POLITECNICO DI TORINO
Repository ISTITUZIONALE

Pathways of mechanical unfolding of FnIII10: Low force intermediates

Original
Pathways of mechanical unfolding of FnIII10: Low force intermediates / Caraglio, M., Imparato, A., Pelizzola, A.. - In: THE JOURNAL OF CHEMICAL PHYSICS. - ISSN 0021-9606. - 133:(2010), p. 065101. [10.1063/1.3464476]

Availability:
This version is available at: 11583/2371647 since:

Publisher:
American Institute of Physics

Published
DOI:10.1063/1.3464476

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

26 November 2026

Pathways of mechanical unfolding of FnIII$_{10}$: Low force intermediates

M. Caraglio,[1,a)] A. Imparato,[2,b)] and A. Pelizzola[1,c)]
[1]*Dipartimento di Fisica and CNISM, Politecnico di Torino, c. Duca degli Abruzzi 24, 10129 Torino, Italy and INFN, Sezione di Torino, Torino, Italy*
[2]*Department of Physics and Astronomy, University of Aarhus, Ny Munkegade, Building 1520, DK-8000 Aarhus C, Denmark*

(Received 12 April 2010; accepted 16 June 2010; published online 12 August 2010)

We study the mechanical unfolding pathways of the FnIII$_{10}$ domain of fibronectin by means of an Ising-like model, using both constant force and constant velocity protocols. At high forces and high velocities our results are consistent with experiments and previous computational studies. Moreover, the simplicity of the model allows us to probe the biologically relevant low force regime, where we predict the existence of two intermediates with very close elongations. The unfolding pathway is characterized by stochastic transitions between these two intermediates. © *2010 American Institute of Physics*. [doi:10.1063/1.3464476]

I. INTRODUCTION

The mechanical unfolding of biopolymers has been the subject of an intense research activity, both experimental and theoretical, in the past 2 decades. For a recent review, see Ref. 1. Innovative single molecule experimental techniques, mainly based on atomic force microscopy (AFM) and optical tweezers, have been used to investigate the response of biopolymers to controlled forces, while theoretical and computational models at different levels of coarse graining have been proposed and investigated.

Among the various molecules studied, fibronectin is particularly important due to its role in tissue elasticity, cell adhesion, and cell migration.[2] Its tenth type III module (FnIII$_{10}$) is known to be crucial for cell adhesion, through the binding of its arginine-glycine-aspartic acid (RGD) motif to transmembrane integrin receptors. The secondary structure of this module consists of two antiparallel β-sheets forming a β-sandwich (Fig. 1). The β-strands are usually denoted with letters from A (the strand closest to the N terminal) to G (the C terminal one). The two sheets are made of strands ABE and DCFG, respectively, and the RGD motif is in the loop separating strands F and G.

The mechanical unfolding of FnIII$_{10}$ has been studied both experimentally[3,4] and by computer simulations.[5–9] Single molecule AFM experiments have shown that FnIII$_{10}$ has a low mechanical stability, compared to other fibronectin type III domains.[3] Furthermore, AFM experiments by the same group[4] showed that FnIII$_{10}$ can unfold according to different pathways. Apparent two-state transitions were observed, as well as unfolding through intermediate states. Experiments on suitable mutants suggested the possible existence of two different intermediate states, which is also consistent with some simulations,[6,8] while other simulations predicted simpler[5,7] or more complex[9] scenarios.

In the present paper we shall study the mechanical unfolding of FnIII$_{10}$ by means of a generalized Ising-like model we have recently proposed.[10–13] The model has already been shown[10,11] to reproduce the general features of mechanical unfolding experiments, such as the force dependence of the average unfolding time in a constant force protocol, or the rate dependence of the unfolding force in a constant rate protocol, together with the corresponding probability distributions. The same model turned out to predict the correct values for the unfolding lengths of a titin domain[10,11] and of ubiquitin.[12] Moreover, it has been used to investigate the unfolding pathways of ubiquitin[12] and of a 236-base RNA fragment,[13] and the resulting pathways turned out to be consistent with both experimental and computational results, where more detailed molecular models were used.

In the case of FnIII$_{10}$ we are particularly interested in exploring the biologically relevant low force regime,[2,14,15] which is thought to be close to the equilibrium unfolding force and cannot be explored by simulations of more detailed, and more computationally expensive, molecular models. Our model can probe forces close to the equilibrium unfolding force, whose value we use to set our force unit. Such value is unfortunately not exactly known. Erickson[14] estimated the equilibrium unfolding force to be at most 5 pN, on the basis of an order of magnitude calculation. On the other hand, an estimate close to 20 pN was reported in Ref. 9. Choosing the value of 20 pN, in order to set our force unit, our results give unfolding forces in very good agreement with the AFM experiments (see Sec. IV).

The paper is organized as follows: In Sec. II we shall describe our model and simulation techniques; some equilibrium results are discussed in Sec. III; in Sec. IV we shall describe our results for the unfolding pathways, using both constant force and constant velocity protocols; finally, in Sec. V we shall draw some conclusions.

II. MODEL AND METHODS

A. Model

We represent the polypeptide chain pulled by an external force through a simple Ising-like model[10–13] that is a gener-

[a)]Electronic mail: michele.caraglio@polito.it.
[b)]Electronic mail: imparato@phys.au.dk.
[c)]Electronic mail: alessandro.pelizzola@polito.it.

 © 2010 American Institute of Physics

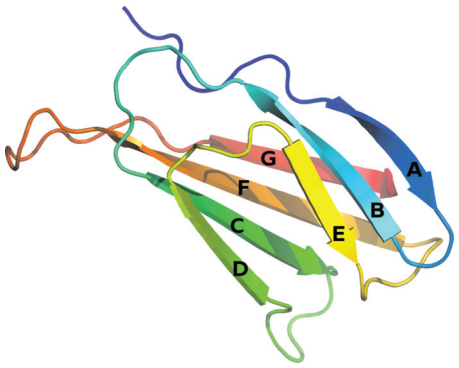


FIG. 1. Sketch of the native structure of $FnIII_{10}$ (Protein Data Bank ID 1ttf) with β-strands labeled A–G in sequence order. Figure generated by PyMOL.

alization of the Wako–Saitô–Muñoz–Eaton (WSME) Gō-type model for protein folding.[16–28] In this model a $(N+1)$ residue polypeptide chain is described by N binary variables m_k $(k=1,\ldots,N)$ associated to the peptide bonds. The variable m_k is equal to 1 if the kth bond is in a native state and 0 if the bond is in an unfolded state.

In the following a *native stretch* will indicate a sequence of consecutive amino acids connected by native bonds and delimited by two non-native bonds. In order to characterize the state of such a stretch, we introduce the quantity

$$S_{ij} \equiv (1-m_i)\left(\prod_{k=i+1}^{j-1} m_k\right)(1-m_j), \tag{1}$$

with $(0 \le i < j \le N+1)$ and the boundary conditions $m_0 = m_{N+1} = 0$. Thus, a native stretch delimited by bonds i and j is characterized by $S_{ij}=1$, while if the sequence is not a native stretch, $S_{ij}=0$. In the case of $j=i+1$ the stretch corresponds to the single $(i+1)$th residue. Therefore, given a certain configuration $m=\{m_k\}$ of the chain, the number M of native stretches is equal to $M=1+\Sigma_{i=1}^{N}(1-m_i)$.

In the WSME model two amino acids can interact only if they are in contact [i.e., if they have at least a pair of non-hydrogen atoms closer than 4 Å in the native structure deposited in the Protein Data Bank (PDB)] and if they belong to the same stretch. The effective Hamiltonian reads

$$H(m,q) = \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \epsilon_{ij}\Delta_{ij}\prod_{k=i}^{j} m_k - k_B T\sum_{i=1}^{N} q_i(1-m_i), \tag{2}$$

where ϵ_{ij} is the energy gain associated to the contact between ith and $(j+1)$th amino acids (see next subsection for details), Δ is called contact matrix and its (i,j) element takes the value $\Delta_{ij}=1$ if such a contact exists in the native structure while $\Delta_{ij}=0$ otherwise or if $j=i+1$ (since if the amino acids are so close in the chain sequence, they have pair of atoms closer than 4 Å even in the unfolded state). The quantity $q_i > 0$ is the entropic cost of ordering bond i. In the following we define $h_{ij}=\epsilon_{ij}\Delta_{ij}$. In Ref. 23 it has been discussed how to compute exactly the partition function of the WSME model through a transfer-matrix formalism.

In our generalized model we substitute the entropic term for the coupling to the force described by a potential energy function V depending on the end-to-end length of the protein $\mathcal{L}$,

$$H(m,\mathcal{L}) = \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} h_{ij}\prod_{k=i}^{j} m_k + V(\mathcal{L}). \tag{3}$$

In our simulations the protein is pulled either by a constant force or at constant pulling velocity. In the first case the force f applied to the protein ends is constant and the potential energy takes the form $V(\mathcal{L})=-f\mathcal{L}$, while in the second case the energy is time-dependent: $V(\mathcal{L})=(k/2)(\mathcal{L}_0+vt-\mathcal{L})^2$, where k is a spring constant, v is the pulling velocity, and $\mathcal{L}_0$ is the initial equilibrium elongation.

In order to define the length $\mathcal{L}$ we assume that each stretch $(S_{ij}=1)$ can be only parallel or antiparallel to the direction of the applied force and we implement such an assumption through a new binary variable σ_{ij} that can take the values $+1$ or -1, respectively. The end-to-end length of the protein is defined as the sum of the lengths l_{ij} of each stretch multiplied by σ_{ij},

$$\mathcal{L}(m,\sigma) = \sum_{i=0}^{N}\sum_{j=i+1}^{N+1} l_{ij}\sigma_{ij}S_{ij}, \tag{4}$$

where the dynamic variable σ is a set of M variables σ_{ij}, each one associated to a stretch. Since the ith amino acid is represented by the sequence of its N_i nitrogen, $C_{\alpha,i}$ central carbon, and C_i carbon atoms, the lengths l_{ij} are obtained from the PDB structure as the native distance between the midpoint of the C_{i-1} and N_i atoms and the midpoint of the C_j and N_{j+1} atoms.

Dealing with the case of constant force, since the variables σ_{ij} do not interact among themselves, it is possible to obtain an effective Hamiltonian which has the same structure of Hamiltonian (2) of the initial model and therefore the equilibrium thermodynamics is exactly solvable also in this case. In fact, given the Hamiltonian

$$H(m,\sigma;f) = \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} h_{ij}\prod_{k=i}^{j} m_k - f\sum_{i=0}^{N}\sum_{j=i+1}^{N+1} l_{ij}\sigma_{ij}S_{ij}, \tag{5}$$

we can perform the sum on the σ variables in the partition function,

$$\mathcal{Z}(f) = \sum_{\{m\}}\sum_{\{\sigma\}} e^{-\beta H(m,\sigma;f)} = \sum_{\{m\}} e^{-\beta H_{\mathrm{eff}}(m;f)}, \tag{6}$$

with

$$H_{\mathrm{eff}}(m;f) = \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} h_{ij}\prod_{k=i}^{j} m_k - \frac{1}{\beta}\sum_{i=0}^{N}\sum_{j=i+1}^{N+1} \ln[2\cosh(\beta f l_{ij})]S_{ij}. \tag{7}$$

In the case of force $f=0$ the last expression reduces to Eq. (2) with $q_i=\ln 2$ for every i.

B. Model parameters, simulation, and analysis

The parameters ϵ_{ij}, as in Refs. 10–12, 20, 23, and 28, are taken equal to $n\epsilon$ where n is an integer such that $5(n-1) < n_{\mathrm{at}} \le 5n$, and n_{at} is the number of pairs of atoms in contact (that is closer than 4 Å) between ith and $(j+1)$th amino acids. The temperature is set to $T=0.768T_m$, where T_m is the equilibrium unfolding temperature at zero force. Since ex-

perimentally $T_m=375$ K,[29] we have $T=288$ K. The force unit is then set in such a way that the equilibrium typical unfolding force at $T=288$ K is 20 pN. Since an experimental measurement of this quantity is missing, it has been chosen on the basis of the estimates reported in Ref. 9. A detailed discussion about the choice of energy and force scales in the model has been reported in Ref. 12.

The nonequilibrium unfolding kinetics have been studied by Monte Carlo (MC) simulations. More precisely, in the framework of a master equation approach,[25] we choose transition rates according to the Metropolis algorithm. Rigorously speaking, this choice cannot be derived from an underlying microscopic dynamics of the molecule. Nevertheless, it has been shown[10–13] that it reproduces many quantitative and qualitative aspects of folding and unfolding of real molecules under an external force. A single MC step consists of a single-bond flip on the variable m_j, chosen with equal probability among the N peptide bond variables, followed by a single-spin flip on the variable σ_{ij}, also chosen with uniform probability among the M stretch orientational variables.[30] In Sec. IV, by comparing our estimated zero-force unfolding time with the corresponding experimental value, we shall find that a MC step corresponds to about 25 ns.

Simulations have been run with nine values of the force (122, 98, 81, 65, 53, 46, 40, 36, and 28 pN) and six constant pulling velocities (0.03, 0.05, 0.1, 0.3, 0.5, and 1 μm/s), for each value of the force or of the velocity 100 different unfolding trajectories have been considered.

Each simulation stops 10^5 MC steps after the protein reaches the value $L_u=\frac{1}{2}L_{\max}=\frac{1}{2}\Sigma_{i=0}^{N}l_{i,i+1}$. An exception is the case $f=28$ pN, where we take $L_u=\frac{2}{3}L_{\max}$, because of the larger length fluctuations and in order to prevent the trajectory ending before a complete unfolding event takes place.

In order to trace unfolding pathways, we use the weighted fraction of native contacts as order parameter,

$$\phi_s=\frac{\Sigma_{i=r_1(s)}^{r_2(s)-2}\Sigma_{j=i+1}^{r_2(s)-1}h_{ij}\Pi_{k=i}^{j}m_k}{\Sigma_{i=r_1(s)}^{r_2(s)-1}\Sigma_{j=i+1}^{r_2(s)}h_{ij}}, \tag{8}$$

where s is the string of bonds we are analyzing, and $r_1(s)$ and $r_2(s)$ are its first and last peptide units. As an example the string containing strands A and B has $r_1(\mathrm{AB})=6$ and $r_2(\mathrm{AB})=23$. A straightforward generalization is necessary for order parameters of strings of non-consecutive strands (i.e., C–F and B–E). ϕ_s turns out to be a better order parameter than the fraction of native bonds used in previous works because of its greater stability with respect to fluctuations. When discussing the folded or unfolded character of an individual β-strand, appropriate order parameters can be identified on the basis of the secondary structure. As an example, strand F appears in a β-sheet between strands C and G, which suggests to use ϕ_{CF} and ϕ_{FG} as order parameters for strand F.

III. EQUILIBRIUM PROPERTIES

As mentioned before, the equilibrium thermodynamics can be solved exactly in our model and we can thus follow the macroscopic state behavior of the protein at different pulling forces. In Ref. 31 the average fraction of native bonds and end-to-end length are plotted as functions of the force f.

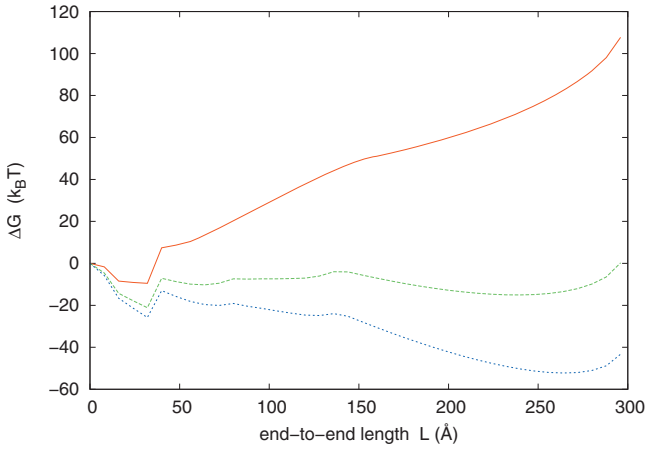


FIG. 2. Free energy landscape at temperature $T=288$ K and for forces $f=0$ pN (red line), $f=20$ pN (green line), and $f=28$ pN (blue line). $\Delta G=G(L)-G(0)$.

To obtain the equilibrium energy landscape of the protein as a function of the reaction coordinate L we expand, following Refs. 11 and 23, the partition function (6) in powers of $e^{\beta f}$ as

$$\mathcal{Z}(f)=\sum_{L=-L_{\max}}^{L=L_{\max}}Z_0(L)e^{\beta fL},$$

where $Z_0(L)$ is a zero force partition function constrained at the length value L,

$$Z_0(L)=\sum_m\sum_\sigma\delta(L-L(m,\sigma))e^{-\beta H(m,\sigma;f=0)},$$

which, as $\mathcal{Z}(f)$, can be computed exactly. The corresponding free energy reads

$$F_0(L)=-k_BT\ln Z_0(L).$$

In presence of a constant force, the free energy landscape is tilted and is given by $G(L)=F_0(L)-fL$. Figure 2 shows the landscape for various forces: At zero force there is just one minimum at about 3.5 nm corresponding to the folded state. By increasing the force three more minima appear: Two of them (end-to-end lengths of about 6 and 13 nm) are always local minima, and will be later associated to intermediate states, while the third one, corresponding to the fully unfolded state, becomes the global minimum when the force exceeds 20 pN.

IV. UNFOLDING PATHWAYS

A. Force clamp

In the force clamp protocol the molecule is first equilibrated in absence of force, then at $t=0$ the force instantaneously jumps to a nonvanishing constant value, which ranges between 28 and 122 pN. Notice that the forces we use are much closer to the equilibrium unfolding force, and hence to *in vivo* conditions, than most previous works since more detailed models can be simulated only for very short

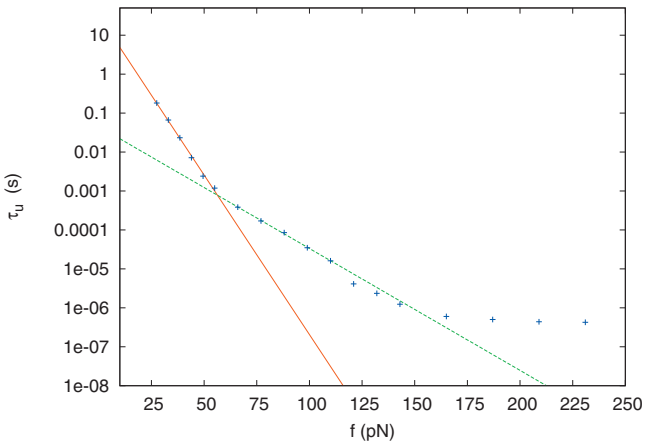


FIG. 3. Mean unfolding time τ_u as a function of the force f applied to the molecule (average over 100 different trajectories). The red line is a fit to Arrhenius' law in the range of forces from 25 to 60 pN. In this range we find from the fit $x_u=3.4\pm0.1$ Å. The green line is a fit from 60 to 115 pN, $x_u=1.3\pm0.1$ Å.

time intervals. The smallest force probed by Karplus and Paci[6] was 69 pN, and they did not observe any unfolding event at this force, while Gao *et al.*[8] used forces not smaller than 400 pN. Only in the all-atom Monte Carlo simulations by Mitternacht *et al.*[9] unfolding events at constant forces as small as 50 pN could be observed.

In Fig. 3 we report the average unfolding time τ_u as a function of force f. Three regimes are clearly distinguishable. In the high force regime the unfolding time saturates to a constant plateau, as observed for several other proteins.[1] In the low force regime (25–60 pN) we have made a fit to Arrhenius' law,

$$\tau_u = \tau_0 \exp\left[-\frac{f x_u}{k_B T}\right],$$

obtaining the unfolding length $x_u=3.4\pm0.1$ Å, which compares well, given the extreme simplicity of our model, with the experimental results $x_u=3.8$ Å.[3] Comparing our zero-force unfolding time τ_0 with its experimental value $\tau_{\text{expt}}=50$ s,[3] we find out that a single MC step in our model corresponds to about 25 ns. In the intermediate force regime (60–115 pN) our fit yields $x_u=1.3\pm0.1$ Å.

The unfolding trajectories can be grouped in four classes according to their main features, i.e., their end-to-end length plateaus (if they exist) and the order parameter behavior for the whole molecule and its various pairs of β-strands.

At large forces we observe simple two-state trajectories, while at smaller forces various intermediates are obtained. A scheme of the possible pathways is shown in Fig. 4.

In trajectories exhibiting intermediate states it turns out, as already pointed out in previous papers,[5,7] that strand G is always the first to break away. In cellular environment such behavior seems to be connected to the function of the RGD motif Arg78-Gly79-Asp80.[5,8] When the module is fully folded, the RGD motif is available for adhesion, while if strand G is pulled and detached from the remainder of the module, the RGD motif gets closer to the surface of the module and is not functional.

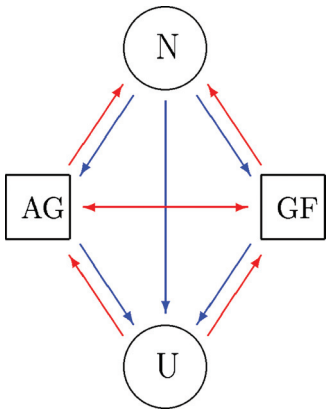
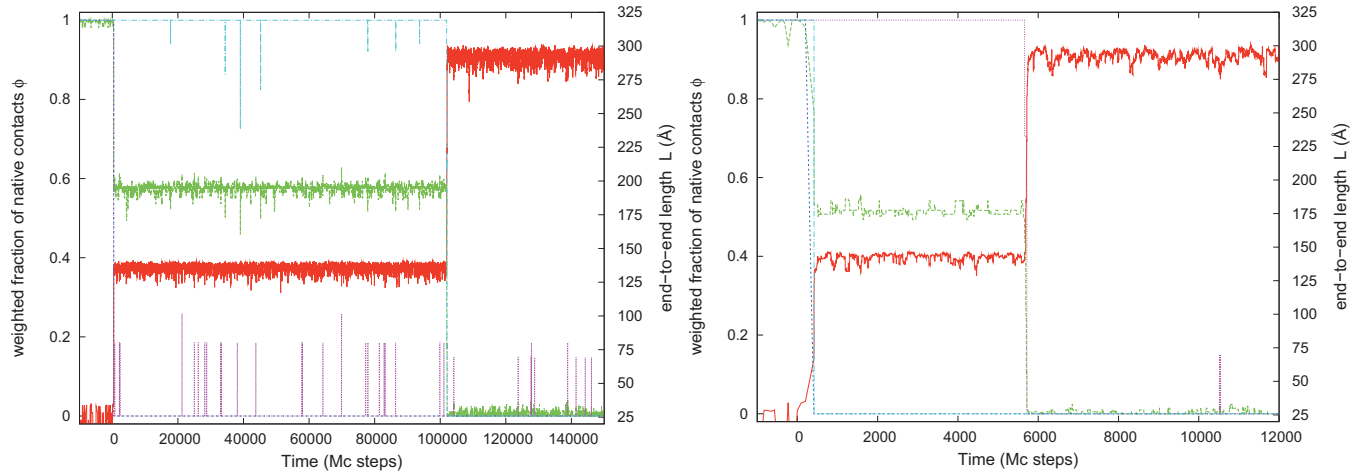


FIG. 4. Unfolding pathways scheme of $FnIII_{10}$ pulled by a constant force. Transitions denoted by red arrows have been observed only at low forces (40, 36, and 28 pN). Oblique red arrows represent refolding transitions.

Strand G detachment may be rapidly followed by complete unfolding or by an intermediate state. A possibility is that strand A detaches almost at the same time of strand G while the remaining part of the molecule stays folded for a certain time before complete unfolding. This kind of unfolding pathway will be labeled with AG, and its intermediate end-to-end length is about 13.5 nm. It may happen that instead of strand A, strand F detaches together with G, such unfolding pathway (intermediate end-to-end length ~14 nm) will be labeled GF.

The last possibility occurs only in the biologically relevant regime of low forces. It is believed[9] that such relevant forces, *in vivo*, are of the same order of magnitude as the equilibrium unfolding force (~20 pN, see Sec. III), although forces as low as 5 pN have been suggested[14] as typical unfolding forces. Our low force unfolding pathway is a mixture of the previous two: Strands A and G are the first to unfold, then, before the molecule completely unfolds, A refolds and F unfolds. This may happen reversibly many times in a single trajectory with consecutive folding (unfolding) of strand A and parallel unfolding (folding) of strand F. Such trajectories will be labeled mixed AG-GF because the molecule is fluctuating between two different intermediates (AG and GF). These intermediates have almost the same end-to-end length, and therefore cannot be distinguished in a simple free energy landscape, as illustrated in Fig. 2, where a single, broad minimum is observed at $L\simeq13$ nm.

Figure 5 shows two typical trajectories at 65 pN constant force. Other typical trajectories are plotted in Ref. 31. During thermalization, before turning the force on at time $t=0$, the length of the polypeptide chain fluctuates around $L=0$ [Fig. 5(a)] since different orientations of the molecule are equally likely. Then, at time $t=0$, a waiting phase starts, which can be easily seen in Figs. 2(a) and 2(b) of Ref. 31. This waiting phase corresponds to a metastable state which is characterized by an end-to-end length ~3.5 nm corresponding to the elongation in the native state. The rise in the end-to-end length to the intermediate value is always associated to the drop in order parameters connected to two different pairs of strands, with the GF pair always involved. The order param-



(a) *Unfolding pathway*: AG (b) *Unfolding pathway*: GF

FIG. 5. Typical MC trajectories: end-to-end length (red line) and a few order parameters as functions of time, with a $f=65$ pN. Green line: fraction of native contacts, whole FnIII$_{10}$. Blue line: fraction of native contacts between strands G and F. Purple line: fraction of native contacts between strands A and B. Cyan line: fraction of native contacts between strands C and F.

eters which have not been plotted go to zero only when the protein reaches the fully elongated configuration (end-to-end length of ~ 29 nm).

In Fig. 6 we report a *mixed* AG-GF trajectory obtained at force $f=28$ pN, slightly larger than the equilibrium unfolding force: After the long waiting phase there are two different intermediate states before the complete unfolding. Despite the large fluctuations in the order parameters associated to the pairs C-F and A-B, it is still possible to see their general behavior and to recognize the first intermediate state as AG and the second as GF. We stress that the GF and AG intermediates have very similar end-to-end length and fraction of native contacts ϕ_s (for the whole chain), making them indistinguishable in simple, one-dimensional, free energy landscapes: Indeed, they are lumped together in the broad minimum at $L \simeq 13$ nm in Fig. 2.

Both the waiting and intermediate states (as the whole unfolding process) are characterized by time lengths varying in a wide range of values for different applied forces and, because of stochasticity, for different trajectories. In Table I we reported the mean life times at various constant forces. The times τ_{AG} and τ_{GF} are obtained by an average of the times occurring between the first and second jumps in the end-to-end length, which have been fixed using the respective threshold values $L=75$ Å and $L=225$ Å. These averages have been calculated only for those trajectories which exhibit the corresponding unfolding pathway, while τ_{AG} and τ_{GF} at force $f=28$ pN and τ_{GF} at force $f=40$ pN are not reported in the table because of vanishing frequencies of the corresponding trajectories, as shown in Table II. The mean waiting phase time τ_{ws} is the average over the 100 trajectories of the time at which the end-to-end length becomes longer than the threshold value $L=75$ Å. For forces $f=98$ and 122 pN it does not make sense to define a waiting phase life time since the protein starts to unravel as soon as the external force is applied at $t=0$. Finally, the unfolding mean time τ_u is the average on all the trajectories of the unfolding time, i.e., the time at which the molecule reaches the unfolding length previously defined.

The probability distributions of intermediate life times for $f=81$ pN have been plotted in Fig. 7, where it can be seen that both distributions can be fitted to the negative exponential function $P(t_s)=(1/\tau_s)\exp\{-t_s/\tau_s\}$ (where s is AG or GF, t_s is the intermediate life time of s, and $\tau_s=\langle t_s \rangle$ its aver-

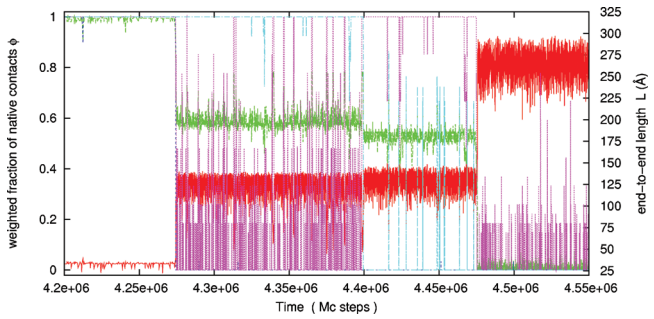


FIG. 6. Mixed AG-GF trajectory: MC time evolution of the end-to-end length (red line) and of some order parameters with a constant force of 28 pN. Colors as in Fig. 5

TABLE I. Unfolding time (τ_u), waiting phase life time (τ_{ws}), AG intermediate life time (τ_{AG}), and GF intermediate life time (τ_{GF}) at different constant forces. Values are in MC steps and are approximated averages on 100 different trajectories at each force.

	τ_u	τ_{ws}	τ_{AG}	τ_{GF}
28 pN	2.2×10^7	8.7×10^6		
36 pN	3.0×10^6	8.8×10^5	4.3×10^5	
40 pN	8.8×10^5	2.8×10^5	6.1×10^5	
46 pN	2.9×10^5	6.9×10^4	2.4×10^5	1.4×10^4
53 pN	1.1×10^5	2.2×10^4	1.2×10^5	9.4×10^3
65 pN	2.7×10^4	3.6×10^3	3.8×10^4	2.9×10^3
81 pN	6.6×10^3	1.1×10^2	1.5×10^4	1.2×10^3
98 pN	1.9×10^3		4.1×10^3	7.4×10^2
122 pN	1.5×10^2		5.7×10^2	1.9×10^2

TABLE II. Relative frequencies of unfolding pathways at constant force. 100 trajectories for each value of the force.

	AG	GF	No intermediates	Mixed AG-GF
28 pN	0	0	0	1
36 pN	0.07	0	0	0.93
40 pN	0.96	0	0	0.04
46 pN	0.93	0.07	0	0
53 pN	0.67	0.32	0.01	0
65 pN	0.59	0.31	0.1	0
81 pN	0.41	0.34	0.25	0
98 pN	0.43	0.23	0.34	0
122 pN	0.2	0.06	0.74	0

age), and that AG has a longer life than GF. Being the unfolding time the sum of the waiting phase time and of the intermediate state time we can naively conclude that if the protein follows the GF pathway, it will reach the unfolded state earlier. For the same reason and since at $f=81$ pN the dominant contribution to the unfolding time comes from τ_{GF} and τ_{AG}, we can argue that at this force the exponential function fits well the unfolding times distribution too.[11,32] Furthermore, at very high forces a lognormal distribution of unfolding times has been proposed.[32] Figure 8 shows this behavior at force $f=150$ pN and the corresponding fit to $P(t_u)=[1/\sqrt{2\pi}\sigma(t_u-t_0)]\exp\{-[\ln^2([t_u-t_0]/m)]/2\sigma^2\}$.

Looking at the data in Table I we can try to interpret the three different force ranges in Fig. 3. In the highest force range there is neither an intermediate state nor a waiting phase and the unfolding time corresponds mainly to the MC time needed for completing the unfolding where every MC move that unravels the molecule and thus increases the length is accepted and every move that reduces the length is refused, that is, an extremely biased random walk, corresponding to the scenario proposed in Ref. 33. Lowering the force, the contributions of τ_{GF} and τ_{AG} to the global unfolding time become important while the waiting phase, if it exists, is still quite short. Finally, in the lowest force interval, also the waiting phase gives its contribution and this matches with the larger slope of the fit line.

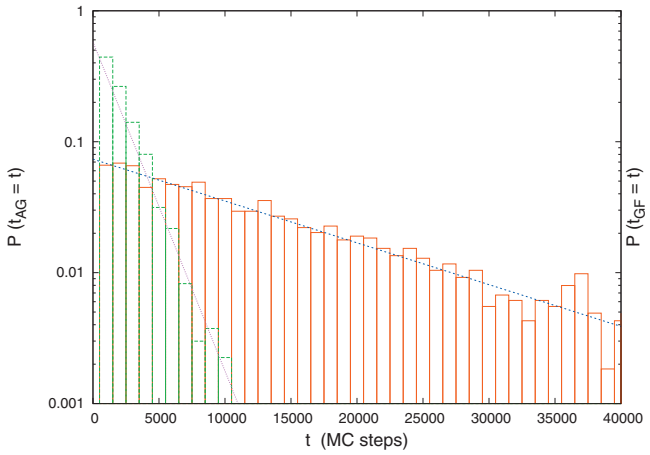


FIG. 7. Histograms of the intermediate life times for AG pathway (red line) and GF pathway (green line) at force $f=81$ pN. Data obtained from 3600 different trajectories. The lines are exponential fits.

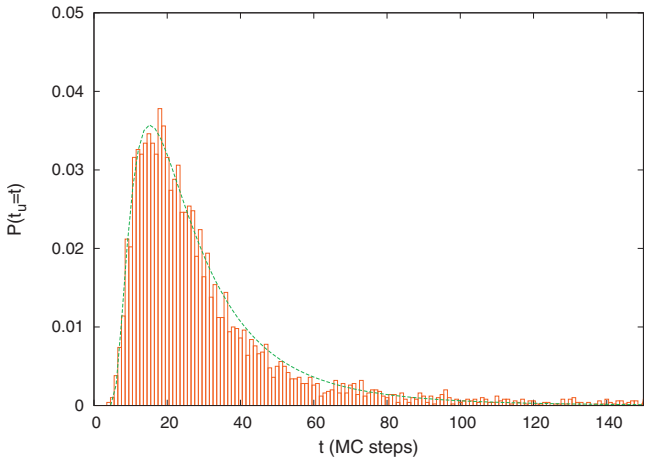


FIG. 8. Histograms of the unfolding times at force $f=150$ pN. Data obtained from 5000 different trajectories and the bin size of the histogram is 1. The fit is to a lognormal distribution.

Table II shows the frequencies of various unfolding pathways. Predictably, as the force increases, the trajectories without any intermediate state become dominant and we expect them to be the only escape route at even higher forces, as already observed in previous all-atom simulations.[9]

At $f=28$ pN, because of long life times and great fluctuations, all the trajectories are of mixed AG-GF type. Furthermore, at such a low force, the molecule can completely refold after it partially unraveled. This can happen many times before complete unfolding.[31]

B. Constant velocity

We run MC simulations at six different pulling velocities (0.03, 0.05, 0.1, 0.3, 0.5, and 1 μm/s) with a spring constant $k=30$ pN/nm and an initial length $\mathcal{L}_0=32$ Å. Once again, our conditions are much closer to experimental ones than most previous simulations. In constant velocity simulations, Vogel *et al.*[5] used $v=50$ m/s (with a spring constant of ~4 nN/nm), Klimov and Thirumalai[7] considered $v=6$ mm/s or faster, while experimental pulling speeds[3,4] were 0.4 and 0.6 μm/s (with spring constants of 45–50 pN/nm) and *in vivo* pulling speeds are believed to be even smaller. Only the all-atom Monte Carlo simulations by Mitternacht *et al.*[9] could probe constant pulling speeds in the same range as we are considering here (with a spring constant of 37 pN/nm).

In Fig. 9 we sketch the possible unfolding pathways scheme in the constant velocity case. Consistent with our constant force results and with previous simulations,[5,7,9] at each value of v most of the trajectories start with the detachment of strand G, giving rise to an intermediate corresponding to the shallow minimum around 6 nm in Fig. 2. In few runs strand A is the first to unravel, but then it refolds, with the consequent detachment of G. Then the unfolding continues through a phase in which strand A is gradually unzipped and when this unzipping is completed the molecule reaches the intermediate AG (end-to-end length of ~13.5 nm). Again we found a mixed AG-GF behavior: Some trajectories do not stay in the AG intermediate until the complete unfolding but they may jump from AG to GF intermediate (end-to-

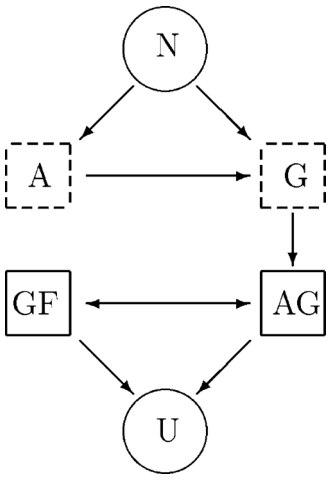
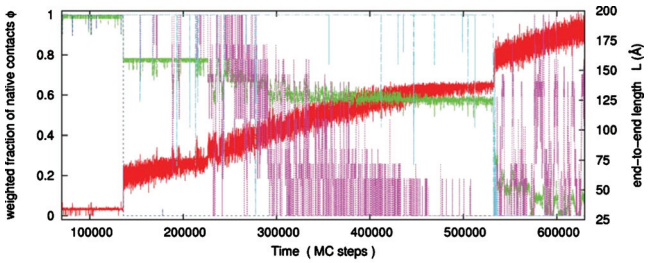
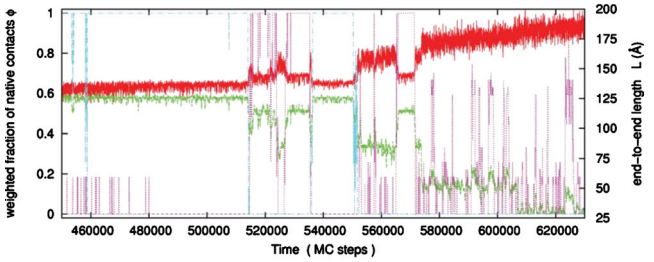


FIG. 9. Unfolding pathways scheme of FnIII$_{10}$ pulled at constant velocity. Intermediate states in the full square boxes have a rupture force remarkably higher than those in dashed boxes.



(a) *Unfolding pathway*: G → AG



(b) *Unfolding pathway*: G → mixed AG-GF (partial)

FIG. 10. MC time evolution of the end-to-end length (red line) and of a few order parameters with a constant velocity of 1 μm/s. Green line: weighted fraction of native contacts, whole FnIII$_{10}$. Blue line: weighted fraction of native contacts between strands G and F. Purple line: weighted fraction of native contacts between strands A and B. Cyan line: weighted fraction of native contacts between strands C and F.

end length of ~14 nm) and back. Table III reports the relative frequencies of various unfolding pathways. It is worth noting that since statistical fluctuations are greater at low pulling rates, the number of mixed AG-GF trajectories and the number of trajectories in which strand A unravels before strand G grows as pulling velocity decreases. Typical trajectories are reported in Figs. 10 and 11, and in Ref. 31.

The average rupture force of the native state depends on the pulling rate and is reported in Table IV. At the pulling speed considered, the average rupture force we obtained for the native state ranges between 80 and 100 pN, which is in remarkable agreement with the AFM results. Fernandez and co-workers reported 75 pN when pulling at 0.6 μm/s (Ref. 3) and 100 pN at 0.4 μm/s.[4] Mitternacht *et al.*[9] reported values from 88 pN at 0.03 μm/s to 114 pN at 0.1 μm/s. Notice that in our results the average rupture force increases with the pulling speed, as predicted by theories[34–38] and verified in experiments.[39] AFM results showed a different behavior, and this was attributed to the interactions with the other modules building up the polyprotein, which is actually pulled in such AFM experiments.[4] In the same work, the average unfolding force of the intermediate states was reported to be 50 pN. Pulling on suitable mutants, two kinds of intermediates were inferred on the basis of experimental results, namely, G and AB. In our model we did not observe intermediate AB, while intermediate G has an average rupture force between 40 and 50 pN. The other intermediates we observed, AG and GF, are more stable, with average unfolding forces around 70 pN.

TABLE III. Relative frequencies of unfolding pathways. 100 trajectories for each value of the velocity.

	G		A→G	
	AG	Mixed AG-GF	AG	Mixed AG-GF
1 μm/s	0.82	0.15	0.03	0.00
0.5 μm/s	0.76	0.20	0.03	0.01
0.3 μm/s	0.49	0.39	0.09	0.03
0.1 μm/s	0.11	0.85	0.01	0.03
0.05 μm/s	0.08	0.87	0.01	0.04
0.03 μm/s	0.07	0.79	0.00	0.14

The distribution of the unfolding forces is well fitted by the theoretical result[34]

$$P(f) = \frac{1}{t_0 r} e^{\beta f x_u} \exp\left[-\frac{k_B T}{r x_u t_0} (e^{\beta f x_u} - 1) \right]. \quad (9)$$

Such an equation corresponds to the rupture force probability distribution of a single molecular bond subject to a force that increases linearly with a rate r.[34] In Fig. 12 we plot the unfolding force histogram at $v=0.5$ μm/s, and fit the data to Eq. (9), with $a=rt_0$ and x_u as fitting parameters. The fit gives $x_u=8.0$ Å, which is larger than the value found for the constant force setup, but it must be kept in mind that the above

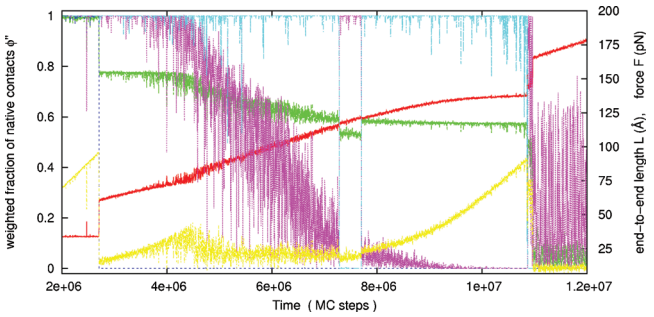


FIG. 11. MC time evolution of end-to-end length (red line), force (yellow line), and of some order parameters with a constant velocity of 0.03 μm/s for a mixed AG-GF trajectory. Green line: weighted fraction of native contacts, whole FnIII$_{10}$. Blue line: weighted fraction of native contacts between strands G and F. Purple line: weighted fraction of native contacts between strands A and B. Cyan line: weighted fraction of native contacts between strands C and F. Bins of 2000 MC steps have been used to reduce fluctuations in the plot.

TABLE IV. Average rupture forces.

Velocity (μm/s)	N→G	G→AG	AG→U	GF→U
1	98.5±6.4	40.8±2.6	99.6±9.9	77.3±7.7
0.5	96.1±6.1	42.2±2.6	96.5±7.3	77.5±4.0
0.3	94.5±6.9	43.4±2.4	92.4±7.8	76.5±5.8
0.1	89.0±6.5	45.4±1.8	86.5±7.9	69.9±5.8
0.05	87.8±5.1	46.1±1.6	81.9±8.4	67.6±5.9
0.03	87.3±5.8	46.9±1.7	81.5±9.7	66.7±5.4

theoretical result was derived for a force which is linear in time with a slope r, while here the force is associated to the harmonic potential which moves at constant velocity v.

V. CONCLUSIONS

We have simulated constant force and constant pulling speed unfolding of $FnIII_{10}$, the tenth type III domain of fibronectin, using an Ising-like model that we have developed and validated in recent years, whose equilibrium thermodynamics is exactly solvable. Force and time units have been determined by comparison with existing estimates of the equilibrium unfolding force and the zero-force average unfolding time. We can probe force and speed ranges close to *in vivo* and experimental conditions, which were not possible in most previous simulations.

At high enough constant force we observed two-state transitions only. At smaller forces and at all pulling speeds considered we observed several intermediates, denoted by A, G, AG, and GF, based on the strands which are unfolded in each intermediate. Possible unfolding pathways are summarized in Fig. 4 for the constant force protocol and in Fig. 9 for the constant pulling speed protocol.

The unfolding pathways depend on the applied force or on the pulling speed, which was already observed in Ref. 9. Such pathways become more complex at low forces and speeds due to the increase in fluctuations. Previous simulations and experiments showed some discrepancies in the unfolding pathways, and our work is not going to resolve such discrepancies, but some general trends are confirmed. In particular, the most frequently observed intermediate in our trajectories was AG, which was observed in all previous simulations.[5–9] In addition, constant pulling speed trajectories always visit intermediate G, which was also observed in most previous simulations[5,7–9] and in AFM experiments.[4] On the other hand, we have never observed intermediate AB, which has been reported in many simulations[6,8,9] and experiments.[4] We have instead observed, at low enough forces and speeds, intermediates A and GF, which were previously reported only by Gao *et al.*[8] (A only) and Mitternacht *et al.*[9] (both A and GF). These intermediates have end-to-end lengths close to G and AG, respectively, and cannot be distinguished in the usual one-dimensional free energy landscape using the end-to-end length as a reaction coordinate. Interestingly, in our trajectories we observe fluctuations between intermediates with similar lengths, that is between A and G or between AG and GF. Fluctuations between AG and GF, in particular, are observed in most trajectories at the lowest forces and pulling speeds we have considered, and therefore one could speculate that they have some biological significance.

From a more quantitative point of view, given the extreme simplicity of our model, it is remarkable that many quantities we can compute agree well with the results from AFM experiments or previous simulations with similar parameters. Our estimate for the native state unfolding length is $x_u=3.4\pm0.1$ Å, to be compared with $x_u=3.8$ Å from AFM results[3] and with $x_u=4$ Å from the simulations by Mitternacht *et al.*[9] The average rupture force we obtained for the native state is in the range of 80–100 pN, to be compared with results from 75 to 100 pN reported by AFM studies,[3,4] and from 88 to 114 pN in the simulations by Mitternacht *et al.*[9] Finally, our intermediate G has an average rupture force between 40 and 50 pN, to be compared with 50 pN found in experiments[4] although it must be mentioned that in such work the intermediate might be an average between the G and AB intermediates.

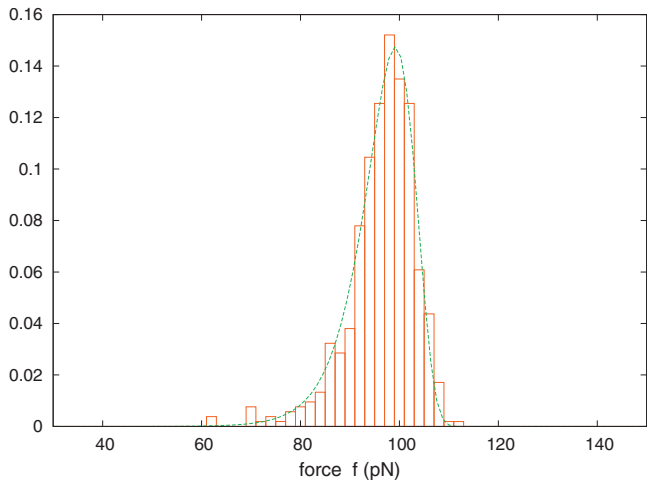


FIG. 12. Distributions of the rupture forces of the native state at pulling velocity $v=0.5$ μm/s. Data obtained from 500 different trajectories; bin size of the histogram is 2. The fit is to Eq. (9).

ACKNOWLEDGMENTS

A.I. gratefully acknowledges financial support from Lundbeck Fonden and from Danish Centre for Scientific Computing (DCSC). A.I. is grateful to Anders Irbäck and to Simon Mitternacht for helpful and stimulating discussions.

[1] S. Kumar and M. Li, Phys. Rep. **486**, 1 (2010).
[2] B. Geiger, A. Bershadsky, R. Pankov, and K. M. Yamada, Nat. Rev. Mol. Cell Biol. **2**, 793 (2001).
[3] A. Oberhauser, C. Badilla-Fernandez, M. Carrion-Vasquez, and J. Fernandez, J. Mol. Biol. **319**, 433 (2002).
[4] L. Li, H.-L. Huang, C. Badilla, and J. Fernandez, J. Mol. Biol. **345**, 817 (2005).
[5] A. Krammer, H. Lu, B. Isralewitz, K. Schulten, and V. Vogel, Proc. Natl. Acad. Sci. U.S.A. **96**, 1351 (1999).
[6] E. Paci and M. Karplus, J. Mol. Biol. **288**, 441 (1999).
[7] D. K. Klimov and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **97**, 7254 (2000).
[8] M. Gao, D. Craig, V. Vogel, and K. Schulten, J. Mol. Biol. **323**, 939 (2002).
[9] S. Mitternacht, S. Luccioli, A. Torcini, A. Imparato, and A. Irbäck, Biophys. J. **96**, 429 (2009).

[10] A. Imparato, A. Pelizzola, and M. Zamparo, Phys. Rev. Lett. **98**, 148102 (2007).
[11] A. Imparato, A. Pelizzola, and M. Zamparo, J. Chem. Phys. **127**, 145105 (2007).
[12] A. Imparato and A. Pelizzola, Phys. Rev. Lett. **100**, 158104 (2008).
[13] A. Imparato, A. Pelizzola, and M. Zamparo, Phys. Rev. Lett. **103**, 188102 (2009).
[14] H. Erickson, Proc. Natl. Acad. Sci. U.S.A. **91**, 10114 (1994).
[15] H. Li, W. Linke, A. Oberhauser, M. Carrion-Vazquez, J. G. Kerkvliet, H. Lu, P. E. Marszalek, and J. M. Fernandez, Nature (London) **418**, 998 (2002).
[16] H. Wako and N. Saitô, J. Phys. Soc. Jpn. **44**, 1931 (1978).
[17] H. Wako and N. Saitô, J. Phys. Soc. Jpn. **44**, 1939 (1978).
[18] V. Muñoz, P. Thompson, J. Hofrichter, and W. Eaton, Nature (London) **390**, 196 (1997).
[19] V. Muñoz, E. Henry, J. Hofrichter, and W. Eaton, Proc. Natl. Acad. Sci. U.S.A. **95**, 5872 (1998).
[20] V. Muñoz and W. Eaton, Proc. Natl. Acad. Sci. U.S.A. **96**, 11311 (1999).
[21] M. Garcia-Mira, M. Sadqi, N. Fischer, J. Sanchez-Ruiz, and V. Muñoz, Science **298**, 2191 (2002).
[22] A. Flammini, J. Banavar, and A. Maritan, Europhys. Lett. **58**, 623 (2002).
[23] P. Bruscolini and A. Pelizzola, Phys. Rev. Lett. **88**, 258101 (2002).
[24] A. Pelizzola, J. Stat. Mech.: Theory Exp. **2005**, P11010.
[25] M. Zamparo and A. Pelizzola, Phys. Rev. Lett. **97**, 068106 (2006).
[26] M. Zamparo and A. Pelizzola, J. Stat. Mech.: Theory Exp. **2006**, P12009.
[27] P. Bruscolini, A. Pelizzola, and M. Zamparo, Phys. Rev. Lett. **99**, 038103 (2007).
[28] P. Bruscolini, A. Pelizzola, and M. Zamparo, J. Chem. Phys. **126**, 215103 (2007).
[29] S. Litvinovich and K. Ingham, J. Mol. Biol. **248**, 611 (1995).
[30] M. Zamparo, Ph.D. thesis, Politecnico di Torino, 2009.
[31] See supplementary material at http://dx.doi.org/10.1063/1.3464476 for additional results on equilibrium properties and unfolding pathways.
[32] P. Szymczak and M. Cieplak, J. Phys.: Condens. Matter **18**, L21 (2006).
[33] S. Luccioli, A. Imparato, S. Mitternacht, A. Irbäck, and A. Torcini, Phys. Rev. E **81**, 010902 (2010).
[34] E. Evans and K. Ritchie, Biophys. J. **72**, 1541 (1997).
[35] E. Evans, Annu. Rev. Biophys. Biomol. Struct. **30**, 105 (2001).
[36] O. Dudko, A. Filippov, J. Klafter, and M. Urbakh, Proc. Natl. Acad. Sci. U.S.A. **100**, 11378 (2003).
[37] I. Derényi, D. Bartolo, and A. Ajdari, Biophys. J. **86**, 1263 (2004).
[38] O. Dudko, G. Hummer, and A. Szabo, Phys. Rev. Lett. **96**, 108101 (2006).
[39] M. Carrion-Vazquez, A. Oberhauser, S. Fowler, P. Marszalek, S. Broedel, J. Clarke, and J. Fernandez, Proc. Natl. Acad. Sci. U.S.A. **96**, 3694 (1999).