Theoretical studies of the kinetics of mechanical unfolding of cross-linked polymer chains and their implications for single-molecule pulling experiments

Kilho Eom,[1] Dmitrii E. Makarov,[2,*] and Gregory J. Rodin[3]
[1]*Department of Aerospace Engineering & Engineering Mechanics, The University of Texas at Austin, Austin, Texas 78712, USA*
[2]*Department of Chemistry & Biochemistry, Institute for Computational Engineering & Science, and Institute for Theoretical Chemistry, The University of Texas at Austin, Austin, Texas 78712, USA*
[3]*Department of Aerospace Engineering & Engineering Mechanics and Institute for Computational Engineering & Science, The University of Texas at Austin, Austin, Texas 78712, USA*
(Received 16 September 2004; published 11 February 2005)

We have used kinetic Monte Carlo simulations to study the kinetics of unfolding of cross-linked polymer chains under mechanical loading. As the ends of a chain are pulled apart, the force transmitted by each cross-link increases until it ruptures. The stochastic cross-link rupture process is assumed to be governed by first order kinetics with a rate that depends exponentially on the transmitted force. We have performed random searches to identify optimal cross-link configurations whose unfolding requires a large applied force (measure of strength) and/or large dissipated energy (measure of toughness). We found that such optimal chains always involve cross-links arranged to form parallel strands. The location of those optimal strands generally depends on the loading rate. Optimal chains with a small number of cross-links were found to be almost as strong and tough as optimal chains with a large number of cross-links. Furthermore, optimality of chains with a small number of cross-links can be easily destroyed by adding cross-links at random. The present findings are relevant for the interpretation of single molecule force probe spectroscopy studies of the mechanical unfolding of "load-bearing" proteins, whose native topology often involves parallel strand arrangements similar to the optimal configurations identified in the study.

DOI: 10.1103/PhysRevE.71.021904 PACS number(s): 87.15.−v, 82.37.Rs

I. INTRODUCTION

A number of proteins exhibit a combination of strength and toughness that cannot be matched by artificial materials [1–4]. Recent single molecule force probe spectroscopy experiments suggest that these remarkable properties are accomplished through the mechanical response of individual protein domains, which are capable of dissipating large energy upon their mechanical unfolding [2,4,5].

In single molecule pulling experiments employing the atomic force microscope (AFM), one end of the protein is attached to a substrate and the other end is attached to a cantilever (see, e.g., Refs. [6–8] for a review); the cantilever then can be displaced at a constant rate. During such an experiment, one measures the pulling force, and then presents the data in the form of the force-displacement curve. The forces generated by different proteins under typical experimental conditions range from a few piconewtons to several hundred piconewtons and generally depend on the pulling rate. If one were to perform an equilibrium, reversible stretching experiment by pulling on the molecule at a sufficiently slow rate then the measured force-vs-displacement curve would become rate independent and the work done by the pulling force would be equal to the free energy difference between the folded and the stretched states of the molecule. In practice, stretching of a molecule is nearly an equilibrium process if the timescale of pulling is longer than that of the molecule's conformational changes. This equilibrium regime is rarely achieved in AFM pulling studies. It further appears that many proteins that perform "load-bearing" functions in living organisms operate far away from equilibrium; as a result their mechanical stability is often uncorrelated with their thermodynamic stability [7,9–12].

For example, the work required to unfold the molecule of the muscle protein titin in a typical AFM pulling experiment is about 2 orders of magnitude higher than its free energy of folding, indicating that this is a highly nonequilibrium process [5]. This property accounts for the high toughness of titin arguably required for its biological function in the muscles. Similarly, the difference between the force-vs-extension curves measured in the course of stretching and subsequent relaxation of spider capture silk proteins [4] reveals that stretching is a nonequilibrium process, in which extra energy is dissipated. In contrast, the work required to unfold of the myosin coiled-coil via pulling on it at similar pulling rates is comparable to the free energy of folding, indicating that this is a nearly equilibrium process [5].

The mechanical resistance of a protein is thus determined both by its structure and by the loading rate. Recently, we have studied a toy model of a cross-linked polymer chain, which we used to identify the chain configurations that lead to its high mechanical resistance [13]. In that model, we considered a Gaussian chain with rigid cross-links. Unfolding of the chain under mechanical loading occurs as a result of rupture of the cross-links. Each cross-link ruptures once its internal force reaches a critical value. Thus, as the chain ends are being pulled apart at a constant rate, the force in each link increases until it ruptures. As the loading proceeds,

*Corresponding author. Email address: makarov@mail.cm.utexas.edu

 ©2005 The American Physical Society

all the cross-links become ruptured and the chain unfolds. The excess work done on the cross-linked chain, as compared to the work done stretching the unconstrained chain, is a measure of the chain toughness. Given the total number of cross-links, one may seek the optimal cross-link configurations that maximize either the excess work or the maximum force during the unfolding process. Our rationale for studying such a simple model was the previous finding [7,10–12,14] that the unfolding mechanism is largely determined by the native topology of the protein. This view is further supported by the success of simplified, Go-like models in predicting the mechanisms of mechanical unfolding [15–18]. Although Gaussian cross-linked chains are merely caricatures of real biopolymers, they may adequately capture the effects of topology on the unfolding mechanism. Indeed, there are good reasons to believe they do. Specifically, the key finding of our previous study is that the optimal configurations that maximize the peak force and the dissipated energy must involve parallel strands. This finding is consistent with experimental studies [7,9,10,19–24] and molecular dynamics simulations [25–29] of the protein domains exhibiting high unfolding forces, such as the I27 domain in titin. Further, this finding has led to the prediction that protein domains with the ubiquitin fold, which features terminal parallel strands similar to those in I27, exhibit superior mechanical properties, despite the fact that they have no apparent mechanical functions in living organisms [30]. This prediction is supported by both experiments [12] and molecular dynamics simulations [30,31].

While providing results that are qualitatively consistent with atomistic scale studies, our model [13] entirely ignored stochastic and rate-dependent aspects of unfolding. This is an unrealistic assumption in many cases because, in general, rupture of a chemical bond is a chemical reaction, i.e., a stochastic process whose rate is affected by the transmitted force [32]. Further, as we mentioned earlier, load-bearing proteins exhibit high toughness and strength precisely because they are loaded at high rates so that unfolding is a nonequilibrium irreversible process accompanied by large energy dissipation.

Models of force-induced rupture of chemical bonds are well known in the contexts of protein unfolding and ligand unbinding [19,20,32–35] and fracture [36]. In those models, rupture of a bond is described by first-order kinetics and its rate depends on the force transmitted by the bond. The main purpose of this paper is to adapt our model of cross-linked Gaussian chains to study how the optimal chain configurations that maximize the excess work and/or the maximum force depend on the loading rate. To this end, we have assumed that rupture of each cross-link is described by first-order kinetics with a force-dependent probability and performed kinetic Monte Carlo studies of the chain unfolding. The main finding of this study is that the parallel-strand arrangements remain optimal even when the stochastic nature of bond breaking is taken into account; While always featuring such parallel strands, the found optimal configurations generally depend on the loading rate.

The rest of this paper is organized as follows. In Sec. II, we describe the model. In Sec. III, we outline the simulation methods. In Sec. IV, we present our simulation results. In Sec. V, we discuss implications of our results for pulling experiments on single molecules.

II. THE MODEL

Consider a polymer chain consisting of $L+1$ beads connected by L links. The chain is assumed to obey Gaussian statistics so that the probability distribution for the distance between beads i and j is given by

$$P(|\mathbf{r}_i - \mathbf{r}_j|) = \left[\frac{3}{2\pi b^2 |i-j|}\right]^{3/2} \exp\left[-\frac{3|\mathbf{r}_i - \mathbf{r}_j|^2}{2|i-j|b^2}\right], \quad (1)$$

where b is the rms length of a single link. One way to construct such a Gaussian chain is to connect neighboring beads by harmonic springs such that its potential energy is given by

$$U = \frac{1}{2}\gamma_0 \sum_{i=1}^{L} |\mathbf{r}_{i+1} - \mathbf{r}_i|^2 \quad \text{with} \quad \gamma_0 = \frac{3k_B T}{b^2}, \quad (2)$$

where k_B is Boltzmann's constant and T is the temperature.

The motion of the chain is constrained by N cross-links. Each link is designated by the indices of its end points, so that the entire set of cross-links is denoted by $C_N = \{\{i_1, j_1\}, \ldots, \{i_N, j_N\}\}$. Each cross-link is regarded as rigid; alternatively, one can model a cross-link as a spring with a spring constant $\gamma_c \gg \gamma_0$. We assume that no bead can be attached to more than one cross-link, so that the maximum number of cross-links is $N=(L+1)/2$.

The chain ends (monomers number 1 and $L+1$) are pulled apart at a constant speed v so that the distance between them grows linearly as a function of time t:

$$|\mathbf{r}_L - \mathbf{r}_0| \equiv e = vt. \quad (3)$$

We suppose that loading is slow compared to a typical timescale of thermal Brownian motion of the chain. In this case, we assume that the value of the pulling force $F(t)$ recorded at any instant t is the force *averaged* over the thermal motion. At the same time, the timescale of cross-link rupture may be comparable with that of loading and so the rupture of a cross-link may result in a measurable change in $F(t)$.

We consider two rupture models for the cross-links. In the first model, which we refer to as *model I*, a cross-link ruptures deterministically once its internal force reaches a critical value f_c. This model has been studied previously [13] but we include it here for comparisons. In the second model, to which we refer as *model II*, rupture of a cross-link is a stochastic process described by first-order kinetics. Specifically, the conditional probability that the cross-link that is intact at time t ruptures in the time interval from t to $t+\Delta t$ depends only on the instantaneous value of the internal force $f(t)$ and is given by [32]

$$k[f(t)]\Delta t = k_0 \exp\left[\frac{f(t)}{f_c}\right]\Delta t, \quad (4)$$

where k_0 is the rupture rate constant at zero force and f_c is a reference force. Equation (4) is a commonly used model, which assumes that the free energy barrier to rupture decreases linearly with the force f [20,32]. Although this equa-

tion is not necessarily quantitative [31,37], it is sufficient for qualitative predictions, as it properly identifies the rapid increase of $k[f(t)]$ once the internal force exceeds f_c.

Because the rate of Eq. (4) is not zero at zero force, then, strictly speaking, any cross-link configuration in model II is unstable and the chain will unfold irreversibly on a timescale of order k_0^{-1} even if no force is applied. This is not realistic since the folded state of a protein at zero force is expected to be thermodynamically more stable than its unfolded state. It is necessary to allow for the recombination of cross-links in order to restore the detailed balance in the system [35,38,39]. At zero force, the rate of recombination for a cross-link would be higher than k_0 thereby rendering it thermodynamically stable. Here, we assume that the time k_0^{-1} is much longer than the timescale of loading. Under this assumption recombination of cross-links during unfolding is unlikely because forces in each cross-link will quickly reach values large enough to destabilize each bond thermodynamically such that the ruptured bond state has lower free energy than that with the bond intact; in other words, once the bond is broken it will be unlikely to reform unless the loading force is removed. For these reasons we did not include cross-link recombination in our model; It would therefore not be applicable to very slow, nearly equilibrium pulling experiments. In this respect, the physical regime explored by the present work is quite different from the reversible stretching conditions assumed in the theoretical studies of RNA and DNA mechanical denaturation [40–44] and in the theories of the reversible stretching of proteinlike heteropolymers [45–47]. Note, however, that nonequilibrium effects have been considered in Ref. [46].

When the ends of a Gaussian chain are pulled apart, its response follows Hooke's law [48,49], which also holds in the presence of cross-links [50]. However, the spring constant of the entire chain changes upon cross-link rupture. Under constant velocity loading conditions, the force-displacement curve $F=F(e)$ is a piece-wise linear function with jumps and different slopes (see Fig. 1). Once all the cross-links are ruptured, the slope is reduced to the effective spring constant of the unconstrained chain, $\Gamma_0=\gamma_0/L$.

The mechanical response of a cross-linked chain is represented by two quantities (cf. Fig. 1): (i) the maximum force F_m and (ii) "toughness," i.e., the excess work done upon unfolding:

$$\Delta W = \int_0^u F(e)de - \frac{1}{2}\Gamma_0 u^2, \tag{5}$$

where u is the distance between the 1st and the $L+1$st beads at the end of the pulling experiment, once all the cross-links have been ruptured.

For model I, rupture is a deterministic process, so that F_m and ΔW are unique for a given set C_N. Further, the force-displacement curve and its parameters F_m and ΔW can be determined upon solving a set of N linear problems that reflect the sequence of the rupture events. In contrast, in model II, rupture is a stochastic process. Accordingly, for a given set C_N, it is necessary to determine the averages of F_m and

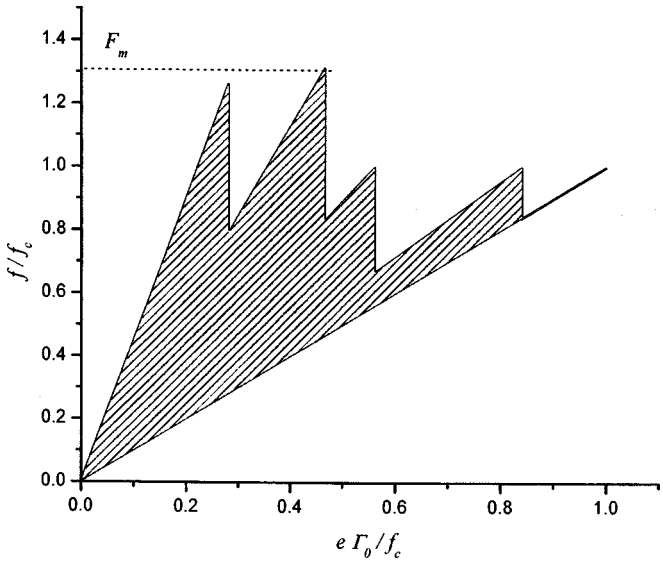

FIG. 1. Unfolding of a cross-linked chain. (a) The configuration of a $L=50$ chain with the cross-links {{7,19},{15,47},{16,42},{21,35},{40,48}}. (b) The force-vs-extension curve of this chain in the case of the deterministic unfolding scenario (model I). Each maximum corresponds to the rupture of one or more cross-links. The mechanical resistance of the chain is characterized by two parameters: The excess work ΔW required to extend the cross-linked chain relative to that for the "denatured" chain (equal to the shaded area) and the maximum force F_m.

ΔW over sufficiently large number of realizations of the stochastic unfolding process; we denote those quantities by $\langle F_m\rangle$ and $\langle \Delta W\rangle$, respectively.

The adopted model will be used in the following settings.

• *Characterization problem*: Given L, C_N, γ_0, k_0, f_c, and v determine $\langle F_m\rangle$ and $\langle \Delta W\rangle$.

• *Optimization problem*: Given L, N, γ_0, k_0, f_c, and v determine the configuration(s) C_N that maximize(s) $\langle F_m\rangle$ and $\langle \Delta W\rangle$.

III. METHODS

A. Elasticity analysis

Between two rupture events, the cross-linked chain responds as a collection of Hookean springs [50]. The springs are identified as follows.

(1) Arrange the $2N$ beads belonging to the cross-links in the ascending order:

$$1 \leqslant i_1 < i_2 < \cdots i_{2N-1} < i_{2N} \leqslant L.$$

(2) Identify each chain segment between two consecutive members of this set as a spring.

(3) Assign to each spring the spring constant γ_0/n, where n is the number of the chain links in the segment.

Once the springs and their spring constants have been identified, the entire assembly can be analyzed using the finite-element method [50]. The results can be expressed as

$$F(t) = \Gamma(t)vt \tag{6}$$

and

$$f_k(t) = \alpha_k(t)F(t), \tag{7}$$

where $\Gamma(t)$ is the instantaneous overall spring constant of the cross-linked chain, $f_k(t)$ is the internal force in the kth cross-link, and $\alpha_k(t)$'s are dimensionless coefficients. The procedure for finding these coefficients is detailed in Ref. [50]. Note that $\Gamma(t)$ and $\alpha_k(t)$ depend on the current configuration of the cross-links and remain constant between rupture events; in general, they are piecewise constant functions of time.

B. Kinetic Monte Carlo method

To simulate the stochastic unfolding process we use the kinetic Monte Carlo method [35,51–53]. Suppose that at time t_0, there are n cross-links. Let us evaluate the probability that the first rupture among those cross-links occurs at a later time, in the time interval between t and $t+\Delta t$. This probability is equal to the probability $S(t,t_0)$ that no cross-link has ruptured in the time interval between t_0 and t, times the sum of the probabilities for each of the cross-link to rupture in the time interval between t and $t+\Delta t$:

$$\Phi(t)\Delta t = S(t,t_0)\sum_{m=1}^{n} k[f_m(t)]\Delta t. \tag{8}$$

Also, in the time interval between t and $t+\Delta t$ the survival probability is reduced by $\Phi(t)\Delta t$, so that

$$-\Phi(t)\Delta t = S(t+\Delta t,t_0) - S(t,t_0) = (dS/dt)\Delta t.$$

This leads to the differential equation for $S(t,t_0)$:

$$dS(t,t_0)/dt = -S(t,t_0)\sum_{m=1}^{n} k[f_m(t)]. \tag{9}$$

Using Eqs. (4), (6), and (7) we have $k[f_m(t)] = k_0 \exp[\alpha_m(t_0)\Gamma(t_0)vt/f_c]$; substituting this into Eq. (9) and integrating we obtain

$$S(t,t_0) = \exp\left\{-k_0\sum_{m=1}^{n}\frac{f_c}{\alpha_m(t_0)\Gamma(t_0)v}\left[\exp\left(\frac{\alpha_m(t_0)\Gamma(t_0)vt}{f_c}\right) - \exp\left(\frac{\alpha_m(t_0)\Gamma(t_0)vt_0}{f_c}\right)\right]\right\} \tag{10}$$

and

$$\Phi(t) = k_0 S(t,t_0)\sum_{m=1}^{n}\exp\left[\frac{\alpha_m(t_0)\Gamma(t_0)vt}{f_c}\right]. \tag{11}$$

The standard method [35,51–53] for generating the time t on a computer is to solve the equation

$$S(t,t_0) = \xi, \tag{12}$$

where ξ is a uniformly distributed random variable in the interval [0,1]. We use modified Newton's method to solve this equation numerically. Once the time t is generated, we need to determine which of the n cross-links ruptures. This is done by computing the weighted probability of rupture for each of the cross-links:

$$w_m = \frac{\exp\left[\frac{f_m(t)}{f_c}\right]}{\sum_{j=1}^{n}\exp\left[\frac{f_j(t)}{f_c}\right]} \quad \text{with} \quad m = 1,\ldots,n. \tag{13}$$

Next, we divide the interval [0,1] into n subintervals whose lengths are w_m. Finally, we generate λ, a realization of a random variable uniformly distributed in the interval [0,1], and identify the subinterval containing λ. The index of this subinterval is equal to the index of the cross-link to be ruptured. This process is followed starting with $t=0$, $n=N$ and until all the cross-links are ruptured.

The quantities $\langle F_m\rangle$ and $\langle \Delta W\rangle$ for a given set C_N are computed by averaging over N_{MC} realizations of the unfolding history; we used $N_{MC}=5000$.

C. Optimization

We used two optimization methods for finding the configurations that maximize $\langle F_m\rangle$ and/or $\langle \Delta W\rangle$. In cases where the search space was sufficiently small, we exhaustively searched over all possible sets C_N. When an exhaustive search was too time-consuming, we resorted to the following "random hill-climbing" procedure [13].

(1) Generate a random set $C_N^{(0)}$ with N cross-links.

(2) Select a cross-link $\{i,j\}$ from the set $C_N^{(0)}$.

(3) Evaluate $\langle F_m\rangle$ (or $\langle \Delta W\rangle$) for $C_N^{(0)}$, and the "adjacent" sets obtained from $C_N^{(0)}$ upon replacing $\{i,j\}$ with $\{i,j\pm 1\}$ or $\{i\pm 1,j\}$. Of course, the sets $\{i,j\pm 1\}$ and $\{i\pm 1,j\}$ must be admissible, in the sense that no bead can be connected to more than one cross-link.

(4) Choose the optimal set among the five sets identified at step (3).

(5) Repeat steps (2)–(4) for all other cross-links to complete the first sweep. This defines a new configuration $C_N^{(1)}$.

(6) Repeat steps (1)–(5) until $C_N^{(i+1)} = C_N^{(i)}$.

(7) Generate new $C_N^{(0)}$ and repeat steps (2)–(6).

IV. RESULTS

A. Single cross-link

Model I. A single cross-link, $\{i,i+l\}$, creates a loop of length l in the chain. The optimal configurations in this case can be found analytically [13]. In particular, $F_m = f_c$ for all i and l, and $\langle \Delta W\rangle$ depends on l only:

$$\Delta W = \frac{f_c^2}{2\Gamma_0}(\tilde{l} - \tilde{l}^2), \tag{14}$$

where we have introduced the dimensionless loop length

$$\tilde{l} = \frac{l}{L}.$$

The excess work reaches its maximum for $\tilde{l} = 1/2$:

TABLE I. Single cross link: The dimensionless loop length $\tilde{l}_{\Delta W}$ that maximizes $\langle \Delta W \rangle$ as a function of the dimensionless pulling velocity $\tilde{v}$.

$\tilde{v}$	0.1	1	5	10	15	20	30	50	100	200	500
$\tilde{l}_{\Delta W}$	0.967	0.84	0.73	0.69	0.675	0.662	0.648	0.63	0.615	0.601	0.587

$$\Delta W = \frac{f_c^2}{8\Gamma_0}.$$

Thus one can regard the configurations with $\tilde{l}=1/2$ as optimal with respect to both F_m and $\langle \Delta W \rangle$.

Model II. The model parameters give rise to the dimensionless time

$$\tau = k_0 t$$

and dimensionless pulling rate

$$\tilde{v} = \frac{\Gamma_0 v}{k_0 f_c}.$$

Following the analysis in Sec. III B, it is straightforward to obtain the probability density function for the dimensionless rupture time τ,

$$\Phi(\tau,\theta) = \exp(\theta\tau)\exp\left\{\frac{1}{\theta}[1-\exp(\theta\tau)]\right\}, \qquad (15)$$

where the parameter θ combines the dimensionless loading rate and geometric parameters,

$$\theta = \frac{\tilde{v}}{1-\tilde{l}}.$$

This combination arises naturally for $N=1$ but not for $N>1$. At the moment of rupture we have

$$F_m(\tau) = f(\tau) = \frac{\Gamma_0}{1-\tilde{l}} v k_0 \tau = f_c \theta \tau \qquad (16)$$

and

$$\Delta W(\tau) = \frac{1}{2}\Gamma_0 (v k_0 \tau)^2 \left(\frac{1}{1-\tilde{l}} - 1\right) = \frac{1}{2}\frac{f_c^2}{\Gamma_0}(1-\tilde{l})\tilde{l}\theta^2\tau^2,$$

and therefore we obtain

$$\langle F_m \rangle = f_c \theta \int_0^\infty \tau \Phi(\tau,\theta) d\tau$$

and

$$\langle \Delta W \rangle = \frac{1}{2}\frac{f_c^2}{\Gamma_0}(1-\tilde{l})\tilde{l}\theta^2 \int_0^\infty \tau^2 \Phi(\tau,\theta) d\tau.$$

The integrals involved in these expressions can be evaluated numerically only. Nevertheless, one can obtain asymptotic approximations valid for $\theta \gg 1$:

$$\langle F_m \rangle \approx f_c \ln\theta = f_c \ln\frac{\tilde{v}}{1-\tilde{l}}, \qquad (17a)$$

$$\langle \Delta W \rangle = \frac{1}{2}\frac{f_c^2}{\Gamma_0}(\tilde{l}-\tilde{l}^2)\ln^2\theta = \frac{1}{2}\frac{f_c^2}{\Gamma_0}(\tilde{l}-\tilde{l}^2)\ln^2\frac{\tilde{v}}{1-\tilde{l}}. \qquad (17b)$$

The meaning of Eq. (17a) is simple: This is the force [Eq. (16)] corresponding to the most probable rupture time that maximizes the probability density of Eq. (15) [33,34]. As expected, this asymptotic expression for $\langle F_m \rangle$ reveals the logarithmic dependence on the loading rate [32–34]. Further, $\langle F_m \rangle$ increases indefinitely as $\tilde{l} \to 1$, i.e., the largest forces are generated by chains with terminal cross-links. The case of $\tilde{l}=1$ is pathological: In this case the ends of a cross-link itself are pulled apart with the speed v. Since in our model the intrinsic spring constant of a cross-link is infinite, this leads to a divergent force in Eq. (17a). This pathology does arise in model I where, by construction, the cross-link ruptures at the force f_c.

The excess work also grows logarithmically with $\tilde{v}$, but in contrast to $\langle F_m \rangle$, its optimization leads to values of $\tilde{l}$ that depend on $\tilde{v}$. In particular, for $\tilde{v} \to \infty$ the optimal value is $\tilde{l} \to 1/2$. In general, for moderately large values of $\tilde{v}$ the optimal value of $\tilde{l}$ is in the range $1/2 < \tilde{l} < 1$ (see Table I). All of these conclusions are straightforward to derive from the asymptotic approximations of Eq. (17) and are confirmed by computing the exact expressions.

It is instructive that the optimal chain configuration maximizing the excess work $\langle \Delta W \rangle$ in model II in the limit of infinitely fast loading is the same as the optimal configuration predicted by model I. The fast pulling limit of model II, where a cross-link rupture is unlikely until the internal force attains a sufficiently large value, $f \geqslant f_c$, can be roughly approximated by model I. The two models however do not become equivalent in this limit: The unfolding force for a single cross-link is independent of the chain configuration and equal to a constant value of f_c in model I while it depends on both on the loading rate and the cross-link location in model II.

B. Small number of cross-links

Model I. This case has been studied in detail in Ref. [13]. The key result is that the same optimal configurations maximize both F_m and ΔW. Those configurations involve "parallel strands" of the form $C_N = \{\{i_1, j_1\}, \{i_2, j_2\}, \ldots, \{i_N, j_N\}\}$ such that $i_1 < i_2 < \cdots i_N < j_1 < j_2 < \cdots < j_N$. For example, for $N=3$ and $L=50$ the optimal configurations have the form $\{\{i, i+l\}, \{i+1, i+l+2\}, \{i+3, i+l+3\}\}$ where $l=26$ (see Fig. 2). Note that the optimal value of l is $l \approx L/2$, which is similar to that found in the case of a single cross-link.

Further, we showed that optimality can be understood in terms of a continuous "super cross-link" (SCL) model. In the

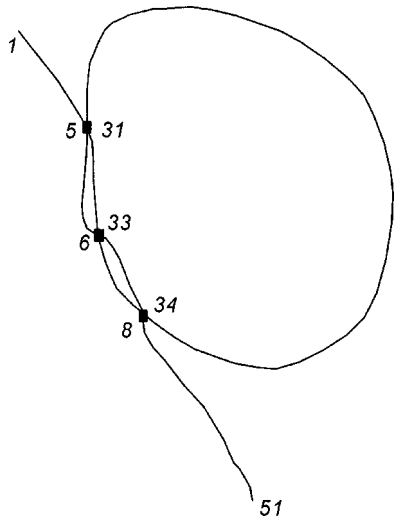

FIG. 2. An optimal NSCL configuration of an $L=50$ chain with $N=3$ cross-links. Within model I, this configuration optimizes both ΔW and F_m. In general, the optimal configurations have the form $\{\{i,i+l\},\{i+1,i+2+l\},\{i+3,i+3+l\}\}$ where l is the loop length. For model II, the loop length l that optimizes $\langle \Delta W \rangle$ is a function of the pulling velocity v while $\langle F_m \rangle$ is optimized by $l=47$ regardless of the pulling velocity.

limit as the chain becomes continuous, that is $L \to \infty$ and $b \to 0$, the topological constraint that any bead can be connected to only one cross-link can be relaxed because, as far as the mechanical response is concerned, neighboring beads become indistinguishable. Therefore, one can create a SCL by placing all N cross-links between the same points, $\{i,i+l\}$. Then the cross-links share the load equally so that the force in each cross-link is F/N, and the SCL acts like a single cross-link that can sustain a maximum force of $F_m = Nf_c$ resulting in an excess work of unfolding equal to [cf. Eq. (14)]

$$\Delta W = \frac{N^2 f_c^2}{2\Gamma_0}(\tilde{l} - \tilde{l}^2).$$

As in the case of $N=1$, the maximum ΔW is achieved when $\tilde{l}=1/2$.

For a discrete chain, we cannot achieve the SCL configurations because of the imposed constraint prohibiting multiple cross-links between the same monomers. Nevertheless, it turns out that the constrained optimal solutions are very close to the SCL's, and they involve parallel strands. We refer to such configurations as "nearly super cross-links" or NSCL's (Fig. 2). The force in each of the cross-links in the NSCL configuration is approximately the same. Further, within model I, rupture of one cross-link in an NSCL configuration results in an increase of the force in each of the remaining cross-links such that NSCL's rupture in an avalanche-like fashion. Because of that the force vs displacement curve $F(e)$ has only a single maximum, similarly to the case of a single cross-link.

Model II. Remarkably, we found that the NSCL configurations appear to be optimal with respect to both $\langle F_m \rangle$ and $\langle \Delta W \rangle$, although the configurations optimal for $\langle F_m \rangle$ are not necessarily optimal for $\langle \Delta W \rangle$, and vice versa. This statement is difficult to verify conclusively, because even for $N=3$ the search space is too large for an exhaustive search. Nevertheless, using the search algorithm described in Sec. III C, we could not find a configuration better than the NSCL of the form $\{\{i,i+l\},\{i+1,i+2+l\},\{i+3,i+3+l\}\}$, where the optimal value of l was determined by the exhaustive search with respect to l. The optimal values of l maximizing $\langle F_m \rangle$ and $\langle \Delta W \rangle$ were different, which is similar to the conclusion reached with model II for $N=1$. Furthermore, the values of $\tilde{l}=l/L$ that optimize $\langle F_m \rangle$ are close to $\tilde{l}=1$ and the optimal values of $\tilde{l}$ that maximize $\langle \Delta W \rangle$ depend on $\tilde{v}$ in a way similar to the case of $N=1$ (see Table II). We also found that $\langle F_m \rangle$ and $\langle \Delta W \rangle$ grow logarithmically with $\tilde{v}$ (Fig. 3).

An attempt to predict the response of NSCL configurations using the rate-dependent SCL model was only partially successful. In particular, the rate-dependent SCL model was able to follow the trends predicted by the simulations but the agreement was mostly qualitative. Furthermore, the predictions of the rate-dependent SCL model were qualitatively similar to those obtained from the analysis for $N=1$. Let us mention that the rate-dependent SCL model was successful in predicting the first but not the last rupture events, especially for intermediate loading rates. In the limit $\tilde{v} \to \infty$, one can use the asymptotic approximations developed for $N=1$, with the provision that k_0 and f_c are replaced with Nk_0 and Nf_c, respectively.

C. Large number of cross-links

For $N \ll L$, F_m and ΔW are proportional to N and N^2, respectively. Preliminary computations [13] have suggested that these scaling rules do not hold for large N, as both F_m and ΔW tend to saturate with increasing N.

Here we study in more detail the case where each bead is connected to another bead so that the total number of cross-links is $N=L/2$ (for an even L) or $(L+1)/2$ (for an odd L). In

TABLE II. The NSCL configuration made of three cross-links: The dimensionless loop length $\tilde{l}_{\Delta W}$ that maximizes $\langle \Delta W \rangle$ and the dimensionless loop length $\tilde{l}_{F_m}$ that maximizes $\langle F_m \rangle$, as functions of the dimensionless velocity $\tilde{v}$.

$\tilde{v}$	0.2	1	2	10	20	30	40	60	100
$\tilde{l}_{\Delta W}$	0.94	0.88	0.84	0.8	0.76	0.72	0.7	0.68	0.66
$\tilde{l}_F$	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94

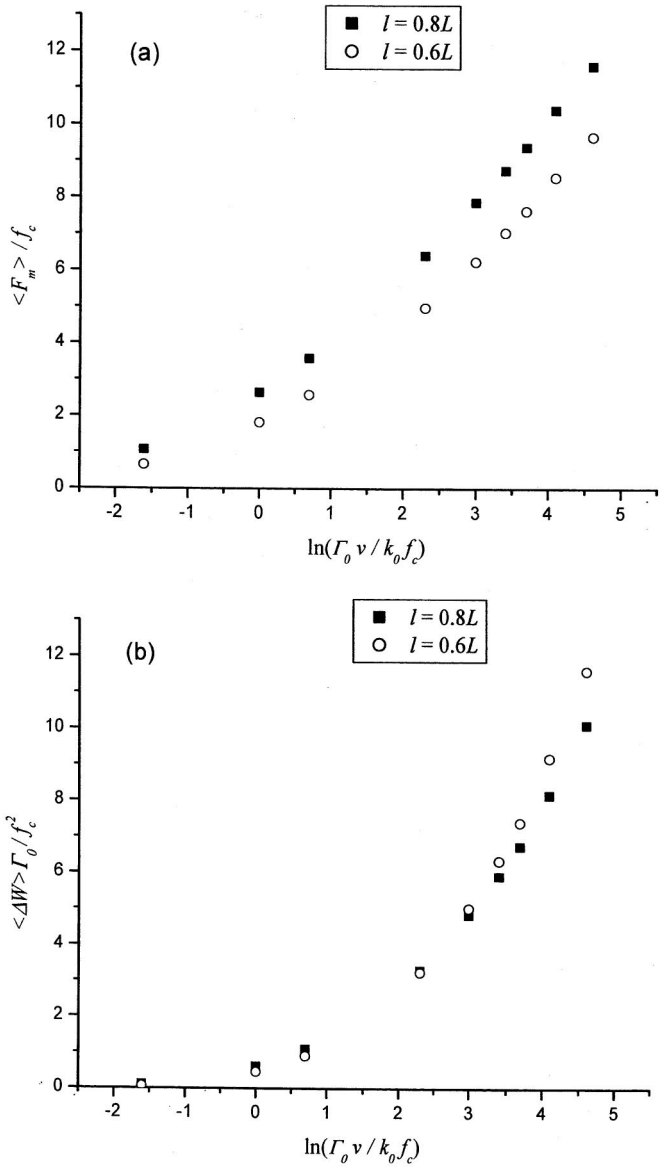

FIG. 3. (a) The maximum force $\langle F_m \rangle$ and (b) the excess work $\langle \Delta W \rangle$ as a function of the pulling rate for NSCL configurations with different values of the loop length l.

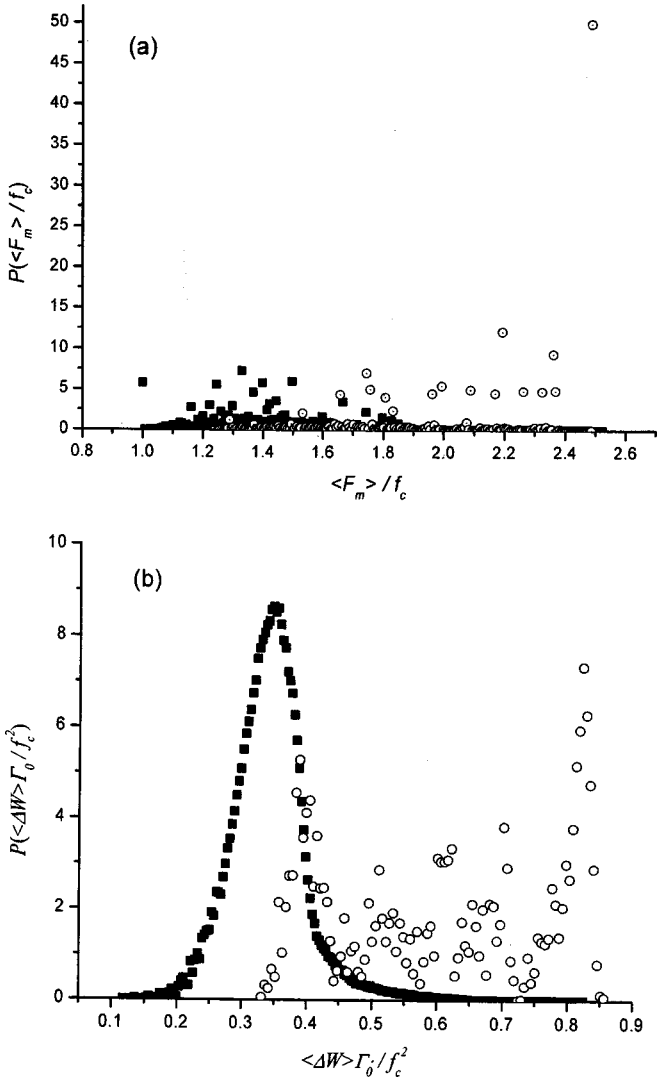

FIG. 4. Probability distributions for (a) F_m and (b) ΔW for randomly generated configurations containing $(L+1)/2$ cross-links ($L=19$), and configurations including the clamp C_3^* with the remaining 7 cross-links generated randomly. The fully random configurations are denoted by the squares and those containing the clamp by the circles.

this case, the search space is large and for this reason we limited our analysis to short chains, $L=19$, and to using model I only. The key result of our computations can be stated as follows.

(a) All optimal configurations contained the subset of three cross-links

$$C_3^* = \{\{i, i+L/2\}, \{i+1, i+3+L/2\}, \{i+4, i+4+L/2\}\},$$

which, again, is a "clamp" of parallel strands. The excess work for the configuration C_3^* in the absence of any other cross-links is equal to $\Delta W^* = 0.79 f_c^2/\Gamma_0$.

(b) By adding seven random cross-links to the clamp one is more likely to reduce than to increase ΔW in comparison to ΔW^*.

(c) The maximum ΔW is $\Delta W_m = 0.93 f_c^2/\Gamma_0$, corresponding to the configuration

$$C_m^{(10)} = \{\{1,15\},\{2,11\},\{3,16\},\{4,14\},\{5,17\},\{6,10\},\{7,9\},$$
$$\{8,18\},\{13,20\},\{12,19\}\},$$

which also maximizes F_m.

(d) The mean value of toughness for randomly generated cross-link configurations is $\Delta \overline{W} \approx 0.35 f_c^2/\Gamma_0$, and only a small fraction of configurations have the toughness close to ΔW_m.

These results are further illustrated in Fig. 4, where we plot the histograms for F_m and ΔW corresponding to randomly generated cross-link configurations and configurations containing the subset C_3^*. The latter, on the average, have larger values of both F_m and ΔW, as compared to random cross-link arrangements. However, adding random cross-links to C_3^* does not necessarily improve the mechanical resistance of

the chain: only a relatively small fraction of such configurations perform better than C_3^*.

V. DISCUSSION: IMPLICATIONS FOR FORCE-INDUCED PROTEIN UNFOLDING EXPERIMENTS

Cohesive interactions in proteins are delocalized and thus rarely can be adequately described as cross-links. For this reason, we expect our model not to make quantitative predictions but rather to provide a guide to the relationship between the overall fold topology and its mechanical response. In certain situations disulphide bonds, hydrogen bonds, or groups of hydrogen bonds in proteins can be modeled as cross-links [26,28]. It may further be possible to synthesize cross-linked polymers, in which cross-links are placed in a controlled fashion. Such polymers could provide an experimental test ground of our theory and also exhibit novel mechanical properties.

Our results can be used to screen the protein databank to identify the proteins that exhibit the topology that may potentially lead to optimal mechanical stability. While this approach has not been pursued systematically yet, there is evidence that it may result in useful predictions. In particular, the mechanical unfolding of the immunoglobulin domain 127, ubiquitin, and protein G—all containing terminal parallel strands—has been observed [12,19,20] and/or predicted via atomistic simulations [30] to require forces much higher than those in the case of "generic," nonmechanical proteins [9]. This is in accord with the conclusion reached here that configurations involving parallel strands are optimal with respect to the unfolding force and work. We have arrived at the same conclusion in our previous study [13] where we used model I thus ignoring the statistical nature of bond rupture [32–34]. The present study demonstrates that rate dependent effects that are well known to be important in force probe spectroscopy pulling experiments [7,8,19,21,24,54–60] do not change the conclusion about the optimality of parallel strands.

In addition, several other observations may be of relevance in the context of mechanical stability of proteins

(1) For sufficiently slow pulling rates, parallel strands formed between the *ends of the chain* (i.e., those with $l \simeq L$) lead to higher values for both F_m and ΔW. In contrast, for very high pulling rates, parallel strands with $l \simeq L/2$ are optimal with respect to ΔW while terminal parallel strands still maximize the unfolding force. Since the loading rates are expected to be slower under physiological conditions as compared to AFM experiments or molecular dynamics simulations (see below), the above observations may explain why most protein domains with superior mechanical properties contain *terminal* parallel β strands [9–12,25–27,30,31,37].

(2) The configurations that include the optimal NSCL configurations are superior to random structures (see Sec. IV C). This result may shed light on the recent finding that proteins with different folds may display similar mechanical resistance. In particular, the unfolding mechanisms of the all-β I27 domain of the muscle protein titin [19,20,25,26] and of the α/β ubiquitin domain [12,30] are very similar and are characterized by a high unfolding force because both of these domains feature the same hydrogen-bond clamp formed by their terminal parallel strands.

(3) Adding random cross-links to an optimal NSCL configuration can be viewed to some extent as a way to mimic the effect of nonnative interactions in our Go-like model. As seen in Fig. 4, these interactions can both reduce and enhance the resistance of the chain to the mechanical unfolding. This suggests that given the native topology, further optimization with respect to the protein's mechanical stability can be achieved via mutations that alter non-native interactions [59].

Our study may also elucidate the effect of the loading regime on the mechanical function of proteins. Under physiological conditions, proteins are subjected to forces that are often quite different from those in AFM studies and/or simulations. Likewise, the timescales at which they are loaded are different from those of pulling experiments. For example, in AFM studies of the muscle protein titin [19,20], individual immunoglobulin-like domains are unfolded in the range of forces $f_u \sim 150$–250 pN, depending on the stretching rate that is typically in the range of $v = 0.1$–10 nm/ms. The rate of loading in these experiments can be roughly estimated as

$$df/dt \simeq f_u/\tau \simeq f_u v/\Delta l,$$

where the domain stretching time τ is estimated as $\Delta l/v$ and Δl is the contour length of the domain. This gives $df/dt \sim 10^{-9}$–10^{-7} (N/s) for $v = 0.1$–10 nm/ms.

By contrast, in the experiments that probe viscoelastic behavior of skeletal myofibrils [61], individual domains are subjected to much lower forces ($f_u \sim 10$ pN) over a timescale of a few seconds and their unfolding events are rare (yet believed to be physiologically important [61]). Using $\tau \sim 1$–10 s, this gives a loading rate of $f_u/\tau \sim 10^{-12}$–10^{-11} (N/s), several orders of magnitude lower than that in AFM experiments.

To make connection to the present study, consider the dimensionless loading rate parameter θ introduced in Sec. IV:

$$\theta = \frac{\Gamma v}{k_0 f_c}, \tag{18a}$$

where Γ is the overall spring constant of the chain prior to rupture. This parameter characterizes the loading timescale relative to that of internal dynamics of the chain. The optimal chain configurations are generally different in the limits $\theta \gg 1$ and $\theta \ll 1$ (cf. Tables I and II). To estimate this parameter for a protein domain undergoing unfolding via the two-state mechanism, we rewrite Eq. (18a) in the form

$$\theta = \frac{df/dt}{k_0 f_c}. \tag{18b}$$

Equation (18b) is more informative because in most AFM studies the domain of interest is part of a longer chain; the velocity at which the ends of the *entire* chain are separated is different from the speed at which the ends of the domain are moved apart and thus the linkage between the domain and the pulling device affects the unfolding dynamics [33,34].

Equation (18b) is written in the form that is independent of the linkage.

Using the values $k_0 = 5 \times 10^{-4}\ \mathrm{s}^{-1}$, $f_c = 16$ pN deduced from the AFM data for the I27 domain [60] we estimate $\theta = 10^2 - 10^3$ for the experiments that probe titin viscoelasticity [61] and $\theta = 10^5 - 10^7$ for AFM experiments.

Given the above difference in the loading rates, what can we learn about biological function of load-bearing proteins from AFM pulling studies? Our study suggests that AFM data can be extrapolated to lower loading rates. Specifically, we have shown the following.

(a) Configurations containing parallel strands are optimal for both slow ($\theta \ll 1$) and fast ($\theta \gg 1$) loading. Further, the configuration optimal with respect to the unfolding force is the same in both regimes.

(b) Unlike the case of generic random configurations, the behavior of optimal configurations is close to that predicted by a simple two-state model over a wide range of loading rates. In particular, the unfolding force exhibits logarithmic dependence on the loading rate, similar to that derived from the two-state model. This supports the use of two-state models [20] to extrapolate AFM data outside the range of experimental loading rates.

ACKNOWLEDGMENTS

This work was supported by grants from the Robert A. Welch Foundation and ACS Petroleum Research Fund and by the National Science Foundation (grants CHE-0347862 and CMS-0219839 to D.E.M. and G.J.R.).

[1] E. Oroudjev, J. Soares, S. Arcdiacono, J. Thompson, S. A. Fossey, and H. G. Hansma, Proc. Natl. Acad. Sci. U.S.A. **99**, 6460 (2002).
[2] B. L. Smith, T. E. Schaffer, M. Viani, J. B. Thompson, N. A. Frederick, J. Kindt, A. Belcher, G. D. Stucky, D. E. Morse, and P. K. Hansma, Nature (London) **399**, 761 (1999).
[3] J. B. Thompson, J. H. Kindt, B. Drake, H. G. Hansma, D. E. Morse, and P. K. Hansma, Nature (London) **414**, 773 (2001).
[4] N. Becker, E. Oroudjev, S. Mutz, J. P. Cleveland, P. K. Hansma, C. Y. Hayashi, D. E. Makarov, and H. G. Hansma, Nat. Mater. **2**, 278 (2003).
[5] I. Schwaiger, C. Sattler, D. R. Hostetter, and M. Rief, Nat. Mater. **1**, 232 (2002).
[6] H. G. Hansma, L. I. Pietrasanta, I. D. Auerbach, C. Sorenson, R. Golan, and P. A. Holden, J. Biomater. Sci., Polym. Ed. **11**, 675 (2000).
[7] R. B. Best, D. J. Brockwell, J. L. Toca-Herrera, A. W. Blake, D. A. Smith, S. E. Radford, and J. Clarke, Anal. Chim. Acta **479**, 87 (2003).
[8] R. Lavery, A. Lebrun, J.-F. Allemand, D. Bensimon, and V. Croquette, J. Phys.: Condens. Matter **14**, R383 (2002).
[9] R. B. Best, B. Li, A. Steward, V. Daggett, and J. Clarke, Biophys. J. **81**, 2344 (2001).
[10] D. J. Brockwell, G. S. Beddard, J. Clarkson, R. C. Zinober, A. Blake, J. Trinick, P. D. Olmsted, D. A. Smith, and S. E. Radford, Biophys. J. **83**, 458 (2002).
[11] D. J. Brockwell, E. Paci, R. C. Zinober, G. S. Beddard, P. D. Olmsted, D. A. Smith, R. N. Perham, and S. E. Radford, Nat. Struct. Biol. **10**, 731 (2003).
[12] M. Carrion-Vazquez, H. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser, and J. M. Fernandez, Nat. Struct. Biol. **10**, 738 (2003).
[13] K. Eom, P.-C. Li, D. E. Makarov, and G. J. Rodin, J. Phys. Chem. B **107**, 8730 (2003).
[14] D. K. Klimov and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **97**, 7254 (2000).
[15] T. Shen, L. S. Canino, and J. A. McCammon, Phys. Rev. Lett. **89**, 068103 (2002).
[16] M. Cieplak, T. X. Hoang, and M. O. Robbins, Proteins: Struct., Funct., Genet. **49**, 104 (2002).
[17] M. Cieplak, T. X. Hoang, and M. O. Robbins, Proteins: Struct., Funct., Genet. **49**, 114 (2002).
[18] M. Cieplak, T. X. Hoang, and M. O. Robbins, Proteins: Struct., Funct., Bioinf. **56**, 285 (2004).
[19] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, Science **276**, 1109 (1997).
[20] M. Rief, J. M. Fernandez, and H. E. Gaub, Phys. Rev. Lett. **81**, 4764 (1998).
[21] A. F. Oberhauser, P. K. Hansma, M. Carrion-Vazquez, and J. M. Fernandez, Proc. Natl. Acad. Sci. U.S.A. **98**, 468 (2001).
[22] M. E. Pennisi, Science **283**, 168 (1999).
[23] H. P. Erickson, Science **276**, 1090 (1997).
[24] T. E. Fisher, A. F. Oberhauser, M. C. Vezquez, P. E. Marsalek, and J. Fernandez, Trends Biochem. Sci. **24**, 379 (1999).
[25] H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten, Biophys. J. **75**, 662 (1998).
[26] H. Lu and K. Schulten, Chem. Phys. **247**, 141 (1999).
[27] B. Isralewitz, M. Gao, and K. Schulten, Curr. Opin. Struct. Biol. **11**, 224 (2001).
[28] P. E. Marsalek, H. Lu, H. Li, M. Carrion-Vazquez, A. F. Oberhauser, K. Schulten, and J. Fernandez, Nature (London) **402**, 100 (1999).
[29] R. Rohs, C. Etchebest, and R. Lavery, Biophys. J. **76**, 2760 (1999).
[30] P.-C. Li and D. E. Makarov, J. Phys. Chem. B **108**, 745 (2004).
[31] P.-C. Li and D. E. Makarov, J. Chem. Phys. **121**, 4826 (2004).
[32] G. I. Bell, Science **200**, 618 (1978).
[33] E. Evans and K. Ritchie, Biophys. J. **72**, 1541 (1997).
[34] E. Evans and K. Ritchie, Biophys. J. **76**, 2439 (1999).
[35] D. E. Makarov, P. K. Hansma, and H. Metiu, J. Chem. Phys. **114**, 9663 (2001).
[36] H. H. Kausch, *Polymer Fracture*, 2nd ed. (Springer-Verlag, Berlin, 1987).
[37] P.-C. Li and D. E. Makarov, J. Chem. Phys. **119**, 9260 (2003).
[38] D. E. Makarov and H. Metiu, J. Chem. Phys. **116**, 5205 (2002).
[39] D. E. Makarov, C. Keller, K. W. Plaxco, and H. Metiu, Proc. Natl. Acad. Sci. U.S.A. **99**, 3535 (2002).
[40] U. Gerland, R. Bundschun, and T. Hwa, Biophys. J. **81**, 1324 (2001).

[41] U. Gerland, R. Bundschun, and T. Hwa, Biophys. J. **84**, 2831 (2003).
[42] D. K. Lubensky and D. R. Nelson, Phys. Rev. Lett. **85**, 1572 (2000).
[43] D. K. Lubensky and D. R. Nelson, Phys. Rev. E **65**, 031917 (2002).
[44] A. Montanari and M. Mezard, Phys. Rev. Lett. **86**, 2178 (2001).
[45] E. I. Shakhnovich and P. L. Geissler, Macromolecules **35**, 4429 (2002).
[46] N. Lee and T. A. Vilgis, Europhys. Lett. **57**, 817 (2002).
[47] E. Jarkova, N.-K. Lee, and S. Obukhov, Macromolecules (to be published).
[48] P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, Ithaca and London, 1953).
[49] P. J. Flory, J. Am. Chem. Soc. **78**, 5222 (1956).
[50] D. E. Makarov and G. J. Rodin, Phys. Rev. E **66**, 011908 (2002).
[51] A. F. Voter, Phys. Rev. B **34**, 6819 (1986).
[52] H. Metiu, Y.-T. Lu, and Z. Zhang, Science **255**, 1088 (1992).
[53] Z. Zhang, K. Haug, and H. Metiu, J. Chem. Phys. **93**, 3614 (1990).
[54] M. Rief, J. Pascual, M. Saraste, and H. E. Gaub, J. Mol. Biol. **286**, 553 (1999).
[55] R. C. Zinober, D. J. Brockwell, G. S. Beddard, A. W. Blake, P. D. Olmsted, S. E. Radford, and D. A. Smith, Protein Sci. **11**, 2759 (2002).
[56] A. F. Oberhauser, P. E. Marszalek, H. Erickson, and J. M. Fernandez, Nature (London) **393**, 181 (1998).
[57] A. F. Oberhauser, C. Badilla-Fernandez, M. Carrion-Vazquez, and J. M. Fernandez, J. Mol. Biol. **319**, 433 (2002).
[58] H. Li, A. F. Oberhauser, S. B. Fowler, J. Clarke, and J. M. Fernandez, Proc. Natl. Acad. Sci. U.S.A. **97**, 6527 (2000).
[59] H. Li, M. Carrion-Vazquez, A. F. Oberhauser, P. E. Marsalek, and J. M. Fernandez, Nat. Struct. Biol. **7**, 1117 (2000).
[60] M. Carrion-Vazquez, A. F. Oberhauser, S. B. Fowler, P. E. Marsalek, S. E. Broedel, J. Clarke, and J. M. Fernandez, Proc. Natl. Acad. Sci. U.S.A. **96**, 3694 (1999).
[61] A. Minajeva, M. Kulke, J. M. Fernandez, and W. A. Linke, Biophys. J. **80**, 1442 (2001).