# A Theoretical Model for the Mechanical Unfolding of Repeat Proteins

Dmitrii E. Makarov*

Department of Chemistry and Biochemistry and Institute for Theoretical Chemistry, University of Texas at Austin, Austin, Texas

ABSTRACT   We consider the mechanical stretching of a polypeptide chain formed by multiple interacting repeats. The folding thermodynamics and the interactions among the repeats are described by the Ising model. Unfolded repeats act as soft entropic springs, whereas folded repeats respond to a force as stiffer springs. We show that the resulting force-extension curve may exhibit a pronounced force maximum corresponding to the unfolding of the first repeat. This event is followed by the unfolding of the remaining repeats, which takes place at a lower force. As the protein extension is increased, the force-extension curve of a sufficiently long repeat protein displays a plateau, where the force remains nearly constant and the protein unfolds sequentially so that the number of unfolded repeats is proportional to the extension. Such a sequential mechanical unfolding mechanism is displayed even by the repeat proteins whose thermal denaturation is highly cooperative, provided that they are long enough. By contrast, the unfolding of short repeat progressions can be cooperative.

## INTRODUCTION

Proteins containing repeated folded units of the same or similar structure are common in nature (1–4). Their folding kinetics and thermodynamics have attracted considerable attention (1–5). As certain repeat proteins are believed to have a mechanical function (6), their mechanical response has also received recent attention. In particular, two experimental groups have used AFM to pull on individual ankyrin repeat proteins (7,8). The purpose of this article is to present a simple theoretical model describing the mechanical response of such proteins. The model is shown in Fig. 1. Each subunit can be in either a folded or an unfolded conformation. Unfolded subunits respond to the force as soft, entropic springs. Folded subunits act like springs that have a higher stiffness. We are particularly interested in the effect of the interactions among the subunits on the mechanical response of the entire chain. To describe those, we adopt an Ising-type model (2,4), which has already been applied to describe the folding of repeat proteins.

Our model is related to those of Buhot and Halperin (9) and Tamashiro and Pincus (10), who considered the stretching of a polymer undergoing a helix-coil transition. This transition was described using the Zimm-Bragg model (11,12), in which each monomer is in either a helical or a coil conformation. The model presented here uses a similar two-state representation for each repeat rather than for each individual monomer. Mathematically, the Zimm-Bragg model is equivalent to the Ising model, and thus, many results reported by others (9,10) can be adapted for our purpose. However, the situation considered there differs from this study in a number of ways. First, whereas in those studies (9,10) helical segments were considered to be rigid rods, it is important in the context of our study to consider

the elasticity of individual folded repeats. Second, those works (9,10) focused on the infinite chain limit, $N \to \infty$. In contrast, typical repeat proteins only contain a finite number of repeats, often $N \sim 5–30$. As will be seen later, finite-size effects are responsible for a pronounced peak in the force-extension curves of certain repeat proteins. This peak would disappear in the limit of an infinitely long chain. Third, individual helical residues are unstable, so a sufficiently long helix has to be nucleated for the transition to the helical state to occur (11,12). In contrast, a more general situation, where individual repeats can be stable on their own, will be considered here.

We show here that if the ends of a repeat protein chain are pulled apart, its unfolding can proceed according to two scenarios. In both scenarios, the chain behaves as a stiff spring at low extensions. In the first scenario, a pronounced peak in the dependence of the force, $f$, on the extension, $z$, is observed when the first repeat becomes unfolded. After this event, the chain yields at a much lower force and the remaining repeats unfold one after another. The second scenario is similar to that found in the studies of the helix-coil transition (9,10): the force-extension curve is monotonic and levels out at a certain plateau value.

We further show that although interactions among repeats can lead to highly cooperative thermal/chemical denaturation (3), mechanical denaturation of the same protein may remain noncooperative and involve sequential unfolding of repeats one at a time. However because short progressions of folded repeats are often thermodynamically unstable, simultaneous unfolding of several repeats may be observed at the end of a pulling experiment.

The rest of this article is organized as follows. In the Model section, we describe the model. This is followed by the Results section and then the Discussion section, where we highlight key findings and discuss their implications for single-molecule AFM pulling experiments. We conclude
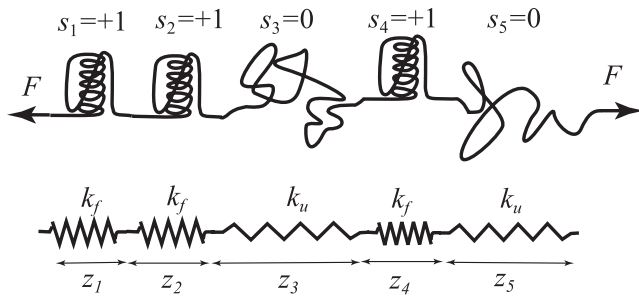
FIGURE 1 Illustration of the model used. The protein is described as a sequence of folded ($s_i = +1$) and unfolded ($s_i = 0$) repeat units. The folded and unfolded repeats behave as Hookean springs with stiffnesses $k_f$ and $k_u$, respectively.

with a summary of the results and their potential generalizations.

## MODEL

We consider a chain of $N$ identical repeats shown in Fig. 1. Each repeat can be in a folded state, f, or unfolded state, u. The (free) energy of the repeats is given by an Ising-like expression:

$$E[\{s_i\}]/k_B T = -\sum_{i=1}^{N} H s_i - J \sum_{i=1}^{N-1} s_i s_{i+1}. \qquad (1)$$

Here, $s_i = 0$ if the $i$th repeat is unfolded and $s_i = 1$ if it is folded. The parameter $H$ is the free energy of folding for an isolated repeat, and the parameter $J > 0$ represents the coupling between two adjacent repeats, which stabilizes the folded conformation of each. We will assume that parameters $H$ and $J$ are such that in the absence of stretching, the repeats are predominantly folded. In other words, the expectation value $\langle s_i \rangle$ is close to 1.

Equation 1 does not contain the free energy associated with the stretching of the chain. To include this, consider the ensemble of the chain conformations with the given value of the overall extension,

$$z = z_1 + z_2 + \dots + z_N, \qquad (2)$$

which is measured in the $z$ direction (here, this direction will coincide with the direction of the stretching force). Then, the free energy of the chain is (9,10,13,14)

$$G = -k_B T \ln p(z), \qquad (3)$$

where the probability distribution of $z$ is given by

$$p(z) = \sum_{\{s_i\}} \exp\left(-\frac{E[\{s_i\}]}{k_B T}\right) p_{\{s_i\}}(z) / \sum_{\{s_i\}} \exp\left(-\frac{E[\{s_i\}]}{k_B T}\right) \qquad (4)$$

Here $p_{\{s_i\}}(z)$ is the probability distribution of $z$ with the variables $\{s_i\}$ fixed. This distribution can be expressed in terms of the probability distributions $p_{s_i}(z_i)$ for the extensions $z_i$ of individual repeats as

$$p_{\{s_i\}}(z) = \int dz_1 dz_2 \dots dz_N p_{s_1}(z_1) p_{s_2}(z_2) \dots p_{s_N}(z_N) \\ \times \delta(z - z_1 - z_2 - \dots z_N). \qquad (5)$$

Note that Eqs. 4 and 5 take advantage of the assumption that the interdomain interaction energy (Eq. 1) is a function of the internal variables, $\{s_i\}$, only and does not depend on $\{z_i\}$, thus allowing the factorization of the integrand in Eq. 5.

We further assume that each unfolded repeat is a Gaussian chain, so that for $s_i = 0$, we have

$$p_0(z) = \frac{1}{\sqrt{2\pi b_u^2}} \exp\left(-\frac{z^2}{2 b_u^2}\right), \qquad (6)$$

where $b_u$ is the root mean-squared extension. Thus, it is mechanically equivalent to a Hookean spring with a spring constant $k_u = k_B T / b_u^2$. Likewise, we will assume that each folded repeat acts as a stiffer Hookean spring with a stiffness of $k_f \equiv k_B T / b_f^2$, so that for $s_i = 1$, we have

$$p_1(z) = \frac{1}{\sqrt{2\pi b_f^2}} \exp\left(-\frac{z^2}{2 b_f^2}\right). \qquad (7)$$

With these assumptions, the integral of Eq. 5 can be evaluated to give

$$p_{\{s_i\}}(z) = \frac{1}{\sqrt{2\pi\left(n_f b_f^2 + n_u b_u^2\right)}} \exp\left(-\frac{z^2}{2\left(n_f b_f^2 + n_u b_u^2\right)}\right), (8)$$

Here $n_f = \sum_{i=1}^{N} s_i$ and $n_u = N - \sum_{i=1}^{N} s_i$ are, respectively, the numbers of folded and unfolded repeats in the chain.

Although the Gaussian chain approximation for unfolded repeats is reasonable as long as their extension is much shorter than their contour length, the assumptions made in Eq. 7 deserve further discussion. This equation effectively ignores that the end-to-end distance for each folded repeat is finite. More generally, we could model those repeats as deformable solid objects. If we were to do so, there would be additional entropic elasticity associated with aligning them in the direction of the force, which would be the dominant contribution to the overall chain elasticity at very low forces (10). Because adjacent folded units in repeat proteins often form superstructures, it may be more realistic in certain cases to model such superstructures, as opposed to individual repeats, as elastic objects, in which case the factorization assumption of Eq. 5 would be violated. These choices would depend on details of the specific protein under study and could be straightforwardly incorporated in our model if desired. Here, we have chosen the simplest possible model that accounts for the difference in the elasticity of folded and unfolded repeats. In an earlier study (13), the extension of folded repeats was neglected altogether. In contrast, two studies (9,10) dealing with polymers that exhibit helix-coil transition treated the helical segments as inextensible rods whose contribution to the overall chain elasticity was only through the aforementioned entropic effect associated with the alignment of the rods along the direction of the force.

## RESULTS

### Analytical approximations

For sufficiently small $N$, Eq. 4 can be computed through exact enumeration. Before showing such exact results, we present here simple analytical approximations, which turn out to capture much of the physics of the problem.

Suppose that of the $N$ repeats, $n_f$ are folded and $n_u = N - n_f$ are unfolded. Further suppose that the number of contiguous blocks of folded repeats is $n$. In other words, the sequence $\{s_i\}$ contains $n$ uninterrupted subsequences of the form $111\dots1$ with the total number of 1s adding up to $n_f$. A single folded repeat counts as a contiguous block, and the number of such repeats is included in $n$. For example, in Fig. 1, $n_f = 3$ and $n = 2$. The free energy of Eq. 1 can then be rewritten in the form

$$E/k_B T = -(H + J)n_f + Jn. \tag{9}$$

If one starts with a fully folded conformation, unfolding an outer repeat ($i = 1$ or $i = N$) will reduce $n_f$ by 1 and thus increase the energy by $J + H$. Likewise, the energy is increased by $J + H$ whenever a repeat at either end of a contiguous folded block (containing more than one repeat) is unfolded. By contrast, the unfolding of an inner repeat within a contiguous folded block will change the energy by $2J + H$, because $n$ will increase from 1 to 2. Using the terminology of Buhot and Halperin (9), we can think of $J$ as the interfacial energy associated with creating a pair of "domain boundaries" between sequences of folded and unfolded repeats, whereas $\Delta G = J + H$ can be regarded as the free energy cost of unfolding a repeat without creating additional domain boundaries (i.e., by propagating an existing domain boundary). For any given $n_f$, $n$, and chain extension, $z$, the free energy of the chain can be written as

$$G(z, n_f, n)/k_B T \approx \frac{z^2}{2\left[n_f b_f^2 + (N - n_f) b_u^2\right]} - (H + J)n_f$$
$$+ Jn - S(n_f, n). \tag{10}$$

Here, the first term is the elastic part corresponding to Eq. 8, and $S(n_f, n) = k_B \ln \Omega(n_f, n)$ is an entropy term that can be calculated by counting the number $\Omega(n_f, n)$ of the sequences $\{s_i\}$ that correspond to the same values $n_f$ and $n$. Equation 10 neglects the logarithmic corrections that result from the $n_f$ dependence of the prefactor in Eq. 8.

Following the arguments of Buhot and Halperin (9), we now neglect the entropy term in Eq. 10. The free energy for the protein, for a given value of the extension $z$, can then be estimated as

$$G(z)/k_B T$$
$$\approx \min_{n, n_f} \left\{ \frac{z^2}{2\left[n_f b_f^2 + (N - n_f) b_u^2\right]} - (H + J)n_f + Jn \right\} \tag{11}$$

If any folded repeats are present, the minimum with respect to $n$ is achieved at $n = 1$ ($n = 0$ would imply that the protein is completely unfolded, which would be inconsistent with a finite value of $n_f$). If $1 \ll n_f \ll N$, then we can treat $n_f$ as a continuous variable, which gives the condition $\partial G(z, n_f, n)/\partial n_f = 0$, or

$$n_f = N \frac{b_u^2}{b_u^2 - b_f^2} - \frac{z}{\sqrt{2(J + H)}} \frac{1}{\sqrt{b_u^2 - b_f^2}}. \tag{12}$$

Substituting this back into Eq. 11 gives

$$G(z)/k_B T = J + z\sqrt{\frac{2(J + H)}{b_u^2 - b_f^2}}. \tag{13}$$

This means that the repeat protein will yield at a force equal to

$$F_{yield} = G'(z) = k_B T \sqrt{\frac{2(J + H)}{b_u^2 - b_f^2}}. \tag{14}$$

The force-extension curve of the protein thus exhibits a plateau where the force $F = F_{yield}$ does not depend on the extension. This situation is similar to fluid-vapor coexistence, which is manifested by a plateau in the pressure-volume diagram. In our case, the coexistence is between the unfolded and folded conformations of the repeats (15).

The condition $0 < n_f \ll N$ results in the following range of acceptable extensions where this regime is valid:

$$z_{min} \ll z < z_{max},$$

$$z_{min} = N b_f^2 \sqrt{\frac{2(J + H)}{b_u^2 - b_f^2}}, \quad z_{max} = N b_u^2 \sqrt{\frac{2(J + H)}{b_u^2 - b_f^2}}. \tag{15}$$

When $z > z_{max}$, all the repeats are unfolded and the protein is an entropic spring with a spring constant equal to $k_B T/N b_u^2$.

Let us now consider the case of a nearly folded protein, where $n_f$ is close to $N$. Within this limit, the continuous approximation for $n_f$ breaks down. For small enough $z$, the minimum of Eq. 11 is achieved when $n = 1$ and $n_f = N$. This means that each repeat will likely remain folded, and thus the protein will respond to stretching as a linear spring with a stiffness equal to $k_f/N$. Thus, for small enough $z$,

$$G(z) = \frac{k_f z^2}{2N} \; (+ \; \text{constant}), \tag{16}$$

and the corresponding force-extension curve is Hook's law:

$$F = G'(z) = (k_f/N)z. \tag{17}$$

As $z$ is increased beyond a certain critical $z_u$ value, the value $n_f = N - 1$ will provide the minimum to Eq. 11. This means that at $z \geq z_u$, one of the repeats will unfold. The value of $z_u$ is readily obtained from Eq. 11:

$$z_u = b_f \sqrt{2N(J + H)} \sqrt{1 + \frac{N b_f^2}{b_u^2 - b_f^2}}. \tag{18}$$

The corresponding rupture force is

$$F_u = k_f z_u/N = \frac{k_B T}{b_f} \sqrt{2(J + H)/N} \sqrt{1 + \frac{N b_f^2}{b_u^2 - b_f^2}}. \tag{19}$$

If we further assume that the discrete nature of $n_f$ is only important for $n_f = N, N - 1$, and that a continuous approximation is good for $n_f < N - 1$, then we arrive at a force-extension law that consists of a linear segment at short extensions, a plateau at longer extensions, and another linear segment corresponding to a fully unfolded protein:

$$F = G'(z) = \begin{cases} (k_f/N)z, & z < z_u \\ F_{yield}, & z_u \leq z < z_{max} \\ (k_u/N)z, & z \geq z_{max} \end{cases} \tag{20}$$
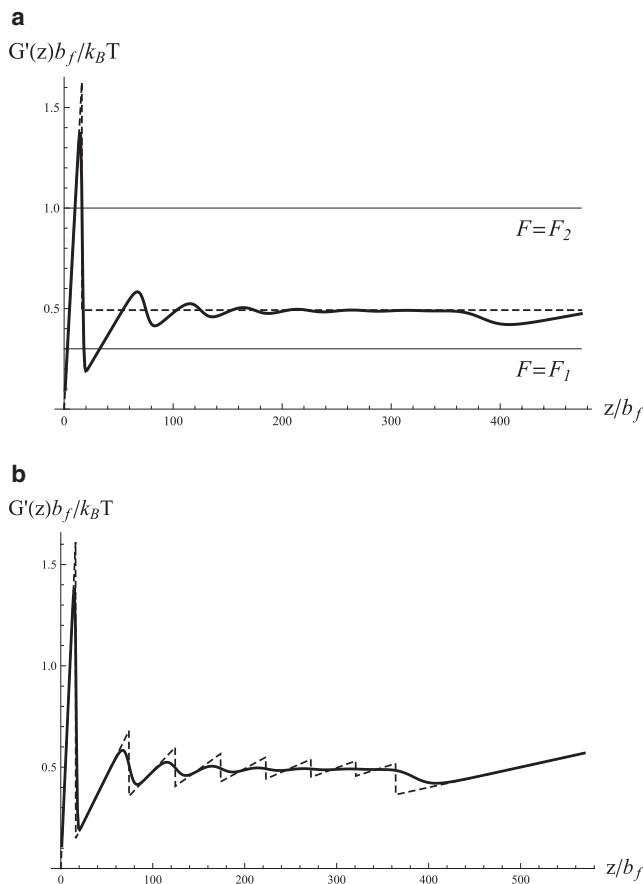
**a**



**b**



FIGURE 2 Force-extension curve of a repeat protein computed numerically and estimated using the approximations described in the first part of the Results section. The parameters of the model used are: $H = 4$, $J = 8$, $N = 10$, and $b_u = 10b_f$. (*a*) Numerical data (*solid line*) versus the approximation using Eq. 20. Two horizontal lines show two different values of the force, $F$. For $F = F_1$, the equation $G'(z) = F$ has three solutions (three crossings of the horizontal line), and for $F = F_2$, there are two solutions. (*b*) Numerical data (*solid line*) compared with the force-extension curve obtained directly from Eq. 11 without invoking a continuous approximation for the number of folded repeats (*dashed line*).

## Numerical results and their comparison with analytical approximations

Fig. 2 shows the dependence of the force $G'(z)$ on the extension $z$ calculated numerically from Eqs. 3 and 4 and by using the simple formulas of Eq. 20. The numerically computed force exhibits a large van der Waals loop, which is well approximated by the peak predicted by the first line of Eq. 20 (see Fig. 2 *a*). Smaller oscillations observed in Fig. 2 are not captured by Eq. 20, since the latter assumes that the number of repeats changes continuously. These oscillations are recovered if one uses Eq. 11 directly, without making a continuous $n_f$ assumption, as shown in Fig. 2 *b*. Consistent with the notion that for larger extensions we can treat the number $n_u$ of unfolded repeats as continuous, and so that the accuracy of Eq. 20 should become better, we observe the oscillations to decay and converge to a plateau value

that is well approximated by $F_{yield}$ given by Eq. 14 (Fig. 2 *a*, *dashed line*). However, when the extension is further increased, a drop in the force is observed before it begins to rise and follow the linear dependence predicted by the last line in Eq. 20. This drop will be explained in the next section.

One has to be careful interpreting the curve $F = G'(z)$ versus $z$ as the "true" force-extension law describing the protein's elasticity. When a constant stretching force, $F$, is applied to the protein, the latter experiences the effective potential $G_F(z) = G(z) - Fz$ (16–18). The points where $G_F'(z) = 0$, i.e., those described by the equation $F = G'(z)$, are the minima (if $G'' > 0$) or the maxima (if $G'' < 0$) of this potential. At low enough forces, there is only one solution satisfying $F = G'(z)$. The situation is different in the van der Waals loop region. At intermediate force values (see, e.g., the case $F = F_1$ in Fig. 2 *a*), this equation has three solutions. The smallest and the largest solutions correspond to free-energy minima and the one in the middle corresponds to the barrier separating those minima. Thus if the protein is held at constant tension $F_1$, it will undergo thermally activated hopping between the minima over the barrier, and the thermodynamic expectation value of $z$ will be an appropriate thermal average over the two basins of attraction. Such thermal hopping can indeed be observed in pulling experiments (19). If we continue to increase the force past the value $F_1$ we will find that more than three solutions, corresponding to multiple minima separated by barriers, are possible at certain force values. Finally, when $F$ becomes larger than a certain value (close to $F_{yield}$ from Eq. 14), there are only two solutions ($G'(z_l) = F$, $G'(z_h) = F$, $z_l < z_h$), as shown in Fig. 2 *a* for the case $F = F_2$. The lower value, $z_l$, corresponds to a metastable minimum, from which the system can escape over the barrier located at $z = z_h$. This means that the folded protein cannot sustain such a high force and will eventually unfold. Whether or not the force peak will be observed then depends on the pulling timescale relative to that of barrier crossing. If, for example, force $F$ is quickly increased, then the system may not have enough time to escape the metastable folded states, and thus the initial peak in the curve $G'(z)$ versus $z$ (or its part) will be observed. When van der Waals loops are traversed in nonequilibrium pulling experiments, the resulting unfolding dynamics are generally stochastic (20,21), exhibiting force peaks and drops corresponding to barrier crossings. The statistics of such events and their dependence on the pulling rate are beyond the scope of our model, which contains no information about kinetics.

## DISCUSSION

### Mechanical unraveling of long repeat proteins is not cooperative. The unfolding of short repeat series can exhibit cooperative behavior

Equations 14, 15, and 18–20 suggest that the mechanical response of the repeat protein does not depend individually
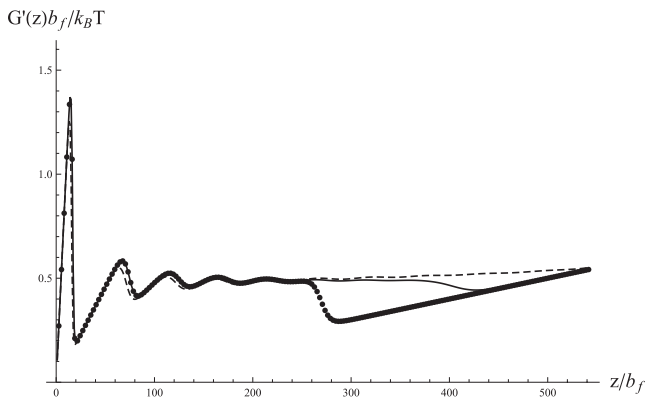
G'(z)b_f/k_BT



FIGURE 3  Force-extension curve of a repeat protein with the same $J + H$ but different $J$ and $H$. Solid line, $J = H = 6$; dashed line, $J = 0$, $H = 12$; points, $H = -15$, $J = 27$. The number of repeats is $N = 10$ and $b_u = 10 b_f$.

on the values of $J$ and $H$ but rather on the sum $J + H$. To test this, we plot in Fig. 3 the dependence of the force $G'(z)$ on $z$ for different values of $J$ and $H$ chosen such that the sum $J + H$ stays constant. As long as the extension $z$ is not too large, this dependence is indeed nearly the same. In particular, at modest extensions, the mechanical response of a protein formed by noninteracting repeats ($H = 12$, $J = 0$) is virtually the same as that in the case $H = -15$, $J = 27$. A negative value of $H$ means that the folding of an individual repeat is unfavorable thermodynamically. As is known in the context of the helix-coil transition theory (11,12), folding in this case requires nucleation of a sufficiently long contiguous block of folded repeats such that the favorable interaction between neighboring units outweighs the initial cost of folding. This situation results in a highly cooperative, sharp transition between the folded and unfolded states, which can, indeed, be observed in thermal or chemical denaturation experiments (3). However, this cooperative behavior with respect to thermal unfolding makes no difference when the same protein is denatured mechanically. The unraveling of the chain in this case starts with the unfolding of the first or last repeat (requiring the lowest free-energy cost) and proceeds sequentially. The protein responds to stretching by unfolding more repeats such that the number of unfolded repeats increases linearly with the extension (cf. Eq. 12). As a result, the tension in the chain exhibits a plateau at a force close to $F_{yield}$. A similar scenario has been previously reported by us for the $J = 0$ case (13), the only difference being that the order in which the repeats unfold does not matter in that case, since there is no interfacial free-energy cost associated with unfolding. The lack of cooperativity in the mechanical unfolding of repeat proteins has already been pointed out in the experimental study by Li et al. (8).

At large extensions, however, the three curves plotted in Fig. 3 begin to diverge. In particular, a significant drop in the force is observed for the case $H = -15$, followed by linear dependence with a slope corresponding to the fully

unfolded protein (cf. the last line of Eq. 20). This implies that complete unraveling of the protein occurs sooner than in the other two cases. To understand this behavior, note that if $H < 0$, then isolated folded repeats are unstable thermodynamically. Consequently, once the number of folded repeats, $n_f$, drops below a certain value, the remaining repeats will lose their stability and unfold spontaneously. Further insight into such spontaneous unfolding is gained by considering the free energy of the chain (Eq. 11). Although at sufficiently small extensions $z$, the free-energy minimum is attained at a finite value of $n_f$ and at $n = 1$, for higher values of $z$, the minimum can be provided by $n_f = n = 0$. When the transition between the two regimes happens at a value $n_f$ such that $n_f < N$, the remaining $N - n_f$ repeats must unfold spontaneously, resulting in a drop in the force. Indeed, Eq. 11 predicts the small drop in force observed in Fig. 2 a and mentioned in the section comparing numerical results with approximations. Moreover, Eq. 11 predicts eight peaks (Fig. 2 a, dashed line), in contrast to $N = 10$ repeats in the chain, indicating that the last three repeats unfold simultaneously. Reducing the stability of individual repeats will promote early spontaneous unfolding of multiple repeats, which is indeed seen in Fig. 3, where the most prominent drop is observed for $H = -15$.

From these considerations, it is also clear that if the length, $N$, of the repeat protein only slightly exceeds the number of repeats needed to nucleate a stable folded structure, its mechanical unfolding will take place in a cooperative fashion, as the unfolding of the first repeat will destabilize the rest of the molecule.

## Conditions for the existence of the force peak

As found in the Results section, the initial rupture event drastically lowers the stiffness of the chain and can lead to a sharp peak in the force followed by subsequent unfolding of the remaining repeats that takes place at much lower forces. This initial peak is pronounced when the rupture force, $F_u$ (Eq. 19), is much higher than the yield force, $F_{yield}$ (Eq. 14). When Eqs. 14 and 19 are compared, these two forces become identical in the $N \to \infty$ limit. Thus, the initial rupture peak is a finite-size effect that disappears for long chains. Numerical calculations using Eq. 4 also confirm this finding. This, of course, is consistent with the notion that van der Waals loops cannot exist in infinite systems. The requirement $F_u \gg F_{yield}$ gives the following condition for the existence of a significant rupture peak:

$$b_f^2 \ll b_u^2 / (N + 1). \qquad (21)$$

This means that stiffer folded repeats (i.e., those with a lower $b_f$) will produce a higher rupture peak. Longer repeat proteins will generally have a less pronounced peak. Since the mean-square end-to-end distance, $b_u^2$, of the unfolded chain grows linearly with its contour length, $L$, longer contour length stored in a repeat will also result in a more pronounced peak.

The validity of Eq. 21 can be verified for a typical AFM protein stretching experiment. Using a typical number, $k_f = 50$ pN/nm, for the stiffness of an individual repeat (6,7), we estimate $b_f \sim \sqrt{k_B T/k_f} \sim 0.2$ nm. Lee et al. (7) found that unfolded repeats behave as wormlike chains with a contour length $L \sim 12$ nm and persistence length $p \sim 2$ nm. For such a chain, we estimate $b_u \sim \sqrt{Lp} \sim 5$ nm. Equation 21 thus holds if $N \ll (b_u/b_f)^2 \sim 300$. This condition is satisfied in Lee et al. (7), where $N \leq 24$. The authors indeed observed a pronounced rupture peak (7), although it was not found in another study, by Li et al. (8). Consistent with the model proposed here, the peak observed in Lee et al. (7) had a shape corresponding to a Hookean spring. The initial high peak was followed by a series of much lower force peaks attributed to the unfolding of individual repeats. This behavior is consistent with the calculated shape of $G'(z)$ (see Fig. 2 b), which shows oscillations reflecting the unfolding of subsequent repeats.

We note that the estimate of Eq. 21 gives, at best, the necessary but not a sufficient condition for the peak to be actually observed. For example, in the idealized scenario discussed in the section about numerical results, above, the force on the chain is raised and the chain extension is measured. It then follows that this peak corresponds to a metastable state of the chain and will only be observed if the pulling is sufficiently fast. The situation is even more complicated in real experiments, where the protein is attached to a cantilever and to a surface. The stiffness of the cantilever and the linker must also be considered and may change the conditions for the peak (22). Differences in the parameters $b_f$, $b_u$, and $N$, as well as in the pulling method, could explain why a large peak appears in some experiments but not in others.

The high strength and toughness exhibited by many load-bearing proteins existing in nature is commonly attributed to the existence of "sacrificial bonds" (23–25), whose rupture releases "hidden length" of the polypeptide chain, allowing the system to dissipate large amounts of energy. Recently, there has been considerable progress in engineering and controlling protein constructs that display such behavior (26–30). For a protein to act as a sacrificial bond when it unfolds, a large free-energy barrier must exist, delaying its unfolding and allowing the folded protein to generate a force that is higher than the force that would destabilize the native state thermodynamically (13,31). The existence of such a barrier is equivalent to having a large peak (van der Waals loop) in the dependence of $G'(z)$ on $z$ (see Results). Equation 21 thus provides a recipe for designing a repeat protein with potentially interesting sacrificial bond properties.

## The yield force is proportional to the square root of the unfolding free energy

Scaling properties of rupture/unfolding forces in AFM pulling experiments have attracted considerable theoretical attention in the past (32–35), although most of the related work has focused on the irreversible unfolding regime, where those forces are controlled by barriers rather than by thermodynamic stability of proteins. Here, we can address the opposite regime of very slow pulling, where we found the yield force to be proportional to $\sqrt{J+H}$. This is a rather general result that arises when the thermodynamic stability of a sacrificial bond system is considered. When a sacrificial bond is broken, a contour length, $L$, becomes liberated. If we assume that the polymer is a Gaussian chain with a Kuhn segment of length $p$, then the entropic cost for stretching it by an amount $z$ is $\Delta S \sim -z^2/(2Lp)$. Thus, if a stretching force $F$ is applied, the corresponding free-energy change is $\Delta G \sim \min_z k_B T z^2/(2Lp) - Fz = -F^2 Lp/2k_B T$. The lowering of the free energy comes at the cost, $\Delta E$, of breaking the bond. Therefore, the bond is destabilized thermodynamically when

$$\Delta E \approx F^2 Lp/2k_B T, \qquad (22)$$

which gives a square-root dependence of the force on the bond energy, $\Delta E$. Setting $Lp = b_u^2$ and $\Delta E/k_B T = J + H$ in Eq. 22, and assuming $b_f = 0$, one recovers the result of Eq. 14 for the yield force. It is also clear from the above arguments that the square root dependence is contingent on Gaussian polymer statistics. For example, if excluded volume effects are included, this would result in a slightly different scaling (13), $F \propto \Delta E^{3/5}$.

## CONCLUSIONS

In this article, we have studied a simple model for the mechanical unfolding of repeat proteins, in which the interactions among individual repeats are described by the Ising model. We find that although thermal denaturation of such a protein may be highly cooperative, its mechanical unfolding is sequential as long as the number of folded repeats is sufficient to ensure that the chain does not unfold spontaneously. We further show that the unfolding of the first repeat may produce a rupture force significantly higher than the force associated with the unfolding of subsequent repeats. These findings can be understood by employing simple analytic approximations for the protein's free energy as a function of its extension and the number of folded repeats.

Structural information about repeat proteins can be incorporated into this model to make it more realistic. Specifically, consider the free energy of a stretched, partially unfolded protein given by Eq. 10. The first term describes the elastic response of the chain. The next two terms describe the interactions among repeats. As long as the elastic part is independent of $n$, the free-energy minimum with respect to $n$ is achieved at $n = 1$. This means that the mechanical unfolding pathway predominantly involves partially unfolded structures that consist of a single folded block flanked by fully unfolded chains. One can use this fact to replace the first term in Eq. 10 by a more realistic description of the protein's elastic response. Such a description could

utilize, e.g., a wormlike chain model for the unfolded tails ([36]) combined with input from atomistic simulations to account for the elasticity of the folded block as a function of its length, $n_f$ ([6]). Thermal unfolding data can further be used to estimate parameters $H$ and $J$.

Finally, note that our model can also be applied to the unfolding of naturally occurring or engineered "polyprotein" constructs containing multiple repeats of the same (or several different) protein domains, which are commonly used in single-molecule AFM pulling studies ([21]). Unlike ankyrin repeat proteins, where individual repeats are often unstable, individual domains within polyproteins can fold independently. Although the assumption of noninteracting, individually stable domains (corresponding to the case $H > 0$, $J = 0$ within the model presented here) is often adequate for the interpretation of AFM pulling experiments ([37–40]), a more general case of interacting domains may be of interest, e.g., in the context of unfolding and refolding of polyubiquitin chains, where a cooperative folding mechanism has been suggested ([41–43]).

## REFERENCES

1. Barrick, D., D. U. Ferreiro, and E. A. Komives. 2008. Folding landscapes of ankyrin repeat proteins: experiments meet theory. *Curr. Opin. Struct. Biol.* 18:27–34.

2. Kajander, T., A. L. Cortajarena, E. R. Main, S. G. Mochrie, and L. Regan. 2005. A new folding paradigm for repeat proteins. *J. Am. Chem. Soc.* 127:10188–10190.

3. Kloss, E., N. Courtemanche, and D. Barrick. 2008. Repeat-protein folding: new insights into origins of cooperativity, stability, and topology. *Arch. Biochem. Biophys.* 469:83–99.

4. Mello, C. C., and D. Barrick. 2004. An experimentally determined protein folding energy landscape. *Proc. Natl. Acad. Sci. USA.* 101:14102–14107.

5. Wolynes, P. G. 1997. Folding funnels and energy landscapes of larger proteins within the capillarity approximation. *Proc. Natl. Acad. Sci. USA.* 94:6170–6175.

6. Sotomayor, M., D. P. Corey, and K. Schulten. 2005. In search of the hair-cell gating spring elastic properties of ankyrin and cadherin repeats. *Structure.* 13:669–682.

7. Lee, G., K. Abdi, Y. Jiang, P. Michaely, V. Bennett, et al. 2006. Nanospring behaviour of ankyrin repeats. *Nature.* 440:246–249.

8. Li, L., S. Wetzel, A. Pluckthun, and J. M. Fernandez. 2006. Stepwise unfolding of ankyrin repeats in a single protein revealed by atomic force microscopy. *Biophys. J.* 90:L30–L32.

9. Buhot, A., and A. Halperin. 2000. Extension of rod-coil multiblock copolymers and the effect of the helix-coil transition. *Phys. Rev. Lett.* 84:2160–2163.

10. Tamashiro, M. N., and P. Pincus. 2001. Helix-coil transition in homopolypeptides under stretching. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 63:021909.

11. Poland, D., and H. Scheraga. 1970. Theory of Helix-Coil Transitions in Biopolymers. Academic Press, New York.

12. Zimm, B. H., and J. K. Bragg. 1959. Theory of the phase transition between helix and random coil in polypeptide chain. *J. Chem. Phys.* 31:526–531.

13. Kirmizialtin, S., L. Huang, and D. E. Makarov. 2005. Topography of the free energy landscape probed via mechanical unfolding of proteins. *J. Chem. Phys.* 122:234915.

14. Jarkova, E., N. K. Lee, and S. Obukhov. 2005. Reading structural information from force-extension curves. *Macromolecules.* 38:2469–2474.

15. Borisov, O. V., and A. Halperin. 1998. Extending polysoaps in the presence of free amphiphiles. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics.* 57:812–822.

16. Makarov, D. E. 2007. Unraveling individual molecules by mechanical forces: theory meets experiment. *Biophys. J.* 92:4135–4136.

17. Li, P. -C., and D. E. Makarov. 2003. Theoretical studies of the mechanical unfolding of the muscle protein titin: bridging the time-scale gap between simulation and experiment. *J. Chem. Phys.* 119:9260–9268.

18. Best, R. B., E. Paci, G. Hummer, and O. K. Dudko. 2008. Pulling direction as a reaction coordinate for the mechanical unfolding of single molecules. *J. Phys. Chem. B.* 112:5968–5976.

19. Liphardt, J., B. Onoa, S. B. Smith, I. J. Tinoco, and C. Bustamante. 2001. Reversible unfolding of single RNA molecules by mechanical force. *Science.* 292:733–737.

20. Evans, E., and K. Ritchie. 1997. Dynamic strength of molecular adhesion bonds. *Biophys. J.* 72:1541–1555.

21. Fisher, T. E., A. F. Oberhauser, M. C. Vezquez, P. E. Marsalek, and J. Fernandez. 1999. The study of protein mechanics with the atomic force microscope. *Trends Biochem. Sci.* 24:379–384.

22. Evans, E., and K. Ritchie. 1999. Strength of a weak bond connecting flexible polymer chains. *Biophys. J.* 76:2439–2447.

23. Becker, N., E. Oroudjev, S. Mutz, J. P. Cleveland, P. K. Hansma, et al. 2003. Molecular nanosprings in spider capture-silk threads. *Nat. Mater.* 2:278–283.

24. Fantner, G. E., E. Oroudjev, G. Schitter, L. S. Golde, P. Thurner, et al. 2006. Sacrificial bonds and hidden length: unraveling molecular mesostructures in tough materials. *Biophys. J.* 90:1411–1418.

25. Smith, B. L., T. E. Schaffer, M. Viani, J. B. Thompson, N. A. Frederick, et al. 1999. Molecular mechanistic origin of the toughness of natural adhesives, fibres and composites. *Nature.* 399:761–763.

26. Cao, Y., C. Lam, M. Wang, and H. Li. 2006. Nonmechanical protein can have significant mechanical stability. *Angew. Chem. Int. Ed. Engl.* 45:642–645.

27. Cao, Y., and H. Li. 2008. Engineered elastomeric proteins with dual elasticity can be controlled by a molecular regulator. *Nature Nanotech.* 3:512–516.

28. Cao, Y., T. Yoo, and H. Li. 2008. Single molecule force spectroscopy reveals engineered metal chelation is a general approach to enhance mechanical stability of proteins. *Proc. Natl. Acad. Sci. USA.* 105:11152–11157.

29. Cao, Y., T. Yoo, S. Zhuang, and H. Li. 2008. Protein–protein interaction regulates proteins' mechanical stability. *J. Mol. Biol.* 378:1132–1141.

30. Li, H., H. -C. Wang, Y. Cao, D. Sharma, and M. Wang. 2008. Configurational entropy modulates the mechanical stability of protein GB1. *J. Mol. Biol.* 379:871–880.

31. Eom, K., D. E. Makarov, and G. J. Rodin. 2005. Theoretical studies of the kinetics of mechanical unfolding of cross-linked polymer chains and their implications for single molecule pulling experiments. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 71:021904.

32. Dias, C. L., M. Dube, F. A. Oliveira, and M. Grant. 2005. Scaling in force spectroscopy of macromolecules. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 72:011918.

33. Dudko, O. K., A. E. Filippov, J. Klafter, and M. Urbakh. 2003. Beyond the conventional description of dynamic force spectroscopy of adhesion bonds. *Proc. Natl. Acad. Sci. USA.* 100:11378–11381.

34. Dudko, O. K., G. Hummer, and A. Szabo. 2006. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.* 96:108101.

35. Maloney, C. E., and D. J. Lacks. 2006. Energy barrier scalings in driven systems. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 73:061106.

36. Rief, M., J. M. Fernandez, and H. E. Gaub. 1998. Elastically coupled two-level systems as a model for biopolymer extensibility. *Phys. Rev. Lett.* 81:4764–4767.

37. Best, R. B., D. J. Brockwell, J. L. Toca-Herrera, A. W. Blake, D. A. Smith, et al. 2003. Force mode atomic force microscopy as a tool for protein folding studies. *Anal. Chim. Acta.* 479:87–105.

38. Makarov, D. E., P. K. Hansma, and H. Metiu. 2001. Kinetic Monte Carlo simulation of titin unfolding. *J. Chem. Phys.* 114:9663–9673.

39. Scott, K. A., A. Steward, S. B. Fowler, and J. Clarke. 2002. Titin: a multidomain protein that behaves as the sum of its parts. *J. Mol. Biol.* 315:819–829.

40. Zinober, R. C., D. J. Brockwell, G. S. Beddard, A. W. Blake, P. D. Olmsted, et al. 2002. Mechanically unfolding proteins: the effect of unfolding history and the supramolecular scaffold. *Protein Sci.* 11:2759–2765.

41. Best, R. B., and G. Hummer. 2005. Comment on ''Force-clamp spectroscopy monitors the folding trajectory of a single protein'', (letter to the editor). *Science.* 308:498.

42. Fernandez, J. M., and H. Li. 2004. Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science.* 303:1674–1678.

43. Schlierf, M., H. Li, and J. M. Fernandez. 2004. The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proc. Natl. Acad. Sci. USA.* 101:7299–7304.