PROTEINS: Structure, Function, and Genetics 35:453–463 (1999)

Steered Molecular Dynamics Simulations of Force-Induced Protein Domain Unfolding

Hui Lu[1] **and Klaus Schulten**[2]*
[1]*Beckman Institute and Department of Nuclear Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois*
[2]*Beckman Institute and Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois*

ABSTRACT **Steered molecular dynamics (SMD), a computer simulation method for studying force-induced reactions in biopolymers, has been applied to investigate the response of protein domains to stretching apart of their terminal ends. The simulations mimic atomic force microscopy and optical tweezer experiments, but proceed on much shorter time scales. The simulations on different domains for 0.6 nanosecond each reveal two types of protein responses: the first type, arising in certain β-sandwich domains, exhibits nanosecond unfolding only after a force above 1,500 pN is applied; the second type, arising in a wider class of protein domain structures, requires significantly weaker forces for nanosecond unfolding. In the first case, strong forces are needed to concertedly break a set of interstrand hydrogen bonds which protect the domains against unfolding through stretching; in the second case, stretching breaks backbone hydrogen bonds one by one, and does not require strong forces for this purpose. Stretching of β-sandwich (immunoglobulin) domains has been investigated further revealing a specific relationship between response to mechanical strain and the architecture of β-sandwich domains. Proteins 1999;35:453–463.**

Key words: protein folding; steered molecular dynamics; domain classification; immunoglobulin domain; fibronectin type III domain

INTRODUCTION

Proteins need to fold efficiently into their native conformations to carry out their functions but also to avoid aggregation of metastable folding intermediates.[1,2] In vivo, proteins need to protect themselves against unfolding for suitably long periods of time to maintain a functional state. Under nonequilibrium conditions the free energy barrier that separates a protein in its native state against unfolding needs to be sufficiently large under physiological conditions; under equilibrium conditions it is the free energy difference between the folded and unfolded state that keeps the protein in its native structure.

Traditionally, experiments have focused on the folding and unfolding behavior of proteins in response to temperature changes and to chemical perturbations through pH changes or denaturants.[3] Correspondingly, theoretical investigations have been centered around the effects of such perturbations.[4] Several molecular dynamics (MD) simulations have studied protein unfolding and have provided atomic level detail of the process. It has been assumed that simulations of unfolding capture the essential aspects of the folding process.[5,6] So far, the simulations carried out could not reproduce well the experimental conditions needed to achieve unfolding due to the short time scales covered in simulations; for example, the temperature in simulations had to be elevated to extremely high values to induce unfolding.[7,8] Despite the highly unphysiological simulation conditions adopted, MD has provided insights into protein folding behavior.[9]

Another factor that can lead to unfolding of proteins is mechanical stress, in particular, stretching. Such stretching arises in muscle, in the extracellular matrix and for cell receptors. The representative proteins usually contain multiple domains linked together in a linear sequence. The response of these proteins to mechanical stress can be studied by investigating the individual domain's response to stretching. Domains arising in the representative systems have evolved to withstand certain levels of mechanical force[10–12] as well as to sense the occurrence of mechanical strain.[13] Examples are domains of titin, a giant protein that provides muscle elasticity[14,15] and controls chromosome shape;[16] the elasticity of titin is due to its 300 immunoglobulin-like (Ig) and fibronectin type III-like (FnIII) domains. Another example is provided by FnIII domains of the extracellular matrix protein tenascin, which is responsible for cell adhesion and control of cell migration with its elasticity originating from its connected FnIII domains.[10]

Atomic force microscopy (AFM) experiments permit one to exert forces to stretch individual proteins and to measure the extensions of the proteins. These experiments demonstrated that Ig/FnIII domains behave like elements of a linearly jointed entropic spring and exhibit sawtooth patterns of measured force-extension profiles.[10–12] The number of force peaks arising in these profiles is equal to the number of Ig/FnIII domains involved in the stretched

Grant sponsor: National Institutes of Health; Grant number: NIH PHS 5 P41 RR05969; Grant sponsor: National Science Foundation; Grant numbers: NSFBIR 94-23827EQ, NSF/GCAG BIR 93-18159, and MCA 93S028; Grant sponsor: Roy J. Carver Charitable Trust.
*Correspondence to: Klaus Schulten, Beckman Institute, University of Illinois at Urbana-Champaign, 405 N. Mathews Avenue, Urbana, Illinois 61801. E-mail: kschulte@ks.uiuc.edu
Received 19 October 1998; Accepted 19 February 1999

protein, the distance between the force peaks corresponding to the length of the fully extended individual Ig/FnIII domains. From the data one can conclude that connected Ig/FnIII domains unfold one by one under the external stretching forces. Optical tweezer experiments provide another means of stretching single proteins through mechanical forces.[17,18] These experiments also demonstrate a stepwise pattern in force-extension profile with each step exhibiting an extension equal to the length of a fully extended individual Ig/FnIII domain. To achieve such one-by-one unfolding, every protein domain needs to provide a resistance to a particular maximum force.

It is noteworthy that Ig and FnIII constitute β sandwich domains with N-terminal and C-terminal strands parallel to each other, but pointing in opposite directions. In the natural function of these domains as well as in AFM and optical tweezer experiments, the forces arising stretch apart the termini. One would like to understand how the architecture of β sandwich domains has served nature to design proteins which can withstand such forces and act as stress sensors. One would also like to resolve at the atomic level the mechanical stretching and unfolding of protein domains which is still out of reach of experiment itself. Steered molecular dynamics (SMD)[19–21] simulations promise to provide answers in this respect.

SMD has been developed to study force-induced biological processes and has already been applied to various ligand unbinding events[21–26] as well as to the stretching of titin's Ig domains.[27] In the latter case SMD simulations have revealed that the maximum force needed to stretch and unfold Ig domains is due to the need to break in a concerted fashion a large number (6–8) of interstrand backbone hydrogen bonds.

In the present study we will investigate the stretching and unfolding of further domains which have been investigated by AFM and other protein domains which have not yet been investigated experimentally. In particular, we want to understand if Ig and FnIII domains show a response to stretching that is qualitatively different from that of domains which do not have to sustain mechanical strain in their function. A comparison of several protein domains may shed light on the design principles of mechanical proteins.

The protein domains investigated in the present study are the first two N-terminal β sandwich domains (cad1 and cad2) of the cell adhesion protein cadherin,[28] the first two N-terminal β sandwich domains (vcam1 and vcam2) of the cell adhesion protein V-CAM,[29] as well as the FnIII domain number 7 (fn7) and number 10 (fn10) of the cell structural protein fibronectin.[30] These are all β sandwich domains with N-terminal and C-terminal strands parallel to each other, but pointing in opposite directions. For comparison, several other protein domains with similar numbers of amino acid residues are included in our study, namely the electron transport protein cytochrome C6 (cc6), a five-helix protein;[31] immunoglobulin binding protein (igb), a three-helix protein;[32] and the C2 domain of synaptotagmin I (c2), an eight-strand β sandwich domain with the two terminal strands parallel to each other and pointing in the same direction.[33]

TABLE I. Characterization of SMD Simulations

System	Type	pdb ID	Force peak (pN)	Peak position (Å)	Peak range (Å–Å)	Class
cad1	β	1edh	1,850	16	8–19	Ia
cad2	β	1edh	1,970	24	16–27	Ia
vcam1	β	1vsc	2,050	10	7–17	Ia
I27[a]	β	1tit	2,040	14	9–17	Ia
vcam2	β	1vsc	1,620	19	15–25	Ib
fn7	β	1fnf	1,630	24	10–29	Ib
fn10	β	1fnf	1,580	17	11–25	Ib
cc6	α	1cyi	no peak	—	—	II
igb	α	1bdd	no peak	—	—	II
c2	β	1rsy	no peak	—	—	II

[a]The data on I27 are from Lu et al.[27]

METHODS

The initial structures of the protein domains studied in this paper were obtained from the Brookhaven protein data bank[34] with pdb IDs listed in Table I. These structures were then solvated, heated up to 300 K and equilibrated through a procedure similar to that described in Lu et al.[27] In the SMD simulations performed, all atoms were modeled explicitly. The total number of atoms in each system varies from 9,300 (α helix domain igb) to 13,700 (β sandwich domain cad2). The MD simulations and data analysis were carried out with the programs NAMD[35] and XPLOR[36] employing the CHARMM22 force field.[37] The simulations were performed with a timestep of 1 fs, a uniform dielectric constant of 1, and a cut-off of Coulomb forces with a switching function starting at a distance of 10 Å and reaching zero at 13 Å.

SMD simulations of constant velocity stretching (SMD-CV protocol) were carried out by fixing one terminus of the domain, and applying external forces to the other terminus. The forces were applied by restraining the pulled end harmonically to a restraint point and moving the restraint point with constant velocity v in the desired direction. The procedure is equivalent to attaching one end of a harmonic spring to the end of a domain and pulling on the other end of the spring, and is similar to the procedure performed on Ig and FnIII in AFM experiments,[10,12] except that the pulling speed adopted in the simulations is six orders of magnitude higher than those in the experiments. The force experienced by the pulled terminal residue is

$$\mathbf{F} = k(\mathbf{v}t - \mathbf{x}). \quad (1)$$

Here $\mathbf{x}$ is the displacement of the pulled atom from its original position, $\mathbf{v}$ is the preselected pulling velocity, and k is the spring constant. Note that in the SMD-CV protocol it is the restraining point, rather than the end of the domain, that is moved at constant velocity. The force calculated from Eq. (1) is a good measure of how difficult it

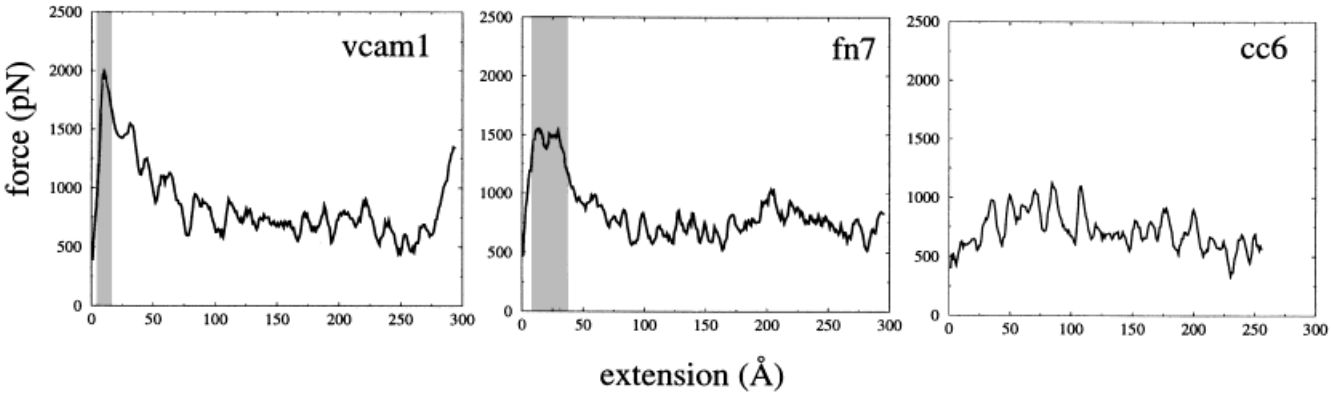

Fig. 1. Force-extension profiles resulting from 600-ps SMD simulations of stretch-induced unfolding of three representative proteins: V-CAM N-terminus domain (vcam1), fibronectin FnIII number 7 domain (fn7), and cytochrome C6 domain (cc6). The force peak regions of vcam1 and fn7 are highlighted (c.f., Fig. 3).

is to cause the domain to extend by "following" the restraining point.

The simulations presented below were carried out with a spring constant of 414 pN/Å and pulling velocity of 0.5 Å/ps, unless specified otherwise. To realize a movement of the restraining point with nearly constant velocity, the position of the restraining point was changed every 100 fs by $v\Delta t$, e.g., 0.05 Å. The 0.05 Å stepwise movement of the restraining point is much smaller than the fluctuation of typical protein atoms in a 300 K MD simulation, 1 Å, so this implementation should not cause an artifact. The value of k [see Eq. (1)] chosen here, 10 $k_BT/Å^2$, corresponds to a spatial (thermal) fluctuation of the constrained C_α atom of $\delta x = \sqrt{k_BT/k} = 0.32$ Å at T = 300 K. The elongation d(t), defined as the increase of the end-to-end distance from that of the native fold, was monitored along with the force F(t). In some cases other properties, e.g., surface area s(t) or bond distances, were also monitored. For the analysis presented below, often the time t was eliminated and force, area, or other properties plotted as a function of extension d such that plots show (F(t),d(t)) or (s(t),d(t)) graphs. The graph (F(t),d(t)) will be referred to as the force-extension profile.

SMD simulations of constant force stretching (SMD-CF protocol) were implemented by fixing one terminus of the domain and applying a constant force to the other terminus. The titin Ig domain I27,[38] which was studied by AFM and SMD previously, was investigated by this protocol.

RESULTS

The protein domains chosen in this study all have similar numbers of amino acids. Applying the SMD-CV protocol as described above, each domain was found to elongate gradually with time t until the polypeptide chain was completely straightened and assumed an extension of about 300 Å at around 0.6 ns when the simulations stopped. The force-extension profiles showed dramatically different features at short extension (within the first ⅛ of the full extension of 300 Å) for the different domains investigated, whereas at longer extension those profiles were similar, the forces fluctuating around nearly identical average values of 750 pN. At extensions near 300 Å, the force increased linearly again for all domains due to bond angle widening and bond length stretching of the completely unfolded polypeptide strands.

Table I lists simulation results of the systems studied. The force-extension profiles fall into two classes: class I, with a dominant force peak (1,500 pN to 2,000 pN) at short extension and weaker forces of average value around 750 pN at long extension; class II, without any dominant force peak, but rather with a gradual increase of the force during stretching to a typical average value of 750 pN at large extensions. The domains cad1, cad2, vcam1, vcam2, fn7, and fn10 belong to class I, the domains cc6, igb, and c2 belong to class II. The titin immunoglobulin-like (Ig) domain I27 previously studied by SMD[27] belongs to class I.

Figure 1 shows the force-extension profiles of three representative domains: vcam1, fn7, and cc6. Vcam1 has a force peak of about 2,000 pN at an extension of 11 Å and fn7 has a force peak of about 1,500 pN at an extension of 23 Å. The force peaks correspond to main energy barriers which separate the folded and the unfolded states of the domain. After the barrier is overcome, the forces for both domains decrease quickly to average values around 750 pN until the domains are completely unfolded. For domain cc6, the force increases quickly to the 750 pN average value without any major force peak arising.

The SMD trajectories provide a detailed view of the conformational transformations during the initial stage of domain stretching. Figure 2 presents three snapshots of the structures of each of the domains, i.e., vcam1, fn7, and cc6. In case of vcam1, the N-terminal β strand A/A′ and the C-terminal β strand G are seen to pass by each other at the extension of the main force peak (9 to 17 Å), during which the force strains and eventually breaks six backbone hydrogen bonds connecting the strands A/A′ and G. In case of fn7, at the extensions (10 to 29 Å) corresponding to the main force peak, the N-terminal β strand A passes by its

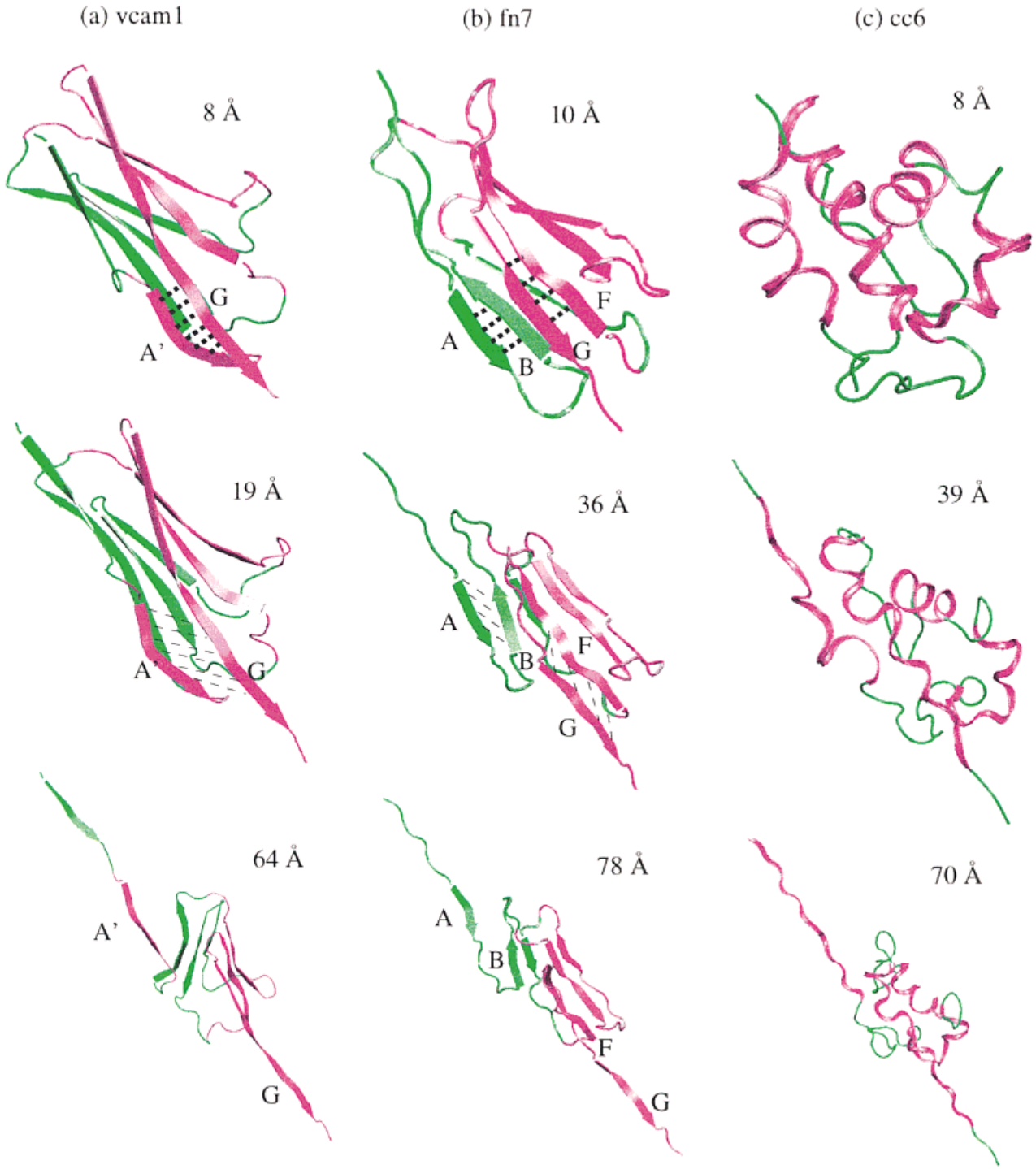

Fig. 2. Snapshots from SMD trajectories of (**a**) vcam1, (**b**) fn7, (**c**) cc6. In each SMD simulation the N terminus of the domains (on the left hand side of the structures as shown) had been held fixed and the force had been applied to the C terminus (on the right hand side of the structures as shown). The extension value for each snapshot is provided. For vcam1 and fn7, the two β sheets of the respective barrel are distinguished through purple and green colors; the β strands involved in the initial unfolding step (see text) are marked with letters and the crucial inter-strand hydrogen bonds are presented as thick dotted lines as long as the bonds exist and as thin dashed lines when the bonds are broken. For cc6, the residues belonging to helices in the folded structure are colored purple, the residues belonging to loops are colored green. (This figure was created with VMD.[46])

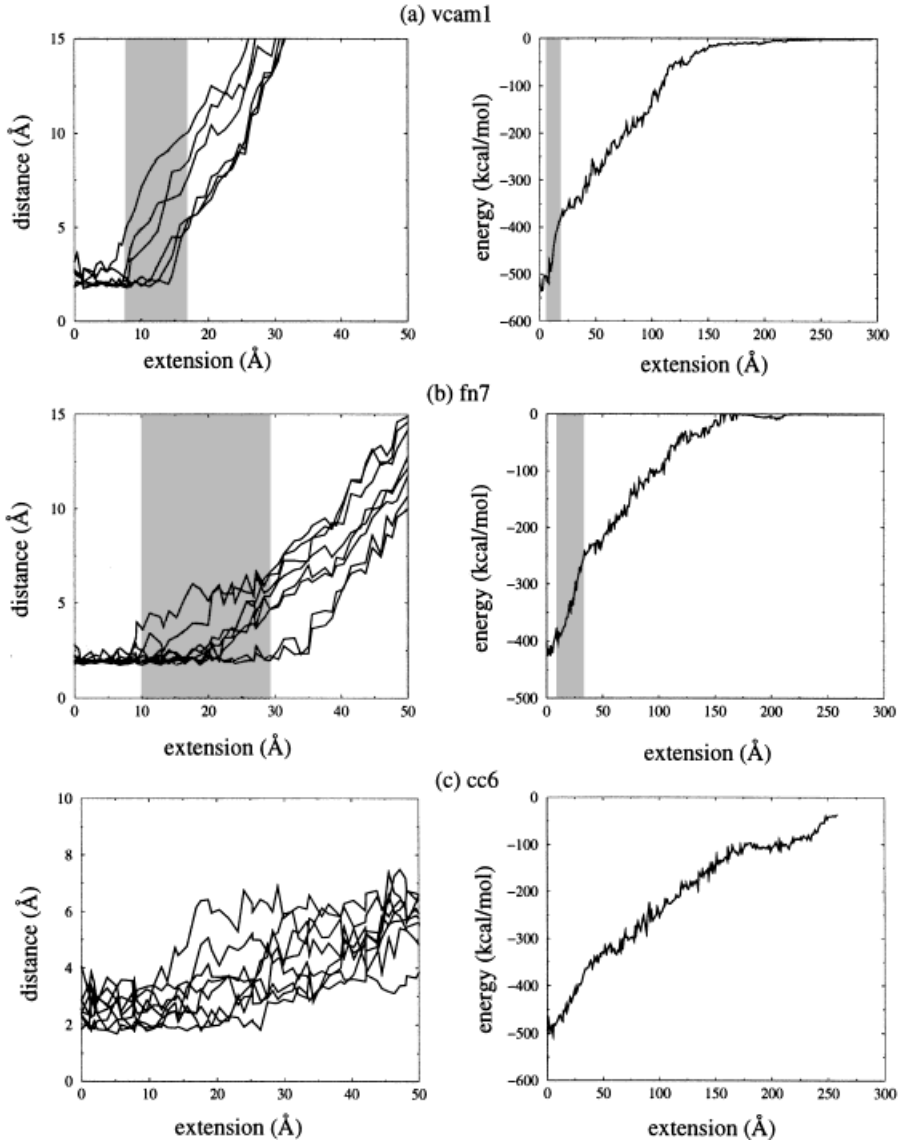

Fig. 3. Distances of individual interstrand hydrogen bonds (left column) and sum of interaction energies of atoms involved in interstrand hydrogen bonds (right column) vs. extension for (**a**) vcam1, (**b**) fn7, (**c**) cc6. The burst regions of vcam1 and fn7 are highlighted.

neighboring strand B at the same time as the C-terminal strand G passes by its neighboring strand F. During the force peak region, the high value of force is needed to break the backbone hydrogen bonds between strands A and B and between strands F and G. In case of the class II domain cc6, the α helix near the N-terminus and the α helix near the C-terminus are both elongated and have moved away from each other; neither interstrand backbone hydrogen bonds nor salt bridges exist in this helical protein such that no large force is required for the separation of the termini.

An analysis of backbone hydrogen bonds along the unfolding trajectories can explain the difference of the recorded force-extension profiles between class I and class II domains. In case of the β sandwich domains vcam1 and fn7, two terminal strands are hydrogen-bonded to at least one other strand. To extend the domain, the external force needs to break these interstrand hydrogen bonds. Figure 3 compares the distances of individual hydrogen bonding participants along the extension for vcam1, fn7, and cc6, actually showing only the hydrogen bonding pairs that have been broken during the first 50 Å extension. The figure presents also the sum of the interaction energies of atoms participating in hydrogen bonds, including in this case contributions of all backbone hydrogen bonding pairs. One can discern that for vcam1 and fn7 hydrogen bonds are being broken in clusters at short extension; this clustered hydrogen bond breaking causes the major force peak in the force-extension profile of these domains as shown in Figure 1. After the mentioned hydrogen bonds are broken concertedly, the forces needed to induce further extension drop quickly and at longer extension the backbone hydrogen bonds in the two domains are broken one-by-one (data not shown). Figure 3 also shows that for cc6 clustered hydrogen bond breaking does not occur, neither at small nor at large extension. As a result, no

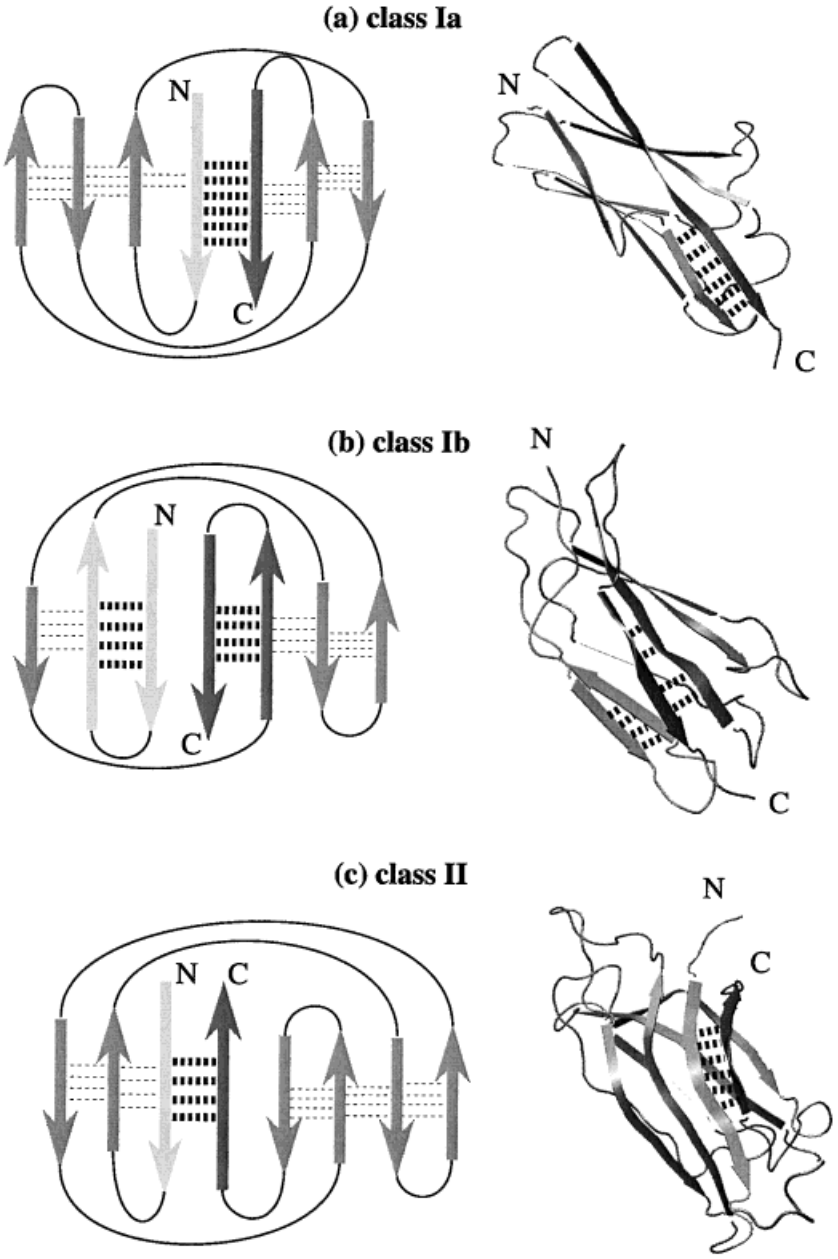

Fig. 4. Classification of protein domains. The schematic presentation on the left depicts the topology of interstrand hydrogen bonding networks of (**a**) vcam1, (**b**) fn7, and (**c**) cc6. On the right are shown the corresponding protein domains rendered with VMD[46] with the essential interstrand hydrogen bonds highlighted as in the schematic representations.

major force peak develops in the force-extension profile of cc6 shown in Figure 1. This is corroborated by the behavior of the hydrogen bond energy shown in Figure 3; for both vcam1 and fn7 this energy exhibits a sharp increase at short extension while for cc6 it exhibits a more gradual increase.

The magnitude of the force needed to stretch and unfold a protein domain depends sensitively on the architecture of the protein. This is shown for vcam1, fn7, and a third β sandwich protein, c2, in Figure 4 which presents the native folds of the domains together with schematic drawings of the respective strand topologies. In case of vcam1, a set of six hydrogen bonds connect directly the N-terminal and C-terminal strands that are pulled apart diametrically by the stretching force (c.f., Fig. 2). This arrangement protects the protein best against unfolding since the simultaneous breaking of the full set of hydrogen bonds is required to significantly stretch the domain. Other domains with equivalent β strand topologies are cad1, cad2, and I27 (c.f., Table I). We define for these proteins a subclass of class I, termed class Ia. In case of fn7, the N- and C-terminal strands are not directly hydrogen-bonded to each other, but rather to non-terminal strands. This permits a greater degree of deformation of the domain under stretching and a somewhat more gradual breaking of hydrogen bonds to initiate unfolding. This is evident from a comparison of the plots of hydrogen bond distances and energies of vcam1 and fn7 presented in Figure 3. One can discern that vcam1 ruptures in a narrower extension window than fn7 does. Figure 1 indicates that weaker

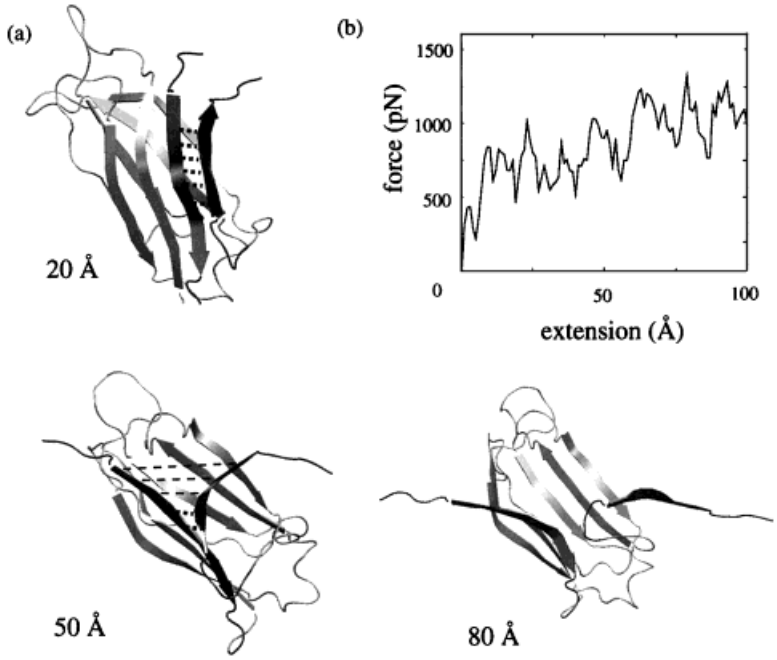

Fig. 5. Sketching of the C2 domain of synaptotagmin I. (**a**) Top left, bottom left, and bottom right are snapshots during the unfolding of the C2 domain at extensions 20 Å, 50 Å, and 80 Å. The H-bonds between the N- and C-terminal strands are highlighted (thick dotted lines for existing bonds and thin dashed lines for broken bonds). (**b**) Top right is the force-extension profile resulting from an SMD simulation of c2. (This figure was created with VMD.[46])

forces (1,600 pN) are needed to rupture fn7 as compared to vcam1 (2,000 pN). Other proteins with the same β strand topology as fn7 and similar rupture behavior are vcam2 and fn10 (c.f., Table I). We define for these proteins another subclass of class I, termed class Ib.

Proteins in classes Ia and Ib are characterized both by a β sandwich topology in which the N-termini and C-termini point in opposite directions. In the case of the β sandwich domain c2 presented in Figure 4 the N and C termini point in the same direction. The structure of this domain along with the force-extension profile and snapshots during the unfolding process are presented in Figure 5. C2 has hydrogen bonds between its terminal β strands. Nevertheless, the forces needed to unfold the domain are relatively small and do not exhibit a major peak at short extension. The unfolding snapshots in Figure 5 demonstrate that the hydrogen bonds between the N- and C-terminal strands break one-by-one during the unfolding and, therefore, do not provide much resistance against stretching. Accordingly, c2 belongs to class II.

In order to test whether the rupture and unfolding of stretched proteins is indeed not governed by influences other than hydrogen bonding we have monitored during the SMD-CV simulations the hydrophobic residue exposure area, a property often implicated as an order parameter for protein folding and a measure of hydrophobic effects. Figure 6 displays the resulting exposure areas versus extension for vcam1, fn7, and cc6 which cover both class I and class II domains. This property does not show distinct features at short extension for either of the three cases. Hence, hydrophobic interaction does not seem to play a dominant role for the magnitude of stretching forces required for domain unfolding.

The stretching forces required to unfold immunoglobulin and fibronectin type III domains by means of atomic

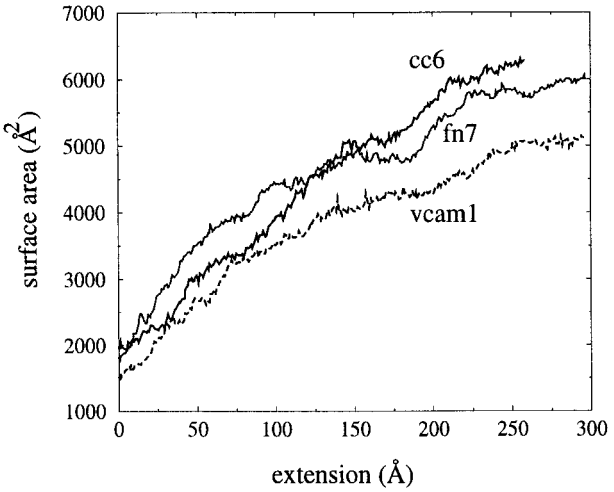

Fig. 6. Areas of solvent exposed hydrophobic residues vs. extension for vcam1, fn7, and cc6.

force microscopy[10,12] are considerably weaker than those required in the respective SMD simulations. The origin of this discrepancy is thought to be the stretching velocity difference of several orders of magnitude between experiments and simulations. In order to test this supposition we have carried out SMD simulations using stretching protocols involving the applications of constant forces to titin I27 domains. Under the SMD-CV protocol adopted in the previous simulations, peak forces of about 2,000 pN were recorded during domain unfolding.[27] In the SMD-CF simulations constant forces of 1,000 and 750 pN were applied to stretching the domain for as long as it took the domain to unfold. In these simulations the domain stretched initially and then halted its extension at values of about 10 Å as

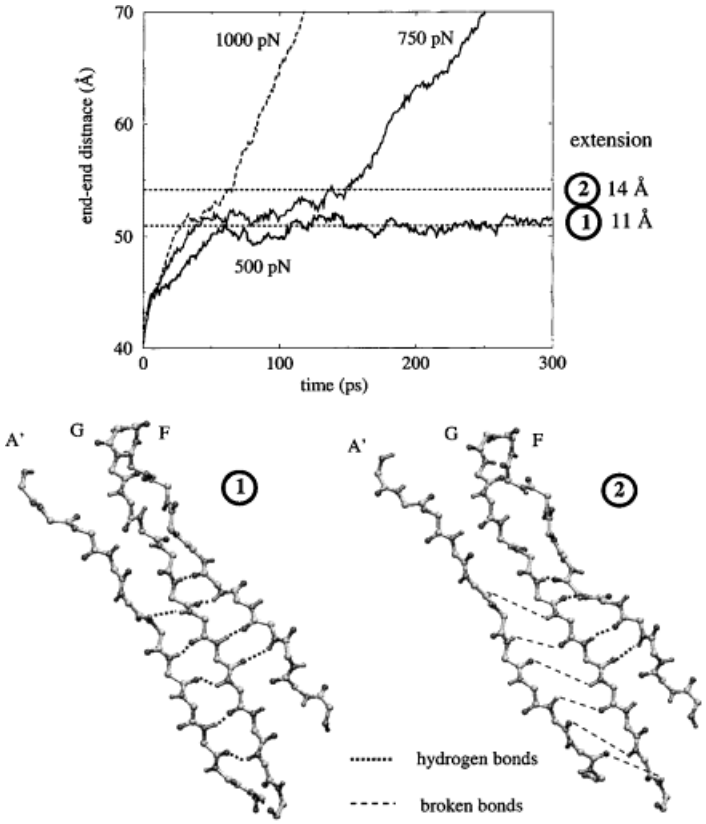

Fig. 7. Time development of the end-end distance resulting from 300-ps SMD simulations of the I27 domain of titin under the influence of constant forces of 1000 pN, 750 pN, and 500 pN. The figure shows the occurrence of an extension plateau. Two typical domain conformations of I27, one at the beginning of the plateau and one at the end of the plateau, are presented through two β strands (A′, G, and F) and the interstrand hydrogen bonds. One can recognize that at the beginning of the plateau all hydrogen bonds are still maintained, but that at the end of the plateau the bonds between strands A′ and G are broken. See also Lu and Schulten.[42] (This figure was created with VMD.[46])

demonstrated by the plateaus in the extension-time profiles of Figure 7. Inspection of the corresponding structures revealed at this point that the residues near the termini are straightened, but the backbone hydrogen bonds between A′ and G are well maintained. The breaking of these six interstrand hydrogen bonds took longer when a weaker force was applied as seen in Figure 7: 30 ps for the 1,000 pN force and 100 ps for the 750 pN force. In all cases a distinct rupture event, involving the breaking of the six hydrogen bonds between strands A and G, preceded the unfolding discernible in Figure 7 through the linear increase of the distance between the domain's termini. This characteristic rapture event has been demonstrated to arise on the pathway of stretched unfolding for both SMD-CV and SMD-CF protocols.

A weaker force, 500 pN, has also been applied to stretch I27. The results presented in Figure 7 show that the extension of the domain reaches 10 Å within 70 ps and remains at that extension for the duration of the 300 ps simulation. The continuation of this simulation, in which the domain was found to unfold at around 950 ps, and simulations with forces as low as 50 pN are presented in Lu and Schulten.[42]

DISCUSSION

Many proteins experience stretching forces in vivo and, hence, it is desirable to understand the responses of protein domains to mechanical strain, as investigated in this paper for 10 cases. Our SMD-CV simulations suggest a classification of proteins based on the domains' response to stretching forces and rationalized by a correlation of the behavior to domain architectures. Class I protein domains exhibit a potential barrier at short extension, termed in the following protection barrier, that furnishes resistance against unfolding. Stretched at a given velocity, the domain remains folded when forces applied to the termini are not strong enough; when forces are larger than a threshold, the protection barrier is overcome and the domain unravels easily.

Typical for the stochastic nature of barrier crossing events, the threshold force (in SMD the peak force observed in the force-extension curve) exhibits a distribution of values. The distributions are narrow as predicted in Izrailev et al.[22] and observed in AFM experiments.[39] For the sake of simplicity, the name threshold force is still used, but refers to the most probable unfolding force.

The protection barrier defined above is due to interstrand hydrogen bonds which need to be broken concurrently before unfolding can occur. Hydrogen bonds near the terminals of class I domains have also been shown to play an important role in chemically induced and temperature-induced unfolding experiments.[40] Class I proteins investigated here are β-sandwich proteins with termini pointing in opposite directions. In the case that the terminal strands are directly hydrogen-bonded to each other (class Ia) the protection is stronger than otherwise (class Ib). Class II protein domains do not have a strong potential

barrier against stretching and the corresponding domains unfold with little resistance.

The dominant force peak resulting in SMD-CV simulations for class I domains is consistent with AFM observations on titin Ig[12] and tenascin FnIII.[10] The SMD results explain the one-by-one unfolding of domains reflected in the sawtooth force-extension profiles of AFM experiments. These experiments do not have a spatial resolution that permits direct comparison to SMD simulations: when the protection barrier of one domain is overcome by the applied force, the domain unfolds easily and contributes about 250 to 280 Å length to the multi-domain's total extension; during that time, other folded domains of the same protein are experiencing a force much lower than the threshold value needed to overcome their own protection barrier; after the unfolded domain is completely stretched, the force increases again and induces similarly the unfolding of another domain. The distance between the consecutive force peaks is the length contributed by the previous domain, i.e., 250 to 280 Å.

The current study suggests that AFM experiments on proteins that consist of several class II domains should show very different force-extension profiles, either without sawtooth pattern or with much lower force peak values than those observed for class I domains. The second possibility, namely a discernible, but low force peak value, may arise from the hydrophobic effect which does not play an important role in force-induced unfolding of class I domains. Both possibilities, i.e., no discernible and discernible, but low, peaks have already been confirmed by recent AFM experiments (J. Fernandez, personal communication).

AFM experiments have not been able to resolve the exact extension at which the initial force peak arises. Estimates of this extension range from 2 to 5 Å for titin Ig domains,[39,41] values which differ from SMD simulation results (14 Å) for single domain stretching reported in Lu et al.[27] and discernible in Figure 1 for vcam1. However, AFM experiments are performed on multi-domain proteins and an unknown extension precedes the occurrence of the first force peak in the sawtooth force-extension curve. Most likely this pre-extension confers a force on each domain, pre-stretching the latter by an unknown amount.

The constant force SMD simulations (SMD-CF) reported here show a domain extension response which seems to explain the mentioned discrepancy. In these simulations (c.f., Fig. 7) the domains stretch linearly in time to reach a plateau of 10 Å at which the extension remains constant, except for an extension fluctuation of 2 to 3 Å; after a certain time, the domains resume a linear extension corresponding to initiation of complete unfolding. This behavior appears to be consistent with the analysis of AFM data if one assumes that the data do not resolve the initial linear stretching. The behavior is analyzed in detail in Lu and Schulten.[42]

The difference between class Ia and class Ib domain unfolding revealed by SMD simulation, namely, smaller rupture forces for class Ib as compared to class Ia, is consistent with three AFM experiments.[10–12] The AFM force peaks recorded for titin I27 (class Ia) unfolding are 40% higher than those recorded for tenascin FnIII (class Ib) domain unfolding. The protection barrier against unfolding in class I domains originates from the necessity of breaking several interstrand backbone hydrogen bonds simultaneously. From the analysis of structures and SMD trajectories one can discern two possibilities for such protection: in the case of titin Ig, vcam1, cad1, and cad2 (Fig. 2) there exist backbone hydrogen bonds between N-terminal strand A or A′ and C-terminal strand G; in the case of FnIII and vcam2 (Fig. 2) such hydrogen bonds connect the N-terminal strand A with its neighboring strand B, and the C-terminal strand G with its neighboring strand F. In the former case (class Ia) the breaking of interstrand hydrogen bonds occurs over a very small extension range, whereas the second case (class Ib) renders the domain more flexible under stretching such that the breaking of the interstrand hydrogen bonds occurs over a wider extension range and, hence, with a reduced rupture force as seen in Figures 1 and 3.

The barrier against unfolding protects class I domains in their native environment where multi-domain proteins experience stretching forces. For example, the extracellular part of cell adhesion proteins cadherin and V-CAM are all composed of class I domains. When adhesion proteins from different cells stick together the proteins need to provide stability against the tendency of the two cells to move away from each other. The extracellular domains of the adhesion proteins need to remain properly folded, but their ability to stretch also provides the system with a necessary degree of flexibility. Another adhesion protein in the extracellular matrix is fibronectin which connects to cell receptors (integrin) and to collagen.[43] One key domain of this protein, fn10, appears to serve two functions, that of detecting and that of sustaining mechanical strain as proposed in Krammer et al.[13] A third example is the muscle protein titin which provides muscle fibers in case of overstretching with an extreme degree of elasticity through its class I immunoglobulin and fibronectin type III domains.

One concern regarding the SMD-CV simulations of class I domains reported here is that the peak force values obtained are much higher than observed in AFM experiments, which is most likely due to the time scale gap.[22] Unfortunately, available computer power limits one presently to μs simulations at best, but at the cost of continuously using a massively parallel computer over several months.[44] The expected simple scaling down of the peak force values without a change in the unfolding scenario does not warrant such effort.[42] A similar observation, that the ligand unbinding scenario does not change when the SMD pulling speed is slowed down has been shown in Izrailev et al.[22] A more suitable way to resolve this discrepancy between SMD and AFM is to reconstruct the potential of mean force along the stretching coordinate using time series analysis methods as suggested in Gullingsrud et al.[45] The resulting potential can be used in a first passage time calculation to determine the rupture force

and its distribution as explained in Izrailev et al.[22] Indeed, measurement of the rupture force distribution[39] agrees with the predicted distribution.[22] In Lu and Schulten[42] the duration of the plateau in extension-time plot in constant force domain stretching has been analyzed in terms of a protection barrier crossing event.

Our SMD simulations indicate that protein domains not functionally implicated in the sustenance of mechanical strain readily unfold when stretched through their termini, obviously lacking a major protection barrier. An analysis of this behavior showed that unfolding requires in this case either only a one-by-one breaking of interstrand hydrogen bonds due to a particular β-strand architecture (c2) or a one-by-one breaking of intrahelix hydrogen bonds which leads to the elongation and separation of α-helices (igb, cc6), neither of which require a strong force.

The stretching and unfolding of protein domains is related in several aspects to protein folding and, hence, promises to contribute to our understanding of the latter. The pathway of unfolding along the stretching coordinates is, at least in a restricted sense, a reversal of natural folding; some folding pathways may reverse the route of stretch-induced unfolding, while many other folding pathways may differ. Simulations of stretching of these domains allows one to investigate the reverse process, as was done in the experiments reported in Carrion-Vazquez et al.[39] The barrier separating folded and unfolded states of class I domains might be representative of similar barriers postulated on the basis of folding studies.[39] To fulfill their functional role, these domains are designed by evolution to sustain stretching and refold easily. The occurrence of unfolding in normal cell function implies that folding intermediates have to be protected against pathological aggregation; hence refolding has to occur quickly.

Finally, SMD simulations may be used to design protein domains with protection barriers of desired strength. Through introduction of disulfide bonds or through cyclization mutants domains may also be designed to unfold only partially.

REFERENCES

1. Pain RH. Mechanisms of protein folding. New York: Oxford University Press; 1994.
2. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem 1997;48:545–600.
3. Gruebele M, Sabelko J, Ballew R, Ervin J. Laser t-jump unduced protein refolding. Accts Chem Res 1998;31:699–707.
4. Pande V, Grosberg A, Tanaka T, Rokhsar D. Pathways for protein folding: is a new view needed. Curr Opin Struct Biol 1998;8:68–79.
5. Brooks C. Simulations of protein folding and unfolding. Curr Opin Struct Biol 1998;8:222–226.
6. Karplus M, Sali A. Theoretical studies of protein folding and unfolding. Curr Opin Struct Biol 1995;5:58–73.
7. Lazaridis T, Karplus M. New view of protein folding reconciled with the old through multiple unfolding simulations. Science 1997;278:1928–1931.
8. Tirado-Rives J, Orozco M, Jorgensen W. Molecular dynamics simulations of the unfolding of barnase in water and 8M aqueous urea. Biochemistry 1997;36:7313–7329.
9. Ladurner A, Itzhaki L, Daggett V, Fersht A. Synergy between simulation and experiment in describing the energy landscape of protein folding. Proc Natl Acad Sci USA 1998;95:8473–8478.
10. Oberhauser AF, Marszalek PE, Erickson H, Fernandez J. The molecular elasticity of tenascin, an extracellular matrix protein. Nature 1998;393:181–185.
11. Rief M, Gautel M, Schemmel A, Gaub H. The mechanical stability of immunoglobulin and fibronectin III domains in the muscle protein titin measured by AFM. Biophys J 1998;75:3008–3014.
12. Rief M, Gautel M, Oesterhelt F, Fernandez JM, Gaub HE. Reversible unfolding of individual titin immunoglobulin domains by AFM. Science 1997;276:1109–1112.
13. Krammer A, Lu H, Isralewitz B, Schulten K, Vogel V. Forced unfolding of the fibronectin type III module reveals a tensile molecular recognition switch. Proc Natl Acad Sci USA 1999;96:1351–1356.
14. Maruyama K. Connectin/titin, a giant elastic protein of muscle. FASEB J 1997;11:341–345.
15. Wang N, Butler JP, Ingber DE. Mechanotransduction across the cell surface and through the cytoskeleton. Science 1993;260:1124–1127.
16. Machado C, Sunkel CE, Andrew DJ. Human autoantibodies reveal titin as a chromosomal protein. J Cell Biol 1998;141:321–333.
17. Kellermayer M, Smith S, Granzier H, Bustamante C. Folding-unfolding transition in single titin modules characterized with laser tweezers. Science 1997;276:1112–1116.
18. Tskhovrebova L, Trinick J, Sleep JA, Simmons RM. Elasticity and unfolding of single molecules of the giant protein titin. Nature 1997;387:308–312.
19. Izrailev S, Stepaniants S, Isralewitz B, et al. Steered molecular dynamics. In: Deuflhard P, Hermans J, Leimkuhler B, et al., editors, Computational molecular dynamics: challenges, methods, ideas, Volume 4 of Lecture notes in computational science and engineering. Berlin: Springer-Verlag; 1998. p 36–62.
20. Leech J, Prins J, Hermans J. SMD: visual steering of molecular dynamics for protein design. IEEE Comp Sci Eng 1996;3:38–45.
21. Grubmüller H, Heymann B, Tavan P. Ligand binding and molecular mechanics calculation of the streptavidin-biotin rupture force. Science 1996;271:997–999.
22. Izrailev S, Stepaniants S, Balsera M, Oono Y, Schulten K. Molecular dynamics study of unbinding of the avidin-biotin complex. Biophys J 1997;72:1568–1581.
23. Isralewitz B, Izrailev S, Schulten K. Binding pathway of retinal to bacterio-opsin: A prediction by molecular dynamics simulations. Biophys J 1997;73:2972–2979.
24. Stepaniants S, Izrailev S, Schulten K. Extraction of lipids from phospholipid membranes by steered molecular dynamics. J Mol Model 1997;3:473–475.
25. Kosztin D, Izrailev S, Schulten K. Unbinding of retinoic acid from its receptor studied by steered molecular dynamics. Biophys J 1999;76:188–197.
26. Wriggers W, Schulten K. Investigating a back door mechanism of actin phosphate release by steered molecular dynamics. Proteins 1999;35:262–273.
27. 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.
28. Nagar B, Overdulin M, Ikura M, Rini J. Structural basis of calcium-induced e-cadherin rigidification and dimerization. Nature 1996;380:360–364.
29. Wang JH, Pepinsky RB, Stehle T, et al. The crystal structure of an n-terminal two-domain fragment of vascular cell adhesion molecule 1 (vcam-1). Proc Natl Acad Sci USA 1995;92:5714–5718.
30. Leahy DJ, Aukhil I, Erickson HP. 2.0 angstroms crystal structure of a four-domain segment of human fibronectin encompassing the rgd loop and synergy region. Cell 1996;84:155–164.
31. Kerfeld C, Anwar H, Interrante R, Merchant S, Yeates T. The structure of chloroplast cytochrome c6 at 1.9 angstroms resolution: evidence for functional oligomerization. J Mol Biol 1996;250:627.
32. Gouda H, Torigoe H, Saito A, Sato M, Arata Y, Shimada I. Three-dimensional solution structure of the b domain of staphylococcal protein a: comparisons of the solution and crystal structures. Biochemistry 1992;31:9665–9672.
33. Brayn Sutton R, Davletov RA, Berghuis AM, Sudhof TC, Sprang SR. Structure of the first C2 domain of synaptotagmin I: a novel Ca2+/phospholipid-binding fold. Cell 1995;80:929–938.

34. Bernstein FC, Koetzle TF, Williams GJ, et al. The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
35. Nelson M, Humphrey W, Gursoy A, et al. NAMD—A parallel, object-oriented molecular dynamics program. J Supercomputing App 1996;10:251–268.
36. Brünger AT. X-PLOR, Version 3.1: a system for x-ray crystallography and NMR. New Haven: The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University; 1992.
37. MacKerell Jr AD, Bashford D, Bellott M, et al. All-hydrogen empirical potential for molecular modeling and dynamics studies of proteins using the CHARMM22 force field. J Phys Chem B 1998;102:3586–3616.
38. Improta S, Politou A, Pastore A. Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. Structure 1996;4:323–337.
39. Carrion-Vazquez M, Oberhauser A, Fowler S, Marszalek P, Broedel S, Clarke J, Fernandez J. Mechanical and chemical unfolding of a single protein: a comparison. Proc Natl Acad Sci USA 1999;96:3694–3699.
40. Meekhof A, Hamill S, Arcus V, Clarke J, Freund S. The dependence of chemical exchange on boundary selection in a fibronectin type III domain from human tenascin. J Mol Biol 1998;282:181–194.
41. Erickson H. Stretching single protein modules: titin is a weird spring. Science 1997;276:1090–1093.
42. Lu H, Schulten K. Steered molecular dynamics simulation of conformational changes of immunoglobulin domain I27 interprete atomic force microscopy observations. Chem Phys 1999: In press.
43. Hynes RO. Fibronectins. New York: Springer-Verlag; 1990.
44. Duan Y, Kollman P. Pathways to a protein folding intermediate observed in a 1 microsecond simulation in aqueous solution. Science 1998;282:740–744.
45. Gullingsrud J, Braun R, Schulten K. Reconstructing potentials of mean force through time series analysis of steered molecular dynamics simulations. J Comp Phys Special Issue on Computational Biophysics, 1999. In press.
46. Humphrey WF, Dalke A, Schulten K. VMD—Visual Molecular Dynamics. J Mol Graph 1996;14:33–38.