Refolding upon Force Quench and Pathways of Mechanical and Thermal Unfolding of Ubiquitin

Mai Suan Li,* Maksim Kouza,* and Chin-Kun Hu†
*Institute of Physics, Polish Academy of Sciences, Warsaw, Poland; and †Institute of Physics, Academia Sinica, Nankang, Taipei, Taiwan

ABSTRACT The refolding from stretched initial conformations of ubiquitin (PDB ID: 1ubq) under the quenched force is studied using the C_α-Gō model and the Langevin dynamics. It is shown that the refolding decouples the collapse and folding kinetics. The force-quench refolding-times scale as $\tau_F \sim \exp(f_q \Delta x_F / k_B T)$, where f_q is the quench force and $\Delta x_F \approx 0.96$ nm is the location of the average transition state along the reaction coordinate given by the end-to-end distance. This value is close to $\Delta x_F \approx 0.8$ nm obtained from the force-clamp experiments. The mechanical and thermal unfolding pathways are studied and compared with the experimental and all-atom simulation results in detail. The sequencing of thermal unfolding was found to be markedly different from the mechanical one. It is found that fixing the N-terminus of ubiquitin changes its mechanical unfolding pathways much more drastically compared to the case when the C-end is anchored. We obtained the distance between the native state and the transition state $\Delta x_{UF} \approx 0.24$ nm, which is in reasonable agreement with the experimental data.

INTRODUCTION

Deciphering the folding and unfolding pathways and free energy landscape of biomolecules remains a challenge in molecular biology. Traditionally, folding and unfolding are monitored by changing temperature or concentration of chemical denaturants. In these experiments, due to thermal fluctuations of initial unfolded conformations it is difficult to describe the folding mechanisms in an unambiguous way. With the help of the atomic force microscopy, mechanical force has been used to prepare well-defined initial states of proteins (1,2). Using the initial force, f_I, which is higher than the equilibrium critical force, f_c, to unfold the tandem of poly ubiquitin (Ub), Fernandez and Li (2) have shown that the refolding can be initiated starting from stretched conformations or force-denaturated ensemble (FDE) and quenching the force to a low constant value, f_q ($f_q < f_c$). Monitoring folding events as a function of the end-to-end distance (R), they have made the following important observations:

1. Contrary to the standard folding from the thermal denaturated ensemble (TDE), the refolding under the quenched force is a multiple stepwise process.
2. The force-quench refolding time obeys the Bell formula (3), $\tau_F \approx \tau_F^0 \exp(f_q \Delta x_F / k_B T)$, where τ_F^0 is the folding time in the absence of the quench force and Δx_F is the average location of the transition state (TS).

Motivated by the experiments of Fernandez and Li (2), we have studied (4) the refolding of the domain I27 of the human muscle protein using the C_α-Gō model (5) and the four-strand β-barrel model sequence S1 (6) (for this sequence the nonnative interactions are also taken into account). Basically, we have reproduced qualitatively the major experimental findings listed above. In addition, we have shown that the refolding is a two-state process in which the folding to the native basin attractor (NBA) follows the quick collapse from initial stretched conformations with low entropy. The corresponding kinetics can be described by the biexponential time dependence, contrary to the single exponential behavior of the folding from the TDE with high entropy.

To make the direct comparison with the experiments of Fernandez and Li (2), in this article we performed simulations for a single domain Ub using the C_α-Gō model (see Materials and Methods for more details). Because the study of refolding of 76-residue Ub (Fig. 1 *a*) by all-atom simulations is beyond present computational facilities, the Gō modeling is an appropriate choice. Most of the simulations have been carried out at $T = 0.85\ T_F = 285$ K. Our present results for refolding upon the force quench are in qualitative agreement with the experimental findings of Fernandez and Li, and with those obtained for I27 and S1 theoretically (4). A number of quantitative differences between I27 and Ub will be also discussed. For Ub, we have found the average location of the transition state $\Delta x_F \approx 0.96$ nm, which is in reasonable agreement with the experimental value 0.8 nm (2).

Experimentally, the unfolding of the polyubiquitin has been studied by applying a constant force (7). The mechanical unfolding of Ub has been investigated previously using Gō-like (8) and all-atom models (8,9). In particular, Irbäck et al. (9) have explored mechanical unfolding pathways of structures *A*, *B*, *C*, *D*, and *E* (see the definition of these structures and the β-strands in the caption to Fig. 1) and the existence of intermediates in detail. We present our results on mechanical unfolding of Ub for the five following reasons:

1. The barrier to the mechanical unfolding has not been computed.
2. Experiments of Schlierf et al. (7) have suggested that Cluster 1 (i.e., strands S1, S2, and the helix *A*) unfolds

Submitted April 26, 2006, and accepted for publication September 22, 2006.
Address reprint requests to Prof. Mai Suan Li, E-mail: masli@ifpan.edu.pl; or Prof. Chin-Kun Hu, E-mail: huck@phys.sinica.edu.tw.

0006-3495/07/01/547/15 $2.00
doi: 10.1529/biophysj.106.087684

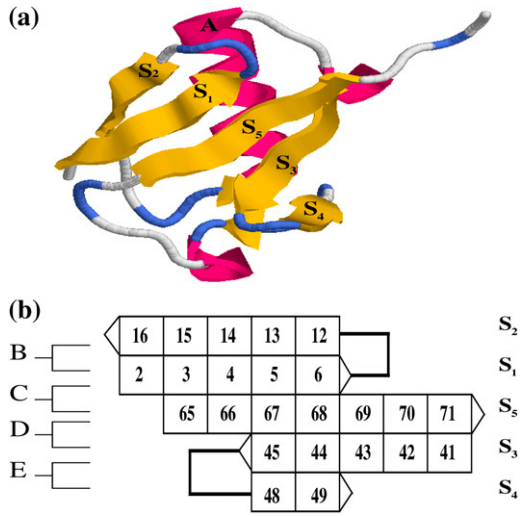

FIGURE 1 (*a*) Native state conformation of ubiquitin taken from the PDB (PDB ID: 1ubq). There are five β-strands: S1 (2–6), S2 (12–16), S3 (41–45), S4 (48–49), and S5 (65–71), and one helix A (23–34). (*b*) Structures *B*, *C*, *D*, and *E* consist of pairs of strands (S1,S2), (S1,S5), (S3,S5), and (S3,S4), respectively. In the text we also refer to helix *A* as structure *A*.

after Cluster 2 (strands S3, S4, and S5). However, this observation has not yet been studied theoretically.

3. Since the structure *C*, which consists of the strands S1 and S5, unzips first, Irbäck et al. (9) pointed out that strand S5 unfolds before S2 (the terminal strands follows the unfolding pathway S1 → S5 → S2). This conclusion may be incorrect, because it has been obtained from the breaking of the contacts within the structure *C*.
4. In pulling and force-clamp experiments, the external force is applied to one end of the proteins, while the other end is kept fixed. Therefore, one important question emerges as to how fixing one terminus affects the unfolding sequencing of Ub. This issue has not been addressed by Irbäck et al. (9).
5. Using a simplified all-atom model, it was shown (9) that mechanical intermediates occur more frequently than in experiments (7). It is relevant to ask if a C_α-Gō model can capture similar intermediates as this may shed light on the role of nonnative interactions.

In this article, from the force dependence of mechanical unfolding times we estimated the distance between the native state and the transition state to be $\Delta x_{UF} \approx 0.24$ nm, which is close to the experimental results of Carrion-Vazquez et al. (10) and Schlierf et al. (7). In agreement with the experiments (7), Cluster 1 was found to unfold after Cluster 2 in our simulations. Applying the force to the both termini, we studied the mechanical unfolding pathways of the terminal strands in detail and obtained the sequencing S1 → S2 → S5, which is different from the result of Irbäck et al. (9). When the N-terminus is fixed and the C-terminus is pulled by a constant force, the unfolding sequencing was found to be very different from the previous case. The unzipping initiates, for example, from the C-terminus but not from the N-terminus. Anchoring the C-end is shown to have a little effect on unfolding pathways. We have demonstrated that the present C_α-Gō model does not capture rare mechanical intermediates, presumably due to the lack of nonnative interactions. Nevertheless, it can correctly describe the two-state unfolding of Ub (7).

It is well known that thermal unfolding pathways may be very different from the mechanical ones, as has been shown for the domain I27 (11). This is because the force is applied locally to the termini while thermal fluctuations have the global effect on the entire protein. In the force case, unzipping should propagate from the termini whereas under thermal fluctuations the most unstable part of a polypeptide chain unfolds first.

The unfolding of Ub under thermal fluctuations was investigated experimentally by Cordier and Grzesiek (12) and by Chung et al. (13). If one assumes that unfolding is the reverse of the refolding process then one can infer information about the unfolding pathways from the experimentally determined ϕ-values (14) and ψ-values (15,16). The most comprehensive ϕ-value analysis is that of Went and Jackson. They found that the C-terminal region, which has very low ϕ-values, unfolds first and then the strand S1 breaks before full unfolding of the α-helix fragment A occurs. However, the detailed unfolding sequencing of the other strands remains unknown.

Theoretically, the thermal unfolding of Ub at high temperatures has been studied by all-atom molecular dynamics (MD) simulations by Alonso and Daggett (17) and Larios et al. (18). In the latter work, the unfolding pathways were not explored. Alonso and Daggett have found that the α-helix fragment *A* is the most resilient toward temperature but the structure *B* breaks as early as the structure *C*. The fact that *B* unfolds early contradicts not only the results for the ϕ-values obtained experimentally by Went and Jackson (14) but also findings from a high resolution NMR (12). Moreover, the sequencing of unfolding events for the structures *D* and *E* was not studied.

What information about the thermal unfolding pathways of Ub can be inferred from the folding simulations of various coarse-grained models? Using a semiempirical approach, Fernandez predicted (19) that the nucleation site involves the β-strands S1 and S5. This suggests that thermal fluctuations break these strands last, but what happens to the other parts of the protein remain unknown. Furthermore, the late breaking of S5 contradicts the unfolding (12) and folding (14) experiments. From later folding simulations of Fernandez et al. (20,21), one can infer that the structures *A*, *B*, and *C* unzip late. Since this information is gained from ϕ-values, it is difficult to determine the sequencing of unfolding events even for these fragments. Using the results of Gilis and Rooman (22), we can only expect that the structures *A* and *B* unfold last. In addition, with the help of a three-bead model it was found (23) that the C-terminal loop structure is the last to

fold in the folding process and most likely plays a spectator role in the folding kinetics. This implies that strands S4, S5, and the second helix (residues 38–40) would unzip first but again the full unfolding sequencing cannot be inferred from this study.

Thus, neither the direct MD (17) nor indirect folding simulations (19–23) provide a complete picture of the thermal unfolding pathways for Ub. One of our aims is to decipher the complete thermal unfolding sequencing and compare it with the mechanical one. The mechanical and thermal routes to the denaturated states have been found to be very different from each other. Under the force, e.g., the β-strand S1, unfolds first, while thermal fluctuations detach strand S5 first. The later observation is in good agreement with NMR data of Cordier and Grzesiek (12). A detailed comparison with available experimental and simulation data on the unfolding sequencing will be presented. The free energy barrier to thermal unfolding was also calculated.

To summarize, in this article we have obtained the following novel results. We have shown that the refolding of Ub is a two-stage process in which the ''burst'' phase exists on very short timescales. The construction of the $T-f$ phase diagram allows us to determine the equilibrium critical force f_c separating the folded and unfolded regions. Using the exponential dependence of the refolding and unfolding times on f, Δx_F and Δx_{UF} were computed. Our results for f_c, Δx_F and Δx_{UF} are in acceptable agreement with the experiments. It has been demonstrated that fixing the N-terminus of Ub has much stronger effect on mechanical unfolding pathways compared to the case when the C-end is anchored. In comparison with previous studies, we provide a more complete picture for thermal unfolding pathways, which are very different from the mechanical ones.

MATERIALS AND METHODS

C_α-Gō model for Ub

We use coarse-grained continuum representation for Ub in which only the positions of C_α-carbons are retained. The interactions between residues are assumed to be Gō-like and the energy of such a model is (5)

$$E = \sum_{\text{bonds}} K_r(r_i - r_{0i})^2 + \sum_{\text{angles}} K_\theta(\theta_i - \theta_{0i})^2 + \sum_{\text{dihedral}} \{K_\phi^{(1)}[1-\cos(\phi_i - \phi_{0i})] + K_\phi^{(3)}[1-\cos 3(\phi_i - \phi_{0i})]\} + \sum_{i>j-3}^{NC} \epsilon_H \left[5\left(\frac{r_{0ij}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{0ij}}{r_{ij}}\right)^{10}\right] + \sum_{i>j-3}^{NNC} \epsilon_H \left(\frac{C}{r_{ij}}\right)^{12} - |\vec{f}.\vec{R}|. \quad (1)$$

Here $\Delta\phi_i = \phi_i - \phi_{0i}$, $r_{i,\,i+1}$ is the distance between beads i and $i+1$, θ_i is the bond angle between bonds $(i-1)$ and i, and ϕ_i is the dihedral angle around the i^{th} bond and r_{ij} is the distance between the i^{th} and j^{th} residues. Subscripts *0*, *NC*, and *NNC* refer to the native conformation, native contacts, and nonnative contacts, respectively. Residues i and j are in native contact if r_{0ij} is less than a cutoff distance d_c taken to be $d_c = 6.5$ Å, where r_{0ij} is the distance between the residues in the native conformation. With this choice of d_c and the native conformation from the PDB (Fig. 1 *a*), we have the total number of native contacts $Q_{max} = 99$.

The first harmonic term in Eq. 1 accounts for chain connectivity and the second term represents the bond-angle potential. The potential for the dihedral angle degrees of freedom is given by the third term in Eq. 1. The interaction energy between residues that are separated by at least three beads is given by 10–12 Lennard-Jones potential. A soft-sphere repulsive potential (the fourth term in Eq. 1) disfavors the formation of nonnative contacts. The last term accounts for the force applied to C- and N-termini along the end-to-end vector $\vec{R}$. We choose $K_r = 100\ \epsilon_H/Å^2$, $K_\theta = 20\ \epsilon_H/\text{rad}^2$, $K_\phi^{(1)} = \epsilon_H$, and $K_\phi^{(3)} = 0.5\ \epsilon_H$, where ϵ_H is the characteristic hydrogen bond energy and $C = 4$ Å. Since $T_F = 0.675\ \epsilon_H$ (see below) and $T_F = 332.5$ K (24), we have $\epsilon_H = 4.1$ kJ/mol $= 0.98$ kcal/mol. Then the force unit $[f] = \epsilon/Å = 68.0$ pN.

We assume the dynamics of the polypeptide chain obeys the Langevin equation. The equations of motion (see (25) for details) were integrated using the velocity form of the Verlet algorithm (26) with the time step $\Delta t = 0.005\tau_L$, where $\tau_L = (ma^2/\epsilon_H)^{1/2} \approx 3$ ps.

Simulations

To obtain the $T-f$ phase diagram we use the fraction of native contacts or the overlap function (27)

$$\chi = \frac{1}{Q_{\text{total}}} \sum_{i<j+1}^{N} \theta(1.2r_{0ij} - r_{ij})\Delta_{ij}, \quad (2)$$

where Δ_{ij} is equal to 1 if residues i and j form a native contact and 0 otherwise, and $\theta(x)$ is the Heaviside function. The argument of this function guarantees that a native contact between i and j is classified as formed when r_{ij} is shorter than 1.2 r_{0ij}. The probability of being in the native state, f_N, which can be measured by various experimental techniques, is defined as $f_N = \langle\chi\rangle$, where $\langle\ldots\rangle$ stands for a thermal average. The $T-f$ phase diagram (a plot of $1-f_N$ as a function of f and T) and thermodynamic quantities were obtained by the multiple histogram method (28), extended to the case when the external force is applied to the termini (29,30). In this case, the reweighting is carried out not only for temperature but also for force. We collected data for six values of T at $f = 0$ and for five values of f at a fixed value of T. The duration of MD runs for each trajectory was chosen to be long enough to get the system fully equilibrated ($9 \times 10^5\tau_L$ from which $1.5 \times 10^5\tau_L$ were spent on equilibration). For a given value of T and f, we have generated 40 independent trajectories for thermal averaging.

For the mechanical unfolding we have considered two cases. In the first case, the external force is applied via both termini N and C. In the second case it is applied to either N- or C-terminus.

To simulate the mechanical unfolding the computation has been performed at $T = 285$ K and mainly at the constant force $f = 70$, 100, 140, and 200 pN. This allows us to compare our results with the mechanical unfolding experiments (7) and to see if the unfolding pathways change at low forces. Starting from the native conformation but with different random number seeds the unfolding sequencing of helix *A* and five β-stands is studied by monitoring fraction of native contacts as a function of the end-to-end extension. In the case of structures *A*, *B*, *C*, *D*, and *E* we consider not only the evolution of the number of intrastructure contacts as has been done by Irbäck et al. (9), but also the evolution of all contacts (intrastructure contacts and the contacts formed by a given structure with the rest of a protein).

In the thermal unfolding case the simulation is also started from the native conformations and it is terminated when all of the native contacts are broken. Due to thermal fluctuations there is no one-to-one correspondence between R and time. Therefore R ceases to be a good reaction coordinate for describing unfolding sequencing. To rescue this, for each i^{th} trajectory we introduce the progressive variable $\delta_i = t/\tau_{UF}^i$, where τ_{UF}^i is the unfolding time. Then we can average the fraction of native contacts over a unique window $0 \le \delta_i \le 1$ and monitor the unfolding sequencing with the help of the progressive variable δ.

RESULTS

Temperature-force phase diagram and thermodynamic quantities

The $T-f$ phase diagram, obtained by the extended histogram method (see Materials and Methods), is shown in Fig. 2 *a*. The folding-unfolding transition, defined by the yellow region, is sharp in the low temperature region but it becomes less cooperative (the fuzzy transition region is wider) as T increases. The weak reentrancy (the critical force slightly increases with T) occurs at low temperatures. This seemingly strange phenomenon occurs as a result of competition between the energy gain and the entropy loss upon stretching. The similar cold unzipping transition was also observed in a number of models for heteropolymers (31) and proteins (29) including the C_α-Gō model for I27 (M. S. Li, unpublished results). As follows from the phase diagram, at $T = 285$ K, the critical force $f_c \approx 30$ pN, which is close to $f_c \approx 25$ pN, is estimated from the experimental pulling data (to estimate f_c from experimental pulling data, we use $f_{max} \approx f_c \ln(v/v_{min})$ (32), where f_{max} is the maximal force needed to unfold a protein at the pulling speed v. From the raw data in Fig. 3 *b* of Carrion-Vasquez et al. (10), we obtain $f_c \approx 25$ pN). Given the simplicity of the model this agreement can be considered satisfactory and it validates the use of the Gō model.

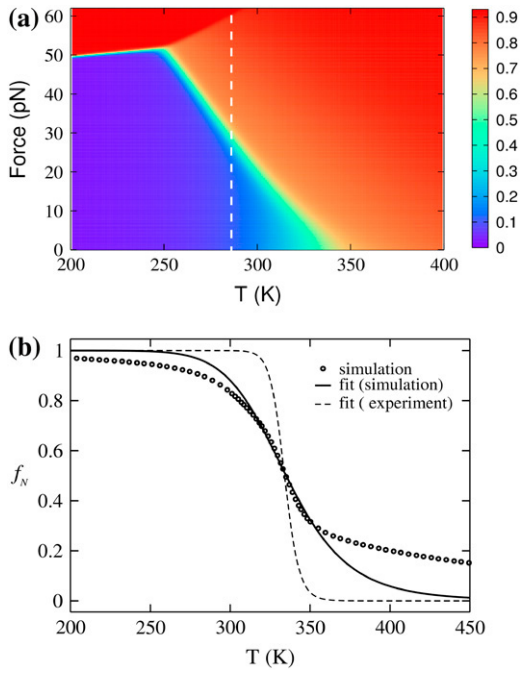

FIGURE 2 (*a*) The $T-f$ phase diagram obtained by the extended histogram method. The force is applied to termini N and C. The color code for $1-\langle\chi(T,f)\rangle$ is given on the right. The blue color corresponds to the state in the NBA, while the red color indicates the unfolded states. The vertical dashed line refers to $T = 0.85$, $T_F \approx 285$ K, at which most of simulations have been performed. (*b*) The temperature dependence of f_N (*open circles*) defined as the renormalized number of native contacts (see Material and Methods). The solid line refers to the two-state fit to the simulation data. The dashed line represents the experimental two-state curve with $\Delta H_m = 48.96$ kcal/mol and $T_m = 332.5$ K (24).

Fig. 2 *b* shows the temperature dependence of population of the native state f_N. Fitting to the standard two-state curve $f_N = 1/(1+\exp[-\Delta H_m(1-(T)/(T_m))/k_B T])$, one can see that it works pretty well (*solid curve*) around the transition temperature but it gets worse at high T due to slow decay of f_N. Such a behavior is characteristic for almost all of theoretical models (25) including the all-atom ones (33). In fitting we have chosen the hydrogen-bond energy $\epsilon_H = 0.98$ kcal/mol in Hamiltonian (1), so that $T_F = T_m = 0.675\epsilon_H/k_B$ coincides with the experimental value 332.5 K (24). From the fit we obtain $\Delta H_m = 11.4$ kcal/mol, which is smaller than the experimental value 48.96 kcal/mol indicating that the Gō model is, as expected, less stable compared to the real Ub. Taking into account nonnative contacts and more realistic interactions between side-chain atoms is expected to increase the stability of the system.

The cooperativity of the denaturation transition may be characterized by the cooperativity index, Ω_c (see (34) and (35) for definition). From simulation data for f_N presented in Fig. 2 *b* we have $\Omega_c \approx 57$, which is considerably lower than the experimental value $\Omega_c \approx 384$ obtained with the help of $\Delta H_m = 48.96$ kcal/mol and $T_m = 332.5$ K (24). The underestimation of Ω_c in our simulations is not only a shortcoming of the off-lattice Gō model (36) but also a common problem of much more sophisticated force fields in all-atom models (33).

Another measure of the cooperativity is the ratio between the van' t Hoff and the calorimetric enthalpy κ_2 (37). For the Gō Ub we obtained $\kappa_2 \approx 0.19$. Applying the base line subtraction (38) gives $\kappa_2 \approx 0.42$, which is still much below $\kappa_2 \approx 1$ for the truly one-or-none transition. Since κ_2 is an extensive parameter, its low value is due to the shortcomings of the off-lattice Gō models but not due to the finite size effects. More rigid lattice models give better results for the calorimetric cooperativity κ_2 (39).

Fig. 3 *a* shows the free energy as a function of Q for several values of force at $T = T_F$. Since there are only two minima, our results support the two-state picture of Ub (7,13). As expected, the external force increases the folding barrier, ΔF_F ($\Delta F_F = F_{TS} - F_D$) and it lowers the unfolding barrier, ΔF_{UF} ($\Delta F_{UF} = F_{TS} - F_N$). From the linear fits in Fig. 3 *b* we obtain the average distance between the TS and D states, $\Delta x_F = \Delta F_F/f \approx 1$ nm, and the distance between TS and the native state, $\Delta x_{UF} = \Delta F_{UF}/f \approx 0.13$ nm. Note that Δx_F is very close to $\Delta x_F \approx 0.96$ nm obtained from refolding times at a bit lower temperature $T = 285$ K (see Fig. 6 below). However, Δx_{UF} is lower than value 0.24 nm followed from mechanical unfolding data at $f > f_c$ (Fig. 8). This difference may be caused by either sensitivity of Δx_{UF} to the temperature, or the determination of Δx_{UF} from the approximate free energy landscape, since a function of a single coordinate Q is not sufficiently accurate.

We have also studied the free energy landscape using R as a reaction coordinate. The dependence of F on R was found

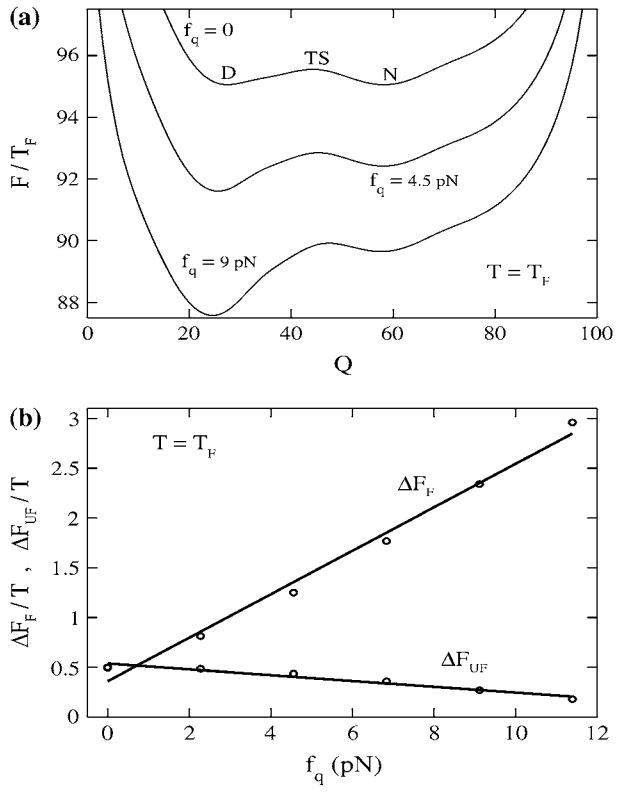

FIGURE 3 (*a*) The dependence of the free energy on Q for selected values of f at $T = T_F$. D and N refer to the denaturated and native states, respectively. (*b*) The dependence of folding and unfolding barriers, obtained from the free energy profiles, on f. The linear fits $y = 0.36 + 0.218x$ and $y = 0.54 - 0.029x$ correspond to ΔF_F and ΔF_{UF}, respectively. From these fits we obtain $\Delta x_F \approx 1$ nm and $\Delta x_{UF} \approx 0.13$ nm.

to be smoother (results not shown) compared to that obtained by Kirmizialtin et al. (40) using a more elaborated model (23) involving the nonnative interactions.

Refolding under quenched force

Our protocol for studying the refolding of Ub is identical to that used in the experiments of Fernandez and Li (2). We first apply the force $f_I \approx 70$ pN to prepare initial conformations (the protein is stretched if $R \geq 0.8\ L$, where the contour length $L = 28.7$ nm). Starting from the FDE we quenched the force to $f_q < f_c$ and then monitored the refolding process by following the time dependence of the number of native contacts $Q(t)$, $R(t)$, and the radius of gyration $R_g(t)$ for typically 50 independent trajectories.

Fig. 4 shows considerable diversity of refolding pathways. In accord with experiments (2) and simulations for I27 (4), the reduction of R occurs in a stepwise manner. In the $f_q = 0$ case (Fig. 4 *a*), R decreases continuously from $\approx$18 nm to 7.5 nm (stage 1) and fluctuates around this value for $\sim$3 ns (stage 2). The further reduction to $R \approx 4.5$ nm (stage 3) until a transition to the NBA. The stepwise nature of variation of $Q(t)$ is also clearly shown up but it is more masked for $R_g(t)$. Although we can interpret another trajectory for $f_q = 0$ (Fig. 4 *b*) in the same way, the timescales are different. Thus, the refolding routes are highly heterogeneous.

The pathway diversity is also evident for $f_q > 0$ (Fig. 4, *c* and *d*). Although the picture remains qualitatively the same as in the $f_q = 0$ case, the timescales for different steps becomes much larger. The molecule fluctuates around $R \approx 7$ nm, e.g., for $\approx$60 ns (stage 2 in Fig. 4 *c*), which is considerably longer than $\approx$3 ns in Fig. 4 *a*. The variation of $R_g(t)$ becomes more drastic compared to the $f_q = 0$ case.

Fig. 5 shows the time dependence of $\langle R(t)\rangle$, $\langle Q(t)\rangle$, and $\langle R_g(t)\rangle$, where $\langle \ldots \rangle$ stands for averaging over 50 trajectories. The left and right panels correspond to the long and short time windows, respectively. For the TDE case (Fig. 5, *a* and *b*), the single exponential fit works pretty well for $\langle R(t)\rangle$ for the whole time interval. A little departure from this behavior is seen for $\langle Q(t)\rangle$ and $\langle R_g(t)\rangle$ for $t < 2$ ns (Fig. 5 *b*). Contrary to the TDE case, even for $f_q = 0$ (Fig. 5, *c* and *d*) the difference between the single and biexponential fits is evident not only for $\langle Q(t)\rangle$ and $\langle R_g(t)\rangle$ but also for $\langle R(t)\rangle$. The timescales, above which two fits become eventually identical, are slightly different for three quantities (Fig. 5 *d*). The failure of the single exponential behavior becomes more and more evident with the increase of f_q, as demonstrated in Fig. 5, *e* and *f*, for the FDE case with $f_q = 6.25$ pN.

Thus, in agreement with our previous results, obtained for I27 and the sequence S1 (4), starting from FDE the refolding kinetics compiles from the fast and slow phase. The characteristic timescales for these phases may be obtained using a sum of two exponentials, $\langle A(t)\rangle = A_0 + A_1 \exp(-t/\tau_1^A) + A_2 \exp(-t/\tau_2^A)$, where A stands for R, R_g, or Q. Here τ_1^A characterizes the burst-phase (first stage) while τ_2^A may be either the collapse time (for R and R_g) or the folding time (for Q) ($\tau_1^A < \tau_2^A$). As in the case of I27 and S1 (4), τ_1^R and $\tau_1^{R_g}$ are almost independent on f_q (results not shown). We attribute this to the fact that the quench force ($f_q^{max} \approx 9$ pN) is much lower than the entropy force (f_e) needed to stretch the protein. At $T = 285$ K, one has to apply $f_e \approx 140$ pN for stretching Ub to $0.8\ L$. Since $f_q^{max} << f_e$, the initial compaction of the chain that is driven by f_e is not sensitive to the small values of f_q. Contrary to τ_1^A, τ_2^A was found to increase with f_q exponentially. Moreover, $\tau_2^R < \tau_2^{R_g} < \tau_F$, implying that the chain compaction occurs before the acquisition of the native state.

Fig. 6 shows the dependence of the folding times on f_q. Using the Bell-type formula (3) and the linear fit in Fig. 6, we obtain $\Delta x_F \approx 0.96$ nm, which is in acceptable agreement with the experimental value $\Delta x_F \approx 0.8$ nm (2). The linear growth of the free energy barrier to folding with f_q is due to the stabilization of the random coil states under the force. Our estimate for Ub is higher than $\Delta x_F \approx 0.6$ nm obtained for I27 (4). One of possible reasons for such a pronounced difference is that we used the cutoff distance $d_c = 0.65$ and 0.6 nm in the Gō model (1) for Ub and I27, respectively. The larger value of d_c would make a protein more stable (more

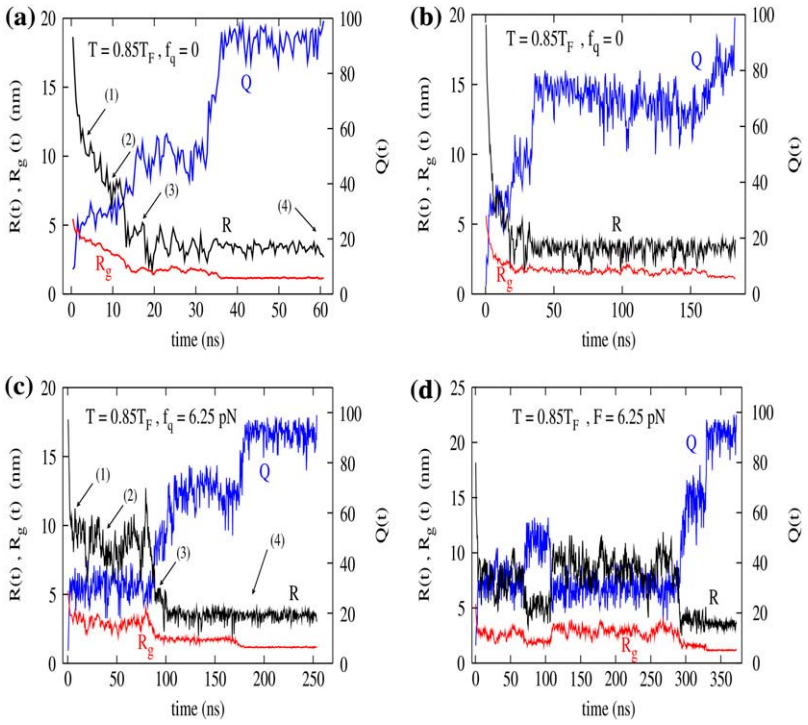

FIGURE 4 (*a*,*b*) The time dependence of Q, R, and R_g for two typical trajectories starting from FDE ($f_q = 0$ and $T = 285$ K). Arrows *1*, *2*, and *3* in panel *a* correspond to time 3.1 ($R = 10.9$ nm), 9.3 ($R = 7.9$ nm), and 17.5 ns ($R = 5$ nm). Arrow *4* marks the folding time $\tau_F = 62$ ns ($R = 2.87$ nm) when all 99 native contacts are formed. Panels *c* and *d* are the same as in panels *a* and *b*, but for $f_q = 6.25$ pN. The corresponding arrows refer to $t = 7.5$ ($R = 11.2$ nm), 32 ($R = 9.4$ nm), 95 ns ($R = 4.8$ nm), and $\tau_F = 175$ ns ($R = 3.65$ nm).

native contacts) and it may change the free energy landscape leading to enhancement of Δx_F. This problem requires further investigation.

Absence of mechanical unfolding intermediates in C_α-Gō model

To study the unfolding dynamics of Ub, Schlierf et al. (7) have performed the AFM experiments at a constant force $f = 100$, 140, and 200 pN. The unfolding intermediates were recorded in ~5% of 800 events at different forces. The typical distance between the initial and intermediate states is $\Delta R = 8.1 \pm 0.7$ nm (7). However, the intermediates do not affect the two-state behavior of the polypeptide chain. Using the all-atom models, Irbäck et al. (9) have also observed the intermediates in the region 6.7 nm $< R <$ 18.5 nm. Although the percentage of intermediates is higher than in the experiments, the two-state unfolding events remain dominating. To check the existence of force-induced intermediates in our model, we have performed the unfolding simulations for $f = 70$, 100, 140, and 200 pN. Because the results are qualitatively similar for all values of force, we present the $f = 100$ pN case only.

Fig. 7 shows the time dependence of $R(t)$ for 15 runs starting from the native value $R_N \approx 3.9$ nm. For all trajectories the plateau occurs at $R \approx 4.4$ nm. As seen below, passing this plateau corresponds to breaking of intrastructure native contacts of structure C. At this stage, the chain ends are almost stretched out, but the rest of the polypeptide chain remains nativelike. The plateau is washed out when we average over many trajectories and $\langle R(t) \rangle$ is well fitted by a single exponential (Fig. 7), in accord with the two-state behavior of Ub (7).

The existence of the plateau observed for individual unfolding events in Fig. 7 agrees with the all-atom simulation results of Irbäck et al. (9), who have also recorded a similar plateau at $R \approx 4.6$ nm at short timescales. However, unfolding intermediates at larger extensions do not occur in our simulations. This is probably related to neglect of the nonnative interactions in the C_α-Gō model. Nevertheless, this simple model provides the correct two-state unfolding picture of Ub in the statistical sense.

Mechanical unfolding barrier

We now try to determine the barrier to the mechanical unfolding from the dependence of the unfolding times τ_{UF} on f. It should be noted that this way of determination of the unfolding barrier is exact and it would give a more reliable estimate compared to the free energy landscape approach in which the free energy profile is approximated as a function of only one order parameter.

We first consider the case when the force is applied via both termini N and C. Since the force lowers the unfolding barrier, τ_{UF} should decrease as f increases (Fig. 8). The present Gō model gives τ_{UF} smaller than the experimental values by approximately eight orders of magnitude. E.g., for $f = 100$ pN, $\tau_{UF} \approx 12$ ns whereas the experiments gives $\tau_{UF} \approx 2.77$ s (7). As seen from Fig. 8, for $f < 140$ pN τ_{UF} depends on f exponentially. In this regime, $\tau_{UF} \approx \tau_{UF}^0 \exp(f x_{UF}/k_B T)$,

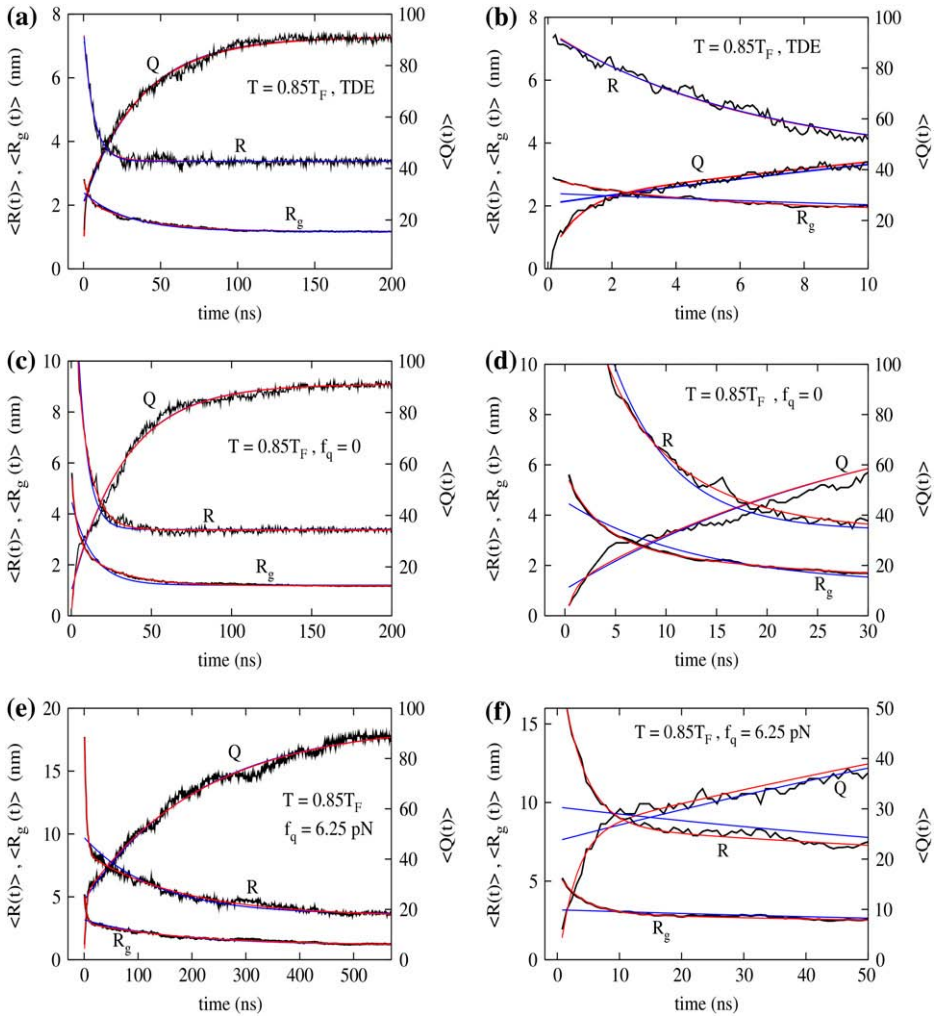

FIGURE 5 (*a*) The time dependence of $\langle Q(t)\rangle$, $\langle R(t)\rangle$, and $\langle R_g(t)\rangle$ when the refolding starts from TDE. (*b*) The same as in panel *a*, but for the short timescale. (*c*,*d*) The same as in panels *a* and *b*, but for FDE with $f_q = 0$. (*e*,*f*) The same as in panels *c* and *d*, but for $f_q = 6.25$ pN.

where Δx_{UF} is the average distance between the N and TS states. From the linear fit in Fig. 8 we obtained $\Delta x_{UF} \approx 0.24$ nm. Using different fitting procedures, Schlierf et al. (7) obtained $\Delta x_{UF} \approx 0.14$ nm and 0.17 nm. The larger value $\Delta x_{UF} \approx 0.25$ nm was reported in the earlier experiments (10). Thus, given experimental uncertainty, the C_α-Gō model provides a reasonable estimate of Δx_{UF} for the two-state Ub.

In the high force regime ($f > 140$ pN), instead of the exponential dependence, τ_{UF} scales with f linearly (*inset* in Fig. 8). The crossover from the exponential to the linear behavior is in full agreement with the earlier theoretical prediction (32). Similar crossover has been also observed (41) for the another Gō-like model of Ub but Δx_{UF} has not been estimated. At very high forces, τ_{UF} is expected to be asymptotically independent of f.

One can show that fixing one terminus of a protein has the same effect on unfolding times no matter whether the N- or C-terminus is fixed. Therefore, we show the results obtained for the case when the N-end is anchored. As seen from Fig. 8, the unfolding process is slowed down nearly by a factor of 2. It may imply that diffusion-collision processes (42) play an important role in the Ub unfolding. Namely, as follows from the diffusion-collision model, the time, required for formation (breaking) contacts, is inversely proportional to the diffusion coefficient, D, of a pair of spherical units. If one of them is idle, D is halved and the time needed to break contacts increases accordingly. Although fixing one end increases the unfolding times, it does not change the distance between the TS and the native state, Δx_{UF} (Fig. 8).

Mechanical unfolding pathways: force is applied to both termini

Here we focus on the mechanical unfolding pathways by monitoring the number of native contacts as a function of the end-to-end extension $\Delta R \equiv R - R_{eq}$, where R_{eq} is the equilibrium value of R. For $T = 285$ K, $R_{eq} \approx 3.4$ nm. Following Schlierf et al. (7), we first divide Ub into two clusters. Cluster 1 consists of strands S1, S2, and the helix *A* (42 native contacts) and cluster 2, strands S3, S4, and S5 (35 native contacts). The dependence of fraction of intracluster native contacts is shown in Fig. 9 for $f = 70$ and 200 pN

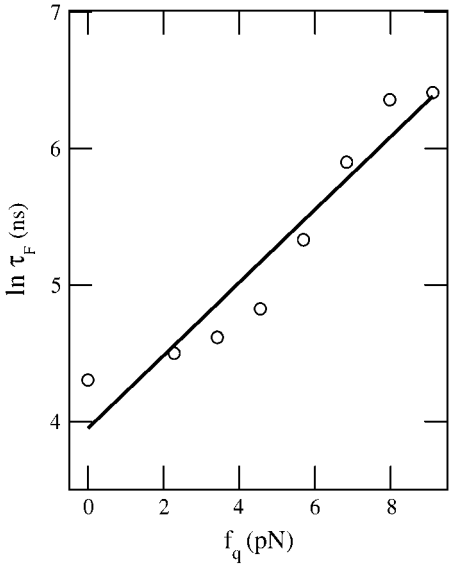

FIGURE 6 The dependence of folding times on the quench force at $T = 285$ K. The value τ_F was computed as the average of the first passage times (τ_F is the same as $\tau^{O}{}_2$ extracted from the biexponential fit for $\langle Q(t)\rangle$). The result is averaged over 30–50 trajectories depending on f_q. From the linear fit, $y = 3.951 + 0.267x$. With correlation level equal to -0.96, we obtain $x_F \approx 0.96$ nm.

(similar results for $f = 100$ and 140 pN are not shown). In agreement with the experiments (7), Cluster 2 unfolds first. The unfolding of these clusters becomes more and more synchronous upon decreasing f. At $f = 70$ pN the competition with thermal fluctuations becomes so important that

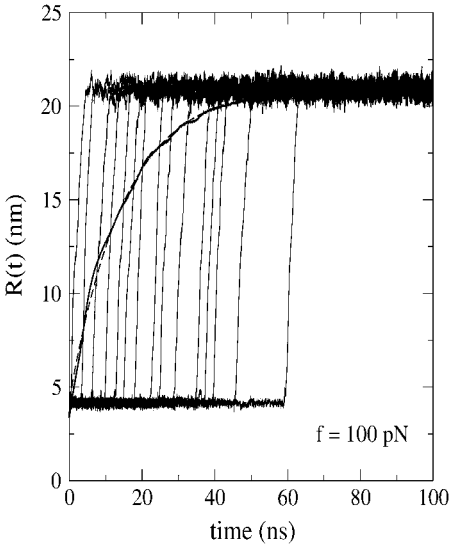

FIGURE 7 Time dependence of the end-to-end distance for $f = 100$ pN. The thin curves refer to 15 representative trajectories. The average over 200 trajectory $\langle R(t)\rangle$ values is represented by the thick line. The dashed curve is the single exponential fit $\langle R(t)\rangle = 21.08 - 16.81 \exp(-x/\tau_{UF})$, where $\tau_{UF} \approx 11.8$ ns.

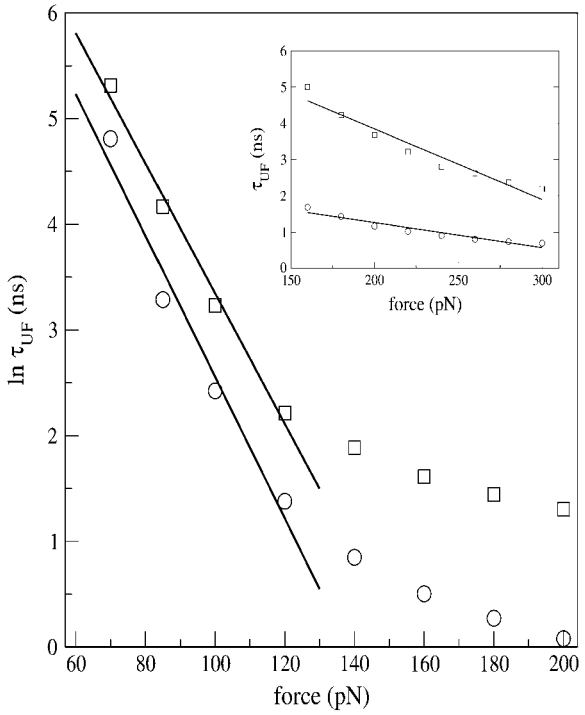

FIGURE 8 Dependence of mechanical unfolding time on the force. Circles refer to the process when the force is applied to both N- and C-termini. Squares signify the case when the N-end is fixed and the C-end is pulled. For the first case the linear fit ($y = 9.247 - 0.067x$) gives the distance between the native state and TS $\Delta x_{UF} \approx 0.24$ nm. In the second case, from the linear fit ($y = 9.510 - 0.062x$) we obtained $\Delta x_{UF} \approx 0.22$ nm. Thus, within error bars, fixing one end does not affect the value of Δx_{UF}. The inset shows the linear dependence of τ_{UF} on f in the high force regime.

two clusters may unzip almost simultaneously. Experiments at low forces are needed to verify this observation.

The arrow in Fig. 9 marks the position $\Delta R = 8.1$ nm, where some intermediates were recorded in the experiments (7). At this point there is intensive loss of native contacts of Cluster 2, suggesting that the intermediates observed on the experiments are conformations in which most of the contacts of this cluster are already broken but Cluster 1 remains relatively structured (≈40% contacts). One can expect that Cluster 1 is more ordered in the intermediate conformations if the side chains and realistic interactions between amino acids are taken into account.

To compare the mechanical unfolding pathways of Ub with the all-atom simulation results (9), we discuss the sequencing of helix A and structures B, C, D, and E in more detail. We monitor the intrastructure native contacts and all contacts separately. The later include not only the contacts within a given structure but also the contacts between it and the rest of the protein. It should be noted that Irbäck et al. have studied the unfolding pathways based on the evolution of the intrastructure contacts. Fig. 10 *a* shows the dependence of the fraction of intrastructure contacts on ΔR at $f = 100$ pN. At $\Delta R \approx 1$ nm, which corresponds to the plateau

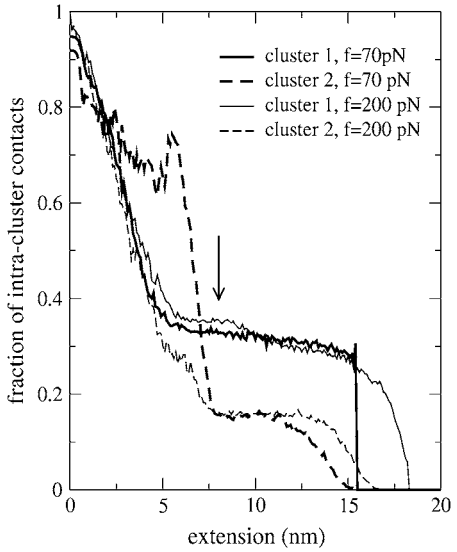

FIGURE 9 The dependence of fraction of the native contacts on the end-to-end extension for Cluster 1 (*solid lines*) and Cluster 2 (*dashed lines*) at $f = 70$ pN and 200 pN. The results are averaged over 200 independent trajectories. The arrow points to the extension $\Delta R = 8.1$ nm.

in Fig. 7, most of the contacts of C are broken. In agreement with the all-atom simulations (9), the unzipping follows $C \rightarrow B \rightarrow D \rightarrow E \rightarrow A$. Since C consists of the terminal strands S1 and S5, it was suggested that these fragments unfold first. However, this scenario may be no longer valid if one considers not only intrastructure contacts but also other possible ones (Fig. 10 *b*). In this case the statistically preferred sequencing is $B \rightarrow C \rightarrow D \rightarrow E \rightarrow A$, which holds not only for $f = 100$ pN but also for other values of f. If it is true, then S2 will unfold even before S5. To make this point more transparent, we plot the fraction of contacts for S1, S2, and S5 as a function of ΔR (Fig. 11 *a*) for a typical trajectory. Clearly, S5 detaches from the core part of a protein after S2 (see also the snapshot in Fig. 11 *b*). So, instead of the sequencing S1 $\rightarrow$ S5 $\rightarrow$ S2 proposed by Irbäck et al., we obtain S1 $\rightarrow$ S2 $\rightarrow$ S5.

The dependence of the fraction of native contacts on ΔR for individual strands is shown in Fig. 12 *a* ($f = 70$ pN) and Fig. 12 *b* ($f = 200$ pN). At $\Delta = 8.1$ nm contacts of S1, S2, and S5 are already broken, whereas S4 and A remain largely structured. In terms of β-strands and A we can interpret the intermediates observed in the experiments of Schlierf et al. (7), as conformations with well-structured S4 and A, and low ordering of S3. This interpretation is more precise compared to the above argument based on unfolding of two clusters because if one considers the average number of native contacts, then Cluster 2 is unstructured in the intermediate state (Fig. 9), but its strand S4 remains highly structured (Fig. 12).

From Fig. 12, we obtain the following mechanical unfolding sequencing:

$$\mathrm{S1} \rightarrow \mathrm{S2} \rightarrow \mathrm{S5} \rightarrow \mathrm{S3} \rightarrow \mathrm{S4} \rightarrow A. \qquad (3)$$

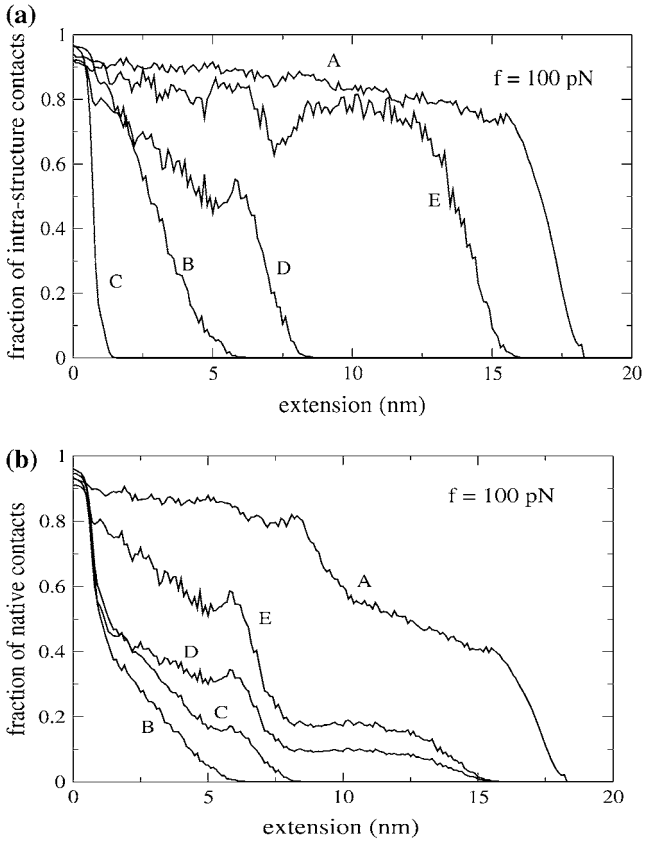

FIGURE 10 (*a*) The dependence of fraction of the intrastructure native contacts on ΔR for structures A, B, C, D, and E at $f = 100$ pN. (*b*) The same as in panel *a*, but for all native contacts. The results are averaged over 200 independent trajectories.

It should be noted that the sequencing (3) is valid in the statistical sense. In some trajectories, S5 unfolds even before S1 and S2 or the native contacts of S1, S2, and S5 may be broken at the same timescale (Table 1). From Table 1, it follows that the probability of having S1 unfolded first decreases with lowering f but the main trend (3) remains unchanged. One has to stress again that the sequencing of the terminal strands S1, S2, and S5 given by Eq. 3 is different from that proposed by Irbäck et al. (9), based on the breaking of the intrastructure contacts of C. Unfortunately, there are no experimental data available for comparison with our theoretical prediction.

Mechanical unfolding pathways: one end is fixed

N-terminus is fixed

Here we adopted the same procedure as in the previous section except the N-terminus is held fixed during simulations. As in the process where both of the termini are subjected to force, one can show that Cluster 1 unfolds after Cluster 2 (results not shown).

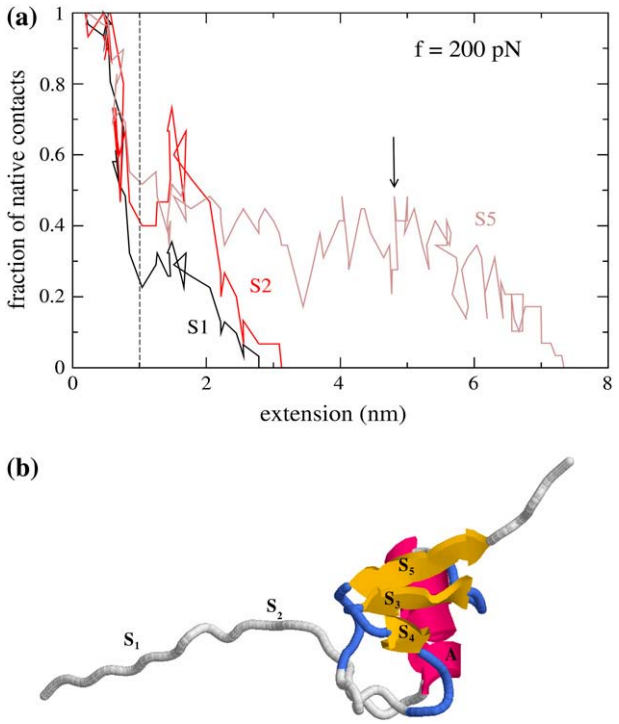

FIGURE 11 (*a*) The dependence of fraction of the native contacts on ΔR for strands S1, S2, and S5 ($f = 200$ pN). The vertical dashed line marks the position of the plateau at $\Delta R \approx 1$ nm. (*b*) The snapshot, chosen at the extension marked by the arrow in *a*, shows that S2 unfolds before S5. At this point, all native contacts of S1 and S2 have already broken, while 50% of the native contacts of S5 are still present.

From Fig. 13, we obtain the unfolding pathways

$$C \rightarrow D \rightarrow E \rightarrow B \rightarrow A, \tag{4a}$$

$$\mathrm{S5} \rightarrow \mathrm{S3} \rightarrow \mathrm{S4} \rightarrow \mathrm{S1} \rightarrow \mathrm{S2} \rightarrow A, \tag{4b}$$

which are also valid for the other values of force ($f = 70$, 100, and 140 pN). Similar to the case when the force is applied to both ends, the structure C unravels first and the helix A remains the most stable. However, the sequencing of B, D, and E changes markedly compared to the result obtained by Irbäck et al. (9) (Fig. 10 *a*).

As evident from Eqs. 3 and 5, anchoring the first terminal has a much more pronounced effect on the unfolding pathways of individual strands. In particular, unzipping commences from the C-terminus instead of from the N-terminus. Fig. 13 *c* shows a typical snapshot where one can see clearly that S_5 detaches first. At the first glance, this fact may seem trivial because S_5 experiences the external force directly. However, our experience on unfolding pathways of the well-studied domain I27 from the human cardiac titin, e.g., shows that it may be not the case. Namely, as follows from pulling experiments (43) and simulations (44), strand A from the N-terminus unravels first, although this terminus is kept fixed. From this point of view, which strand of Ub actually detaches first is, a priori, not clear. In our opinion, it depends on the interplay between the native topology and the speed

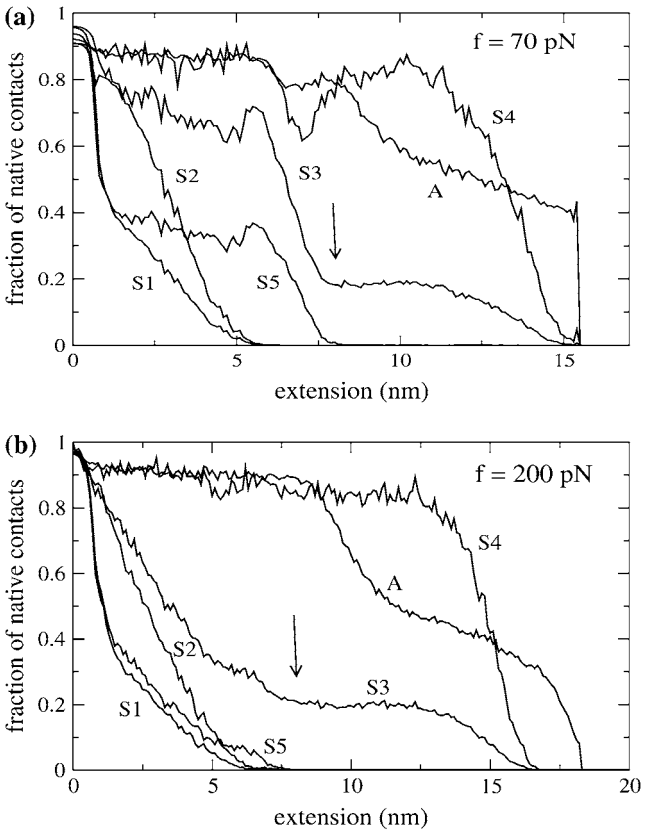

FIGURE 12 (*a*) The dependence of fraction of the native contacts on extension for A and all β-strands at $f = 70$ pN. (*b*) The same as in panel *a*, but for $f = 200$ pN. The arrow points to $\Delta R = 8.1$ nm where the intermediates are recorded on the experiments (7). The results are averaged over 200 trajectories.

of tension propagation. The latter factor probably plays a more important role for Ub, while the opposite situation happens with I27. One possible reason for it is related to the high stability of helix A, which does not allow either for the N-terminal to unravel first or for servility in unfolding starting from the C-end.

C-terminus is fixed

One can show that unfolding pathways of structures A, B, C, D, and E remain exactly the same as in the case when

TABLE 1 Dependence of unfolding pathways on the external force

Force (pN)	S1 → S2 → S5(%)	S5 → S1 → S2(%)	(S1,S2,S5) (%)
70	81	8	11
100	76	10	14
140	53	23	24
200	49	26	25

There are three possible scenarios: S1 → S2 → S5; S5 → S1 → S2; and three strands unzipping almost simultaneously (S1,S2,S5). The probabilities of observing these events are given in percentage.

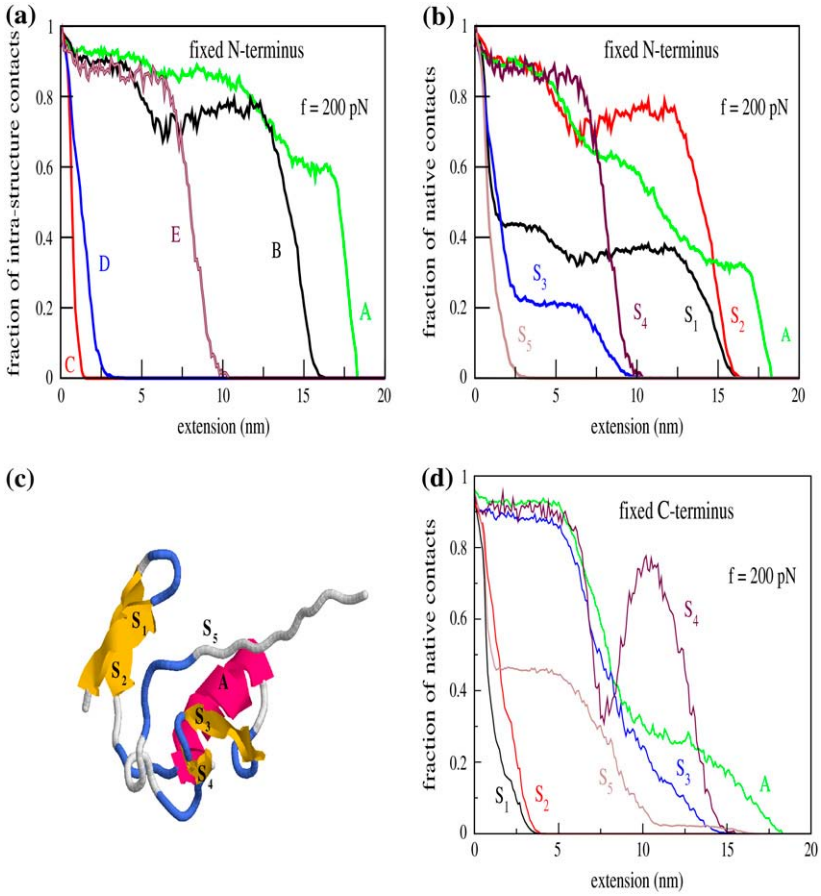

FIGURE 13 (*a*) The dependence of fraction of the intrastructure native contacts on extension for all structures at $f = 200$ pN. The N-terminus is fixed and the external force is applied via the C-terminus. (*b*) The same as in panel *a*, but for the native contacts of all individual β-strands and helix *A*. The results are averaged over 200 trajectories. (*c*) A typical snapshot to show that S_5 is fully detached from the core while S_1 and S_2 still have ≈50% and 100% contacts, respectively. (*d*) The same as in panel *b*, but the C-end is anchored and N-end is pulled. The strong drop in the fraction of native contacts of S_4 at $\Delta R \approx 7.5$ nm does not correspond to the substantial change of structure as it has only three native contacts in total.

Ub has been pulled from both termini (see Fig. 10). Concerning the individual strands, a slight difference is observed for S_5 (compare Fig. 13 *d* and Fig. 12). Most of the native contacts of this domain break before S_3 and S_4, except the long tail at extension $\Delta R \gtrsim 11$nm due to high mechanical stability of only one contact between residues 61 and 65 (the highest resistance of this pair is probably due to the fact that among 25 possible contacts of S_5 it has the shortest distance $|61 - 65| = 4$ in sequence). This scenario holds in ~90% of trajectories, whereas S_5 unravels completely earlier than S_3 and S_4 in the remaining trajectories. Thus, anchoring C-terminus has much less effect on unfolding pathways than in the case when the N-terminus is immobile.

It is worthwhile to note that, experimentally, one has studied the effect of extension geometry on the mechanical stability of Ub fixing its C-terminus (10). The greatest mechanical strength (the longest unfolding time) occurs when the protein is extended between N- and C-termini. This result has been supported by Monte Carlo (10) as well as MD (8) simulations. However, the mechanical unfolding sequencing has not been studied yet. It would be interesting to check our results on the effect of fixing one end on Ub mechanical unfolding pathways by experiments.

Thermal unfolding pathways

To study the thermal unfolding we follow the protocol described in Materials and Methods. Two-hundred trajectories were generated starting from the native conformation with different random seed numbers. The fractions of native contacts of helix *A* and five β-strands are averaged over all trajectories for the time window $0 \leq \delta \leq 1$. The unfolding routes are studied by monitoring these fractions as a function of δ. Above $T \approx 500$ K, the strong thermal fluctuations (entropy-driven regime) make all strands and helix *A* unfold almost simultaneously. Below this temperature, the statistical preference for the unfolding sequencing is observed. We focus on $T = 370$ and 425 K. As in the case of the mechanical unfolding, Cluster 2 unfolds before Cluster 1 (results not shown). However, the main departure from the mechanical behavior is that the strong resistance to thermal fluctuations of Cluster 1 is mainly due to the stability of strand S2 but not of helix *A* (compare Fig. 14, *c* and *d*, with Fig. 12). The unfolding of Cluster 2 before Cluster 1 is qualitatively consistent with the experimental observation that the C-terminal fragment (residues 36–76) is largely unstructured while nativelike structure persists in the N-terminal fragment (residues 1–35) (45–47). This is also consistent with the data

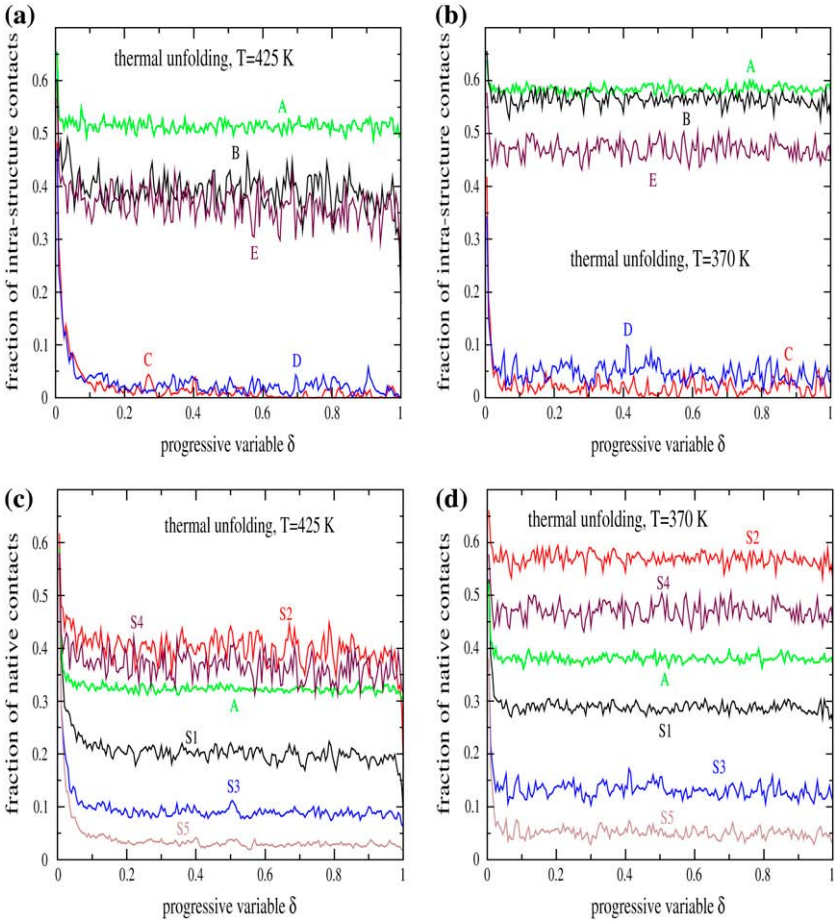

FIGURE 14 (*a*) The dependence of fraction of intrastructure native contacts on the progressive variable δ for all structures at $T = 425$ K. (*b*) The same as in panel *a*, but for $T = 370$ K. (*c*) The dependence of the all native contacts of the β-strands and helix *A* at $T = 425$ K. (*d*) The same as in panel *c*, but for $T = 370$ K.

from the folding simulations (23) as well as with the experiments of Went and Jackson (14), who have shown that the ϕ-values ≈ 0 in the C-terminal region. However, our finding is at odds with the high ϕ-values obtained for several residues in this region by all-atom simulations (48) and by a semiempirical approach (19). One possible reason for high ϕ-values in the C-terminal region is the force fields. For example, Marianayagam and Jackson have employed the GROMOS 96 force field (49) within the GROMACS software package (50). It would be useful to check whether the other force fields give the same result.

The evolution of the fraction of intrastructure contacts of *A*, *B*, *C*, *D*, and *E* is shown in Fig. 14 *a* ($T = 425$ K) and *b* ($T = 370$ K). Roughly we have the unfolding sequencing, given by Eq. 5, which strongly differs from the mechanical one. The large stability of the α-helix fragment *A* against thermal fluctuations is consistent with the all-atom unfolding simulations (17) and the experiments (14). The N-terminal structure *B* unfolds even after the core part *E*, and at $T = 370$ K its stability is comparable with helix *A*. The fact that *B* can withstand thermal fluctuations at high temperatures agrees with the experimental results of Went and Jackson (14) and of Cordier and Grzesiek (12), who used the notation β_1/β_2 instead of *B*. This also agrees with the results of Gilis and Rooman (22), who used a coarse-grained model, but disagrees with results from all-atom simulations (17). This disagreement is probably because Alonso and Daggett studied only two short trajectories and *B* did not completely unfold (17). The early unzipping of the structure *C* (Eq. 5a) is consistent with the MD prediction (17). Thus our thermal unfolding sequencing (Eq. 5a) is more complete compared to the all-atom simulation, and it gives reasonable agreement with the experiments.

We now consider the thermal unstability of individual β-strands and helix *A*. At $T = 370$ K (Fig. 14 *d*), the trend that S2 unfolds after S4 is more evident compared to the $T = 425$ K case (Fig. 14 *c*). Overall, the simple Gō model leads to the sequencing given by:

$$(C, D) \rightarrow E \rightarrow B \rightarrow A \tag{5a}$$

$$\mathrm{S5} \rightarrow \mathrm{S3} \rightarrow \mathrm{S1} \rightarrow A \rightarrow (\mathrm{S4}, \mathrm{S2}). \tag{5b}$$

From Eqs. 3, 4b, and 5b, it is obvious that the thermal unfolding pathways of individual strands differ markedly from the mechanical ones. This is not surprising, because the force should unfold the termini first, while under thermal fluctuations the most unstable part is expected to detach first. Interestingly, for these structures the thermal and mechanical

pathways (compare Eqs. 5a and 4a) are almost identical, except that the sequencing of *C* and *D* is less pronounced in the former case. This coincidence is probably accidental.

The fact that S5 unfolds first agrees with the high-resolution NMR data of Cordier and Grzesiek (12), who studied the temperature dependence of hydrogen bonds of Ub. However, using the ψ-value analysis, Krantz et al. (15) have found that S5 (B3 in their notation) breaks even after S1 and S2. As pointed out by Fersht (51), one possible reason would be if there is any plasticity in the transition state that can accommodate the crosslink between the metal and bihistidines, then ψ-values would be significantly greater than zero even for an unstructured region, leading to an overestimation of structure in the transition state. In agreement with our results, the ϕ-value analysis (14) yields that S5 breaks before S1 and *A*, but it fails to determine whether S5 breaks before S3. By modeling the amide I vibrations, Chung et al. (13) argued that S1 and S2 are more stable than S3, S4, and S5. Equation 5b shows that the thermal stability of S1 and S2 is indeed higher than S3 and S5 but S4 may be more stable than S1. The reason for only partial agreement between our results and those of Chung et al. (13) remains unclear. It may be caused either by the simplicity of the Gō model or by the model proposed in Chung et al. (13). The relatively high stability of S4 (Eq. 5b) is supported by the ψ-value analysis (15).

Thermal unfolding barrier

Fig. 15 shows the temperature dependence of the unfolding time τ_{UF}, which depends on the thermal unfolding barrier, ΔF^{T}_{UF}, exponentially, $\tau_{UF} \approx \tau^{0}_{UF} \exp(\Delta F^{T}_{UF}/k_B T)$. From the linear fit in Fig. 15 we obtain $\Delta F^{T}_{UF} \approx 10.48\epsilon_h \approx 10.3$ kcal/mol. It is interesting to note that ΔF^{T}_{UF} is compatible with

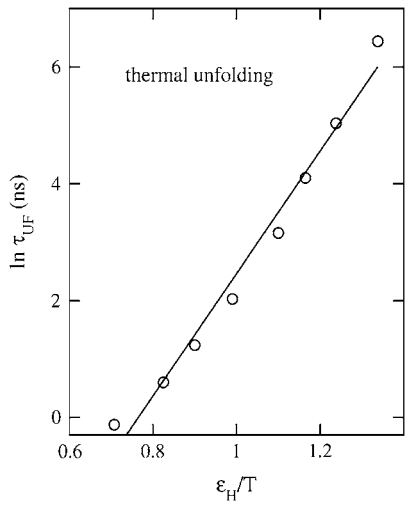

FIGURE 15 Dependence of thermal unfolding time τ_{UF} on ϵ_H/T, where ϵ_H is the hydrogen-bond energy. The straight line is a fit $y = -8.01 + 10.48x$.

$\Delta H_m \approx 11.4$ kcal/mol obtained from the equilibrium data (Fig. 2 *b*). However, the latter is defined by an equilibrium constant (the free energy difference between native and denatured states) but not by the rate constant (see, for example, (52)).

DISCUSSION

We have studied the refolding of Ub following the same protocol as in the force-clamp experiments of Fernandez and Li (2). Under the low quenched force the refolding is a two-stage process characterized by two different timescales τ^{A}_1 and τ^{A}_2, where $\tau^{A}_1 \ll \tau^{A}_2$. This result further strengthens our previous prediction (4) that the nature of the folding starting from the FDE does not depend on the details of the models. The simple C_α-Gō model provides reasonable estimates for the equilibrium critical force f_c as well as the averaged distance between the D and TS states, Δx_F, and the distance between the N and TS states, Δx_{UF}. We have also obtained ΔH_m from the two-state fit of the population of the NBA, f_N, and the thermal unfolding barrier ΔF^{T}_{UF}. It would be interesting to measure ΔF^{T}_{UF} experimentally and compare it with ΔH_m.

The shortcoming of the Gō model we used is its failure to capture seldom-unfolding intermediates observed in the experiments (7) as well as in the all-atom simulations (9). However, it mimics the overall two-state behavior of Ub. Our simulations suggest that the nonnative interactions, neglected in the C_α Gō model, may be the cause of mechanical unfolding intermediates.

Due to thermal fluctuations, the thermal unfolding pathways are not well defined as in the mechanical case. Nevertheless, at $T < 500$ K the statistical preference in the sequencing of unfolding events is evident. In accord with the experiments, Cluster 2 unfolds before Cluster 1 in the mechanical as well as in the thermal cases. However, in terms of individual strands we predict that mechanical and thermal unfolding follows very different pathways (Eq. 3 and Eq. 5b). Mechanically, strand S1 is the most unstable, whereas the thermal fluctuations break contacts of S5 first. If we consider only breaking of intrastructure native contacts, then our mechanical sequencing agrees with the all-atom simulation results (9). It is probably not unexpected because mechanical unfolding pathways may depend largely on the topology of the native conformation and in some cases the Gō-like models may give results comparable with experimental ones (8). However, contrary to Irbäck et al. (9), we predict that the terminal strands follow the mechanical unfolding sequencing: S1 → S2 → S5. It would be very exciting to perform the AFM experiments to verify this prediction and the whole unfolding sequencing (Eq. 3).

We have considered the effect of fixing one end on unfolding kinetics and found that it delays the unfolding by nearly a factor of 2 regardless to what end is anchored. We argue that this general result may be understood, using the

diffusion-collision model developed by Karplus and Weaver (42). However, fixing one terminus does not affect the distance between the native state and TS. One of the most interesting results is that which terminus we keep fixed matters for the unfolding sequencing. Namely, anchoring the N-end changes it dramatically (see Eq. 3 and Eq. 4b), whereas fixing the C-end has only a minor effect.

As evident from Eqs. 5a and 5b and the detailed discussion in the Introduction, our thermal unfolding sequencing is more complete compared with previous theoretical studies (17,19–23). We have obtained some agreement with the experimental data (12–15) on the instability of the structures and β-strands. However, the picture for thermal unfolding pathways is still incomplete. More experiments are needed to check our prediction given by Eqs. 5a and 5b.

We have also shown that refolding from FDE and folding from TDE have the same pathways, which are not sensitive to the quenched force. The refolding/folding sequencing is the same as for the thermal unfolding (see Eqs. 5a and 5b) but in the inverse order, implying that the protein folding is a reversible process.

Note added in proof: After acceptance of this manuscript, we became aware of Irbäck and Mitternacht (53) where the similar result on thermal unfolding pathways has been obtained using the all-atom simulations.

M.S.L. thanks A. Irbäck, D. K. Klimov, S. Mitternacht, E. P. O'Brien Jr., and D. Thirumalai for very useful discussions.

This work was supported by Komitet Badan Naukowych grant No. 1P03B01827, the National Science Council in Taiwan under grant Nos. NSC 93-2112-M-001-027 and 95-2119-M-002-001, and Academia Sinica in Taiwan under grant Nos. AS-92-TP-A09 and AS-95-TP-A07.

REFERENCES

1. Fisher, T. E., A. F. Oberhauser, M. Carrion-Vazquez, P. E. Marszalek, and T. M. Fernandez. 1999. The study of protein mechanics with the atomic force microscope. *Trends Biochem. Sci.* 24:379–384.
2. Fernandez, J. M., and H. Li. 2004. Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science*. 303:1674–1678.
3. Bell, G. I. 1978. Models for the specific adhesion of cells to cells. *Science*. 100:618–627.
4. Li, M. S., C. K. Hu, D. K. Klimov, and D. Thirumalai. 2006. Multiple stepwise refolding of immunoglobulin domain i27 upon force quench depends on initial conditions. *Proc. Natl. Acad. Sci. USA*. 103:93–98.
5. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: what determines the structural detail of transition state ensemble and en-route intermediates for protein folding? *J. Mol. Biol.* 298:937–953.
6. Klimov, D. K., and D. Thirumalai. 2000. Native topology determines force-induced unfolding pathways in globular proteins. *Proc. Natl. Acad. Sci. USA*. 97:7254–7259.
7. 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.
8. West, D. K., D. J. Brockwell, P. D. Olmsted, S. E. Radford, and E. Paci. 2006. Mechanical resistance of proteins explained using simple molecular models. *Biophys. J.* 90:287–297.
9. Irback, A., S. Mitternacht, and S. Mohanty. 2005. Dissecting the mechanical unfolding of ubiquitin. *Proc. Natl. Acad. Sci. USA*. 102: 13427–13432.
10. Carrion-Vazquez, M., H. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser, and J. M. Fernandez. 2003. The mechanical stability of ubiquitin is linkage dependent. *Nat. Struct. Biol.* 10:738–743.
11. Paci, E., and M. Karplus. 2000. Unfolding proteins by external forces and temperature: the importance of topology and energetics. *Proc. Natl. Acad. Sci. USA*. 97:6521–6526.
12. Cordier, F., and S. Grzesiek. 2002. Temperature-dependence of protein hydrogen bond properties as studied by high-resolution NMR. *J. Mol. Biol.* 315:739–752.
13. Chung, H. S., M. Khalil, A. W. Smith, Z. Ganim, and A. Tomakoff. 2005. Conformational changes during the nanosecond-to-millisecond unfolding of ubiquitin. *Proc. Natl. Acad. Sci. USA*. 102:612–617.
14. Went, H. M., and S. E. Jackson. 2005. Ubiquitin folds through a highly polarized transition state. *Protein Eng. Des. Sel.* 18:229–237.
15. Krantz, B. A., R. S. Dothager, and T. R. Sosnick. 2004. Discerning the structure and energy of multiple transition states in protein folding using ψ-analysis. *J. Mol. Biol.* 337:463–475.
16. Sosnick, T. R., B. A. Kratz, R. S. Dothager, and M. Baxa. 2006. Characterizing the protein folding transition state using ψ analysis. *Chem. Rev.* 106:1862–1876.
17. Alonso, D. O. V., and V. Daggett. 1998. Molecular dynamics simulations of hydrophobic collapse of ubiquitin. *Protein Sci.* 7:860–874.
18. Larios, E., J. S. Li, K. Schulten, H. Kihara, and M. Gruebele. 2004. Multiple probes reveal a native-like intermediate during low-temperature refolding of ubiquitin. *J. Mol. Biol.* 340:115–125.
19. Fernandez, A. 2001. Conformation-dependent environments in folding proteins. *J. Phys. Chem.* 114:2489–2502.
20. Fernandez, A. 2002. Time-resolved backbone desolvation and mutational hot spots in folding nucleus. *Proteins*. 47:447–457.
21. Fernandez, A., A. Colubri, and R. Berry. 2002. Three-body correlations in protein folding: the origin of cooperativity. *Physica A*. 307:235–259.
22. Gilis, D., and M. Rooman. 2001. Identification and ab initio simulations of early folding units in proteins. *Proteins Struct. Funct. Gen.* 42: 164–176.
23. Sorenson, J. M., and T. Head-Gordon. 2002. Toward minimalist models of larger proteins: a ubiquitin-like protein. *Proteins Struct. Funct. Gen.* 46:368–379.
24. Thomas, S. T., V. V. Loladze, and G. I. Makhatadze. 2001. Hydration of the peptide backbone largely defines the thermodynamic propensity scale of residues at the C'position of the C-capping box of α-helices. *Proc. Natl. Acad. Sci. USA*. 98:10670–10675.
25. Kouza, M., C. F. Chang, S. Hayryan, T. H. Yu, M. S. Li, T. H. Huang, and C. K. Hu. 2005. Folding of the protein domain HBSBD. *Biophys. J.* 89:3353–3361.
26. Swope, W. C., H. C. Andersen, P. H. Berens, and K. R. Wilson. 1982. Computer simulation method for the calculation of equilibrium constants for the formation of physical clusters and molecules: application to small water clusters. *J. Chem. Phys.* 76:637–649.
27. Camacho, C. J., and D. Thirumalai. 1993. Kinetics and thermodynamics of folding in model proteins. *Proc. Natl. Acad. Sci. USA*. 90:6369–6372.
28. Ferrenberg, A. M., and R. H. Swendsen. 1989. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* 63:1195–1198.
29. Klimov, D. K., and D. Thirumalai. 1999. Stretching single-domain proteins: phase diagram and kinetics of force-induced unfolding. *Proc. Natl. Acad. Sci. USA*. 96:6166–6171.
30. Klimov, D. K., and D. Thirumalai. 2001. Lattice model studies of force-induced unfolding of proteins. *J. Phys. Chem. B*. 105:6648–6654.
31. Geissler, P. L., and E. I. Shakhnovich. 2002. Reversible stretching of random heteropolymers. *Phys. Rev. E*. 65:056110–056113.
32. Evans, E., and K. Ritchie. 1997. Dynamics strength of molecular adhesion bonds. *Biophys. J.* 72:1541–1555.
33. Nguyen, P. H., G. Stock, E. Mittag, C. K. Hu, and M. S. Li. 2005. Free energy landscape and folding mechanism of β-hairpin in explicit water: a replica exchange molecular dynamics study. *Proteins Struct. Funct. Bioinformat.* 61:795–808.

34. Klimov, D. K., and D. Thirumalai. 1998. Cooperativity in protein folding: from lattice models with sidechains to real proteins. *Fold. Des.* 3:127–139.

35. Li, M. S., D. K. Klimov, and D. Thirumalai. 2004. Finite size effects on thermal denaturation of globular proteins. *Phys. Rev. Lett.* 93: 268107–268110.

36. Kouza, M., M. S. Li, E. P. O'Brien, Jr., C. K. Hu, and D. Thirumalai. 2006. Effect of finite size on cooperativity and folding rates of proteins. *J. Phys. Chem. A.* 110:671–676.

37. Kaya, H., and H. S. Chan. 2000. Energetic components of cooperative protein folding. *Phys. Rev. Lett.* 85:4823–4826.

38. Chan, H. S., S. Shimizu, and H. Kaya. 2004. Cooperativity principles in protein folding. *Methods Enzymol.* 380:350–379.

39. Li, M. S., D. K. Klimov, and D. Thirumalai. 2005. Finite size effects on calorimetric cooperativity of two-state proteins. *Physica A.* 350:38–44.

40. Kirmizialtin, S., L. Huang, and D. E. Makarov. 2005. Topography of the free energy landscape probed via mechanical unfolding of proteins. *J. Chem. Phys.* 122:234915–234926.

41. Szymczak, P., and M. Cieplak. 2006. Stretching of proteins in a force-clamp. *J. Phys. Condens. Matt.* 18:L21–L28.

42. Karplus, M., and D. L. Weaver. 1976. Protein-folding dynamics. *Nature.* 260:404–406.

43. Marszalek, P. E., H. Lu, H. Li, M. Carrion-Vazquez, A. F. Oberhauser, K. Schulten, and J. M. Fernandez. 1999. Mechanical unfolding intermediates in titin modules. *Nature.* 402:100–103.

44. Lu, H., and K. Schulten. 1999. Steered molecular dynamics simulations of force-induced protein domain unfolding. *Proteins Struct. Funct. Gen.* 35:453–463.

45. Bofill, R., and M. S. Searle. 2005. Engineering stabilising β-sheet interactions into a conformationally flexible region of the folding transition state of ubiquitin. *J. Mol. Biol.* 353:373–384.

46. Cox, J. P. L., P. A. Evans, L. C. Packman, D. H. Williams, and D. N. Woolfson. 1993. Dissecting the structure of a partially folded protein. CD and NMR studies of peptides from ubiquitin. *J. Mol. Biol.* 234: 483–492.

47. Jourdan, M., and M. S. Searle. 2000. Cooperative assembly of a native-like ubiquitin structure through peptide fragment complexation: energetics of peptide association and folding. *Biochemistry.* 39:12355–12364.

48. Marianayagam, N. J., and S. E. Jackson. 2004. The folding pathway of ubiquitin from all-atom molecular dynamics simulations. *Biophys. Chem.* 111:159–171.

49. van Gunsteren, W., S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. Scott, and I. Tironi. 1996. Biomolecular Simulation: The GROMOS96 Manual and User Guide. Vdf Hochschulverlag AG an der ETH, Zurich.

50. Berendsen, H., D. van der Spoel, and R. van Drunen. 1995. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Comm.* 91:43–56.

51. Fersht, A. R. 2004. ϕ-Value versus ψ-analysis. *Proc. Natl. Acad. Sci. USA.* 101:17327–17328.

52. Nothonha, M., J. C. Lima, M. Bastos, H. Santos, and A. L. Macanita. 2004. Unfolding of ubiquitin studied by picosecond time-resolved fluorescence of the tyrosine residue. *Biophys. J.* 87:2609–2620.

53. Irbäck, A., and S. Mitternacht. Thermal versus mechanical unfolding of ubiquitin. *Proteins.* 65:759–766.