Protein mechanical unfolding: Importance of non-native interactions

Maksim Kouza,[1] Chin-Kun Hu,[2] Hoang Zung,[3] and Mai Suan Li[1,a)]
[1]*Institute of Physics, Polish Academy of Sciences, Al. Lotnikow 32/46, Warsaw 02-668, Poland*
[2]*Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan and Department of Physics and Center for Nonlinear and Complex Systems, Chung Yuan Christian University, Chungli 32023, Taiwan*
[3]*Computational Physics Laboratory, Vietnam National University, 227 Nguyen Van Cu, Dist. 5, Ho Chi Minh City, Vietnam*

(Received 22 May 2009; accepted 14 November 2009; published online 4 December 2009)

Mechanical unfolding of the fourth domain of *Distyostelium discoideum* filamin (DDFLN4) was studied by all-atom molecular dynamics simulations, using the GROMOS96 force field 43a1 and the simple point charge explicit water solvent. Our study reveals an important role of non-native interactions in the unfolding process. Namely, the existence of a peak centered at the end-to-end extension $\Delta R \sim 22$ nm in the force-extension curve is associated with breaking of non-native hydrogen bonds. Such a peak has been observed in experiments but not in Go models, where non-native interactions are neglected. We predict that an additional peak occurs at $\Delta R \sim 2$ nm using not only GROMOS96 force field 43a1 but also Amber 94 and OPLS force fields. This result would stimulate further experimental studies on elastic properties of DDFLN4. © *2009 American Institute of Physics*. [doi:10.1063/1.3272275]

I. INTRODUCTION

The last ten years witnessed an intense activity in single molecule force spectroscopy experiments in detecting inter- and intramolecular forces of biological systems to understand their functions and structures. Much of the research focused on elastic properties of proteins, DNA, and RNA, i.e., their response to an external force, following the seminal papers by Rief *et al.*[1] and Tskhovrebova *et al.*[2] The main advantage of this technique is its ability to separate out fluctuations of individual molecules from the ensemble average behavior observed in traditional bulk biochemical experiments. This allows one for studying unfolding pathways in detail using the end-to-end distance as a good reaction coordinate. Moreover, the single molecule force spectroscopy can be used to decipher the unfolding free energy landscape (FEL) of biomolecules.[3,4]

As cytoskeletal proteins, large actin-binding proteins play a key role in cell organization, mechanics, and signaling.[5] During the permanent cytoskeleton reorganization, all involved participants are subject to mechanical stress. One of them is the fourth domain of *Distyostelium discoideum* filamin (DDFLN4), which binds different components of actin-binding protein. Therefore, understanding the mechanical response of DDFLN4 protein to an external force is of great interest. In recent Atomic Force Microscopy (AFM) experiments, Schwaiger *et al.*[6,7] have shown that DDFLN4 (Fig. 1), which has seven β-strands in the native state (NS), unfolds via intermediates. In the force-extension curve, they observed two peaks at the end-to-end extension $\Delta R \approx 11$ and 22 nm (Sec. II). However, using a Go model,[8] Li *et al.*[4] have also obtained two peaks but they are located at $\Delta R \approx 1.5$ and 11 nm (see also Sec. II). A natural question to ask is if the disagreement between experiments and theory is due to oversimplification of the Go modeling, where non-native interactions between residues are omitted. To answer this question, we performed all-atom molecular dynamics simulations, using the GROMOS96 force field 43a1 (Ref. 9) and the simple point charge (SPC) explicit water solvent.[10] In order to check robustness of our results, limited all simulations have been carried out with the help of the Amber 94 (Ref. 11) and OPLS (Ref. 12) force fields.

We have shown that two peaks do appear at almost the same positions as in the experiments[6,7] and, more importantly, the peak at $\Delta R \approx 22$ nm comes from the non-native interactions. This explains why it has not been seen in the previous Go simulations.[4] In our opinion, this result is very important as it opposes the common belief[13,14] that mechanical unfolding properties of proteins are governed by their native topology. In addition to two peaks at large ΔR, in agreement with the Go results,[4] we also observed a maximum at $\Delta R \approx 2$ nm. Because such a peak was not detected by the AFM experiments,[6,7] further experimental and theoretical studies are required to clarify this point.

II. MATERIALS AND METHODS

Simulation details. We used the GROMOS96 force field 43a1 (Ref. 9) to model DDFLN4 which has 100 amino acids, and the SPC water model[10] to describe the solvent. The GROMACS version 3.3.1 has been employed. The protein was placed in an orthorhombic box with the edges of 4.0, 4.5, and 43 nm, and with 25 000–26 000 water molecules [Fig. 1(c)].

In all simulations, the GROMACS program suite[15,16] was employed. The equations of motion were integrated by using a leap-frog algorithm with a time step of 2 fs. The LINCS (Ref. 17) was used to constrain bond lengths with a relative geometric tolerance of 10^{-4}. We used the particle-mesh Ewald method to treat the long-range electrostatic

a)Electronic mail: masli@ifpan.edu.pl.

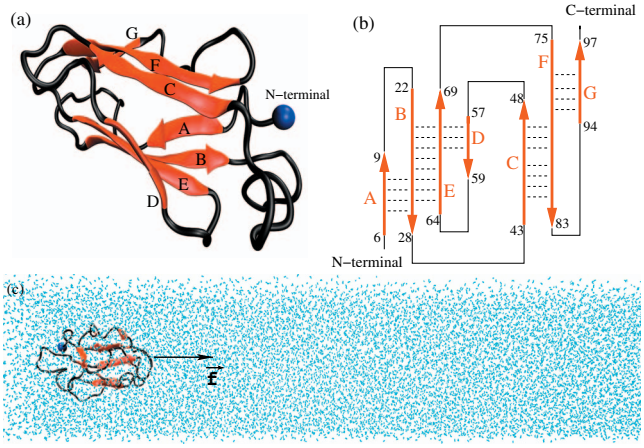

FIG. 1. (a) The native conformation of DDFLN4 (PDB code: 1KSR) with seven β-strands (A–G). (b) Plot which shows backbone contacts (dotted lines) between β-strands. All adjacent β-strands are antiparallel to each other. A, B, E, and D belong to the first β-sheet, whereas the second one contains C, F, and G. (c) The solvated system in the orthorhombic box of water (cyan). VMD software (Ref. 27) was used for a plot.

interactions.[18] The nonbonded interaction pair-list were updated every 10 fs, using a cutoff of 1.2 nm.

The protein was minimized using the steepest decent method. Subsequently, unconstrained molecular dynamics simulation was performed to equilibrate the solvated system for 100 ps at constant pressure (1 atm) and temperature $T=300$ K with the help of the Berendsen coupling procedure.[19] The system was then equilibrated further at constant temperature $T=300$ K and constant volume. Afterward, the N-terminal was kept fixed and the force was applied to the C-terminal through a virtual cantilever moving at the constant velocity v along the biggest z-axis of simulation box. We also performed limited simulations for the case when the N-terminal is pulled.

During the simulations, the spring constant was chosen as $k=1000$ kJ/(mol nm^2) ≈ 1700 pN/nm which is an upper limit for k of a cantilever used in AFM experiments. Movement of the pulled termini causes an extension of the protein and the total force can be measured by $F=k(vt-x)$, where x is the displacement of the pulled atom from its original position. The resulting force is computed for each time step to generate a force extension profile, which has peaks showing the most mechanically stable places in a protein.

Overall, the simulation procedure is similar to the experimental one, except that pulling speeds in our simulations are several orders of magnitude higher than those used in experiments. In the N-terminal fixed case, we performed simulations for $v=1\times10^9$, 5×10^9, 1.2×10^{10}, and 2.5×10^{10} nm/s, while in the AFM experiments one used $v\sim100-1000$ nm/s.[6] For each value of v we generated four trajectories. In the C-fixed case, the simulations were carried out for $v=5\times10^9$ nm/s and three trajectories.

A backbone contact between amino acids i and $j(|i-j|>3)$ is defined as formed if the distance between two corresponding C_α-atoms is smaller than a cutoff distance $d_c=6.5$ Å. With this choice, the molecule has 163 native contacts. A hydrogen bond is formed provided the distance between donor D (or atom N) and acceptor A (or atom O) ≤ 3.5 Å, and the angle D-H-A $\geq 145°$.

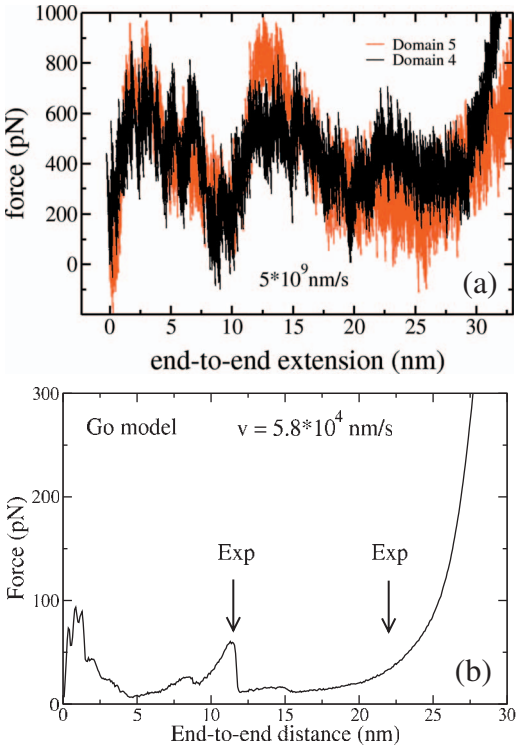

FIG. 2. (a) The force-extension profiles for domain 4 (black) and domain 5 (red). The results were obtained by the all-atom simulations for the pulling speed $v=5\times10^9$ nm/s. (b) The force-extension curve for DDFLN4 and pulling speed $v=5.8\times10^4$ nm/s was obtained by using the Go model (Ref. 8). Results were averaged over 40 trajectories. Arrows refer to positions of two peaks ($\Delta R=11$ and 22 nm) which are expected to be the same as observed on the experiments (Refs. 6 and 7).

The unfolding process was studied by monitoring the dependence of numbers of backbone contacts and hydrogen bonds (HBs) formed by seven β-strands enumerated as A–G [Fig. 1(a)] on the end-to-end extension. In the NS, backbone contacts exist between seven pairs of β-strands P_{AB}, P_{AF}, P_{BE}, P_{CD}, P_{CF}, P_{DE}, and P_{FG} as shown in Fig. 1(b). Additional information on unfolding pathways was also obtained from the evolution of contacts of these pairs. To understand the nature of the third peak in the force-extension curve, we also monitored the evolution of contacts formed by non-native β-strands which emerge as the unfolding progresses.

Validity of GROMOS96 force field 43a1. It should be noted that the Gromos force field proved useful for studying structural dynamics and kinetics of peptides and proteins.[20] It cannot only decipher folding mechanisms but also provide reasonable estimates for folding times.[21] While the steered molecular dynamics with NAMD (Ref. 22) is widely used for studying mechanical properties of biomolecules for a decade, the pulling option has been recently implemented in the GROMACS software. Therefore, we want to check its validity for stretching proteins first. Having used speed $v=5\times10^9$ nm/s, we can demonstrate (results not shown) that the force-extension profile for the domain I27 is similar to the result obtained by NAMD.[23] We also carried out limited runs for domain 5 (DDFLN5) for comparison with DDFLN4. In agreement with the experiments,[6,7] domain 4 is less stable and has one peak more than DDFLN5 [Fig. 2(a)]. Thus, the GROMACS software with the GROMOS96 force field 43a1

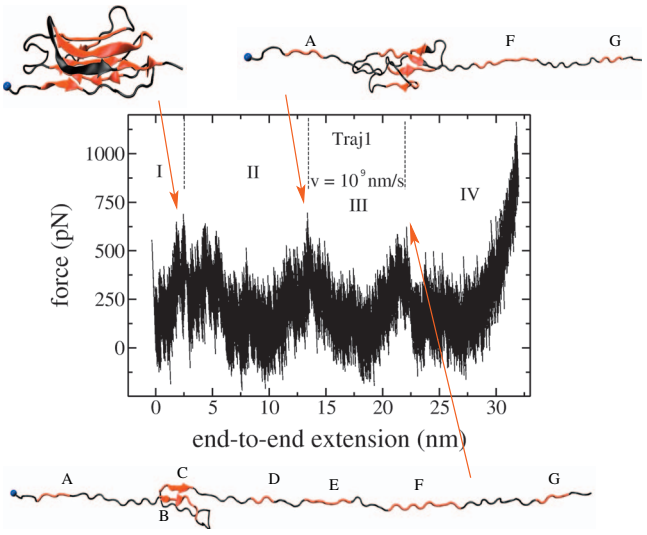

FIG. 3. Force-extension profile for trajectory 1 for $v=10^9$ nm/s. Vertical dashed lines separate four unfolding regimes. Shown are typical snapshots around three peaks. Heights of peaks (from left) are $f_{max1}=695$ pN, $f_{max2}=704$ pN, and $f_{max3}=626$ pN.

can serve as a reliable tool for studying protein mechanical unfolding. To ensure that this conclusion is correct, we also performed a limited set of simulations using all-atom Amber 94 (Ref. 11) and OPLS (Ref. 12) force fields. The simulations were carried out within the GROMACS program suite.[15,16]

A short survey on previous results obtained by experiments and Go simulations. In order to make the presentation transparent, let us briefly discuss the previous experimental[6,7] and simulation results.[24] Experimentally one observed two peaks in the force-extension curve.[6,7] However, determining their precise location as a function of the end-to-end distance is not an easy task, because, on experiments, unfolding lengths were measured by fitting data to the wormlike chain model but not by direct peak spacing. Schwaiger *et al.*[6] reported that the first peak occurs due to unfolding of strands A and B and the loop between B and C. In this first unfolding event, the length changes by $\Delta L_1=14-15$ nm, where L is the length parameter in the wormlike chain model. Assuming the distance between to neighboring amino acids is $a=0.36$ nm, they have shown that this length gain corresponds to full stretching of ~40 first residues. Since L and R have different physical meanings, this does not mean that the first peak is located at the extension $\Delta R=\Delta L_1$. Here, we propose to estimate its location by comparing the experimental data with our simulation results.[24] Note that such a trick has been already used to locate the position of the hump in the force-extension curve for the titin domain I27.[23,25] Using the Go model[8] and the pulling speed $v=5.8\times10^4$ nm/s, we obtain the force-extension curve shown in Fig. 2(b). The first peak at $\Delta R\approx1.5$ nm in this theoretical curve cannot correspond to the first peak observed in the experiments for two reasons. First, this value of ΔR is too small to describe unfolding of 40 first residues. Second, the first experimental peak occurs due to unfolding of strands A and B,[6] while the theoretical peak corresponds to breaking of native contacts between strands A and F.[24] The second theoretical peak at $\Delta R=11\pm1$ nm (the error bar comes from averaging over different trajectories) could be identified as the first experimental peak because it also corresponds to unfolding of strands A and B.[24] Thus, assuming that the Go model correctly captures the first experimental peak, one can infer that this peak is located at $\Delta R=11\pm1$ nm. From the experimental data,[6,7] we can estimate the distance between two peaks $\sim11\pm1$ nm. Therefore the second experimental peak is expected to locate at $\Delta R=22\pm2$ nm. This peak was not observed in our previous Go simulations [Fig. 2(b)].[24] But we will show that the all-atom models, where the non-native interactions are taken into account, can reproduce it.

III. RESULTS

A. Existence of three peaks in force-extension profile

Since the results obtained for four pulling speeds (Sec. II) are qualitatively similar, we will focus on the smallest $v=10^9$ nm/s case. The force extension curve, obtained at this speed, for the trajectory 1, can be divided into four regions (Fig. 3).

Region I ($0\lesssim\Delta R\lesssim2.4$ nm). Due to thermal fluctuations, the total force fluctuates a lot, but, in general, it increases and reaches the first maximum $f_{max1}=695$ pN at $\Delta R\approx2.42$ nm. A typical snapshot before the first unfolding event (Fig. 3) shows that structures remain nativelike. Dur-

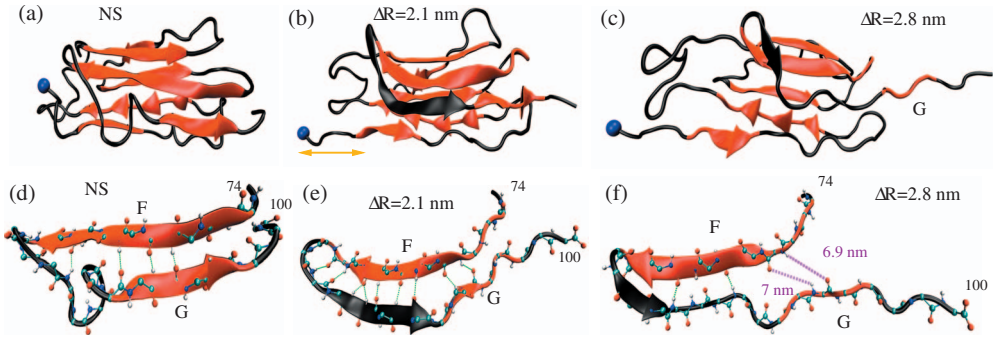

FIG. 4. (a) The NS conformation is shown for comparison with the other ones. (b) A typical conformation before the first unfolding event takes place ($\Delta R\approx2.1$ nm). The yellow arrow shows a part of protein which starts to unfold. An additional non-native β-strand between amino acids 87 and 92 is marked by black color. (c) A conformation after the first peak, at $\Delta R\approx2.8$ nm, where strand G has already detached from the core. (d) The same as in (a) but four HBs (green color) between β-strands are displayed. (e) The same as in (b) but all nine HBs are shown. (f) The same as in (c) but broken HBs (purple) between F and G are displayed.

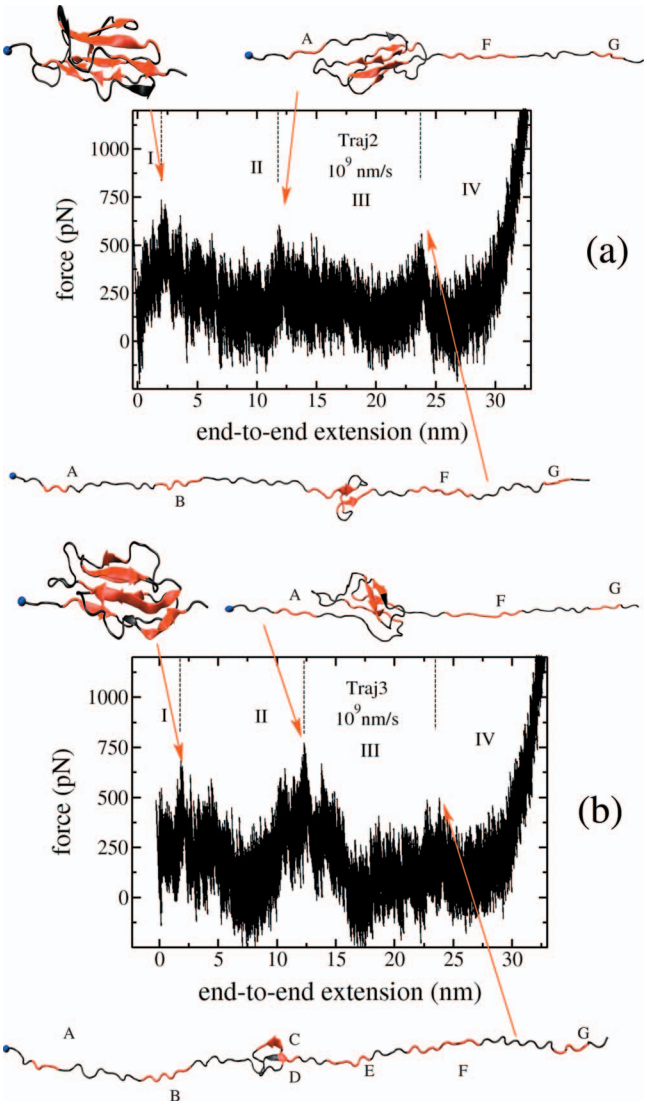

FIG. 5. The same as in Fig. 3 but for trajectories 2 (a) and 3 (b). Heights of peaks (from left) are $f_{max1}=740$ pN, $f_{max2}=614$ pN, and $f_{max3}=563$ pN for trajectory 2 and $f_{max1}=685$ pN, $f_{max2}=773$ pN, and $f_{max3}=500$ pN for trajectory 3.

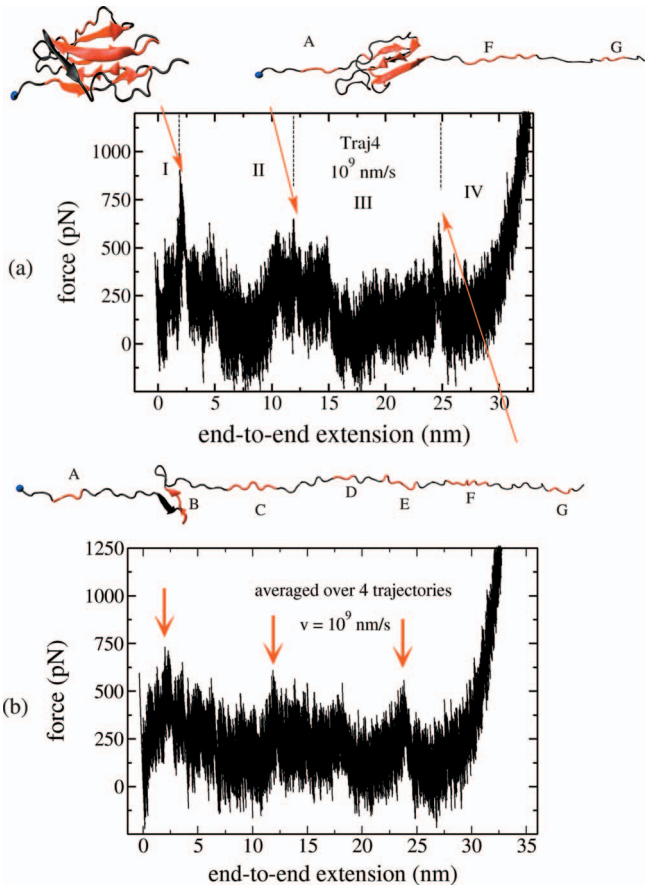

FIG. 6. (a) The same as in Fig. 3 but for trajectory 4. Heights of peaks (from left) are $f_{max1}=906$ pN, $f_{max2}=652$ pN, and $f_{max3}=630$ pN. (b) The averaged over four trajectories force-extension profile at $v=10^9$ nm/s. The corresponding peaks are $f_{max1}=597$ pN, $f_{max2}=530$ pN, and $f_{max3}=398$ pN.

The distance between the second and third peaks is about 12 nm.

ing the first period, the N-terminal part is being extended, but the protein maintains all β-sheet secondary structures [Fig. 4(b)]. Although, the unfolding starts from the N-terminal [Fig. 4(b)], after the first peak, strand G from the C-termini got unfolded first [Figs. 4(c) and 4(f)]. In order to understand the nature of this peak on the molecular level, we consider the evolution of HBs in detail. As a molecule departs from the NS, non-native HBs are created and at $\Delta R=2.1$ nm, e.g., a non-native β-strand between amino acids 87 and 92 [Fig. 4(b)] is formed. This leads to an increase in the number of HBs between F and G from 4 [Fig. 4(d)] to 9 [Fig. 4(e)]. Structures with the enhanced number of HBs should show strong resistance to the external perturbation and the first peak occurs due to their unfolding [Fig. 4(b)]. It should be noted that this maximum was also observed in the Go simulations,[4,24] but not in the experiments.[6,7] Both all-atom and Go simulations reveal that the unfolding of strand G is responsible for its occurrence.

Region II (2.4 nm$\lesssim\Delta R\lesssim$13.4 nm). After the first peak, the force drops rapidly from 695 to 300 pN and secondary structure elements begin to break down. During this period, strands A, F, and G unfold completely, whereas the B, C, D, and E strands remain structured (see Fig. 3 for a typical snapshot).

Region III (13.4 nm$\lesssim\Delta R\lesssim$22.1 nm). During the second and third stages, the complete unfolding of strands D and E takes place. Strands B and C undergo significant conformational changes, losing their equilibrium HBs. Even though a core formed by them remains compact (see bottom of Fig. 3 for a typical snapshot). Below we will show in detail that the third peak is associated with breaking of non-native HBs between strands B and C.

Region IV ($\Delta R\gtrsim$22.1 nm). After breaking of non-native HBs between B and C, the polypeptide chain gradually reaches its rod state. For the pulling speed $v=10^9$ nm/s, the existence of three peaks is robust as they are also observed in all three remaining trajectories [Figs. 5 and 6(a)]. It is also clearly evident from Fig. 6(b), where the force-extension curve, averaged over four trajectories, is displayed. The positions of peaks fluctuate from trajectory to trajectory but within error bars the locations of first two peaks are in the reasonable agreement with the Go simulations.[24] The

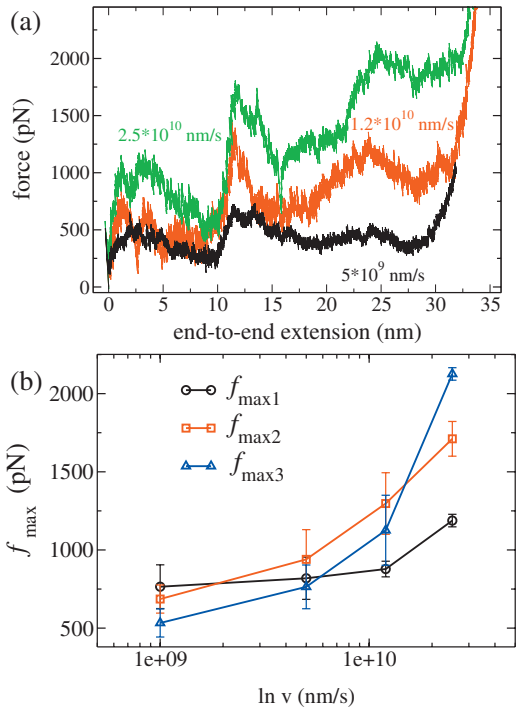

FIG. 7. (a) Force-extension profiles for three values of v shown next to the curves. (b) Dependence of heights of three peaks on v. Results are averaged over four trajectories for each value of v.

distance between the second and third peaks is about 12 nm [Fig. 6(b)] and this is in accord with the experiments.[6,7]

B. Robustness of three peaks against pulling speeds

The question we now ask is if the existence of three peaks is independent of pulling speeds. To this end, we performed simulations at various loading speeds and the results are shown in Fig. 7(a). In accordance with the kinetic theory,[26] the heights of maxima decrease as v is lowered, but three peaks exist for all values of pulling speeds. Since the first peak was not observed in the experiments,[6,7] a natural question emerges is whether it is an artifact of high pulling speeds used in our simulations. Except data at the highest value of v [Fig. 7(b)], within error bars three maxima are compatible. Therefore, in agreement with the coarse-grained simulation results,[24] the peak centered at $\Delta R \approx 2$ nm is expected to remain at experimental loading rates. It remains unclear if this is a shortcoming of theory or of experiments because it is hard to imagine that a β-protein like DDFLN4 displays the first peak at such a large extension $\Delta R \approx 11$ nm. The force-extension curve of the titin domain I27, which has a similar native topology, for example, displays the first peak at $\Delta R \approx 0.8$ nm.[25] From this prospect, our theoretical result for DDFLN4 is more favorable. One of possible reasons of why the experiments did not detect this maximum is related to a strong linker effect as a single DDFLN4 domain is sandwiched between Ig domains I27–30 and domains I31–34 from titin.[6] Remember that in the case of the domain I27, the hump in the force-extension curve was not observed in the first experiments.[1] It was experimentally detected[25] only after the theoretical prediction for its existence.[23] It would be very interesting if the first peak predicted by our simulations will be confirmed by experiments.

C. Importance of non-native interactions

As mentioned above, the third peak at $\Delta R \approx 22$ nm was observed in the experiments but not in Go models,[4,24] where non-native interactions are omitted. In this section, we show, at molecular level, that these very interactions lead to its existence. To this end, for the first trajectory of the lowest pulling speed ($v = 10^9$ nm/s), we plot the number of native contacts formed by seven strands and their pairs as a function of ΔR. The first peak corresponds to unfolding of strand G [Fig. 8(a)] as all (A,F) and (F,G) contacts are broken just after passing it [Fig. 8(b)]. Thus, the structure of the first intermediate state, which corresponds to this peak, consists of six ordered strands A–F [a typical snapshot is given in Fig. 4(c)].

The second unfolding event is associated with full unfolding of A and F and the drastic decrease in native contacts of B and C [Fig. 8(a)]. After the second peak only (B,E), (C,D), and (D,E) native contacts survive [Fig. 8(b)]. The structure of the corresponding second intermediate state contains partially structured strands B, C, D, and E. A typical snapshot is displayed in top of Fig. 3.

Remarkably, for $\Delta R \gtrsim 18$ nm, none of the native contacts exist, except very small fluctuations of a few contacts of strand B around $\Delta R \approx 22.5$ nm [Fig. 8(a)]. Such fluctuations are negligible as they are not even manifested in existence of native contacts between corresponding pairs (A,B) and (B,E) [Fig. 8(b)]. Therefore, we come to a very interesting conclusion that the third peak centered at $\Delta R \approx 22.5$ nm is not related to native interactions. This explains why it was not detected by simulations[4,24] using the Go model.[8]

The mechanism underlying occurrence of the third peak may be revealed using the results shown in Fig. 8(c), where the number of all backbone contacts (native and non-native) is plotted as a function of ΔR. Since, for $\Delta R \gtrsim 18$ nm, native contacts vanish, this peak is associated with an abrupt decrease in non-native contacts between strands B and C. Its nature may be also understood by monitoring the dependence of HBs on ΔR [Fig. 8(d)], which shows that the last maximum is caused by loss of HBs of these strands. More precisely, five HBs between B and C, which were not present in the native conformation, are broken [Fig. 9(a)]. Interestingly, these bonds appear at $\Delta R \gtrsim 15$ nm, i.e., after the second unfolding event [Fig. 9(a)]. Thus, our study cannot only reproduce the experimentally observed peak at $\Delta R \approx 22$ nm, but also shed light on its nature on the molecular level. From this perspective, all-atom simulations are superior to experiments.

One corollary from Fig. 8 is that one cannot provide a complete description of the unfolding process based on the evolution of only native contacts. It is because, as a molecule extends, its secondary structures change and new non-native secondary structures may occur. Beyond the extension of 17–18 nm (see snapshot at bottom of Fig. 3), e.g., the protein lost all native contacts, but it does not get a extended state

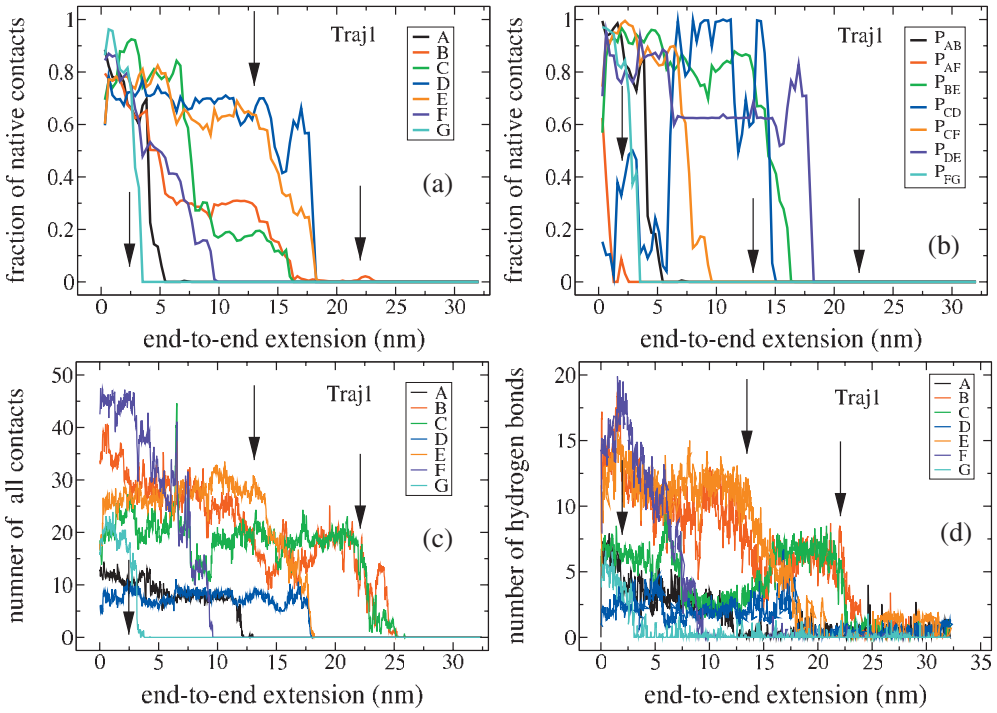

FIG. 8. (a) Dependence of the number of native backbone contacts formed by individual strands on ΔR for $v=10^9$ nm/s. Arrows refer to positions of three peaks in the force-extension curve. (b) The same as in (a) but for pairs of strands. (c) The same as in (a) but for all contacts (native and non-native). (d) The same as in (c) but for HBs.

without any structures. Therefore, a full description of mechanical unfolding may be obtained by monitoring either all backbone contacts or HBs, as these two quantities give the same unfolding picture [Figs. 8(c) and 8(d)].

As said above, for trajectory 1, the third peak occurs due to breaking non-native HBs between strands B and C, but this mechanism is not universal for all trajectories we have studied. In the case of second trajectory, it is associated with breaking of the non-native HBs between strands C and E [Fig. 9(b)], which have been created during stage III [Fig. 5(a)]. After the rupture of the non-native HBs, a partial recovery of the D and E strands at $\Delta R \sim 24$ nm has been observed. However, the further unfolding of refolded pieces of these strands does not affect the force-extension profile significantly and its influence is much less pronounced compared non-native HBs between C and E [Fig. 9(b)].

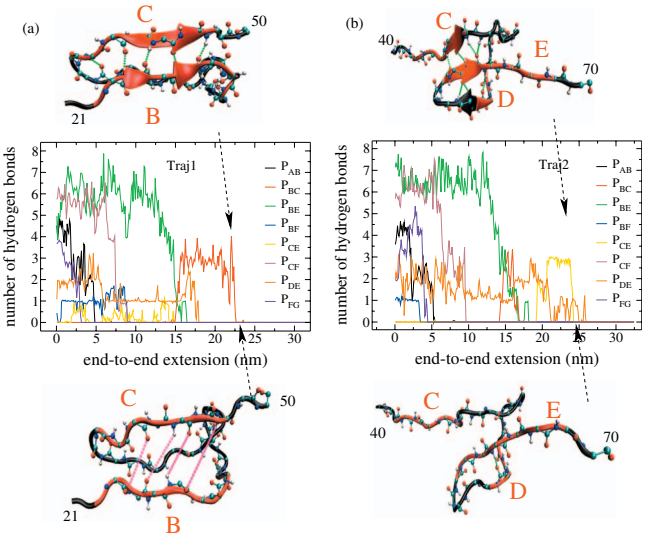

FIG. 9. (a) Dependence of the number of HBs between pairs of strands for the first trajectory and $v=10^9$ nm/s. Non-native HBs between strands B and C start to appear at $\Delta R \approx 15$ nm (red curve). Their rupture leads to the maximum centered at $\Delta R \approx 22.4$ nm. Upper snapshot shows five HBs between B an C (green) before the third unfolding event. Lower snapshot is a fragment after the third peak, where all HBs are already broken (purple dotted lines). (b) The same as in (a), but for trajectory 2. The upper snapshot shows the HBs between strands C and E before their rupture. Just after the last unfolding event, one HB between D and E also survives (lower snapshot).

For the first two trajectories, it was enough to consider the evolution of native and non-native HBs between the strands formed in the native state. However, for the third and fourth trajectories, determining the nature of the third peak requires consideration of additional β-strands that appear during the unfolding process. In the case of trajectory 3, the native strand D has been extended for three amino acids more (from amino acid 54 to 59) and this extension is labeled as D^+. In the upper snapshot of Fig. 10(a), this non-native part is shown in black. The red curve in this figure clearly shows that the third peak appears due to the breaking of four non-native HBs between the C and D^+ strands. After this unfolding event HBs between all pairs are ruptured.

For trajectory 4, HBs between all native strands were broken before the protein reaches $\Delta R \approx 15$ nm [Fig. 10(b)]. The nature of the third peak is purely related to the rupture of non-native HBs between strand B and the non-native strand B^+ (from amino acid 17 to 21) which is described by the black arrow [upper snapshot of Fig. 10(b)]. Thus, the non-native interactions are responsible for occurrence of the last peak, but individual pathways show a rich diversity.

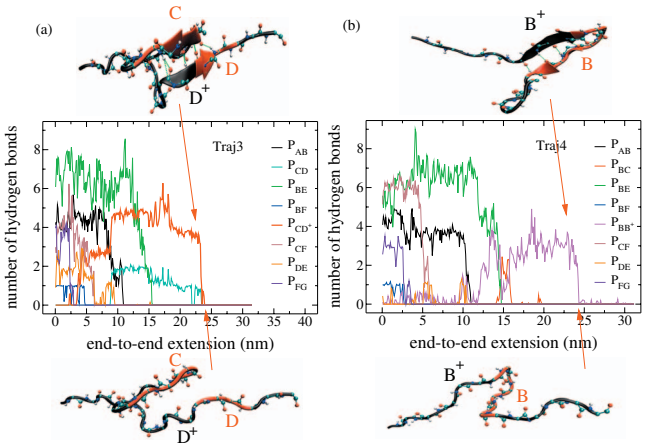

FIG. 10. (a) The same as in Fig. 9, but for trajectory 3. The upper snapshot shows HBs between C and the non-native strand D^+ (black) which includes amino acids 54–59. The rupture of these bonds (lower snapshot) leads to the occurrence of the third peak. (b) The same as in (a), but for trajectory 4. Here the nature of the last peak is related to breaking of HBs between B and the non-native strand B^+ (black, amino acids 17–21) as shown in the upper snapshot.

D. Unfolding pathways: N-fixed case

To obtain sequencing of unfolding events at $v=10^9$ nm/s, we use dependencies of the number of HBs on ΔR. From Figs. 8(d) and 11, we have the following unfolding pathways for four trajectories:

$$\begin{aligned} &G \to F \to A \to (D,E) \to (B,C), \quad \text{trajectory } 1, \\ &G \to F \to A \to B \to C \to (D,E), \quad \text{trajectory } 2, \\ &G \to F \to A \to E \to B \to D \to C, \quad \text{trajectory } 3 \\ &G \to F \to A \to (D,E) \to C \to B, \quad \text{trajectory } 4. \end{aligned} \tag{1}$$

Although these pathways are different, they share a common feature that the C-terminal unfolds first. This is consistent with the results obtained by Go simulations at high pulling speeds $v \sim 10^6$ nm/s,[24] but contradicts the experiments,[6,7] which showed that strands A and B from the N-termini unfold first. On the other hand, our Go simulations[24] revealed that the agreement with the experimental results is achieved if one performs simulations at relatively low pulling speeds $v \sim 10^4$ nm/s. Therefore, one can expect that the difference in sequencing of unfolding events between present all-atom results and the experimental ones is merely due to large values of v we used. In order to check this, one has to carry out all-atom simulations, at least, at $v \sim 10^4$ nm/s, but such a task is far beyond current computational facilities.

Schwaiger *et al.*[6,7] suggested that conformations of the single unfolding intermediate which corresponds to the first peak in the force-extension curve contain five strands C–G with A and B fully unstructured. As follows from our all-atom simulations, there exist two intermediates related to peaks located at $\Delta R \sim 2$ and 12 nm. Since at $\Delta R \sim 2$ nm, only G unfolds [Figs. 4(c), 8(d), and 11], the structure of the first intermediate consists of six structured strands A–F which is more ordered than the experimental one. The second intermediate, located at $\Delta R \sim 12$ nm, is less ordered having four strands B–E structured [Figs. 3, 5, 6, 8(d), and 11]. Again, a huge difference in pulling speeds is a reason for differences in the nature of experimentally and theoretically predicted intermediates.

It is very important to note that unfolding pathways depend on pulling speeds but the number of peaks in the force-extension curve remains unchanged. If non-native interactions are neglected as in the Go model then one has two peaks,[24] but the inclusion of these interactions leads to occurrence of the third maximum at $\Delta R \sim 22$ nm (Fig. 7).

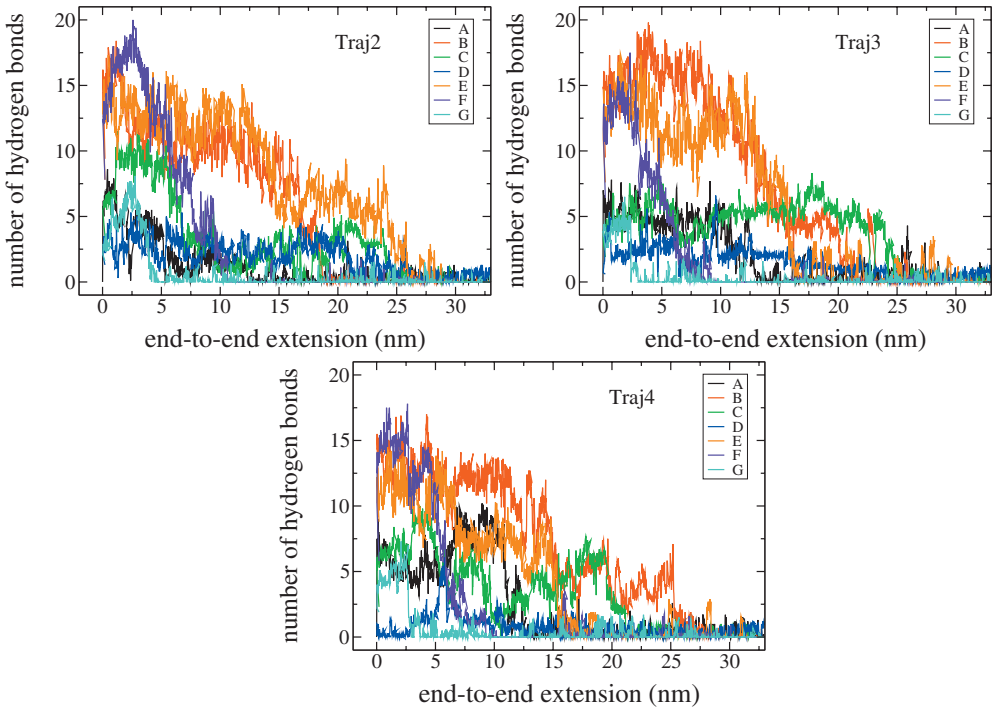

FIG. 11. The end-to-end extension dependence of the number of HBs, formed by seven strands, for trajectories 2, 3 and 4. $v=10^9$ nm/s.

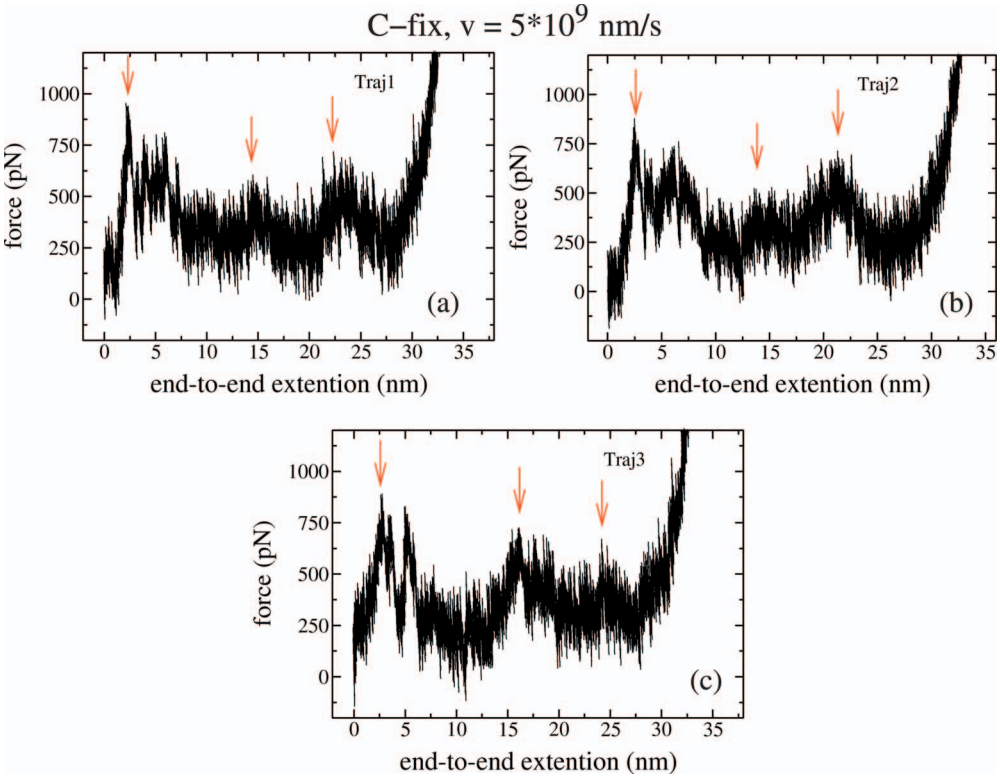

FIG. 12. The force-extension profiles, obtained in three molecular dynamics runs with the C-terminal kept fixed. $v=5\times10^{9}$ nm/s. Arrows refer to positions of maxima.

Thus, the main meaning of the present work is not in describing unfolding pathways but in capturing this peak.

E. Unfolding pathways: C-fixed case

It is known that unfolding pathways may depend on what end of a protein is kept fixed. For ubiquitin, pulling at the C-terminal gives pathways different from those obtained in the case when the N-end is pulled.[28] However, unfolding pathways of the domain I27 do not depend on what terminal is fixed.[28] To study the effect of terminal fixation on unfolding pathways of DDFLN4, we performed limited simulations at $v=5\times10^{9}$ nm/s for three trajectories. As in the N-fixed case, the force-extension curves display three peaks (Fig. 12). Thus, the number of peaks does not depend on what terminal is immobilized.

Figure 13 shows the dependence of HBs on ΔR for trajectory 1. Before the last unfolding event, there exist only few non-native HBs between strands C and E (upper snapshot). The lower snapshot shows that after the third peak these bonds got ruptured. One can show that the third peak in the second and third trajectories also comes from the rupture of non-native HBs (results not shown). Thus, the nature of the third peak remains essentially the same as in the N-fixed case.

From the dependencies of the total number of backbone contacts formed by secondary structures (Fig. 14), we obtain the following unfolding pathways:

$$\begin{aligned} &\mathrm{A}\rightarrow\mathrm{G}\rightarrow\mathrm{F}\rightarrow\mathrm{B}\rightarrow(\mathrm{C},\mathrm{E})\rightarrow\mathrm{D}, \quad \text{trajectory } 1,\\ &\mathrm{A}\rightarrow\mathrm{G}\rightarrow\mathrm{B}\rightarrow\mathrm{F}\rightarrow(\mathrm{C},\mathrm{D},\mathrm{E}), \quad \text{trajectory } 2, \\ &\mathrm{A}\rightarrow\mathrm{B}\rightarrow\mathrm{G}\rightarrow(\mathrm{C},\mathrm{F})\rightarrow(\mathrm{E},\mathrm{D}), \quad \text{trajectory } 3. \end{aligned} \tag{2}$$

In three cases, the strand A is unfolds first [Eq. (2)]. For the third trajectory A and B unfold first and this scenario is in accord with the experiments of Schwaiger *et al.*[6,7] However, such an agreement is fortuitous as it happens just in one case. As evident from Eqs. (1) and (2), unfolding pathways depend on what terminal is pulled. Intuitively, it is clear because at high pulling speeds, a terminal, which the external force is applied to, should unfold first as it unfolds before the force propagates to the opposite end.

One can infer that the first intermediate, which corresponds to the first peak, consists of conformations with only strand A unfolded (Figs. 12 and 14). This is valid for all three trajectories. Structures of the second intermediate, related to the second unfolding event, show more diversity. For the first trajectory, it consists of four structured strands B–E [Fig. 14(a)]. After the second peak, A, F, and G become unstructured. The second and third trajectories have the same second intermediate which contains four strands C–F [Figs. 14(b) and 14(c)]. Overall, as in the N-fixed case, due to high pulling speeds, the simulations give different unfolding pathways compared to the experimental ones. Nevertheless, the third peak, which is absent in the Go modeling, occurs in both N-fixed and C-fixed all-atom simulations.

The question we now ask is whether the dependence of unfolding pathways on the terminal fixation is intrinsic for

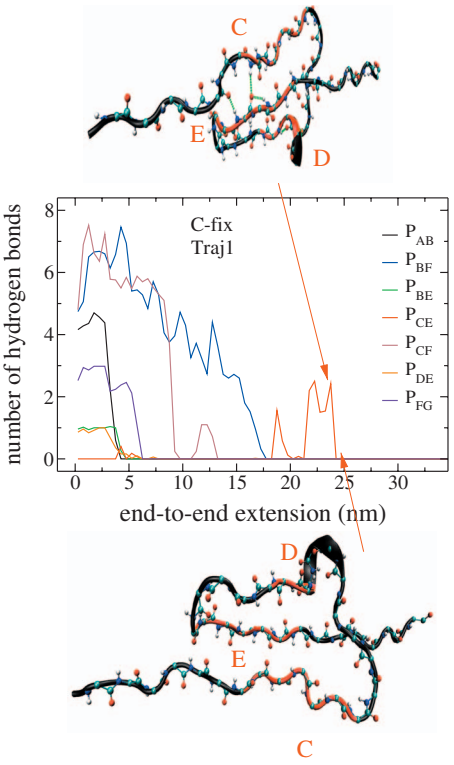

FIG. 13. Dependence of the number of HBs between pairs of strands for the first trajectory. The pulling speed $v=5\times10^9$ nm/s and the C-end is hold fixed. The upper snapshot shows non-native HBs between strand C and E before their rupture. After passing the third peak they are broken (lower snapshot).

DDFLN4 or this is an artifact of high loading rates. Since it is impossible to carry out all-atom simulations at small v, we performed the simulations at $v\sim10^4$ nm/s using the Go model[8] with the C-end hold fixed. The results, obtained for the N-fixed case, were reported previously.[24] It turns out that at low pulling speeds unfolding pathways are the same as in the N-fixed case (results not shown). Therefore, we speculate that, similar to the domain I27,[28] at low loading rates unfolding pathways of DDFLN4 do not depend on the way one pulls it. Our speculation is not unreasonable because for small v, it does not matter what end is pulled as the external force is uniformly felt along a chain. Then, a strand, which has a weaker link with the core, would unfold first.

F. Unfolding intermediates and pathways at low pulling speeds

Because all-atom simulations at low v are prohibited, in this section, we try to infer information on unfolding intermediates and pathways using the energetic argument and results followed from the Go model.[24] As shown in our previous work,[24] the peak, centered at $\Delta R\approx2$ nm, results from breaking of contacts between A and F. Moreover, within the Go modeling, contacts between these strands break first not only at low but also at high loading rates. Our high-v all-atom results [Fig. 8(b) and similar results for other cases are not shown] also confirm this. In order to see if (A,F) native contacts break first at low loading rates, we calculated the interaction energies between terminal pairs (A,B), (A,F), and (F,G) using the Gromos force field 43a1 and equilibrium conformations. The latter are those conformations which have been obtained after the equilibration step and used as initial conformations for pulling. At the room temperature, they are very close to the native conformation (RMSD ≈0.7 Å). Since the interaction energies $E_{AB}\approx-120.4$, $E_{AF}\approx-12.2$, and $E_{FG}\approx-114.4$ kcal/mol, one can expect that (A,F) contacts always break first. Therefore, in the first intermediate state, all strands remain structured but the contacts between A and F are lost. As mentioned above, the absence of the first peak in experiments may be due to the strong linker effect. Another possible reason is that one needs a relatively small amount of energy to break (A,F) couplings. This subtle effect may be overlooked in the experiments.

At low v, the Go simulations[24] show that A and B unfold first and this pathway agrees very well with the experiments of Schwaiger *et al.*[6,7] Consequently, the corresponding theoretical intermediate, which consists of five strands C–G, is also identical to the experimental one. Since the second peak at $\Delta R\approx11$ nm is caused by native interactions, its nature would be the same as in the Go model. Thus, we expect that at low pulling speeds, the second intermediate provided by the Gromos force field coincides with the experimental one.

G. Robustness of the results against other force fields

To make sure that the results obtained by the GROMOS96 force field 43a1 are robust, we also performed a limited set of simulations for $v=5\times10^9$ nm/s using the Amber 94 (Ref. 11) and OPLS (Ref. 12) force fields. Figure 15 shows two typical force-extension curves obtained by these force fields (other similar curves are not shown) for the pulling speed $v=5\times10^9$ nm/s and the N-terminus is hold fixed. In accordance with the GROMOS96 force field 43a1 case, the Amber 94 and OPLS force fields provide three peaks and the nature of the third peak is also associated with non-native interactions (results not shown). Thus, the existence of three peaks and their nature do not depend on the force fields we used and one can expect that this conclusion remains valid for other existing force fields.

IV. CONCLUSIONS

We used the GROMOS96 force field 43a1 with explicit water to understand the mechanical unfolding of the protein domain DDFLN4. The validity of this approach was carefully checked not only by considering the well-studied titin domain I27 and the domain DDFLN5 which is mechanically more stable than DDFLN4, but also by comparison with results obtained by Amber 94 and OPLS force fields. We reproduced the experimental result on existence of two peaks located at $\Delta R\approx11$ and 22 nm. Our key result is that the later maximum occurs due to breaking of non-native HBs. It cannot be encountered by the Go models in which non-native interactions are neglected.[4,24] Thus, our result points to the importance of these interactions for the mechanical unfolding of DDFLN4. For the first time we have shown that the description of elastic properties of proteins may be not complete ignoring non-native interactions. This finding is valu-

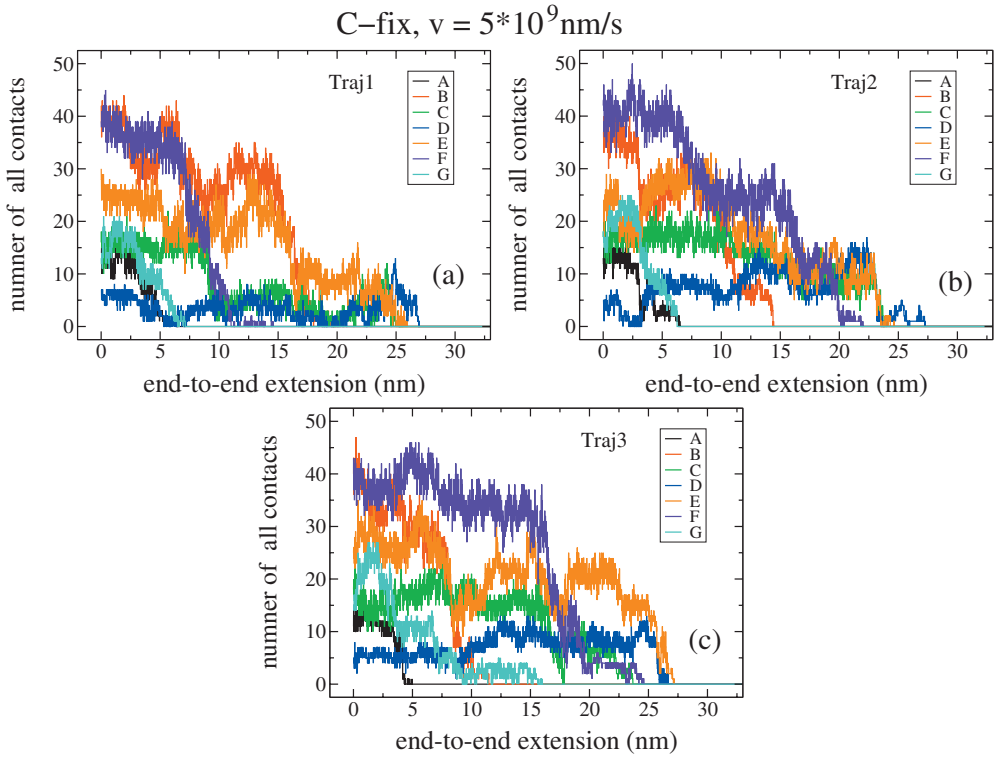

FIG. 14. Dependence of the total number of all backbone contacts (native and non-native), formed by seven strands, on ΔR. $v=5\times10^{9}$ nm/s and the C-end is immobile.

able as the unfolding by an external force is widely believed to be solely governed by native topology of proteins.

Our all-atom simulation study supports the result obtained by the Go model[4,24] that an additional peak occurs at $\Delta R\approx2$ nm. However, it was not observed by the AFM experiments of Schwaiger *et al.*[6,7] In order to solve this controversy, one has to carry out not only additional experiments, but also all-atom simulations at lower pulling speeds which are nowadays beyond computational facilities.

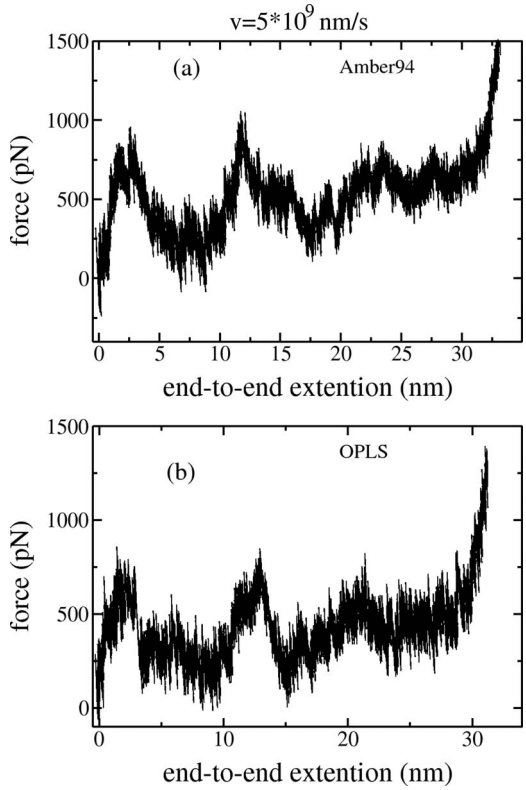

FIG. 15. Shown is the force-extension profile obtained by using the Amber 94 (upper panel) and OPLS (lower panel) force fields. The N-end is kept fixed and the pulling speed $v=5\times10^{9}$ nm/s.

ACKNOWLEDGMENTS

The work was supported by the Ministry of Science and Informatics in Poland (Grant No. 202-204-234), National Science Council in Taiwan under Grant No. NSC 96-2911-M 001-003-MY3, National Center for Theoretical Sciences in Taiwan, and Academia Sinica (Taiwan) under Grant No. AS-95-TP-A07. M.K. is very grateful to the Polish committee for UNESCO for the financial support.

[1] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, Science **276**, 1109 (1997).
[2] L. Tskhovrebova, K. Trinick, J. A. Sleep, and M. Simons, Nature (London) **387**, 308 (1997).
[3] C. Bustamante, Y. R. Chemla, N. R. Forde, and D. Izhaky, Annu. Rev. Biochem. **73**, 705 (2004).
[4] M. S. Li, A. M. Gabovich, and A. I. Voitenko, J. Chem. Phys. **129**, 105102 (2008); S. Kumar and M. S. Li, Phys. Rep. (2009), doi: 10.1016/j.physrep.2009.11.001.
[5] T. P. Stossel, J. Condeelis, L. Cooley, J. H. Hartwig, A. Noegel, M. Schleicher, and S. Shapiro, Nat. Rev. Mol. Cell Biol. **2**, 138 (2001).
[6] I. Schwaiger, A. Kardinal, M. Schleicher, A. A. Noegel, and M. Rief, Nat. Struct. Mol. Biol. **11**, 81 (2004).
[7] I. Schwaiger, M. Schlierf, A. Noegel, and M. Rief, EMBO Rep. **6**, 46 (2005).
[8] C. Clementi, H. Nymeyer, and J. N. Onuchic, J. Mol. Biol. **298**, 937 (2000).
[9] W. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. Scott, and I. Tironi, Vdf Hochschulverlag AG an der ETH, Zurich, 1996.

[10] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, 1981), Vol. 331.
[11] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
[12] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).
[13] D. K. West, D. J. Brockwell, P. D. Olmsted, S. E. Radford, and E. Paci, Biophys. J. **90**, 287 (2006).
[14] M. S. Li, Biophys. J. **93**, 2644 (2007).
[15] H. Berendsen, D. van der Spoel, and R. van Drunen, Comput. Phys. Commun. **91**, 43 (1995).
[16] E. Lindahl, B. Hess, and D. van der Spoel, J. Mol. Model. **7**, 306 (2001).
[17] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, J. Comput. Chem. **18**, 1463 (1997).
[18] T. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
[19] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Dinola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[20] J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, and D. E. Shaw, Curr. Opin. Struct. Biol. **19**, 120 (2009).
[21] P. H. Nguyen, G. Stock, E. Mittag, C. K. Hu, and M. S. Li, Proteins **61**, 795 (2005).
[22] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, J. Comput. Chem. **26**, 1781 (2005).
[23] H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten, Biophys. J. **75**, 662 (1998).
[24] M. S. Li and M. Kouza, J. Chem. Phys. **130**, 145102 (2009).
[25] P. E. Marszalek, H. Lu, H. Li, M. Carrion-Vazquez, A. F. Oberhauser, K. Schulten, and J.-M. Fernandez, Nature (London) **402**, 100 (1999).
[26] E. Evans and K. Ritchie, Biophys. J. **72**, 1541 (1997).
[27] W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graphics **14**, 33 (1996).
[28] M. S. Li, M. Kouza, and C. K. Hu, Biophys. J. **92**, 547 (2007).