



# Mechanics of forced unfolding of proteins

Tianxiang Su, Prashant K. Purohit \*

*Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia, PA 19104, USA*

Received 14 November 2008; received in revised form 21 January 2009; accepted 26 January 2009

## Abstract

We describe and solve a two-state kinetic model for the forced unfolding of proteins. The protein oligomer is modeled as a heterogeneous, freely jointed chain with two possible values of Kuhn length and contour length representing its folded and unfolded configurations. We obtain analytical solutions for the force–extension response of the protein oligomer for different types of loading conditions. We fit the analytical solutions for constant-velocity pulling to the force–extension data for ubiquitin and fibrinogen and obtain model parameters, such as Kuhn lengths and kinetic coefficients, for both proteins. We then predict their response under a linearly increasing force and find that our solutions for ubiquitin are consistent with a different set of experiments. Our calculations suggest that the refolding rate of proteins at low forces is several orders larger than the unfolding rate, and neglecting it can lead to lower predictions for the unfolding force, especially at high stretching velocities. By accounting for the refolding of proteins we obtain a critical force below which equilibrium is biased in favor of the folded state. Our calculations also suggest new methods to determine the distance of the transition state from the energy wells representing the folded and unfolded states of a protein.

© 2009 Acta Materialia Inc. Published by Elsevier Ltd. All rights reserved.

*Keywords:* Proteins; AFM; Kinetic model

## 1. Introduction

Over the last two decades atomic force microscopy (AFM) has established itself as a valuable experimental technique to probe the structure and energetics of proteins [1–11]. Force–extension measurements obtained via AFM have shown that the mechanics and chemistry of proteins are intimately linked [3,12]. The data emerging from AFM experiments are interpreted using steered molecular dynamics (SMD) and Monte Carlo (MC) simulations [1,10,13,14]. The SMD simulations complement AFM experiments by providing information about short-lived and metastable intermediate states that could not be gleaned from experiments alone. However, the unfolding forces predicted by SMD simulations are much larger than those obtained in AFM experiments since the rates of pulling in the SMD simulations are constrained (due to limitations on the time step) to be orders of magnitude larger

than realistic AFM pulling rates [13]. The MC simulations are based on the two-state model of Rief et al. [14] and they reproduce the AFM data quite well [1,9,10]. These methods, however, suffer from the limitation that the kinetic parameters have to be determined by trial and error. In general, the refolding rate is set to zero in these simulations and only one persistence length is used for both the folded and unfolded states of the protein. This reduces the dimensionality of the parameter space to be searched [1,3] but is unrealistic since refolding is dominant at low forces and unfolded proteins are expected to be floppier than their folded counterparts. Furthermore, MC simulations have been used primarily to fit the data from AFM experiments where the protein is pulled at a constant velocity, and we are not aware of any attempt to use MC methods to determine the response of proteins under other pulling conditions, such as a force linearly increasing with time.

Our goals in this paper are: (i) to unify the description of protein unfolding under different types of pulling conditions within a single model; (ii) to account for refolding and explain the consequences of neglecting it; and (iii) to

\* Corresponding author. Tel.: +1 215 898 3870; fax: +1 215 573 6334.  
E-mail address: [purohit@seas.upenn.edu](mailto:purohit@seas.upenn.edu) (P.K. Purohit).

predict the response of proteins under different pulling conditions from a knowledge of their kinetic and mechanical properties. We begin with a version of the two-state model of Rief et al. [14] and obtain analytical expressions for the force–extension profile that can be directly fitted to the experimental data. Unlike many of the MC simulations, we do not set the refolding rate to zero. In fact, we show that there is a critical value of the applied force under which the equilibrium in the folding/unfolding reaction is biased towards folding. We demonstrate the applicability of our model by fitting it to published AFM experimental data on two different proteins: ubiquitin and fibrinogen. For ubiquitin we show that the kinetic and mechanical parameters obtained from fitting our expressions to a constant-velocity pulling experiment can be used to predict its response in an experiment where the force is linearly increasing with time. After validating our model with ubiquitin, which has been extensively studied under various types of loading conditions, we apply the same procedure to fibrinogen and obtain predictions for its response under a linearly increasing force.

## 2. Results

### 2.1. Three equations governing the forced unfolding of proteins

In the problem of sequentially unfolding a protein oligomer, we have three unknown functions of time  $t$ :  $x(t)$ ,  $F(t)$  and  $N_f(t)$ , which are, respectively, the extension, force and the number of folded proteins. The total number of proteins  $N$  is a constant throughout the experiment, so that,  $N_f + N_u \equiv N$ . Therefore,  $N_u(t)$ , the number of unfolded proteins, is not viewed as an unknown function. The three equations that close the system are: (1) the equilibrium force–extension relation,  $x = x(F, N_f)$ , obtained from either the freely jointed chain (FJC) model or the worm-like chain (WLC) model of polymer elasticity [15,16]; (2) the kinetic equation, which measures the rates of unfolding and refolding, obtained from either Bell's model [17], or other more sophisticated kinetic models based on Kramer's rate theory [18]; (3) the equation which determines the manner of applying the external constraint, e.g., for constant-velocity pulling, it is  $dx/dt = v_c$ , where  $v_c$  is the pulling speed, and for constant-force pulling, it is  $dF/dt = 0$ . This system of equations unifies problems of protein unfolding under different kinds of loading conditions into a single mathematical framework. By merely changing the last equation, one can obtain the unfolding behavior of proteins under constant-velocity pulling, constant-force pulling, pulling with a force linearly increasing with time, etc.

### 2.2. Equilibrium force–extension relation

The mechanical properties of a protein in the folded and unfolded states are expected to be different—a folded protein is stiff, whereas an unfolded protein is floppy. Hence,

we model the protein oligomer as a heterogeneous FJC with two possible values of Kuhn length  $l_f$  and  $l_u$  for the folded and unfolded states, respectively (Fig. 1a). For such a heterogeneous FJC, the total extension can be rigorously shown to be the sum of the extensions of the homogeneous subchains [19,20]:

$$x = N_f L_{fs} \left[ \coth \left( \frac{Fl_f}{k_B T} \right) - \frac{k_B T}{Fl_f} \right] + N_u L_{us} \left[ \coth \left( \frac{Fl_u}{k_B T} \right) - \frac{k_B T}{Fl_u} \right] \quad (1)$$

where  $L_{fs}$  and  $L_{us}$  are the contour lengths of a single folded and unfolded protein, respectively,  $k_B$  is the Boltzmann constant,  $T$  is the absolute temperature and the meanings of the symbols  $x$ ,  $F$ ,  $N_f$ ,  $N_u$ ,  $l_f$  and  $l_u$  have been discussed above. Note that, in reality, there is only one contour length  $L_{us}$  that associates with the fully unfolded proteins. The other contour length  $L_{fs}$  in Eq. (1) is meant to represent the maximum length of the proteins if unfolding is somehow prevented.

This model implies that the equilibrium force–extension behavior of a protein chain is governed only by four parameters,  $L_{fs}$ ,  $l_f$ ,  $L_{us}$  and  $l_u$ , regardless of the number of copies in the oligomer. However, if the chain is modeled as homogeneous, we need  $N$  Kuhn lengths and  $N$  contour lengths to fit the  $N$  curves obtained in a constant-velocity pulling experiment (e.g., [21]).

In order to demonstrate the applicability of our model, we consider force–extension measurements on ubiquitin (N-C linked) and fibrinogen. For ubiquitin (experimental

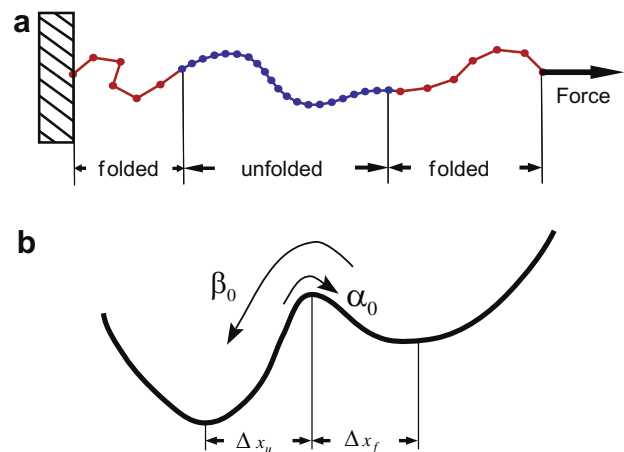


Fig. 1. Illustration of the two-state kinetic model. (a) A chain of mixed folded and unfolded proteins is modeled as a heterogeneous freely jointed chain. A single folded (unfolded) protein is represented by a  $N_f$ -segment ( $N_u$ -segment) subchain with Kuhn length  $l_f$  ( $l_u$ ). In this illustration, two folded and one unfolded proteins are represented by the two red and one blue subchains, respectively. Note that, in reality, the actual number of segments in each subchain may be much larger. Also,  $l_u$  is expected to be smaller than  $l_f$  since an unfolded protein is expected to be floppier than a folded protein. (b) Energy landscape of the two-state model. The ordinate is the Gibbs free energy and the abscissa is the reaction coordinate. The two wells, representing the folded and unfolded states of a protein, are separated by an energy barrier with transition distances  $\Delta x_u$  and  $\Delta x_f$ . At zero force, the folding rate and the unfolding rate are  $\beta_0$  and  $\alpha_0$ , respectively, with  $\beta_0 \gg \alpha_0$ . An applied force can lower the energy barrier and thus change the folding and unfolding rates.

data from Ref. [7]), we know that the last curve in the force–extension profile (see Fig. 2a) corresponds to six unfolded proteins and zero folded proteins, so we apply the homogeneous FJC model to fit this curve and obtain  $L_{us} = 25.37$  nm,  $l_u = 0.33$  nm. The other two parameters can be obtained by fitting one other curve using the heterogeneous FJC model. Here we use the first curve and get  $L_{fs} = 6.29$  nm and  $l_f = 0.60$  nm. We similarly determine the parameters for fibrinogen using the first and last curves (Fig. 2b, experimental data from Ref. [9]) and get  $L_{fs} = 11.39$  nm,  $L_{us} = 44.62$  nm,  $l_f = 0.57$  nm and  $l_u = 0.31$  nm. These results are not significantly different if we use any other two curves for the fitting. Without any more free parameters, we then predict the intermediate curves for both proteins using Eq. (1) and compare the predictions with the experimental data. The results are shown in Fig. 2 (red curves) and the predictions of the heterogeneous FJC model match the experimental data quite well for both proteins.

Our results show that  $l_u \approx l_f/2$  for both proteins, which agrees with our intuition that unfolded proteins should be floppier than folded ones. Furthermore, our estimates for the contour lengths of the fully unfolded proteins  $L_{us}$  agree well with published results (27.4 nm for ubiquitin [5,22] and 40 nm for fibrinogen [11]), which shows that our fitted parameters are indeed physically relevant. However, it is worth pointing out that  $L_{fs}$  is the maximum length of a single folded protein if unfolding is somehow prevented, therefore, it will be different from the end-to-end distance of the protein in its native state. In fact, the contour length of ubiquitin in its native state is about 3.8 nm [5], and simulations have shown that this number increases to 4.7 nm under a constant force of 200 pN while the protein remains in a native-like state [22].

### 2.3. Kinetic equation

It has been shown by both experiments and simulations that, at least for ubiquitin, most ( $\sim 95\%$ ) of the unfolding events follow a two-state pathway [8,22]. Therefore, following Bell's theory [17], we propose that the change in the number of folded proteins is given by:

$$\frac{dN_f^*}{dt} = -k_u N_f + k_f N_u \quad (2)$$

where  $k_u = \alpha_0 \exp(F\Delta x_u/k_B T)$ ,  $k_f = \beta_0 \exp(-F\Delta x_f/k_B T)$ ,  $\alpha_0$  and  $\beta_0$  are the unfolding and refolding rates when no force is applied,  $\Delta x_u$  and  $\Delta x_f$  are the distances to the transition state (Fig. 1b), and  $N_f^*$ , set to be a real number, is a continuous version of the integer  $N_f$ . Its initial value is set to be equal to  $N_f$  and it evolves according to Eq. (2). On the other hand,  $N_f$ , the number of folded proteins, evolves in such a way that it jumps by  $\pm 1$  whenever  $N_f^*$  reaches an integer. Note that for simplicity we assume here that  $\Delta x_f$  and  $\Delta x_u$  are unaffected by the external force. This assumption is valid when the local curvatures of the potential wells are large [23,24]. More sophisticated kinetic models (e.g., [18]) can be easily incorporated into our description. We stick with Bell's model here because the goal of this paper is to apply the three-equation mathematical framework to understand the unfolding behaviors of proteins under different pulling conditions, and Bell's model is simple enough to give analytic or semi-analytic solutions for all the conditions discussed below, and at the same time captures most of the physics reported in the experiments. The condition for unfolding or refolding events to happen is:

$$\int dN_f^* = \int (-k_u N_f + k_f N_u) dt = \pm 1 \quad (3)$$

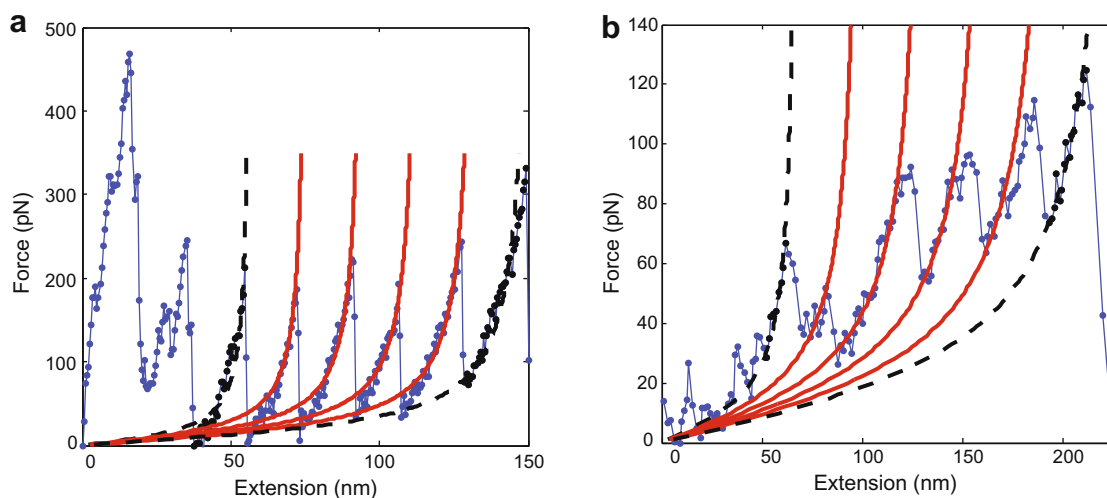


Fig. 2. Predictions of the force–extension profiles using the heterogeneous FJC model with only four free parameters: (a) ubiquitin; (b) fibrinogen. Blue curves are the experimental data—ubiquitin data from Ref. [7] and fibrinogen data from Ref. [9]. We use two of the experimental curves (black dots) to fit the Kuhn length and the contour length of the folded and unfolded proteins (black dashed lines are the fitting results). Then, without any more free parameters, we use the heterogeneous FJC model (Eq. (1)) to predict all the other curves. The predictions (red curves) match well with the experimental data for both proteins. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

where +1 represents refolding and –1 represents unfolding of a protein.

The advantage of the present method over MC simulations is that we can solve exactly for all the four kinetic parameters ( $\alpha_0$ ,  $\beta_0$ ,  $\Delta x_u$  and  $\Delta x_f$ ) from the experimental data (discussed below) instead of guessing which parameter values fit the data best. Guessing the best-fit kinetic parameters is especially difficult for the MC simulations when taking the refolding rate  $\beta_0$  into account since the parameter space is large. In fact, in many cases  $\beta_0$  is simply set to zero by taking advantage of the fact that refolding is negligible at large forces [1,3]. A caveat of our deterministic model is that it ignores the randomness of the unfolding/folding events, but it is really meant to represent the average behavior of a large ensemble of experiments. In fact, we show in what follows that the kinetic parameters obtained from the deterministic model assuming  $k_f = 0$  are close to those obtained previously by MC simulations. Furthermore, the kinetic parameters obtained from the deterministic model can be used in the MC simulations to obtain particular instances of the unfolding pathway, thus providing information about higher moments, e.g., the variance of the unfolding force in a constant-velocity pulling experiment.

#### 2.4. Constant-velocity pulling

For the case of constant-velocity pulling, the external constraint equation is  $dx/dt = v_c$ . Using this relation, the unfolding and refolding condition (Eq. (3)) can be rewritten as:

$$\int_{x_1}^{x_2} (k_u N_f - k_f N_u) dx \pm v_c = 0 \quad (4)$$

where the positive (negative) sign represents refolding (unfolding) of one protein,  $x_1$  is the initial extension of a particular continuous force–extension curve,  $x_2$  is the extension when unfolding/refolding is imminent,  $k_u(F(x))$  and  $k_f(F(x))$  are functions of the extension, and  $N_f$  and  $N_u$  are the number of folded and unfolded proteins which remain constants for each curve.

Each of the unfolding events in the force–extension profile should satisfy Eq. (4) and we can use any four of them to solve for the four free kinetic parameters. This results in four algebraic equations that we solved numerically using Newton’s method for both ubiquitin and fibrinogen (pulling velocity for both the proteins is  $v_c = 1000 \text{ nm s}^{-1}$  [7,9]). The result for ubiquitin is:  $\alpha_0 = 3.75 \text{ s}^{-1}$ ,  $\beta_0 = 1293.65 \text{ s}^{-1}$ ,  $\Delta x_u = 0.08 \text{ nm}$  and  $\Delta x_f = 0.31 \text{ nm}$ . The refolding rate  $\beta_0$  found here is comparable to that obtained previously from MD simulations [25]. The result for fibrinogen is:  $\alpha_0 = 3.19 \text{ s}^{-1}$ ,  $\beta_0 = 7691.16 \text{ s}^{-1}$ ,  $\Delta x_u = 0.10 \text{ nm}$  and  $\Delta x_f = 0.67 \text{ nm}$ . Similar transition distances have been reported for many other protein domains and are suggested to reflect the critical breaking of hydrogen bonds or hydrophobic interactions in the process of protein unfolding [26]. Using the solved kinetic parameters, we can

predict other breaking points and thus the whole force–extension profile using Eqs. (1) and (4) and  $dx/dt = v_c$ . The predictions match well with the experimental data and are shown in Fig. 3. Our solutions for both the proteins imply that  $\beta_0$  is several orders larger than  $\alpha_0$ . Intuitively, this is expected because when no external force is applied, the proteins tend to rapidly fold into their native state. The fact that  $\beta_0 \gg \alpha_0$  suggests that the refolding rate cannot be ignored, at least when the force is small.

If we assume that the refolding rate  $\beta_0$  is zero as is commonly done in MC simulations, the solution for the kinetic parameters for ubiquitin is:  $\alpha_0 = 0.05 \text{ s}^{-1}$ ,  $\Delta x_u = 0.17 \text{ nm}$ . These results are quite close to those obtained previously by other experiments that also assumed  $k_f = 0$  [8,27,28], and suggest that our model is consistent with the MC simulations used before. In fact, the authors in Ref. [8] assumed  $k_u = 0$  and found  $\alpha_0 = 0.015 \text{ s}^{-1}$ ,  $\Delta x_u = 0.17 \text{ nm}$  using the constant-force pulling data and  $\alpha_0 = 0.0375 \text{ s}^{-1}$ ,  $\Delta x_u = 0.14 \text{ nm}$  using the linearly increasing force pulling data on ubiquitin. Our results show that the unfolding rate  $\alpha_0$  obtained by ignoring the refolding rate is significantly lower than that obtained by taking the refolding rate into account. In fact, setting  $\beta_0 = 0$  should always lead to an underprediction of  $\alpha_0$ . The reason for the underprediction is that when the refolding rate is ignored in the kinetic equation, the unfolding rate predicted is in fact a “net rate” for the proteins to change from the folded state to the unfolded state. This calculated “net rate” should be smaller than the true unfolding rate because it includes the contribution of the refolding rate which is high at low forces.

It has been shown that the average breaking force is approximately linear with respect to the logarithm of pulling velocity [5,7]. We use the parameters calculated above and predict the force–extension profiles under different pulling speeds ranging from  $v_c = 10^3 \text{ nm s}^{-1}$  to  $v_c = 10^{11} \text{ nm s}^{-1}$  for both proteins (six copies, force–extension profiles not shown). The linear relation between the average breaking force and the logarithm of pulling velocity is found using either set ( $k_f = 0$ ) and ( $k_f \neq 0$ ) of the kinetic parameters obtained previously (Fig. 4a). For ubiquitin, our calculations show that the slope is 63.1 (pN per 10-fold change of the velocity in  $\text{nm s}^{-1}$ ) if we use the set of parameters that assumes  $k_f = 0$ , and 135.3 if we use another set of parameters that takes refolding into account. Recent MD results on fibrinogen, with pulling velocity  $2.5 \times 10^9 \text{ nm s}^{-1}$ , show that the unfolding events happen at force  $\sim 10^3 \text{ pN}$  [11]. This result is close to our predicted average breaking force for fibrinogen at similar pulling velocity (see Fig. 4a, blue solid line). However, Fig. 4a shows that if we set  $k_f = 0$ , the predicted unfolding force is much smaller than the one that takes the refolding rate into account. The reason the unfolding force is higher with non-zero  $k_f$  is as follows. Recall that for given  $k_u$  and  $k_f$ , the breaking extension  $x_2$  (and hence the breaking force) is calculated from the integral over the entire force–extension curve (see Eq. (4)) including the low force regime

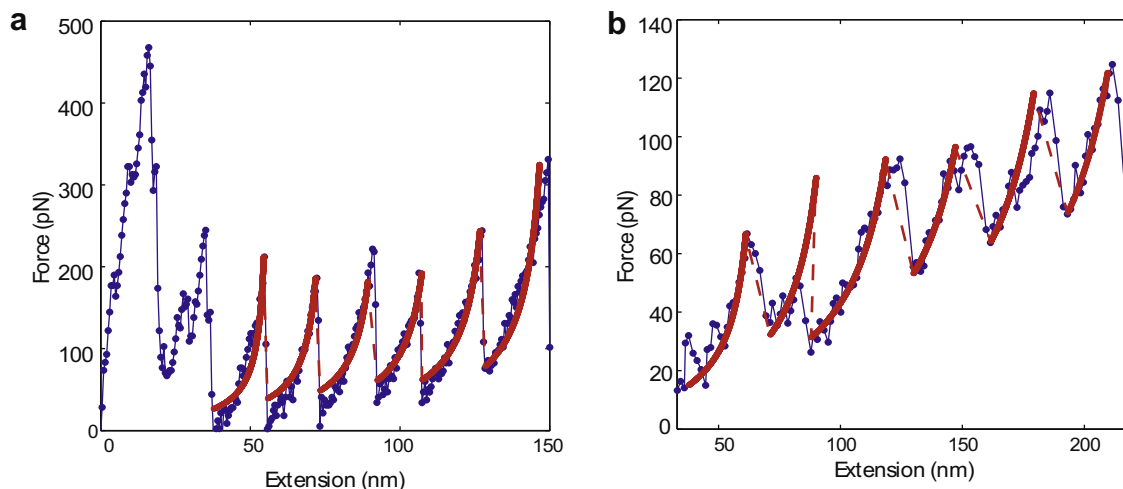


Fig. 3. Force–extension profiles of six copies of (a) ubiquitin and (b) fibrinogen ( $v_c = 1000 \text{ nm s}^{-1}$ ) in constant-velocity pulling. Blue curves, experimental data (ubiquitin data from Ref. [7] and fibrinogen data from Ref. [9]); red curves, prediction using our two-state kinetic model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

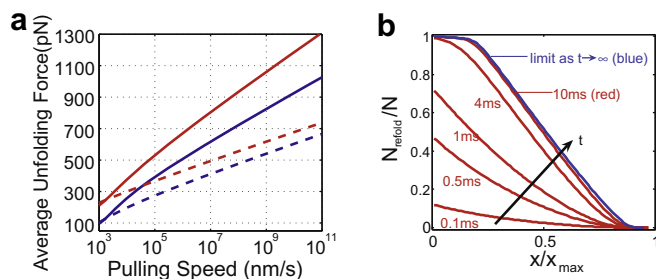


Fig. 4. Constant-velocity pulling. (a) Dependence of the average unfolding force on pulling speed. Red, ubiquitin; blue, fibrinogen; solid line, accounting for refolding ( $k_f \neq 0$ ); dashed line, ignoring refolding ( $k_f = 0$ ). For both proteins, the predicted unfolding force using the set of parameters that assume  $k_f = 0$  is much smaller, especially at high pulling velocities, than the one that takes the refolding rate into account. (b) Relaxation profiles ( $v_c = 0 \text{ nm s}^{-1}$ ). The evolution of the number of refolded proteins is shown as a function of the relaxation extension  $x_r$  (normalized by  $x_{max} = NL_{us}$ ). A limit profile (blue) is approached as time approaches infinity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

where refolding is dominant. The integral in Eq. (4) is initially negative because the force is low and  $k_f \gg k_u$ , so for the case  $k_f \neq 0$ , the force–extension curve should go higher in order for the integral to reach the positive value  $v_c$ . This suggests that a poor prediction of the unfolding rate at low pulling velocities leads to large errors in the predictions for the breaking events at high pulling velocities.

We also note here that if we set  $N_f$  and  $N_u$  in both Eqs. (1) and (2) to be real numbers that change continuously, then together with the constraint equation,  $dx/dt = v_c$ , we can analytically reproduce the continuous force–extension profile for the coiled-coil proteins with a shallow force plateau as observed experimentally [4,11]. The solutions will be discussed elsewhere since we mainly focus in this paper on the unfolding of globular proteins. Furthermore, our model can be applied to study the relaxation behavior of

the protein chain after all the proteins are fully stretched as has been done in experiments [29]. We keep the relaxation extension  $x_r$  fixed by setting  $v_c = 0$ , let the initial  $N_f$  be 0 (i.e., all the proteins are unfolded initially), and obtain the solution for  $N_f(t, x_r)$ . Fig. 4b shows how the number of refolded proteins evolves as a function of time and also its dependence on  $x_r$ . The trend in the solution agrees with recent experimental observations on a different protein [29]. We further note that the correct limit profile as  $t \rightarrow \infty$  (Fig. 4b, blue line) can be obtained only by taking both the unfolding and refolding rates into account.

## 2.5. Pulling with a force linearly increasing with time

For a protein oligomer stretched under a linearly increasing force, we have  $dF/dt = v_f$ , and the unfolding/refolding condition (Eq. (3)) becomes:

$$W(F_2) = \pm v_f + W(F_1) \quad (F_2 > F_1) \quad (5)$$

Eq. (5) can be used to determine the unfolding force  $F_2$ . The double exponential function  $W(F)$  is given by  $W(F) = k_B T [N_f k_u(F)/\Delta x_u + N_u k_f(F)/\Delta x_f]$ , and  $F_1$  is the initial force at each step (note that the force–extension profile is stepwise). Depending on the sign (positive/negative) of  $v_f$ , one protein unfolds/refolds when  $F$  linearly increases to reach  $F_2$ .

We use Eqs. (1) and (5) together with  $dF/dt = v_f$  ( $v_f = 300 \text{ pN s}^{-1}$ ) to generate the stepwise extension–time profile for both ubiquitin and fibrinogen (all the kinetic parameters have been obtained in the constant-velocity pulling section). The results are shown in Fig. 5a and b (red). For ubiquitin, the unfolding events occur around  $\sim 100 \text{ pN}$  and the breaking extension is nearly linear in time, with a slope predicted as  $680.9 \text{ nm s}^{-1}$ . Both these results are consistent with experimental observations [8]. For fibrinogen, the result shown here constitutes a falsifiable prediction from our model and can be easily tested

using current AFM techniques. For both proteins, some refolding events are observed at small force (inset of Fig. 5a and b), which would not be predicted if the refolding rate is ignored (Fig. 5c).

Further analysis of Eq. (5) shows that the initial force  $F_1$  at each step should be larger than a critical force in order that one protein unfolds at the end of the step (otherwise one protein will refold at the end of the step). This critical force  $F_{cl}$  is determined by:

$$W(F_{cl}) = v_f + W(F_s) \quad (F_{cl} < F_s) \quad (6)$$

where  $F_s$  is the unique stationary point of the function  $W(F)$ :  $F_s = [k_B T / (\Delta x_u + \Delta x_f)] \ln[\beta_0 N_u / (\alpha_0 N_f)]$ . In general, since  $W(F)$  depends on  $N_f$ , the critical force increases as more and more proteins unfold (Fig. 6). For example, consider the unfolding of a chain of nine ubiquitins; the critical force computed using Eq. (6) increases from  $\sim 30$  to  $\sim 80$  pN during the unfolding process.

If we let  $N_f$  and  $N_u$  in both Eqs. (1) and (2) be continuous real numbers, then the solution for  $N_f(t)$  satisfying  $N_f(0) = N$  and  $F = v_f t$  is:

$$\frac{N_f(t)}{N} = Q^{-1}(t) \left[ \int_0^t \beta_0 Q(t) e^{C_2 t} dt + Q(0) \right] \quad (7)$$

where the function  $Q(t)$  and the two constants  $C_1$  (appearing in  $Q(t)$ ) and  $C_2$  are given by:

$$Q(t) = \exp[\alpha_0 \exp(C_1 t) / C_1 + \beta_0 \exp(C_2 t) / C_2] \quad (8)$$

$$C_1 = \frac{v_f \Delta x_u}{k_B T}, \quad C_2 = \frac{-v_f \Delta x_f}{k_B T} \quad (9)$$

An analytic expression of  $x(t)$  can be obtained by plugging Eq. (7) into  $x = x(N_f(t), F(t))$  (Eq. (1)). The results for both the proteins are shown in Fig. 5a and b (blue). The curves agree quite well with the predictions of the discrete model, implying that whether we take  $N_f$  and  $N_u$  as integers or real numbers does not greatly affect the results for this kind of experiments. The advantage of assuming

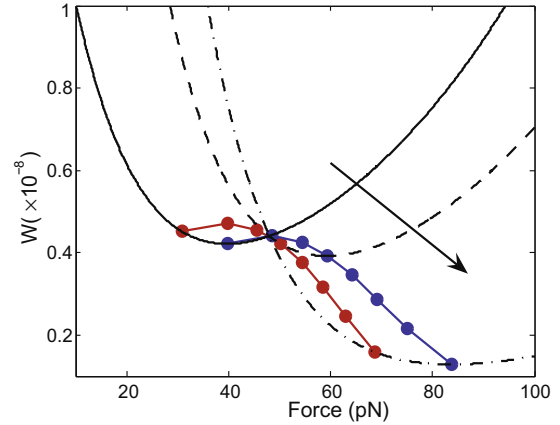


Fig. 6. The double exponential function  $W(F)$  together with the stationary points (blue) and the critical forces (red) for pulling with a force linearly increasing in time. Assume that there are nine copies of ubiquitin in the chain. The double exponential function when  $N_f = 8$ ,  $N_u = 1$  is shown as a black solid line (when  $N_u = 0$ , the function is only a single exponential). As more and more proteins unfold,  $N_f$  decreases and the curve  $W(F)$  shifts to the right along the  $F$  axis (see the arrow). We plot  $W(F)$  as a black dashed line for  $N_f = 5$ ,  $N_u = 4$  and as a black dashed-dotted line for  $N_f = 1$ ,  $N_u = 8$ . The movement of the curve results in an increase in the critical force  $F_{cl}$  (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

continuous  $N_f$  and  $N_u$  is that we have an analytical expression for  $x = x(N_f(t), F(t))$ . This can be directly fitted to the experimental data for linearly increasing force.

### 2.6. Constant-force pulling

Stretching proteins under a constant force produces a staircase-like extension–time profile, in which the extension remains piecewise constant over each step. Since the force is a constant, the unfolding/refolding condition Eq. (3) leads to the dwell time  $\Delta t$  for one unfolding ( $N_f$  decreases by 1) or refolding ( $N_f$  increases by 1) event:

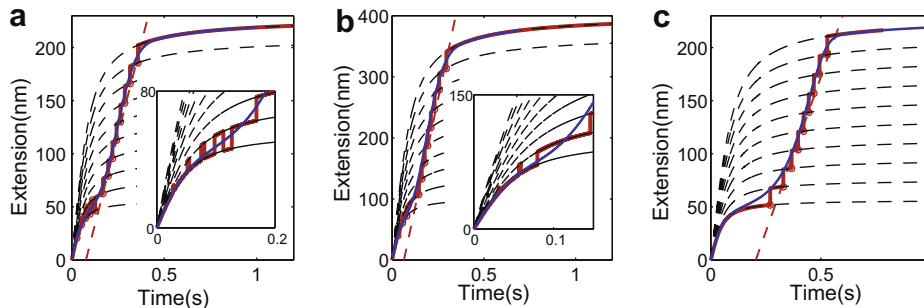


Fig. 5. Extension–time profiles for pulling using a linearly increasing force ( $v_f = 300$  pN  $s^{-1}$ , nine copies of proteins): (a) ubiquitin and (b) fibrinogen. Red curves, predicted stepwise profile, using integer  $N_f$  and  $N_u$  in the governing equations (Eqs. (1) and (2)). Each of the steps obeys the equilibrium force–extension relation, which is shown as dashed black curves (Eq. (1)). Red circles and dashed line, unfolding events and the fitting results. The fitting equation is  $x = 680.9t - 52.12$  for ubiquitin and  $x = 1419t - 92.01$  for fibrinogen ( $x$  is the extension in units of nm). Blue lines, solutions obtained by using continuous  $N_f$  and  $N_u$  in the governing equations, which match well with the discrete model shown in red. The profiles for ubiquitin obtained here are consistent with those measured in Ref. [8]. Insets, if we take the refolding rate into account, some refolding events are observed at small force. (c) Extension–time profiles for nine copies of ubiquitin assuming  $k_f = 0$  shows no refolding events at small force even though it reproduces the overall trend. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

$$\Delta t = \frac{1}{|k_u N_f - k_f N_u|}. \quad (10)$$

If  $k_u N_f - k_f N_u > 0$ , then  $dN_f^*/dt < 0$  (Eq. (2)), so one protein unfolds after the dwell time  $\Delta t$ , otherwise, one protein refolds after  $\Delta t$ . Note that usually  $N_u$  is small or zero at the beginning of the experiment, so that  $k_u N_f - k_f N_u > 0$  results in sequential unfolding of the proteins. As more and more proteins unfold,  $(k_u N_f - k_f N_u)$  decreases and thus the unfolding time increases, which is indeed observed in experiments [8]. If the force is large enough that the term  $(k_u N_f - k_f N_u)$  remains positive before  $N_f$  decreases to 0, then all the proteins can unfold (Fig. 7a). On the other hand, if the force is small (e.g.,  $\sim 80$  pN for ubiquitin),  $(k_u N_f - k_f N_u)$  becomes negative at some time before  $N_f$  reaches 0, then unfolding events cease, and periodic refolding and unfolding events ensue with  $(k_u N_f - k_f N_u)$  switching sign each time an event occurs (inset of Fig. 7a). Although it is difficult to do the constant-force experiment on copies of proteins at low force ( $< 70$  pN for ubiquitin) using AFM [8], the unfolding and refolding “hopping” was indeed found in a simulation on an  $\alpha$ -helix with a similar force (78.2 pN) [23] as well as in experiments on RNA hairpin using optical tweezers [24].

By setting  $k_u N_f - k_f N_u = 0$ , we can obtain the critical force below which refolding/unfolding “hopping” will occur:

$$F_{cc} = \frac{k_B T}{\Delta x_u + \Delta x_f} \ln \left( \frac{\beta_0 N_u}{\alpha_0 N_f} \right) \quad (11)$$

This critical force  $F_{cc}$  keeps increasing as more and more proteins unfold.

Using the parameters obtained from constant-velocity pulling, we find that an applied force larger than 84 pN is required for all the ubiquitins to unfold if there are nine copies of the protein in the chain. Note that the critical

force is different for constant-force pulling and pulling with a linearly increasing force unless  $v_f = 0$  (Eqs. (11) and (6)). Also note that there is no such critical force if we assume  $k_f = 0$ , and therefore a model ignoring refolding unrealistically predicts that all the proteins in the chain should unfold and no “hopping” events should occur no matter how small the force is.

If we assume that  $N_f$  and  $N_u$  are continuous real numbers in the model, then the analytic solution for  $N_f(t)$  is:

$$N_f(t) = N_{f\infty} + (N_{f0} - N_{f\infty})e^{-t/\tau} \quad (12)$$

where  $N_{f\infty} = Nk_f/(k_f + k_u)$  is the number of folded proteins that remain in the chain as  $t \rightarrow \infty$ ,  $N_{f0}$  is the number of folded proteins at  $t = 0$  and  $\tau = 1/(k_f + k_u)$  is the time constant.

We plot  $N_{f\infty}/N$  vs. a dimensionless force  $\Pi_F = \frac{F(\Delta x_u + \Delta x_f)}{k_B T}$  in Fig. 7b and find that the transition of  $N_{f\infty}/N$  from 90% to 10% occurs over a narrow range of  $\Pi_F$ , a result also reported in an earlier simulation work on an  $\alpha$ -helix and suggested to have some relation to mechanotransduction [23]. From our expression for  $N_{f\infty}$ , we further show that this transition range  $\delta F$  (Fig. 7b) can be calculated analytically and the value turns out to be a universal constant  $\ln 81 \approx 4.4$ , independent of all the kinetic parameters. In other words, for any protein that obeys the two-state model, the transition always occurs over 4.4 units of the dimensionless force. Also, it can be shown that the dimensionless force  $\Pi_F^*$  that unfolds half the proteins is related to the kinetic parameters by  $\Pi_F^* = \ln(\beta_0/\alpha_0)$ . This suggests a new way to determine  $\beta_0$  experimentally. Note that all the discussions here for  $N_{f\infty}$  and  $\delta F$  are the results of taking the refolding rate into account, otherwise  $N_{f\infty}/N \equiv 0$  and is not a function of  $F$ .

Eq. (12) combined with Eq. (1) further leads to an analytic expression for the relative extension as a function of time:

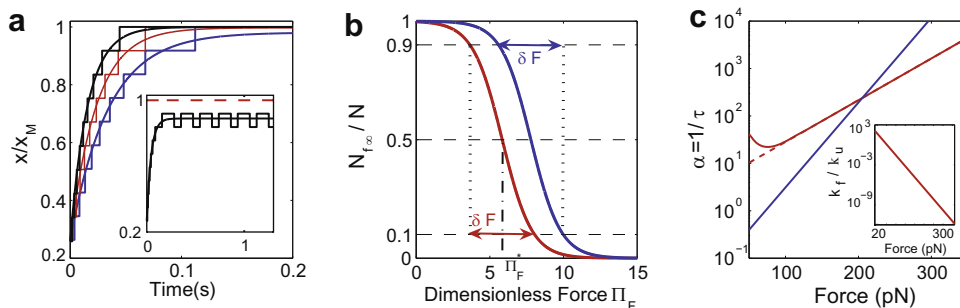


Fig. 7. Constant-force pulling profiles. (a) Extension vs. time profile for nine copies of ubiquitin. Blue,  $F = 100$  pN; red,  $F = 120$  pN; black,  $F = 140$  pN. Inset,  $F = 80$  pN. Stepwise curves, solution when assuming  $N_f$  and  $N_u$  to be integers in both Eqs. (1) and (2). Continuous curves, analytic solution obtained by letting  $N_f$  and  $N_u$  be real numbers in the equations. The dwell time for the stepwise solution, which is obtained analytically in the text, increases as the protein unfolds. (b)  $N_{f\infty}/N$  as a function of the dimensionless force  $\Pi_F$ . Red, ubiquitin; blue, fibrinogen.  $\delta F$ , the range of the dimensionless force over which  $N_{f\infty}/N$  changes from 90% to 10%, is a universal constant  $\ln 81$  for all proteins.  $\Pi_F^*$  is the dimensionless force for half the proteins to unfold. It is shown in the text that  $\Pi_F^* = \ln(\beta_0/\alpha_0)$ . (c) Unfolding rate  $\alpha$  as a function of the applied force  $F$ . The red solid line is the result of taking the refolding rate into account. When the force is large enough,  $\log \alpha$  is a linear function of  $F$  (red broken line). The blue solid line is the prediction assuming  $k_f = 0$ . Inset,  $k_f/k_u$  as a function of the applied force. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

$$\frac{x}{x_M} = \left(1 - \frac{\Delta x N_{f\infty}}{x_M}\right) - \frac{\Delta x (N_{f0} - N_{f\infty})}{x_M} e^{-t/\tau} \quad (13)$$

where  $x_M = NL_{us}[\coth(FL_u/(k_B T)) - k_B T/(FL_u)]$  is the extension of the chain when all proteins are unfolded, and  $\Delta x = L_{us}[\coth(FL_u/(k_B T)) - k_B T/(FL_u)] - L_{fs}[\coth(FL_f/(k_B T)) - k_B T/(FL_f)]$  is the difference in length between a single unfolded and folded protein. We plot the analytic solution from the continuous model (Eq. (13)) together with the stepwise solution from the discrete model in Fig. 7a. The two results agree quite well.

Moreover, Eq. (13) leads to the conclusion that the unfolding rate  $\alpha = 1/\tau$  is a double exponential of the force  $F$  (Fig. 7c):

$$\alpha = \frac{1}{\tau} = \alpha_0 \exp\left(\frac{F\Delta x_u}{k_B T}\right) + \beta_0 \exp\left(-\frac{F\Delta x_f}{k_B T}\right) \quad (14)$$

When the applied force is large enough so that  $k_f \ll k_u$  (inset of Fig. 7c), then  $\ln \alpha \approx (\Delta x_u F/k_B T) + \ln \alpha_0$  is linear with respect to the force  $F$ , as has been shown experimentally [8]. For small forces, the unfolding rate decreases with increasing force, which is consistent with earlier works [23,27]. Note that decreasing the unfolding rate at small force is due to the decreasing  $k_f$ , which again suggests that  $k_f$  cannot be ignored, especially at small forces. Eq. (14) further gives a way to fit all the four kinetic parameters using the constant-force pulling experimental data.

### 3. Conclusion

We have obtained solutions to a kinetic two-state model for protein unfolding based on a heterogeneous FJC model and Bell's model. This model describes the forced unfolding of a chain of proteins under various kinds of loading conditions. Using this model we have obtained analytic solutions that can predict the response of a chain of proteins under a linearly increasing force or a constant force. The model can also be used to fit the experimental data from constant-velocity pulling experiments, as we have demonstrated for ubiquitin and fibrinogen. In particular, we have used the experimental data to solve directly for all the four kinetic parameters and predict the response of the proteins under a linearly increasing or constant force. Our solutions show that the refolding rate is much larger than the unfolding rate at zero force and that interesting physics is revealed if we account for the refolding rate at low forces.

### Acknowledgments

The authors thank Andre Brown and Dennis Discher for providing the experimental data on fibrinogen, and Guoliang Yang for the experimental data on ubiquitin. This work was partially supported by a seed grant from the Nano-Bio Interface Center at the University of Pennsylvania.

### References

- [1] Rief M, Gautel M, Oesterhelt F, Fernandez JM, Gaub HE. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* 1997;276:1109–12.
- [2] Oberhauser AF, Marszalek PE, Erickson HP, Fernandez JM. The molecular elasticity of the extracellular matrix protein tenascin. *Nature* 1998;393:181–5.
- [3] Carrion-Vazquez M, Oberhauser AF, Fowler SB, Marszalek PE, Broedel SE, Clarke J, et al. Mechanical and chemical unfolding of a single protein: a comparison. *Proc Natl Acad Sci USA* 1999;96:3694–9.
- [4] Schwaiger I, Sattler C, Hostetter DR, Rief M. The myosin coiled-coil is a truly elastic protein structure. *Nat Mater* 2002;1:232–5.
- [5] Carrion-Vazquez M, Li H, Lu H, Marszalek PE, Oberhauser AF, Fernandez JM. The mechanical stability of ubiquitin is linkage dependent. *Nat Struct Biol* 2003;10:738–43.
- [6] Fernandez JM, Li H. Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science* 2004;303:1674–8.
- [7] Chyan CL, Lin FC, Peng H, Yuan JM, Chang CH, Lin SH, et al. Reversible mechanical unfolding of single ubiquitin molecules. *Biophys J* 2004;87:3995–4006.
- [8] Schlierf M, Li H, Fernandez JM. The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proc Natl Acad Sci USA* 2004;101:7299–304.
- [9] Brown AE, Litvinov RI, Discher DE, Weisel JW. Forced unfolding of coiled-coils in fibrinogen by single-molecule AFM. *Biophys J* 2007;92:L39–41.
- [10] Bornschlog T, Rief M. Single-molecule dynamics of mechanical coiled-coil unzipping. *Langmuir* 2008;24:1338–42.
- [11] Lim BBC, Lee EH, Sotomayor M, Schulten K. Molecular basis of fibrin clot elasticity. *Structure* 2008;16:449–59.
- [12] Bhasin N, Carl P, Harper S, Feng G, Lu H, Speicher DW, et al. Chemistry on a single protein, VCAM-1, during forced unfolding. *J Biol Chem* 2004;279:45865–74.
- [13] Sotomayor M, Schulten K. Single-molecule experiments in vitro and in silico. *Science* 2007;316:1144–8.
- [14] Rief M, Fernandez JM, Gaub HE. Elastically coupled two-level systems as a model for biopolymer extensibility. *Phys Rev Lett* 1998;81:4764–7.
- [15] Bustamante C, Marko JF, Siggia ED, Smith S. Entropic elasticity of lambda-phage DNA. *Science* 1994;265:1599–600.
- [16] Marko JF, Siggia ED. Stretching DNA. *Macromolecules* 1995;28:8759–70.
- [17] Bell GI. Models for specific adhesion of cells to cells. *Science* 1978;200:618–27.
- [18] Dudko OK, Hummer G, Szabo A. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys Rev Lett* 2006;96(1–4):108101.
- [19] Nelson P. *Biological physics: energy, information, life*. New York: W.H. Freeman and Company; 2008.
- [20] Su T. A model for forced unfolding of proteins. Technical Report, Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania; 2008. p. 1–30.
- [21] Fisher TE, Oberhauser AF, Carrion-Vazquez M, Marszalek PE, Fernandez JM. The study of protein mechanics with the atomic force microscope. *Trends Biochem Sci* 1999;24:379–84.
- [22] Irback A, Mitternacht S, Mohanty S. Dissecting the mechanical unfolding of ubiquitin. *Proc Natl Acad Sci USA* 2005;102:13427–32.
- [23] Karcher H, Lee SE, Kaazempur-Mofrad MR, Kamm RD. A coarse-grained model for force-induced protein deformation and kinetics. *Biophys J* 2006;90:2686–97.
- [24] Bustamante C, Chemla YR, Forde NR, Izhaky D. Mechanical processes in biochemistry. *Annu Rev Biochem* 2004;73:705–48.
- [25] Best RB, Hummer G. Protein folding kinetics under force from molecular simulation. *J Am Chem Soc* 2008;130:3706–7.

- [26] Schlierf M, Rief M. Temperature softening of a protein in single-molecule experiments. *J Mol Biol* 2005;354:407–503.
- [27] Best RB, Paci E, Hummer G, Dudko OK. Pulling direction as a reaction coordinate for the mechanical unfolding of single molecules. *J Phys Chem B* 2008;112(19):5968–76.
- [28] Imparato A, Pelizzola A. Mechanical unfolding and refolding pathways of ubiquitin. *Phys Rev Lett* 2008;100:158104.
- [29] Cao Y, Li H. Polyprotein of GB1 is an ideal artificial elastomeric protein. *Nat Mater* 2007;6:109–14.