Lattice Model Studies of Force-Induced Unfolding of Proteins†

D. K. Klimov* and D. Thirumalai*

Institute for Physical Science and Technology and Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742

Received: January 16, 2001; In Final Form: April 19, 2001

We probe the general characteristics of force-induced unfolding of proteins using lattice models. The computations show that the experimental observations, such as the shape of the force-extension curves and hysteresis, are qualitatively reproduced using the coarse-grained models. Force hysteresis, which occurs because the structural relaxation times are longer than the time scales for dissipation of stored mechanical energy, strongly depends on the rate of application of force or the pulling speed. As a result, refolding is not spontaneous, when force is decreased after fully extending the polypeptide chain. Most importantly, we show that the distribution of unfolding free energy barriers in the *absence of force* can be obtained using the dynamics of force-induced unfolding. This key result immediately suggests that dynamic single molecule force spectroscopy can be used to directly measure the folding free energy landscape of proteins.

I. Introduction

A remarkable series of single molecule experiments have shown that proteins unfold when are subject to force.[1−6] These nanomanipulation techniques, in which proteins are pulled using atomic force microscopy (AFM) or optical tweezers, can be used to map the free energy landscape of biomolecules at the single molecule level. To date, these experiments have provided measurements of the forces required to stretch the molecule by a certain distance. This, in turn, has led to a fundamental microscopic understanding of the elastic behavior of muscle proteins. The initial experiments showed that, upon application of force, the multimodular protein titin unfolds one domain at a time.[1,3] Similar experiments on a number of proteins have verified that unfolding at the single molecule level can be monitored by subjecting the folded proteins to tension.[4,7−10]

The AFM experiments, which showed the characteristic saw-tooth pattern in the force-extension ($f - z$) curves, were interpreted using the wormlike chain (WLC) model for the domains.[1−3] The complete unraveling of the weakest domains was inferred from the spacing between the saw-tooth patterns. Adaptation of Flory estimate of the size of the polypeptide chain can be used to predict the expected spacing assuming that complete unfolding of individual domains occurs as long as sufficient force is applied. In the folded state, the radius of gyration R_g of the immunoglobulin Ig domains made up of about $N = 90$ amino acid residues is $R_g \approx aN^{1/3} \approx 2$ nm assuming that a, the distance between the alpha-carbon atoms, is 3.8 Å. In the fully stretched conformation, $R_g = (N - 1)a \approx 33$ nm. Thus, the expected spacing is about 31 nm, which is in fair agreement with the experimental value.[1,3,8] This elementary argument supports the experimental interpretation.

Besides providing a microscopic basis for elasticity in biomolecules the stretching experiments can, in principle, contribute to a richer picture of the routes navigated by proteins compared to the classical biochemical experiments, in which folding (or unfolding) is initiated by altering the concentration of denaturants or temperature. In this paper, we show, using lattice models of proteins, that certain features of the folding landscape of proteins in the absence of force may be obtained using the kinetic and thermodynamic properties of proteins that are subject to force. Because the detailed folding mechanism and pathways for these model proteins can be computed with zero and nonzero force values a critical assessment of the utility of single molecule stretching experiments can be made.

Because lattice models are, at best, coarse-grained caricature of proteins they cannot be used to address the details of force-induced unfolding pathways. This level of description would require detailed all atom simulations and other theoretical approaches[11−13]. Nevertheless, it is possible to obtain general features of unfolding of proteins under tension using lattice models. Their detailed numerical studies also suggest novel experiments.

II. Model and Computational Methods

A. Protein Lattice Model. We model a polypeptide chain as a self-avoiding walk on a cubic lattice with the spacing $a = 3.8$ Å. A conformation of a chain is given by the vectors $\{\vec{r}_i\}$, $i = 1, 2...N$. A contact is formed, when two nonbonded beads become neighbors on a lattice. The energy of the chain is the sum of interaction energies assigned to the contacts

$$E_p = \sum_{i<j} \Delta(|r_i - r_j|)B_{ij} \qquad (1)$$

where Δ is unity, when $|r_i - r_j| = a$ and zero, otherwise, and B_{ij} is the contact energy between beads i and j. For B_{ij}, we used contact potentials matrix introduced by Kolinski, Godzik, and Skolnick (KGS),[14] in which B_{ij} are expressed in the units of RT. We have performed exhaustive force-induced unfolding simulations for a number of sequences with $N = 15$, 27, and 36. For clarity, most of the results presented here are for $N = 36$ sequence used in our previous papers.[15,16]

† Part of the special issue "Bruce Berne Festschrift". It is a pleasure to dedicate this paper to Bruce J. Berne on the occasion of his 60th birthday.

* To whom correspondence should be addressed.

10.1021/jp0101561 CCC: $20.00
Published on Web 06/14/2001

We apply force f toward the ends of a sequence. The stretching energy E_s arising due to unraveling a sequence by force f, is taken in the form

$$E_s = -fr \tag{2}$$

where r is the end-to-end distance ($= |r_1 - r_N|$). In writing eq 2, we assume that relaxation of rotational degrees of freedom takes place on a much faster time scale than folding/unfolding time scales.[15,17] Apart from logarithmic factors, the low temperature limit of the rotationally averaged contribution to the free energy due to tension obtained in the context of rubber elasticity coincides with eq 2.

The total energy of a conformation E is given by a sum of E_p and E_s terms. Force tends to unravel protein up to a completely stretched (rodlike) state with the energy $E_s = -f(N-1)$. It is clear that when f is small, the energy of the stretched state is also small with respect to the energy of native (folded) state E_n. Therefore, up to a certain critical value $f = f_c$, the native (folded) state is still a ground state, i.e., the state with the lowest energy. However, when $f \geq f_c$, the native (folded) state is no longer stable and unfolded ground state is observed. Depending on f and sequence these new ground states may be partially or completely stretched.[15]

B. Thermodynamic Functions

The thermodynamic properties are computed using the multiple histogram technique.[18,19] In most folding simulations, temperature T is the sole external parameter, which specifies external conditions. When the polypeptide chain is subject to tension the external conditions are characterized by two parameters, T and f. Standard implementation of the multiple histogram technique,[19] i.e., when histograms are obtained at different T, would not produce reliable results for forced unfolding because highly stretched conformations are almost never sampled. This is not a problem in the calculation of thermodynamic properties for thermal folding, since such states have negligible Boltzmann weight. By contrast, upon tension induced unfolding stretched states (including completely stretched rodlike state) become thermodynamically important. With this in mind, we modified the multiple histogram technique and collected histograms at fixed temperature T_s, but at different values of force f. This can be readily done since the Ferrenberg−Swendsen formulation of multiple histogram method is quite general.[18] We chose T_s to ensure frequent sampling of the nativelike states at small values of f. It must also be not too high to allow sampling of highly stretched states when force is applied. In what follows, we describe the details of the implementation of the method.

The data to be analyzed by the multiple histogram technique are produced by standard Monte Carlo method.[20] At first, we need to establish the range of force values over which force induced unfolding actually takes place. For each sequence we generated 20 trajectories at $T_s \simeq T_F$ with the native state as an initial condition. In the course of these simulations, which correspond to slow stretching (compare to slow cooling), we gradually increase force from zero to $f_{max} = 4.0$. At $f_{max} = 4.0$, all sequences adopt highly stretched conformations with occasional contacts present due to thermal fluctuations. We generated $M = 75$ production trajectories with the native state as the initial condition. For 36-mer sequence the temperature of simulation $T_s = 0.53$, which is close to T_F, the folding temperature at zero force. The force f was incremented by Δf every t Monte Carlo steps (MCS). Each trajectory starts with $f_{min} = 0.0$ and ends at $f_{max} = 4.0$. In the course of a trajectory, the force changes periodically according to $f_k = f_{min} + k\Delta f$, where $k = 0,...,(f_{max} - f_{min})/\Delta f$ and $\Delta f = 1.0$. Thus, we collected histograms at five values of force $f = 0.0$, 1.0, 2.0, 3.0, and 4.0. Histograms were collected separately at the force values f_k using all M trajectories. In total, we have $K = (f_{max} - f_{min})/\Delta f + 1$ histograms. Three variables were monitored in our multiple histogram simulations. These are the potential energy E_p (eq 1), the end-to-end distance r, and the overlap function[20] defined as

$$\chi = 1 - \frac{1}{N^2 - 3N + 2} \sum_{i \neq j,\, j \pm 1} \delta(r_{ij} - r_{ij}^0) \tag{3}$$

where r_{ij} is the distance between the beads i and j, r_{ij}^0 is the corresponding distance in the native conformation, and $\delta(x)$ is the Kronecker delta function.

The length of the trajectory interval generated for a given fixed force f_k was set to 10^7 MCS. A portion of trajectory ($\lesssim 10^6$ MCS) immediately after the force increment must be excluded in order to allow a sequence to equilibrate at a new f_k. While collecting states in the histograms we used 0.1 grid interval for the energy E_p and discrete exact values for χ and r.

A thermodynamic average of a quantity A being the function of, say, E_p, r, χ is calculated using

$$\langle A \rangle = \frac{\sum_{E_p}\sum_{r}\sum_{\chi} A(E_p, r, \chi) e^{-((E_p - fr)/T)} \dfrac{\sum_{k=1}^{K} h(E_p, r, \chi)(f_k)}{\sum_{k=1}^{K} n_k e^{F_k - ((E_p - f_k r)/T_s)}}}{\sum_{E_p}\sum_{r}\sum_{\chi} e^{-((E_p - fr)/T)} \dfrac{\sum_{k=1}^{K} h(E_p, r, \chi)(f_k)}{\sum_{k=1}^{K} n_k e^{F_k - ((E_p - f_k r)/T_s)}}} \tag{4}$$

where K is the number of histograms, T_s is the temperature of simulations, f_k is the force, at which kth histogram is collected, n_k is the number of states in the kth histogram, and F_k is the scaled free energy to be calculated self-consistently from the following equation

$$e^{-F_l} = \sum_{E_p}\sum_{r}\sum_{\chi} e^{-((E_p - f_l r)/T_s)} \frac{\sum_{k=1}^{K} h(E_p, r, \chi)(f_k)}{\sum_{k=1}^{K} n_k e^{F_k - ((E_p - f_k r)/T_s)}} \tag{5}$$

The values of F_k can be determined from an iterative scheme within about 500−1000 iterations with excellent accuracy.

We have calculated the overlap function $\langle \chi\ (T, f) \rangle$, total (conformational plus stretching) energy $\langle E\ (T, f) \rangle$, and the end-to-end distance $\langle r\ (T, f) \rangle$ as functions of temperature and force.

C. Probing Unfolding Kinetics

The unfolding times were calculated from Monte Carlo simulations performed at the simulation temperature T_s and simulation force f_s. The kinetic simulation temperature is sequence dependent and is obtained from the condition $\langle \chi\ (T = T_s, f = 0) \rangle = 0.15$, which gives $T_s = 0.49$. The first passage unfolding time $\tau_{u,1i}$ for a particular initial condition i is associated with the first instance when the chain does not make any contacts. From this we calculate the function $P_f(t)$, which is the fraction of trajectories which do not reach the stretched conformation at time t

$$P_f(t) = 1 - \int_0^t P_{fp}(s)\mathrm{d}s \tag{6}$$

where $P_{fp}(s)$ is the distribution of first passage times to the stretched state

$$P_{fp}(s) = \frac{1}{M}\sum_{i=1}^{M}\delta(s - \tau_{u,1i}) \tag{7}$$

It may be shown that the mean unfolding time τ_U is given by

$$\tau_U = \int_0^\infty tP_{fp}(t)\mathrm{d}t = \int_0^\infty P_f(t)\mathrm{d}t \tag{8}$$

The fit of $P_f(t)$ may be used for calculating τ_U. We found that a reasonable fit to function $P_f(t)$ may be achieved using exponentials. The mean unfolding time can also be calculated using the equation

$$\tau_U = \frac{1}{M}\sum_{i=1}^{M}\tau_{u,1i} \tag{9}$$

To gain a more detailed picture of tension-induced unfolding it is instructive to calculate the meantimes of breaking up individual contacts. Because of thermal fluctuations it is not informative to define the breaking time $\tau_{b,i}^k$ of the contact labeled k for trajectory i as the time when the contact is disrupted for the first time. Instead, it is useful to introduce a waiting time interval Δt and assume that the contact k is broken at the time $\tau_{b,i}^k$, if it does not occur over the time interval Δt, immediately preceding $\tau_{b,i}^k$. We found qualitatively similar results with $\Delta t = 100-1000$ MCS. However, if $\Delta t \ll 100$ MCS, the breaking time $\tau_{b,i}^k$ may not reflect the actual breakage of respective contact, but merely instantaneous fluctuations. We set Δt to 100 MCS.

An approach to stretched state is monitored using several quantities, such as the end-to-end distance $\langle r(t)\rangle$

$$\langle r(t)\rangle = \frac{1}{M}\sum_{i=1}^{M} r_i(t) \tag{10}$$

where $r_i(t)$ is end-to-end distance at the time t for the trajectory i. Similarly, the time dependence of the overlap function $\langle\chi(t)\rangle$[15] can be computed.

III. Results and Discussion

A. Phase Diagrams. The native conformation for 36-mer sequence, which has 40 contacts, is maximally compact. The multiple histogram method suitably modified for the present study (see Methods) allows us to construct phase diagrams (Figure 1). The thermodynamic properties and folding kinetics at $f = 0$ show that the sequence is a two-state folder. This is not unexpected because the folding transition temperature $T_F = 0.53$ (determined from the peak in the fluctuations of the overlap $\Delta\chi$) coincides with T_θ, the collapse temperature obtained from the maximum in the specific heat C_v.[20] The phase diagram represented in terms of the surface plot for $\langle\chi(T, f)\rangle$ shows that upon application of force $f \gtrsim f_c \approx 1.0$ there is an abrupt unfolding. However, the native state is stable at low temperatures $T \lesssim T_F = 0.53$ and weak force $f \lesssim f_c$. There are no intermediates present in the transition from the folded state to the fully stretched conformations. This is a direct evidence of two-state unfolding transition induced by force. This behavior is not universal, but depends on the folding characteristics in the absence of force.[15,17,21]

The boundary between folded and unfolded states is well-defined (Figure 2) and resembles the boundary of type I superconducting state on the B, H plane. Using the diagram $\langle\chi(T, f)\rangle$, we define the boundary of folded state as the curve on (T, f) plane, for which the overlap with the native state $\langle\chi\rangle$ is equal to $\langle\chi(T_F, f = 0)\rangle$, where T_F is the folding transition temperature. Keeping the analogy to the superconductors in mind, we fit the native state boundary with the function

$$f_c \sim f_0\left(1 - \left(\frac{T}{T_F}\right)^\alpha\right) \tag{11}$$

where f_0 is the critical value of force at $T = 0$. Using eq 11 to fit the curve enclosing the folded state, we found that $f_0 = 0.98$ and the exponent $\alpha = 6.0$ (Figure 2). The large value of α may suggest that the force-induced unfolding transition is weakly first order. This possibility is further supported in Figure 3,

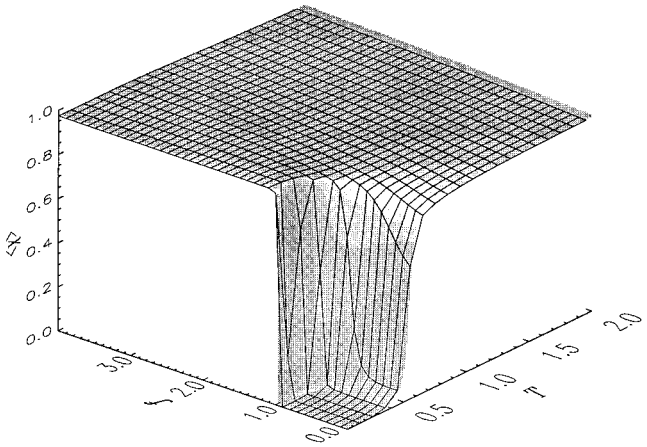

Figure 1. Temperature−force (T, f) phase diagram for the 36-mer sequence is presented in terms of the surface plot of the thermal average of the overlap function $\langle\chi(T, f)\rangle$. The plot shows that the native state (small $\langle\chi\rangle$) is stable at low temperatures $T < T_F$ and small forces $f < f_0$. A highly cooperative unfolding of the native structure occurring when T or f increase suggests two-state transition. Note that force-induced unfolding at f_0 (along f-axis) is more cooperative than the temperature-induced unfolding at T_F (along T-axis).

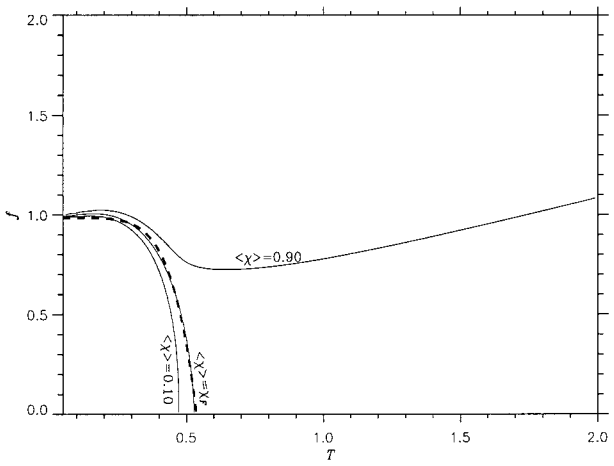

Figure 2. Contour plot for several overlap value levels $\langle\chi\rangle$ in the $T-f$ plane. The boundary of the native state is given by the contour level of $\langle\chi(T, f)\rangle = \langle\chi(T_F, 0)\rangle = \chi_F$, which determines the dependence of the critical force f_c on temperature. Fitting f_c with eq 11 (thick dashed line) gives $f_0 = 0.98$ and $\alpha = 6.0$. The value of f_0 measures the critical value of force, at which the native state becomes unstable at low T, whereas α controls the nature of force-induced unfolding transition. Large α as obtained for the 36-mer sequence suggests highly cooperative first order like transition.

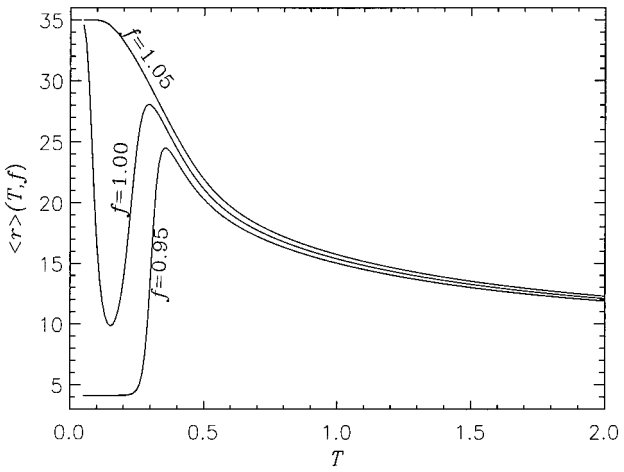

Figure 3. The dependence of the end-to-end extension $\langle r \rangle$ on temperature for several fixed values of force f (marked in the plot) in the vicinity of the critical force $f_0 = 0.98$. The plot reveals dramatic conformational changes within relatively narrow temperature interval that may be interpreted as an evidence for all-or-none nature of unfolding transition.

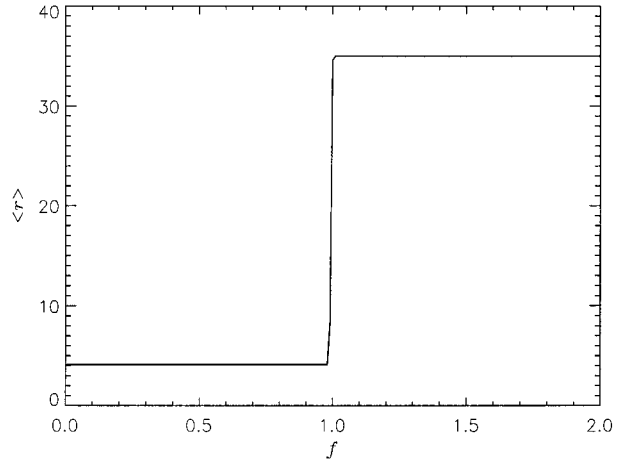

Figure 4. The profile of the end-to-end extension $\langle r \rangle$ as a function of force at fixed low temperature $T = 0.1$. An extremely sharp increase in $\langle r \rangle$ is observed at $f \approx f_0$. At smaller forces the extension $\langle r \rangle$ is at the native value 4.12, whereas at $f > f_0$ the sequence is fully stretched ($\langle r \rangle = 35$). The plot explicitly shows a two-state nature of force-induced unfolding.

which shows the end-to-end distance $\langle r(T, f) \rangle$ as a function of temperature for $f = 0.95$, 1.00, and 1.05, which are in the neighborhood of f_0. Strikingly, in this narrow transition region $\langle r \rangle$ does not change monotonically with T. At $f = 1.00 \gtrsim f_0$ the fully stretched state stable at low temperature is destabilized over a narrow T interval. Within this narrow interval the native folded state is almost completely recovered ($\langle \chi \rangle$drops below 0.2). This implies that over a very narrow temperature range ($\Delta T \lesssim 0.03$) a coexistence between the rod state and the folded conformation is possible.

The conformational changes (Figure 4) which the sequence undergoes at f_c are very dramatic (even spectacular). This figure displays $\langle r(T = 0.1, f) \rangle$ as a function of f. At $f_c \approx 0.98$ the sequence abruptly unfolds from the intact native conformation with $\chi = 0.0$ and $r = |r_1 - r_N| = 4.12$ to a completely stretched state with $r = 35$. This highly cooperative transition occurs in an extremely narrow interval of force of about 0.01. It is also interesting to analyze the dependence of the energy $\langle E \rangle$ upon such a dramatic conformational change. This dependence is represented by a combination of two linear functions, each corresponding to the native and completely stretched states. At $f < f_c \approx f_0$ we have $E \simeq E_0 - f|r_1 - r_N|$, whereas at $f > f_c \approx f_0$ E is just $-(N-1)f$. From this, we may get an estimation for f_0 as follows

$$E_0 - f_0|r_1 - r_N| \simeq -(N-1)f_0 \tag{12}$$

In the native state $|r_1 - r_N| = \sqrt{17}$ and $E_0 = -30.4$, from which we get $f_0 = 0.98$, which is in perfect agreement with the fit based on eq 11.

B. Kinetics of Unfolding at Constant Force. The kinetic unfolding simulations were performed at the force f_s fixed at the value of 4.0. The native structure was used as initial condition in these simulations. The unfolding rate increases as f becomes larger. This is also borne out in the free energy profiles calculated as a function of the number of native contacts. The average unfolding barrier $\langle \Delta F_U^{\ddagger}/T_s \rangle = 4.6$ at $f = 0$ and decreases to 2.8 at $f = 0.7$. The kinetics of approach to the fully stretched state (for details see ref 15) shows that unfolding occurs in a stepwise fashion. There is considerable heterogeneity in the times needed to reach the rodlike state. We show below that the stretch times of individual molecules in the presence of force can be used to calculate the distribution of unfolding barrier heights when $f = 0$.

C. Distribution of Unfolding Free Energy Barriers at $f = 0$. One of the potential, but unexploited, uses of force spectroscopy is the possibility of measuring the distribution of barrier heights at $f = 0$. This idea can be demonstrated using lattice model. For the 36-mer sequence, the refolding time $\tau_F(f = 0)$ at $T_s = 0.49$ is 5.7×10^5 MCS. Because this sequence is a two-state folder we can write $\tau_F = \tau_0^F \exp(\langle \Delta F_N^{\ddagger} \rangle / T_s)$, where $\langle \Delta F_N^{\ddagger} \rangle$ is the average refolding barrier between the unfolded state **U** and the folding transition state. Using the computed value for τ_F and the estimate $\langle \Delta F_N^{\ddagger} \rangle / T_s \approx 0.5$ from the calculated free energy profiles (see Figure 3b of ref 15) we obtain $\tau_0^F \approx 349\,338$ MCS. To compute the distribution of the unfolding free energy barriers for $f = 0$, $P(\Delta F_U^{\ddagger})$, we assume that the prefactor for unfolding $\tau_0^U \approx \tau_0^F$ and that their values do not change significantly in the presence of f. For each individual trajectory, we compute the dynamics of approach to the stretched state starting from the native conformation. Previously, we showed that there is a large variability in the stretch times $\tau_{u,1i}$ for individual trajectories.[15] In the presence of force we can write

$$\tau_{u,1i} = \tau_0^U \exp(\beta\Omega) \tag{13}$$

where $\beta\Omega = \beta(F_{U,i}^{\ddagger} - f\Delta x)$ with $\beta = 1/T_s$ and $\Delta F_{U,i}^{\ddagger}$ being the unfolding free energy barrier for a trajectory i. The parameter Δx is the distance, by which the polypeptide chain is extended before unfolding spontaneously occurs. For our simulation conditions, we showed that $\Delta x = 0.02L$ ($L = (N-1)a$).[15] From the computation of the stretch times for each unfolding molecule, $P(\Delta F_U^{\ddagger})$ can be readily calculated. A plot of $P(\Delta F_U^{\ddagger})$ is shown in Figure 5. The average unfolding barrier height computed using $\int \Delta F_U^{\ddagger} P(\Delta F_U^{\ddagger}) d\Delta F_U^{\ddagger} = 4.1T_s$ is consistent with our previous estimate using the free energy profile for this sequence.[15]

The procedure described here can be used in experimental studies provided the dynamics of unfolding upon stretching can be monitored. This might require, perhaps, a two-dimensional experiment, in which force and fluorescence are simultaneously monitored. It is likely that single molecule fluorescence resonant

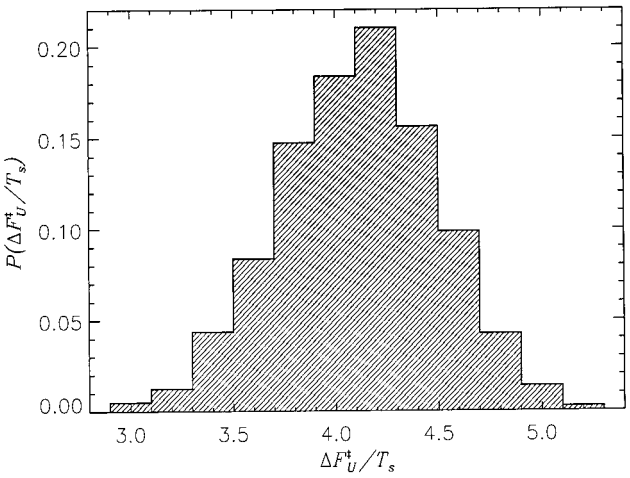

Figure 5. The distribution of the unfolding free energy barriers $P(\Delta F^{\ddagger}_U)$ at zero force and $T_s = 0.49 < T_F$. The variations in $\Delta F^{\ddagger}_U$ computed, as described in the text, using 800 individual unfolding trajectories are due to the heterogeneity of unfolding pathways. The average unfolding free energy barrier is 4.1, which is in a fair agreement with $\langle \Delta F^{\ddagger}_U \rangle / T_s = 4.6$ estimated from the free energy profile calculated as a function of the number of native contacts, see ref 15.

energy transfer experiments alone can be used to measure $P(\Delta F^{\ddagger}_U)$. However, the combination of the two can give a richer picture of the underlying free energy landscape.

D. Hysteresis during Stretch-Release Cycles. The stretching of titin by optical tweezers[2,3] showed that when the unfolded protein is released, the molecule does not spontaneously refold. Upon decreasing the force, its extension initially decreases by about 50% compared to the fully stretched state. After this event, it may take several seconds (depending on the pulling speed) for the native state to recover. The release part of the force-extension ($f - z$) curve does not coincide with the stretch part unless the force is very small, i.e., there is a pronounced kinetic hysteresis in the stretch-release cycle. Kinetic hysteresis occurs because the rate of dissipation of the stored elastic energy in the molecule is much higher than the rate of gain in the free energy upon folding. The latter is a measure of the structural relaxation of the molecule. In the experiment of Kellermayer et al.[2], titin was stretched at the rate $\upsilon_s \approx 60$ nm/s. If we assume that refolding occurs domain by domain upon release the characteristic structural relaxation rate $\lambda_s \sim R_g/\tau_F$, where R_g is the size of the domain and τ_F is the refolding time at equilibrium. Because hysteresis is observed only when the extension exceeds a threshold value, $R_g \simeq L \simeq 28$ nm for Ig or FnIII domains.[1,7] The condition for hysteresis is $R_z = \upsilon_s/\lambda_s > 1$. In the optical tweezers experiments $R_z \approx 1.1$ using $\tau_F \approx 0.5s$.[22] We explored the possibility of observing the hysteresis in lattice models. In the experiments, the force required to stretch the molecule to a given extent is measured. In our computations, it is easier to fix the force f_s and calculate the amount by which the chain is stretched. The stretch-release cycles were performed by periodically applying force to the ends of the chain. In the stretch mode of the cycle, we started with $f = 0$ at time $t = 0$ and increased f to 2.0 or 4.0 at a rate $\delta f = 0.1 \times 10^{-4}$ MCS^{-1}. By normalizing the force to the threshold value, a dimensionless variable $R_f = \delta f \tau_F / f_0 \approx 5.8$ emerges (see previous section for τ_F). For this value of R_f, we expect pronounced hysteresis when the force is decreased after fully extending the polypeptide chain. The rate of release was also set to $\delta f = 0.1 \times 10^{-4}$ MCS^{-1}.

The 36-mer sequence shows hysteresis during stretch-release cycles. In Figure 6 we present $\langle r(f) \rangle$ and $\langle \chi(f) \rangle$ hysteresis curves

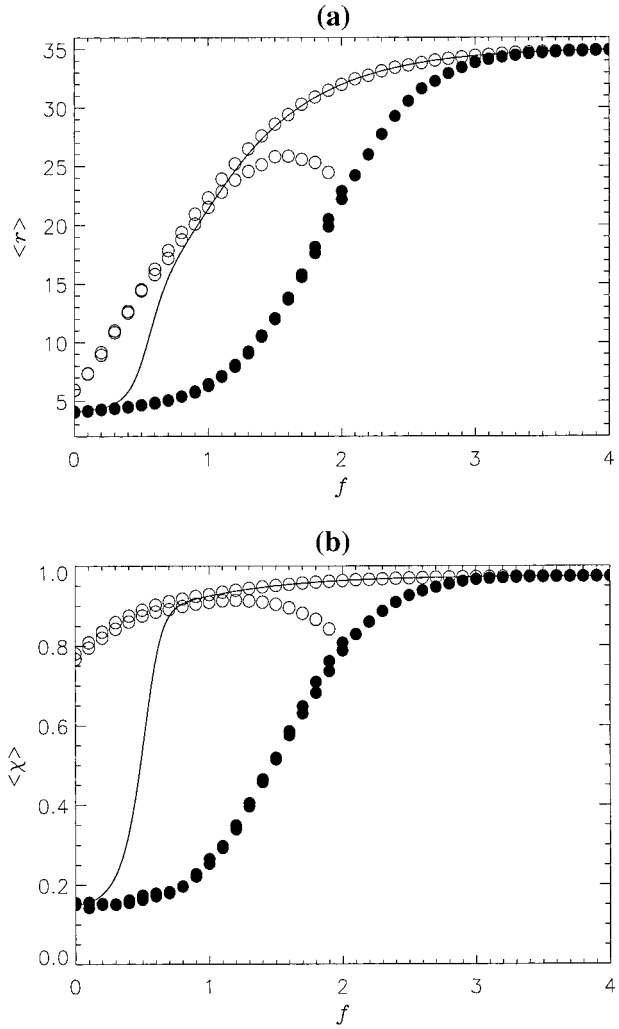

Figure 6. Hysteresis observed during stretch-release cycles is displayed in terms of the extension $\langle r \rangle$ (upper panel) and the overlap $\langle \chi \rangle$ (lower panel). Both quantities are averaged over 400 individual trajectories. The full and empty circles represent stretch and release modes in the cycle, respectively. There are two sets of data on each plot. The first corresponds to the cycle, in which the force reaches maximum of 2.0 (i.e., the cycle is switched from stretch to release mode at $f = 2.0$). In the second cycle the maximum force reached is 4.0. The solid lines indicates the equilibrium values for $\langle r \rangle$ and $\langle \chi \rangle$. The rate of changing the force within both cycles is $\delta f = 0.1 \times 10^{-4}$ MCS^{-1}. Hysteresis is observed because the rate of dissipation of the elastic energy is much higher than refolding rate. Note that after a switch in the cycle mode "inertial" effect develops, in which unfolding continues for a certain time even though the force already started to decrease.

calculated for two maximum values of force, 2.0 and 4.0. The polypeptide chain cannot equilibrate fast enough to accommodate the rapid changes in f and hence the structural relaxation considerably lags during the stretch mode of the cycle. However, eventually at $f > 3.0$, the molecule reaches equilibrium as $\langle r(f) \rangle$ and $\langle \chi(f) \rangle$ start to coincide with the equilibrium values shown by solid lines (Figure 6). In the release mode of the cycle, when f decreases hysteresis appears indicating that the pathways of stretching and releasing do not coincide. Even at $f = 0.0$ sequence is not fully equilibrated because its overlap function is still very large. The very large relaxation time needed to reach equilibrium in the release mode is due to the rapid decrease in f.

The equilibrium force-extension curves can be fit using the standard WLC model. The elastic behavior of a number of muscle proteins including tandem repeats of single domain

proteins is interpreted using the semiflexible chain model.[1−3] In our studies, we find that WLC fit holds at $T > T_F$, whereas significant deviations from WLC model are observed at $T \lesssim T_F$. This is not surprising because the WLC model is likely to describe the extension of random coillike conformations, which the model sequences adopt only at high temperatures ($> T_F$).

E. Dependence of Unfolding Events on the Relative Location of the Amino and Carboxy Termini. The unfolding pathways induced by force depend sensitively on the structure of the native state. The order of unfolding events in the proteins belonging to the immunoglobin family, which share very similar native folds, depends on the precise packing of the β-strands.[21] In most of the proteins for which stretching experiments have been done (e.g., refs 1,4,9), the amino and carboxy termini are on the opposite sites so that upon application of force the strands (helices) slide past each other in a shearing-type motion. The manner in which the tertiary contacts are disrupted when the ends of the chain are close to each other has not been experimentally examined. To address this, we consider two sequences with $N = 27$, whose ground states are maximally compact. The contact energies B_{ij} in eq 1 are drawn from the Gaussian distribution according to

$$P(B_{ij}) = \frac{1}{\sqrt{2\pi}B}\exp\left(-\frac{(B_{ij} - B_0)^2}{2B^2}\right) \qquad (14)$$

where $B_0 = -0.1$ is the average interaction that specifies the strength of the drive toward forming compact structures at low temperatures and the dispersion B (set to unity) gives the extent of diversity of the interactions among beads. In one of these sequences, referred to as 61 in refs 15 and 20, the termini of the chain are on the same facet of the cube, whereas in the other (sequence 63) they are on the opposite facets. The values of $\sigma_T = (T_\theta - T_F)/T_\theta$ (where T_θ and T_F are the collapse and folding temperatures, respectively) for both sequences are small so that they fold by two-state kinetics when $f = 0$[20].

The order, in which the tertiary contacts break up, is probed by computing the average breaking times τ_b^k ($k = 1, 2, 3...28$ for 27-mer) upon applying constant force to the ends of the chain. The dependence of τ_b^k on the position of contact k is established using the quantity $s = (i/N)((N - j)/N)$, which measures how close the contact k, formed between residues i and j, is to the sequence ends. For both sequences τ_b^k correlate well with s (the correlation factor ranges from 0.6 to 0.9). Parenthetically, we note that there is no correlation between τ_b^k and the contact order or the energetics of the contact. We find different scenarios of unfolding for the two sequences. For sequence, 61 contacts, which are closer to one or the other sequence ends, unravel nearly concurrently. In contrast, disruption of contacts localized near different ends in sequence 63 takes place in a dramatically distinct manner. All the contacts positioned near one terminus break up nearly simultaneously and early, whereas the contacts closer to the opposite end survive for a much longer time. Because both of the sequences are structurally similar, we attribute this difference in the unfolding to the relative positions of the terminal residues in the native structure.

IV. Conclusions

Some of the general principles that determine the way proteins unfold when subject to tension can be gleaned using lattice models.[15,17] A number of features that are typically observed in experimental studies of stretching of proteins are reproduced in this study. They are as follows: (a) The $(f - z)$ curves typically exhibit the behavior expected for WLC model[1,2] especially at $T \gtrsim T_F$. This is particularly interesting given the finite size of the polypeptide chain, (b) Just as in the unfolding of individual domains in engineered $(Ig_{27})_n$ constructs[6,8] we also observe an abrupt unfolding without any detectable intermediates, and (c) The stretch-release cycle shows kinetic hysteresis as seen in the experiments on titin using optical tweezers.[2,3] Upon decrease of force, refolding does not commence for a period of time far exceeding τ_F, the refolding time at $f = 0$. Because the rate of release of force in the computations is much greater than the experimental value, the hysteresis effect observed in our computations is much more dramatic. This is seen in Figure 6 which shows that even with $f = 0$ the chain has not reached the native state.

Our calculations also suggest potential applications of force spectroscopy that have not yet been exploited. We have suggested that, by a suitable dynamic, experiments involving fluorescent labeling, the distribution of barrier heights *in the absence of force* can be measured. If the proposed experiments in conjunction with our theoretical analysis is successful many unresolved questions in the folding of biomolecules (proteins and RNA) can be unambiguously answered. We believe that this is the greatest potential of such experiments—a point that clearly emerges in our calculations. We also predict that such dynamic measurements of unfolding of $(Ig_{27})_n$ should exhibit stepwise unraveling just as was observed in the coil−stretch transition in DNA under strong elongational flow. The verification of this prediction would require higher spatiotemporal resolution than is currently available.

Acknowledgment. Useful correspondence with Peter G. Wolynes, nearly 2 years ago, on subtle aspects of probing tension-induced unfolding in lattice models of proteins is greatly appreciated. This work was supported in part by a grant from the National Science Foundation (Grant No. CHE 99-75150).

References and Notes

(1) Rief, M.; Gautel, M.; Oesterhelt, F.; Frenandez, J. M.; Gaub, H. E. *Science* **1997**, *276*, 1109.
(2) Kellermayer, M. S. F.; Smith, S. B.; Granzier, H. L.; Bustamante, C. *Science* **1997**, *276*, 1112.
(3) Tskhovrebova, L.; Trinick, J.; Sleep, J. A.; Simmons, R. M. *Nature* **1997**, *387*, 308.
(4) Oberhauser, A. F.; Marszalek, P. E.; Erickson, H. P.; Fernandez, J. M. *Nature* **1998**, *393*, 181.
(5) Yang, G.; Cecconi, C.; Baase, W. A.; Vetter, I. R.; Breyer, W. A.; Haack, J. A.; Matthews, B. W.; Dahlquist, F. W.; Bustamante, C. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 139.
(6) Li, H.; Oberhouser, A. F.; Fowler, S. B.; Clarke, J.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 6527.
(7) Rief, M.; Gautel, M.; Schemmel, A.; Gaub, H. E. *Biophys. J.* **1998**, *75*, 3008.
(8) Carrion-Vazquez, M.; Oberhauser, A. F.; Fowler, S. B.; Marszalek, P. E.; Broedel, S. E.; Clarke, J.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3694.
(9) Rief, M.; Pascual, J.; Saraste, M.; Gaub, H. E. *J. Mol. Biol.* **1999**, *286*, 553.
(10) Oberhauser, A. F.; Marszalek, P. E.; Carrion-Vazquez, M.; Fernandez, J. M. *Natur. Struct. Biol.* **1999**, *6*, 1025.
(11) Lu, H.; Isralewitz, B.; Krammer, A.; Vogel, V.; Schulten, K. *Biophys. J.* **1998**, *75*, 662.
(12) Krammer, A.; Lu, H.; Izralewitz, B.; Schulten, K.; Vogel, V. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 1351.
(13) Lu, H.; Schulten, K. *Biophys. J.* **2000**, *79*, 51.
(14) Kolinski, A.; Godzik, A.; Skolnick, J. *J. Chem. Phys.* **1993**, *98*, 7420.
(15) Klimov, D. K.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6166.

(16) Klimov, D. K.; Thirumalai, D. *J. Chem. Phys.* **1998**, *109*, 4119.
(17) Socci, N. D.; Onuchic, J. N.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 2031.
(18) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195.
(19) Klimov, D. K.; Thirumalai, D. *Folding & Design* **1998**, *3*, 127.
(20) Klimov, D. K.; Thirumalai, D. *Proteins Struct. Funct. Gen.* **1996**, *26*, 411.
(21) Klimov, D. K.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7254.
(22) Fong, S.; Hamill, S. J.; Proctor, M.; Freund, S. M. V.; Benian, G. M.; Chothia, C.; Bycroft, M.; Clarke, J. *J. Mol. Biol.* **1996**, *264*, 624.