Folding and Stretching in a Go-like Model of Titin

Marek Cieplak,[1,2,*] **Trinh Xuan Hoang,**[3] **and Mark O. Robbins**[1]

[1]*Department of Physics and Astronomy, The Johns Hopkins University, Baltimore, Maryland*
[2]*Institute of Physics, Polish Academy of Sciences, Warsaw, Poland*
[3]*International School for Advanced Studies (SISSA), Trieste, Italy*

ABSTRACT **Mechanical stretching of the I27 domain of titin and of its double and triple repeats are studied through molecular dynamics simulations of a Go-like model with Lennard-Jones contact interactions. We provide a thorough characterization of the system and correlate the sequencing of the folding and unraveling events with each other and with the contact order. The roles of cantilever stiffness and pulling rate are studied. Unraveling of tandem titin structures has a serial nature. The force-displacement curves in this coarse-grained model are similar to those obtained through all atom calculations. Proteins 2002;49:114–124.** © 2002 Wiley-Liss, Inc.

INTRODUCTION

The giant protein molecule titin, also known as connectin, is responsible for the elastic and extension properties of smooth, skeletal, and cardiac muscles.[1–4] Titin spans half of the sarcomere, the repeat segment in muscle fibers, and has been implicated in certain heart diseases. The biological importance of titin and the nature of its function make it an ideal candidate for studies of the mechanical properties of proteins.

Titin consists of about 30,000 amino acids that are organized into about 300 domains that form the so-called A-band followed by the I-band. Domains that are similar to fibronecti type III are located only in the A band, whereas immunoglobulin-like (Ig) domains appear throughout the length of titin. In the middle of the I-band there is one special domain[5] called PEVK that consists of between 163 (in a cardiac titin) and 2174 (in a skeletal titin) residues. The number of Ig domains in the I-band ranges from 37 in cardiac titin to 90 in skeletal titin. The native structure of only one of these Ig domains, I27 (which is especially stable), has been resolved through NMR spectroscopy [2] and found to have the architecture of a β-sandwich. The remaining domains are believed to be similar in structure, even though the sequence similarity is only of order 25%.

The first mechanical studies of titin concluded that small stretching forces affect primarily the PEVK domain[3,6] and still larger forces induce extension of the Ig domains.[7] Recent experiments on stretching of single titin molecules by optical tweezers[8,9] and atomic force microscopy[10] showed relevance of both PEVK and Ig domains to the mechanics of titin. The overall picture is that the I-band accommodates stretch by straightening of the Ig domains and by unfolding of the PEVK domain. Unraveling of the 25 consecutive Ig domains resulted in a train of sawtooth force patterns as a function of tip displacement that repeated in a periodic manner,[10] which made Ericson[3] call titin a "weird spring." Initiation of stretching of the Ig domains was found to require a very high force sustained for only a short tip displacement.[10] The force drops once the domain is destabilized, and later force peaks are not as high.

Natural titins are inhomogeneous. However, protein engineering has recently allowed production of tandem repeats of identical Ig modules.[11–13] Studies of proteins made of 8 and 12 Ig domains have indicated[14] that the hydrogen bonds allow the domains of titin to stretch reversibly only up to some limiting extension. Beyond that limit, misfolding takes place. This is consistent with the studies of mechanical unfolding of secondary structures in the companion article,[15] where proteins refolded to the proper sequence until an irreversibility length was exceeded.

Understanding of the unraveling of titin has been facilitated by all atom computer simulations with[4,16] and without[17,18] explicit water molecules. These studies indicated the existence of a large bottleneck to unfolding of a single domain at small end-to-end extensions and pointed to a serial character of the many-domain unraveling in titin.[18] Klimov and Thirumalai have considered simplified coarse-grained lattice[19] and off-lattice[20] models. The latter were based on a model[21] that contained three kinds of amino acids and Lennard-Jones interactions between pairs of hydrophobic residues. Two model sequences with four-stranded β-barrel topology were considered. Klimov and Thirumalai[20] noted that thermal unfolding appeared to proceed along pathways that were distinct from the stretching trajectories. These studies led them to suggest that a natural way to characterize constant force-induced unfold-
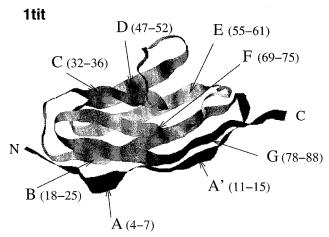
1tit

Fig. 1. The ribbon representation of the domain 1tit. The symbols indicate β-strands together with the sequence position of the amino acids involved.

ing is in terms of a phase diagram on a plane defined by the force and the concentration of denatured fragments.

In this article, we also consider coarse-grained models because these models allow for a thorough characterization of mechanical, equilibrium, and folding properties, which is essential to an understanding of the system. Specifically, we extend our analysis of the unfolding of secondary structures[15] to Go-like models[22] of one, two, and three Ig domains. These models are defined through the experimentally determined native structures, and they capture essential aspects of the important role played by the native geometry on the folding mechanism.[23–25]

We first focus on a single domain and examine its folding characteristics. We determine the characteristic folding time as a function of temperature, establish the succession of folding events, and relate it to the contact order—the separation of two amino acids along the chain. We then determine the succession of events in mechanical unraveling and again relate it to the contact order. We explore cross-correlations between the thermal and mechanical event sequencing and single out long-ranged contacts as providing a link between the two.

The protein is stretched by attaching its ends to harmonic springs to model a Hookean cantilever. One end of the cantilever is displaced with a constant speed v_p. We study the dependence of the force-displacement curves on cantilever stiffness and pulling rate. Finally, we study tandem arrangements of several domains and show that they unravel in a serial manner that is in sharp contrast to the parallel unraveling found for two α-helices in tandem.[15]

MODEL AND METHOD

The Go-like[22,23] coarse-grained model we use is explained in references.[26–28] The native structure of I27 is taken from the PDB[29] data bank where it is stored under the name 1tit, which we shall use as an alternative to I27. The coarse-grained picture of 1tit is shown in Figure 1. 1tit consists of N = 98 residues that are organized into eight

β-strands and connecting turns. There are no α-helices in 1tit.

In the Go model, amino acids are represented by point particles, or beads, which are located at the positions of the C^α atoms. Consecutive beads in the chain are tethered by an anharmonic potential: $V_{BB} = \Sigma_{i=1}^{N-1} U_{BB}(d_i)$, where d_i is the distance between beads i and $i+1$,

$$U^{BB}(d) = k_1(d - d_0)^2 + k_2(d - d_0)^4, \qquad (1)$$

where $d_0 = 3.8$ Å is the equilibrium bond length, $k_1 = 1\epsilon/\text{Å}^2$ and $k_2 = 100\epsilon/\text{Å}^4$. The resulting bonds are more easily stretched to small displacements than realistic atomic potentials. This speeds folding, but increases the length of the molecule slightly too much in unfolding simulations. The potentials between nonadjacent beads are chosen so that the native structure minimizes the energy. The interaction between beads that form native contacts (defined as C^α atoms separated by <7.5 Å) are of the Lennard-Jones type, whereas the interactions for nonnative contacts are purely repulsive. The Lennard-Jones couplings are scaled by a uniform energy parameter ϵ and the characteristic lengths, σ_{ij}, are contact dependent. Tandem structures of two or three domains are constructed by repeating 1tit domains in series with one extra peptide link between the domains.

The model of the pulling cantilever is as in the preceding article.[15] Both ends of the protein are attached to harmonic springs of spring constant k. The outer end of one spring is held fixed, and the outer end of the other is pulled at constant speed v_p. The stretching is implemented parallel to the initial end-to-end position vector. This corresponds to stretching the protein with a cantilever of stiffness $k/2$ at a constant rate. The net force acting on the bead attached to the moving end is denoted by F, the cantilever displacement is denoted by d, and the end-to-end distance of the protein by L. The case of constant force was also considered, but unfolding then has an all-or-nothing character that yields little information.

Two cantilever stiffnesses and two velocities are considered. The case of stiff cantilevers corresponds to $k = 30$ ϵ/σ^2, and k = 0.12 ϵ/σ^2 for soft cantilevers. The case of slow pulling corresponds to a cantilever velocity of 0.005 Å/τ. Here $\tau = \sqrt{m\sigma^2/\epsilon} \approx 3ps$ is the characteristic time for the Lennard-Jones potentials, where $\sigma = 5$ Å is a typical value of σ_{ij}, and m is the average mass of the amino acids. Results begin to become rate dependent just below the fast pulling rate of $v_p = 0.5$ Å/τ.

To control the temperature, T, of the system and mimic the effect of solvent molecules, the equations of motion for each bead include Langevin noise and damping terms.[30] A damping constant of $\gamma = 2$ m/τ is used. It has been argued[21] that realistic values of the solvent damping are 25 times larger. However, using a smaller damping rate decreases the required simulation time without affecting the sequencings of events.[26,27,31] Because the protein relaxes more rapidly, our effective pulling velocities are substantially slower than their nominal values. Almost all data correspond to using the same average mass for all
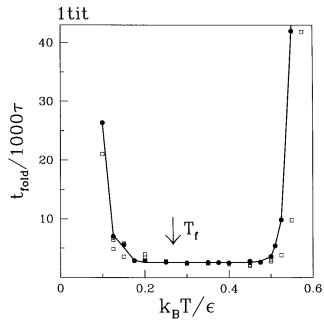
Fig. 2.   Median folding time for the Go-like model of 1tit. Solid symbols, simulations with uniform masses; open squares, simulations using the actual masses of the amino acids in the sequence. The arrow indicates the value of the folding temperature.
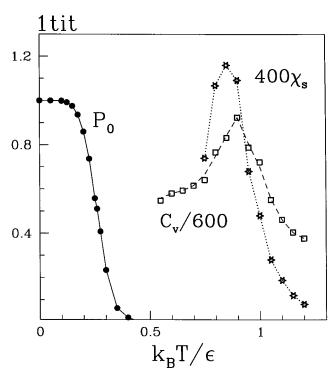

Fig. 3.   Equilibrium properties of 1tit. Circles, the probability for the protein to be in the native state; asterisks, the dimensionless structural susceptibility; squares, the specific heat in Lennard-Jones units.

amino acids, but we also studied the effect of amino acid-dependent masses on folding times.

All folding times were determined by considering the median times, over 201 trajectories, needed to form all native contacts. The criterion for forming a contact is that the distance between the corresponding beads is $<1.5\sigma_{ij}$. As explained in the companion article,[15] our studies of the mechanics of stretching are performed at $T = 0$ in order to eliminate the need for extensive averaging over trajectories and to get signals that depend primarily on the energy landscape. The precise role of the temperature on the force-extension curves remains to be elucidated.

RESULTS AND DISCUSSION
Folding Properties of 1tit

The sequencing of folding events depends on temperature and becomes smoothest and most natural at the temperature of fastest folding $T_{\min}$.[27] For most proteins, the characteristic folding time t_{fold} rises rapidly on either side of $T_{\min}$. Figure 2 shows that the dependence of t_{fold} on T for 1tit exhibits an unusually broad basin of optimality that extends from ~0.2 to 0.5 ϵ/k_B. Another important temperature is the folding temperature, T_f, at which the equilibrium probability, P_0, of finding the system in its native conformation is one half. Figure 3 shows P_0 versus T, as determined from ~20 trajectories that lasted for 60,000τ each. The value of $T_f = 0.26$ ϵ/k_B is in the basin of optimal folding times, implying that the system is a good folder. This is also confirmed by studies of the T-dependence of the specific heat C_v and structural susceptibility per bead χ_s as defined by Camacho and Thirumalai.[32] Figure 3 shows that the peaks in our calculated C_v

and χ_s almost coincide. This coincidence has been identified as another signature of good foldability.[33]

The values of the masses of the amino acids affect the equations of motion for the individual beads. Figure 2 demonstrates that taking into account the amino acid-dependent masses of the beads does not affect the folding times in any noticeable manner. This is consistent with the overall coarse-grained character of the Go model. In more realistic models, however, the nonuniformity of the masses, and, more importantly, the nonuniformity of the amino acid shapes and chemical functions, are expected to have an impact on the kinetics of folding.

The broad minimum in folding time makes it difficult to define a precise value of T_{min}. We chose to study the sequencing of folding events at $T = 0.25$ ϵ/k_B since this temperature is in the bottom of the basin of optimal folding times and close to the folding temperature $T_f = 0.26$ ϵ/k_B. The folding process at this T is characterized in two ways. The first, shown in Figure 4, is a plot of the average time to establish a native contact versus the contact order $|j-i|$, defined as the distance along the backbone between amino acid j and amino acid i. The second characterization, shown in Figure 5, is through a plot of the contact matrix that indicates which beads make a contact.

Figure 4 shows that folding of I27 takes place in stages separated by substantial time gaps. All short range contacts ($|j-i| < $ ~20) are established within the first 600τ. In the next stage, occurring between 1400 and 1700τ, most of the intermediate- and some of the long-range contacts form. Then, around 2100τ, most of the longest-ranged
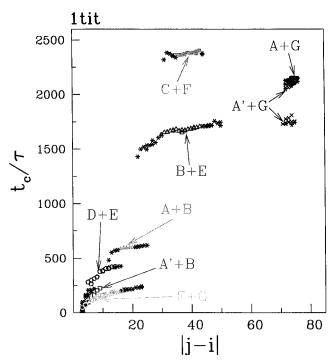
Fig. 4. Sequencing of the folding events in the Go model of 1tit as represented by the time needed to establish a contact versus the contact order. Open symbols, contacts that form the β-sheets, i.e., contacts between strands. The labels indicate the pairs of strands associated with those contacts. The stars correspond to other contacts.

Fig. 5. The contact map of 1tit without the contacts of the $i,i+2$ type. The correspondence between i and β-sheets is given in Figure 1. The symbols are divided into four groups (the division is different than in Fig. 4) to illustrate the average flow of contact formation in folding. Solid squares show contacts that are established in the first stage and are thus present at time 1000τ, the number shown next to the symbols. These are the short range contacts, the turns and some antiparallel β-sheets, which consist of members of the first two groups of Figure 4. The stars correspond to contacts established between 1000 and 1750τ, i.e., in the second stage of the evolution. These are primarily the antiparallel β-sheets and some of the longest-ranged parallel β-sheets. Triangles indicate the formation of the remaining longest-ranged parallel β-sheets in the third stage, up to the time of 2250τ. Open squares show the intermediate range contacts, which are formed last.

contacts are established. This is followed by the completion of the structure through building up of the remaining intermediate-range contacts. Thus, the sequencing of the folding events takes place in stages that are governed, to a large extent, by the contact order. The latter conclusion is consistent with the findings in References 34–37. In addition to the 217 contacts shown in Figure 4, there are also 87 contacts with $|j-i| = 2$. These contacts of shortest range are established rapidly and are not displayed for the sake of clarity.

The kinetics of folding events can also be gleaned from the contact map shown in Figure 5. Here, the symbols indicate the stage at which a given contact is formed. This representation allows one to infer details of the secondary structure formation that are only implicit in the index that defines the contact order. The last to form are the intermediate-ranged antiparallel β-sheets, which cannot be established before contacts of the longest range lock the overall topology in place.

In the notation of Figure 1, the average folding trajectory proceeds according to the scenario: $F + G \rightarrow A' + B \rightarrow D + E \rightarrow A + B \rightarrow B + E \rightarrow A' + G \rightarrow A + G \rightarrow C + F$, i.e., first F connects with G, then A' with B, and so on. Although this average pattern is the most common, it was followed by only 24 of 100 individual trajectories. In contrast, more than half of the individual trajectories agreed with the average succession for the proteins studied in References 26 and 27 (crambin, CI2, and the SH3 domain). This difference is related to the longer sequence length of 1tit and the greater number of events needed to
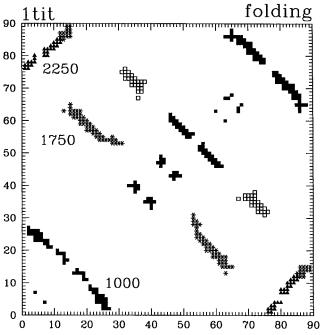
fold it to the native state. Other common trajectories correspond to permutations in contact making. For instance, 15 trajectories have the same sequence of events as the average succession except that the order of $A' + B$ and $F + G$ is reversed. In 45 trajectories the last four stages are identical to the average succession. A total of 90 trajectories establish β-sheets with low contact order ($F + G$, $A' + B$, $D + E$, $A + B$) before the sheets with high contact order are formed ($B + E$, $A' + G$, $A + G$, $C + F$).

Stretching of 1tit

The two snapshots of the mechanically unfolded model of 1tit shown in Figure 6 indicate that stretching affects the short- and long-range contacts simultaneously. Both ends of the protein straighten over longer and longer length scales, but the central "knot" gets unraveled as well, starting first at the longest-ranged contacts that pin the structure. These longest ranged contacts are *not* those that were established at the last stage of thermal folding, but they do arise toward the end of folding. In contrast to the cases considered in the previous article,[15] the orientation of the bonds relative to the pulling force changes as the protein unravels. The vectorial nature of the pulling
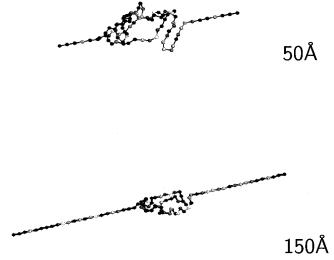
Fig. 6.   Snapshot pictures of the Go model of 1tit during stretching by a stiff cantilever at a pulling velocity of 0.005 Å/τ. The numbers indicate the displacement of the cantilever.
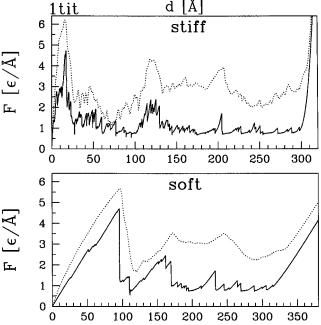

Fig. 7.   Force versus cantilever displacement for the Go model of 1tit for the slow (solid lines) and fast (dotted lines) pulling rates.

force is one of the key factors that leads to differences between folding and unfolding.[38] Before we look into the issues of event sequencing in more detail, we discuss the force-extension curves.

Figure 7 shows the force as a function of cantilever displacement for the two values of stiffness. The curves are terminated when the protein is fully extended, and the sharp rise in force at the end of the curves reflects stretching of covalent bonds along the backbone. As in the companion article,[15] the force curves show a series of upward ramps followed by rapid drops where contacts break. The slope of the upward ramps is the combined

stiffness of the protein and cantilever. The protein is softer than the stiff spring, and its elastic properties dominate in this case. The opposite applies for soft springs, and the slope of the upward ramps nearly coincides with $k/2 = 0.06$ $\epsilon/\text{Å}^2$. The ramps end when one or more contacts break. This allows the protein to extend and the force drops. As the cantilever stiffness decreases, the force drops more slowly with increases in the length L of the protein. If the extension due to breaking one set of contacts is not large enough, the force may remain above the threshold for breaking the next set of contacts. This leads to large avalanches where many contacts break in a single extended event. The stiff spring is able to resolve nearly all independent contact ruptures, whereas they coalesce into a much smaller number of large events in the soft spring case. Increasing the velocity 100-fold (Figure 7, dotted lines) limits the ability of tension to equilibrate along the chain and causes further merging of events.

For all force curves, the largest maximum occurs near the beginning of stretching. This peak represents the main bottleneck to mechanical unfolding. Subsequent peaks are visible, but are less than half as large. A few small drops are also visible on the way up to the main peak at which a total of 28 contacts break. Similar curves for mechanical unfolding of secondary structures showed a very different pattern. In the cases studied,[15] the force needed to break bonds tended to increase or remain constant until nearly all contacts had failed.

In the subsequent unraveling of 1tit there are further bursts of contact rupture, but none of them is as significant as the first one. When the cantilever is displaced by ∼300 Å (the top panel of Fig. 7), the domain is fully stretched and the force starts to increase rapidly, indicating an incipient rupture. At this value of the displacement the end-to-end distance L is ∼342 Å which corresponds to an almost 8-fold stretch relative to the native value of 43.19 Å. The experimental data show the immunoglobulin domain unraveling on extension from 40 to 300 Å,[39,40] which is consistent with the range in our model.

Figure 8 shows the number of native contacts, n_{NAT}, and the energy, E, of the model protein as a function of cantilever displacement at small pulling rates. For the stiff cantilever the dependence on d appears to be nearly continuous and monotonic, but a closer inspection reveals the presence of small jumps at certain values of d. These correlate with the bursts in the F versus d curve at the top of Figure 7. For the soft cantilever the steps involve much larger changes in both n_{NAT} and E, and the synchronization with drops in the force curve is more evident.

To illustrate the unfolding sequence, we first plot the cantilever displacement where each bond opens, d_u, as a function of contact order. Results for stiff and soft springs are shown in Figures 9 and 10, respectively. Note the presence of clear clusters in the d_u versus $|j-i|$ plane. The same set of bonds are clustered in the stiff and soft plots; however, the shapes of the clusters are different. The more horizontal character of the clustering in the soft case is due to coalescence of multiple bonds into coherent breaking events.
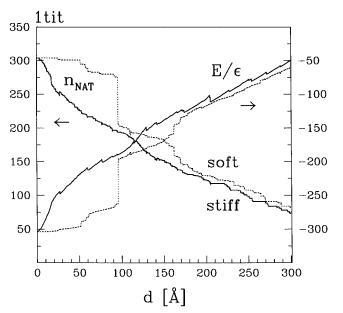
Fig. 8. Energy (right axis) and number of native contacts still present (left axis) as a function of cantilever displacement for the slow pulling rate and for the two cantilever stiffnesses. The number of contacts also includes those of the $i, i+2$ type.

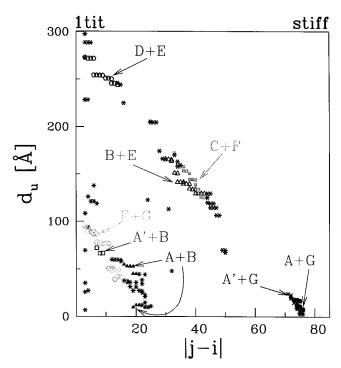Fig. 10. Same as in Figure 9 but for the soft cantilever.

Fig. 9. Sequencing of the stretching events in the Go model of 1tit as represented by the cantilever displacement at which a contact breaks. The contacts of the $i, i+2$ type are not shown here. This is the case of the stiff cantilever that is pulled at low speed. The symbols have the same meaning as in Figure 4.

Several differences between mechanical unfolding and thermal folding are evident when one compares Figures 9 and 10 with Figure 4. First, in thermal folding, the intermediate range ($|j-i|$ of ~40) and long-range contacts

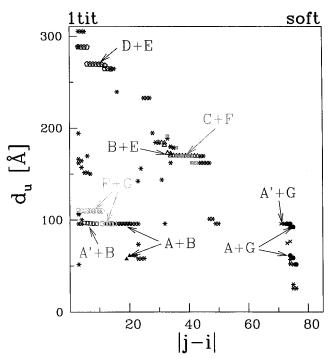are each divided into two time-separated groups. In contrast, in mechanical unraveling all long-range contacts cluster together and (except for several outliers) the intermediate contacts also form a single cluster. Second, in thermal folding the short-range contacts get established rapidly, whereas in stretching they continue to rupture throughout almost the entire unfolding process. Notice though that stretching leads to breaking of nearly all the long-range bonds before the short-range contacts begin to fail.

These differences are also evident when the contact map for unfolding with a stiff cantilever shown in Figure 11 is compared with the thermal folding map shown in Figure 5. In both cases the symbols indicate the stage at which the event occurred, and local clusters tend to evolve in the same stage. However, the order of unraveling shows no simple correlation with that of folding. This is particularly true for the contacts of short range that lie along the diagonal in Figures 5 and 11. All of these bonds form at early times during folding, but they break in the second and fourth stages of mechanical unfolding. The intermediate range bonds all unravel in the third stage of unfolding, but form in the second and fourth stages of unfolding. Only the long-range bonds act together in both cases, breaking in the first stage of unfolding and forming in the third stage of folding.

To illustrate the cross-correlation between stretching and folding, the breaking distance for each contact is plotted against folding time for stiff and soft cantilevers in Figures 12 and 13, respectively. Different symbols are used to indicate different ranges of bond order. The low order contacts (3-11) span the full vertical range but are confined to short times. This same lack of correlation is
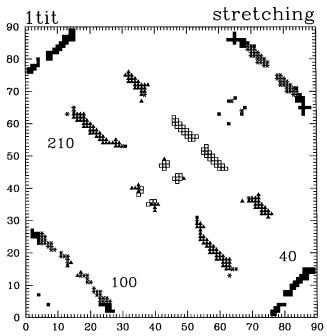
Fig. 11. The contact map that represents the evolution of contact breakage during unfolding with the stiff cantilever. Solid squares, contacts that are broken by the time d is 40 Å; asterisks, the additional contacts that are broken when d is 100 Å; triangles, further broken contacts when d is 210 Å; open squares, contacts that break at still larger d. The correspondence between i and β-sheets is given in Figure 1.
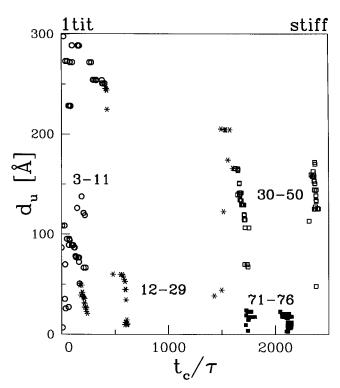

Fig. 12. Stretching distances at which a bond rupture takes place plotted versus average time needed to establish contact on folding. This is the case of a stiff cantilever that is being pulled slowly. The numbers indicate the range of the contact order that is associated with the symbol shown.

seen in the case of two α-helices connected together.[15] The long-range bonds show a clear anticorrelation, occurring at short distances and long times. Intermediate-range bonds are localized in a narrow range of d_u, but clustered into two different time intervals.

Comparison to Previous Simulations

It is remarkable that room temperature all atom simulations (CHARMM-based) by Lu et al.[4] for a stiff cantilever and v_p = 0.5 Å/ps yield a pattern (see Fig. 14) that is very similar to the one shown in the top panel of Figure 7. Their simulations place the main "burst" of contact rupture as occurring between 10 and 17 Å, with a peak at 14 Å. Although their forces are plotted as a function of L rather than d, the two differ by <1 Å for the stiff spring of Figure 7. Our force peak is at a slightly larger displacement (~17 vs. ~14 Å), mainly because the anharmonic potential (Eq. 1) used to describe the bonds between consecutive C^α's is artificially soft at small displacements. Each of the roughly 16 bonds that carry most of the stress is overextended by ~0.2 Å compared with a more realistic bond potential. The same effect leads to a greater final extension of the molecule in our simulations (~310 Å) than in Figure 14 (~280 Å). These disparities are removed if a larger quadratic term is used in the anharmonic potential, but the unfolding sequence that we now describe is not affected by this minor change in potential.

In both Figures 7 and 14 there are a few bond-breaking events that lead to a shoulder about half way up the first peak. In this stage the N end of the A strand (Fig. 1) pulls

away from the B strand, allowing the A strand to straighten out. This places almost all the stress across the bonds connecting A' and G. The main burst corresponds to rupture of these $A'G$ bonds, allowing the protein to partially unravel. At a separation of 50 Å, the G strand is only connected to F, forming a β-hairpin that can be seen on the righthand side of Figure 6. As in the companion article,[15] this β-hairpin unzips from the last remaining "rung" until it is completely stretched. This straightening is also the next stage of failure in the all atom simulations. Both models then show simultaneous rupture of CF and BE linkages. At the separation of 150 Å shown in Figure 6, the last remaining BE and CF bonds are at the left and right ends of the unstraightened portion of the protein, respectively. Once they rupture, the molecule unravels and reorients. As in the all-atom simulations, the last remaining bonds are DE bonds. This β-hairpin unzips at a displacement of ~250 Å in our simulations. Note that these DE bonds are not intrinsically stronger than other bonds in the system. They are just screened from stress by other bonds until the remainder of the protein has unraveled.

The data of Figure 14 were obtained from all-atom simulations at v_p = 0.5 Å/ps. The nominal value of the highest velocity in Figure 7 (dashed line) is ~0.16 Å/ps, but the effective velocity is lower because of the artificially low solvent damping. Nonetheless, there is a substantial reduction in the pulling force when the velocity is dropped by a
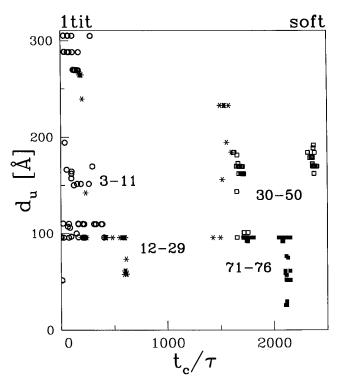
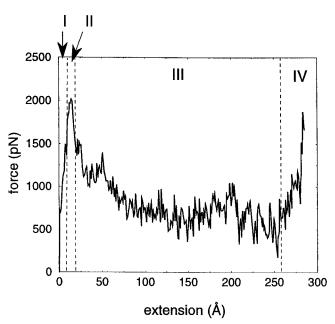Fig. 13. Similar to Figure 12 but for the case of the soft cantilever.

Fig. 14. Force-extension profile obtained in all-atom simulations[4] of the I27 domain of titin. The pulling speed is 0.5 Å/ps, and the pulling spring constant is $10k_BT/Å^2$, where T corresponds to room temperature so that k is about 414 pN/Å. The pulling force increases until the extension reaches ~14 Å, where the force peaks at ~2000 pN, and then it drops to 1500 pN within 3 Å. The extension domain is divided into four regions: preburst (I), major burst (II), postburst (III), and pulling of fully extended chain (IV). This image was made by the Theoretical Biophysics group at the Beckman Institute, University of Illinois at Urbana-Champaign, and it is available on the web at http://www.ks.uiuc.edu/Publications/Papers/abstract.cgi?tbcode=LU98. (Reprinted from Lu et al., Biophys J 1998;75:662–671, with permission from the Biophysical Society.)

factor of 100 in our simulations (solid line). In later work, Lu and Schulten[16] extended their studies down to $v_p = 0.1$ Å/ps. As in Figure 7, they found that lowering the velocity shifted the force curves downward. The peak force dropped nearly a factor of two, and they used a thermal activation model to extrapolate their results to experimental velocities, which are roughly six orders of magnitude lower. However, our results indicate that viscous dissipation may still be important at their lowest velocity. Figure 7 shows there is still substantial viscous dissipation at $v_p \sim 0.16$ Å/ps, even though the damping is artificially reduced from that in the all-atom model. As we now discuss, even bigger shifts in the force may be produced by the use of a finite drop of water about the protein in the all-atom calculation.

The CHARMM-based simulations include water to incorporate its effect on bond-strength.[4,16] However, they remove water molecules outside a spherical drop to limit the number of atoms in their simulations. This spherical drop gets stretched into a cigar shape as the simulation progresses, and part of the pulling force on the cantilever results from the surface tension of the water. A rough estimate of this force is $2\pi r\gamma_w \sim 400$ pN, where $r \sim 10$ Å is the radius of the cigar normal to the pulling direction and $\gamma_w = 0.07$ N/m is the surface tension of water. This estimate is comparable to the entire pulling force in the range of extensions between 100 and 250 Å in Figure 14, indicating that it should be carefully subtracted to determine the actual tension carried by the protein's backbone.

It is also interesting to estimate the degree to which viscous forces have relaxed in the CHARMM-based calculations. The key dimensionless ratio is the capillary number $Ca = \mu v_p/\gamma_w$ where μ is the viscosity. Typically Ca

must be much smaller than unity for most of the stress to be carried by the surface tension rather than viscous stresses. How much smaller depends on the exact geometry. For the lowest velocity used by Lu and Schulten ($v_p = 0.1$ Å/ps), one finds $Ca = 0.14$, indicating that viscous stresses are still significant. These stresses have two important consequences. One is to artificially increase the pulling force. The second is to distribute stress through the fluid rather than directly along the bonds within the protein. This could lead to a change in the order of bond breaking.

Our results also agree qualitatively with those obtained by Paci and Karplus.[18] They have performed a controlled biased force calculation, without an explicit cantilever but also CHARMM-based, and found that the large bottleneck to unfolding arises at end-to-end extensions of order 6 Å. As above, the main burst was identified with breaking of the hydrogen bonds between β-strands A' and G. The lower extension in their simulation may be due to the different ensemble used to implement the pulling force.[18] Marszalek et al.[12] have obtained a much smaller critical extension of ~2.5 Å by fitting a two-state model with worm-like-chain elasticity to experimental results. However, all the simulations described above reveal an energy landscape with extra minima before the main force peak, and the elasticity is also more complicated than that of a worm-like-chain.
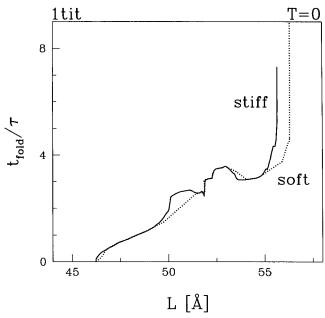
Fig. 15. Refolding times for 1tit after stretching to an end-to-end distance L. To the right of the data points shown, the protein does not return to its native state but instead misfolds. Solid lines, a stiff cantilever; dotted lines, a soft cantilever. The corresponding threshold values for the cantilever displacement are $d_{ir} = 12.6$ and 61.7 Å for the stiff and soft cantilevers respectively.

Fig. 16. Snapshot pictures of the unraveling of two 1tit domains connected in tandem. The numbers indicate the cantilever displacement.

A qualitatively similar profile has been found by Klimov and Thirumalai[20] in a model four-stranded Lennard-Jones system called S1. A quite distinct pattern, however, with a major maximum at about half of the full extension, has been determined for another four-stranded model system called S2. The difference between the two has been attributed to the presence of a robust intermediate state in the unraveling of S2. This would suggest lack of such an intermediate state in titin.

Irreversibility Length

Figure 2 suggests that the folding time is infinite at $T = 0$, i.e., when one starts with a typical open conformation it will never find its way to the native state. However, we find that when one stretches the protein slowly by less than some irreversibility length L_{ir},[15] the protein will fold back after release. Figure 15 shows the non-monotonic dependence of refolding time on the end-to-end length L of the protein at the point of release. The plots are terminated when the protein begins to misfold. This limiting length corresponds to L_{ir}, and is ~56 Å. The precise value of L_{ir} depends on the properties of the cantilever; however, the variations in the cantilever displacement at the onset of irreversibility, d_{ir}, are much larger. For the stiff and soft cases considered here the values of d_{ir} are 12.6 and 61.7 Å, respectively. As for simpler proteins in the companion article,[15] the onset of irreversibility is associated with the same set of broken contacts. From Figure 7, we see that both values of d_{ir} lie about half way up the ramp to the first large force peak. From Figure 8, we find that 20 contacts are broken at d_{ir} for both cantilever stiffnesses.
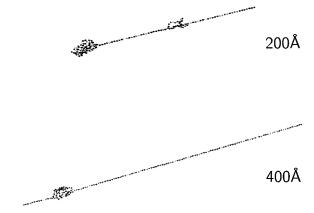
Several Domains in Tandem

Titin consists of many different domains of globular proteins. Because the structure of one of these domains is known and because the structure is the primary experimental input to Go modeling, we consider tandem structures made of repeats of 1tit. Figure 16 illustrates unraveling of two domains. It indicates that the unraveling proceeds basically in series, as in Reference 18, whereas unraveling of two α-helices has been found to proceed in parallel.[15] On closer inspection, it appears that the two domains start unraveling together, but after a small number of contacts get broken in both domains, only the forward domain continues to unfold and only after this process is completed does the backward domain engage in action. This is clearly seen in Figure 17, which shows the displacement at which contacts break versus their contact order for the case of the slowly pulled stiff cantilever. The data points are marked to differentiate between the forward and backward domains. The closed symbols from the backward domain lie almost entirely in the upper half of the d_u versus $|j-i|$ plane and are a nearly perfect repeat of the pattern formed by the open symbols representing the forward domain. Simulations with three domains show another periodic repeat of the same pattern.

The serial character of unraveling is also seen from the force-displacement curves shown in Figures 18 and 19 for two and three domains, respectively. Independent of the cantilever stiffness, there is a nearly periodic repetition of the events that take place in one domain. The reason for the serial character of the unraveling is the existence of the high force peak at the beginning of the unfolding process. Once this peak is past, a domain unfolds completely at a lower force, which is not sufficient to initiate unfolding of other domains. Only when the first domain is completely unraveled can the force rise and initiate unfolding of another domain.

Experimental data[10] also show a periodically repeated sawtooth-like pattern, but there is an overall upward trend in the curves as one unravels successive domains. The primary reason for this trend is the fact that the domains in series are not identical, and the most weakly
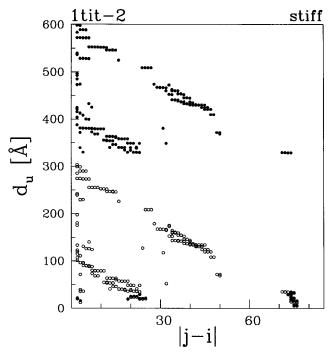
Fig. 17. Sequencing of the stretching events in the Go model of two 1tit domains as represented by d_u. The contacts of the $i,i+2$ type are not shown here. Open symbols, the forward domain; solid symbols, the backward domain.

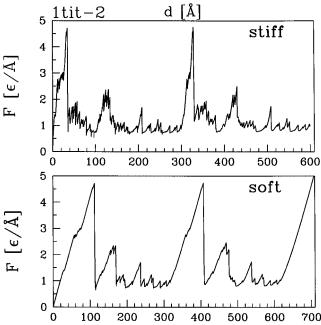Fig. 18. Force versus cantilever displacement for two 1tit domains in series.

Fig. 19. Force versus cantilever displacement for three 1tit domains in series.

bound of them all unravels first. However, some upward trend remains when the domains are made identical.[14] The existence of this upward trend has been claimed to be due to an entropic elasticity,[41] but may also reflect an increasing fluctuation of the angle of the individual domains relative to the pulling force. The serial character of
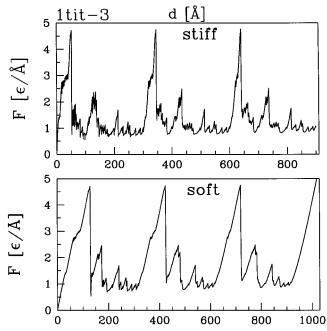
unraveling is also present in many natural adhesives[42] and has been described as analogous to the story of Sisyphus of the Greek mythology: "The case of extending a modular fibre is analogous. One needs to pull hard, and do work, but before the breaking point ("the summit") is reached, a domain unfolds or a loop opens, and the energy stored in the fiber is dissipated as heat. Then, the fiber has to be pulled on again, until the next domain breaks and so on."[42]

CONCLUSIONS

Force spectroscopy is a useful tool for obtaining information about the strength of modules in a protein and for inferring relationships between structure and function. However, inferring information about folding pathways from mechanical data turns out to be far from straightforward. The companion article[15] showed that folding of the simple secondary structures considered is uniquely related to their mechanical unraveling, but that the sense of the correlation in α-helices is opposite to that in β-hairpins. For the more complicated geometry of titin considered here, any correlation between stretching and folding appears to be restricted to the long-range contacts. These contacts tend to form last and unravel at the beginning, although how soon depends on the nature of the cantilever. It would be useful to study other proteins using similar techniques to determine possible systematics in behavior. The use of simplified models, such as the Go-like model considered here, is encouraged because we have found the mechanical results to be strikingly similar to those obtained through all atom simulations. The irreversibility length may be a useful parameter to determine in experimental studies of mechanical misfolding.

REFERENCES

1. Labeit S, Barlow DP, Gautel M, Gibson T, Holt J, Hsieh CL, Francke U, Leonard K, Wardale J, Whiting A, Trinick J. A regular pattern of 2 types of 100-residue motif in the sequence of titin. Nature 1990;345:273–276.
2. Improta S, Politou AS, Pastore A. Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. Structure 1996;15:323–327.
3. Erickson HP. Stretching single protein molecules: titin is a weird spring. Science 1997;276:1090–1090.
4. Lu H, Isralewitz B, Krammer A, Vogel V, Schulten K. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. Biophys J 1998;75:662–671.
5. Labeit S, Kolmerer B. Titins—giant proteins in charge of muscle ultrastructure and elasticity. Science 1995;270:293–296.
6. Linke WA, Ivemeyer M, Olivieri N, Kolmerer B, Ruegg JC, Labeit S. Towards a molecular understanding of the elasticity of titin. J Mol Biol 1996;261:62–71.
7. Granzier H, Helmes M, Trombitas K. Nonuniform elasticity of titin in cardiac myocytes: a study using immunoelectron microscopy and cellular mechanics. Biophys J 1996;70:430–442.
8. Tskhovrebova L, Trinick K, Sleep JA, Simmons M. Elasticity and unfolding of single molecules of the giant muscle protein titin. Nature 1997;387:308–312.
9. Kellermayer MSZ, Smith SB, Granzier HL, Bustamante C. Folding-unfolding in single titin molecules characterized with laser tweezers. Science 1997;276:1112–1116.
10. Rief M, Gautel M, Oesterhelt F, Fernandez JM, Gaub HE. Reversible unfolding of individual titin immunoglobulin domains by AFM. Science 1997;276:1109–1112.
11. Marszalek PE, Lu H, Li HB, Carrion-Vazquez M, Oberhauser AF, Schulten K, Fernandez JM. Mechanical unfolding intermediates in titin modules. Nature 1999;402:100–103.
12. Carrion-Vasquez M, Oberhauser AF, Fowler SB, Marszalek PE, Broedel SE, Clarke J, Fernandez JM. Mechanical and chemical unfolding of a single protein: a comparison. Proc Natl Acad Sci USA 1999;96:3694–3699.
13. Oberhauser AF, Hansma PK, Carrion-Vazquez M, Fernandez JM. Stepwise unfolding of titin under force-clamp atomic force microscopy. Proc Natl Acad Sci USA 2001;98:468–472.
14. Oberhauser AF, Marszalek PE, Carrion-Vazquez M, Fernandez JM. Single protein misfolding events captured by atomic force microscopy. Nat Struct Biol 1999;6:1025–1028.
15. Cieplak M, Hoang TX, Robbins MO. Sequencing of events in mechanical stretching of secondary structures of proteins and the folding pathways. Submitted for publication.
16. Lu H, Schulten K. Steered molecular dynamics simulation of conformational changes of immunoglobulin domain I27 interprete atomic force microscopy observations Chem Phys 1999;247:141–153.
17. Paci E, Karplus M. Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. J Mol Biol 1999;288:441–459.
18. Paci E, Karplus M. Unfolding proteins by external forces and temperature: the importance of topology and energetics. Proc Natl Acad Sci USA 2000;97:6521–6526.
19. Klimov DK, Thirumalai D. Stretching single-domain proteins: phase diagram and kinetics of force-induced unfolding. Proc Natl Acad Sci USA 1999;96:6166–6170.
20. Klimov DK, Thirumalai D. Native topology determines force-induced unfolding pathways in globular proteins. Proc Natl Acad Sci USA 2000;97:7254–7259.
21. Veitshans T, Klimov D, Thirumalai D. Protein folding kinetics: time scales, pathways and energy landscapes in terms of sequence-dependent properties. Folding Des 1997;2:1–22.
22. Abe H, Go N. Noninteracting local-structure model of folding and unfolding transition in globular proteins. II. Application to two-dimensional lattice proteins. Biopolymers 1981;20:1013–1031.
23. Takada S. Go-ing for the prediction of protein folding mechanism. Proc Natl Acad Sci USA 1999;96:11698–11700.
24. Micheletti C, Banavar JR, Maritan A, Seno F. Protein structures and optimal folding from a geometrical variational principle. Phys Rev Lett 1999;82:3372–3375.
25. Maritan A, Micheletti C, Banavar JR. Role of secondary motifs in fast folding polymers: A dynamical variational principle. Phys Rev Lett 2000;84:3009–3012.
26. Hoang TX, Cieplak M. Molecular dynamics of folding of secondary structures in Go-like models of proteins. J Chem Phys 2000;112:6851–6862.
27. Hoang TX, Cieplak M. Sequencing of folding events in Go-like proteins. J Chem Phys 2001;113:8319–8328.
28. Cieplak M, Hoang TX. Kinetics non-optimality and vibrational stability of proteins. Proteins Struct Funct Genet 2001;44:20–25.
29. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
30. Grest GS, Kremer K. Molecular dynamics simulation for polymers in the presence of a heat bath. Phys Rev A 1986;33:3628–3631.
31. Klimov DK, Thirumalai, D. Viscosity dependence of the folding rates of proteins. Phys Rev Lett 1997;79:317–320.
32. Camacho CJ, Thirumalai D. Kinetics and thermodynamics of folding in model proteins. Proc Natl Acad Sci USA 1993; 90:6369–6372.
33. Klimov DK, Thirumalai D. A criterion that determines the foldability of proteins. Phys Rev Lett 1996;76:4070–4073.
34. Unger R, Moult J. Local interactions dominate folding in a simple protein model. J Mol Biol 1996;259:988–994.
35. Chan HS. Matching speed and locality. Nature 1998;392:761–763.
36. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol 1998;277:985–994.
37. Plaxco KW, Simons KT, Ruczinski I, Baker D. Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. Biochemistry 2000;39:11177–11183.
38. Socci ND, Onuchic JN, Wolynes PG. Stretching lattice models of protein folding. Proc Natl Acad Sci USA 1999;96:2031–2035.
39. Soteriou A, Clarke A, Martin S, Trinick J. Titin folding energy and elasticity. Proc R Soc London Ser B 1993;254:83–86.
40. Erickson HP. Reversible unfolding of fibronectin type-III and immunoglobulin domains provides the structural basis for stretch and elasticity of titin and fibronectin. Proc Natl Acad Sci USA 1994;91:10114–10118.
41. Fisher TE, Oberhauser AF, Carrion-Vaquez M, Marszalek PE, Fernandez JM. The study of protein mechanics with atomic force microscopy. Trends Biochem Sci 1999; 24:379–384.
42. Smith BL, Schaeffer TE, Viani M, Thompson JB, Frederic NA, Kindt J, Belcher A, Stucky GD, Morse DE, Hansma PK. Molecular mechanistic origin of the toughness of natural adhesives, fibres and compsites. Nature 1999;399:761–763.