioinformatics **22**, 2693 (2006).

[19] J. I. Sulkowska, A. Kloczkowski, T. Sen, M. Cieplak, and R. Jernigan, Proteins **71**, 45 (2008).

[20] M. Schlierf, Z. T. Yew, M. Rief, and E. Paci, Biophys. J. **99**, 1620 (2010).

[21] Z. T. Yew, S. Krivov, and E. Paci, J. Phys. Chem. B **112**, 16902 (2008).

[22] M. S. Li, M. Kouza, and C. K. Hu, Biophys. J. **92**, 547 (2007).

[23] M. S. Li and M. Kouza, J. Chem. Phys. **130**, 145102 (2009).

[24] A. Irback, S. Mittetnacht, and S. Mohanty, Proc. Natl. Acad. Sci. U.S.A. **102**, 13427 (2005).

[25] M. Kouza, C. K. Hu, H. Zung, and M. S. Li, J. Chem. Phys. **131**, 215103 (2009).

[26] P. C. Li and D. E. Makarov, J. Chem. Phys. **119**, 9260 (2003).

[27] S. Mitternacht, S. Luccioli, A. Torcini, A. Imparato, and A. Irback, Biophys. J. **96**, 429 (2009).

[28] M. Carrion-Vazquez, A. F. Obserhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, and J. M. Fernandez, Proc. Natl. Acad. Sci. U.S.A. **96**, 3694 (1999).

[29] A. S. Politou, M. Gautel, M. Pfuhl, S. Labeit, and A. Pastore, Biochemistry **33**, 4730 (1994).

[30] M. P. Allen and D. J. Tildesley, *Computer Simulations of Liquids* (Oxford Science, Oxford, 1987).

[31] M. Kouza, C. F. Chang, S. Hayryan, T. H. Yu, M. S. Li, T. H. Huang, and C. K. Hu, Biophys. J. **89**, 3353 (2005).

[32] T. Veitshans, D. K. Klimov, and D. Thirumalai, Folding Des. **2**, 1 (1997).

[33] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).

[34] D. K. Klimov and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **97**, 7254 (2000).

[35] D. Gront, S. Kmiecik, and A. Kolinski, J. Comput. Chem. **28**, 1593 (2007).

[36] G. S. Krivov, M. V. Shapovalov, and R. L. J. Dunbrack, Proteins: Struct., Funct., Bioinf. **77**, 778 (2009).

[37] S. Kmiecik, D. Gront, M. Kouza, and A. Kolinski, J. Phys. Chem. B **116**, 7026 (2012).

[38] P. L. Geissler and E. I. Shakhnovich, Phys. Rev. E **65**, 056110 (2002).

[39] M. S. Li, A. Gabovich, and A. Voitenko, J. Chem. Phys. **129**, 105102 (2008).

[40] GI Bell, Science **200**, 618 (1978).

[41] E. Evans and K. Ritchie, Biophys. J. **72**, 1541 (1997).

[42] M. Schlierf and M. Rief, Biophys. J. **90**, L33 (2006).

[43] O. K. Dudko, G. Hummer, and A. Szabo, Phys. Rev. Lett. **96**, 108101 (2006).

[44] G. Hummer and A. Szabo, Biophys. J. **85**, 5 (2003).

[45] O. K. Dudko, A. E. Filippov, J. Klafter, and U. Urbakh, Proc. Natl. Acad. Sci. U.S.A. **100**, 11378 (2003).

[46] P. M. Williams, S. B. Fowler, R. B. Best, J. L. Toca-Herrera, K. A. Scott, A. Steward, and J. Clarke, Nature (London) **422**, 446 (2003).

[47] M. Gao and H. L. K. Schulten, J. Muscle Res. Cell Motil. **23**, 513 (2002).

[48] Z. T. Yew, M. Schlierf, M. Rief, and E. Paci, Phys. Rev. E **81**, 031923 (2010).

[49] Y. Suzuki and O. K. Dudko, Phys. Rev. Lett. **104**, 048101 (2010).

[50] Y. Suzuki and O. K. Dudko, J. Chem. Phys. **134**, 065102 (2011).

[51]C. Kappel, N. Dolker, R. Kumar, M. Zink, U. Zachariae, and H. Grubmuller, Phys. Rev. Lett. **109**, 118304 (2012).

[52]J. I. Sulkowska and M. Cieplak, J. Phys.: Condens. Matter **19**, 283201 (2007).

[53]M. S. Li, Biophys. J. **93**, 2644 (2007).

[54]D. J. Brockwell, E. Paci, R. Zinober, G. Beddard, P. Olmsted, D. Smith, R. Perham, and S. Radford, Nat. Struct. Biol. **10**, 731 (2003).

[55]M. Cieplak and P. E. Marszalek, J. Chem. Phys. **123**, 194903 (2005).

[56]M. S. Li, C. K. Hu, D. K. Klimov, and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **103**, 93 (2006).

[57]A. Valbuena, J. Oroz, R. Hervas, A. M. Vera, D. Rodriguez, M. Menendez, J. I. Sulkowska, M. Cieplak, and M. Carrion-Vazquez, Proc. Natl. Acad. Sci. U.S.A. **106**, 13791 (2009).

[58]G. Arad-Haase, S. Chuartzman, S. Dagan, R. Nevo, M. Kouza, B. Mai, H. Nguyen, M. S. Li, and Z. Reich, Biophys. J. **99**, 238 (2010).

[59]S. Kumar and M. S. Li, Phys. Rep. **486**, 1 (2010).

[60]S. Miyazawa and R. L. Jernigan, Macromolecules **18**, 534 (1985).

[61]A. Kolinski, A. Godzik, and J. Skolnick, J. Chem. Phys. **98**, 7420 (1993).

[62]A. Kolinski, Acta Biochim. Pol. **51**, 349 (2004).

[63]See supplementary material at http://dx.doi.org/10.1063/1.4817773 for all-atom simulation results in explicit solvent.