doi:10.1016/S0022-2836(02)01001-X available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2002) 323, 939–950

JMB

Identifying Unfolding Intermediates of FN-III$_{10}$ by Steered Molecular Dynamics

Mu Gao[1], David Craig[2], Viola Vogel[2] and Klaus Schulten[1]*

[1]*Department of Physics and Beckman Institute, University of Illinois, Urbana, IL 61801 USA*

[2]*Department of Bioengineering University of Washington Seattle, WA 98195, USA*

**Corresponding author*

Experimental studies have indicated that FN-III modules undergo reversible unfolding as a mechanism of elasticity. The unfolding of FN-III modules, including the cell-binding FN-III$_{10}$ module, has further been suggested to be functionally relevant by exposing buried cryptic sites or modulating cell binding. While steered molecular dynamics (SMD) simulations have provided one tool to investigate this process, computational requirements so far have limited detailed analysis to the early stages of unfolding. Here, we use an extended periodic box to probe the unfolding of FN-III$_{10}$ for extensions longer than 60 Å. Up to three plateaus, corresponding to three metastable intermediates, were observed in the extension-time profile from SMD stretching of FN-III$_{10}$. The first and second plateaus correspond to a twisted and an aligned state prior to unraveling FN-III$_{10}$ β-strands. The third plateau, at an extension of ~100 Å, follows unraveling of FN-III$_{10}$ A and B-strands and precedes breaking of interstrand hydrogen bonds between F and G-strands. The simulations revealed three forced unfolding pathways of FN-III$_{10}$, one of which is preferentially selected under physiological conditions. Implications for fibronectin fibrillogenesis are discussed.

Keywords: fibronectin; unfolding; steered molecular dynamics; intermediate; atomic force microscopy

Introduction

The extracellular matrix protein fibronectin (FN) not only provides a substrate for cell anchorage, but also regulates cell adhesion by transmitting environmental signals to cells.[1,2] Cells bind fibronectin through transmembrane heterodimeric proteins, termed integrins, that mechanically couple the cytoskeleton to the extracellular matrix. Clustering of integrins at focal contacts has been shown to result in significant mechanical tension on fibronectin fibrils. A number of remarkable findings suggest that these cell-derived mechanical forces, in addition to biochemical cues, play a role in regulating the functional state of fibronectin.[3,4]

Fibronectin is a multimodular extracellular protein composed of more than 20 modules per monomer of three types: type I (FN-I), type II (FN-II), and type III (FN-III) (shown schematically in Figure 1). Cells assemble fibronectin dimers into fibrillar networks that provide mechanical stability to the extracellular matrix and connective tissue.[1,2] Integrin binding to fibronectin RGD-loop on FN-III$_{10}$ results in the formation of a cytoskeletal apparatus that mechanically couples actin contractility to extracellular fibronectin fibers.[3] Fluorescent microscopy studies using chimeric fibronectin with green fluorescent protein have demonstrated that cells stretch fibronectin fibers up to four times their equilibrium length.[5] Recent studies using fluorescent resonance energy transfer to measure fibronectin fibrils stretched into non-equilibrium, suggest that cells not only integrate fibronectin in an extended confirmation into fibrils, but also hyperextend many fibrils so that FN-III modules become unfolded.[6,7] It has been proposed that the unfolding of FN-III modules is functionally relevant to providing both the observed elasticity in fibronectin fibrils and a mechanism for exposing the so-called cryptic sites.[3,4,8,9] Functionally relevant cryptic sites have been specifically proposed to exist on a number of FN-III modules including FN-III$_1$,[10] FN-III$_2$,[11] FN-III$_7$,[12] FN-III$_9$,[13] FN-III$_{EDB}$,[14] FN-III$_{10}$,[15] FN-III$_{13\text{-}14}$,[16] and FN-III$_{15}$.[12]

Abbreviations used: AFM, atomic force microscopy; FN, fibronectin; FN-III, fibronectin type III; RGD, arginine-glycine-aspartic acid; SMD, steered molecular dynamics.

E-mail address of the corresponding author: kschulte@ks.uiuc.edu; http://www.ks.uiuc.edu/

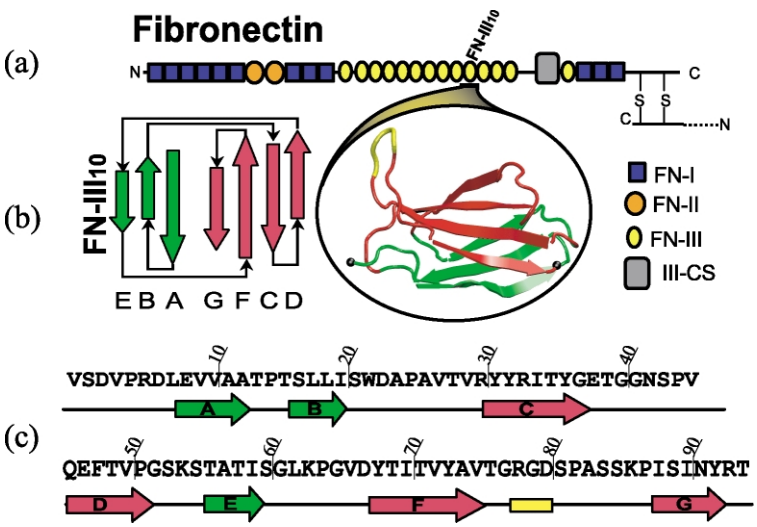

Figure 1. Schematic representation of the fibronectin monomer. (a) The modular structure. FN-I modules are shown as blue squares, FN-II modules are shown as orange circles, and FN-III modules are shown as yellow ovals. Two fibronectin monomers join through disulfide bonds. (b) FN-III$_{10}$ is illustrated through two representations highlighting its secondary (left) and tertiary (right) structure. β-strands from the upper β-sheet GFCD are shown in red, while β-strands from the lower β-sheet ABE are shown in green. The RGD loop between the F and G-strands is shown in yellow. (c) Primary sequence of FN-III$_{10}$ and corresponding β-strands (indicated below the sequence). The RGD loop is shown in yellow.

These experimental studies on fibronectin and its FN-III modules suggest that the force-induced mechanical unfolding pathway and corresponding intermediates of FN-III modules play important physiological roles. For example, the computationally verified straightening of the RGD-loop in stretched FN-III$_{10}$ may control binding to membrane receptor integrins.[17] By swapping β-strands of FN modules, intermediates in the unfolding pathway may serve as nucleation sites[13] for the formation of fibrils (fibrillogenesis) during assembly of the extracellular FN matrix.

FN-III$_{10}$ has become a model for our understanding of the unfolding pathway of FN-III modules, partly because of its importance in cell binding. Chemical and thermal denaturant studies demonstrate that FN-III$_{10}$ is one of the most thermostable FN-III modules.[18] These studies suggest that the A and G-strands are the last to form during folding of FN-III$_{10}$.[19,20] While chemical and thermal unfolding methods have provided significant insight into the folding pathway, they may not be well suited to investigate the mechanical unfolding pathway. Comparisons between mechanical and chemical unfolding pathways show that there exist few correlations between mechanical stability and thermodynamic stability[21] and that the respective unfolding pathways may differ.[22–24] To investigate the mechanical unfolding pathway of FN-III modules, other approaches such as atomic force microscopy (AFM) have been used. One of the most significant findings from these investigations is that FN-III modules vary significantly in mechanical stability, suggesting the order of FN-III unraveling is important to its function.[24–27]

AFM studies alone cannot provide a complete atomic-level view of the mechanical unfolding process. Because of this, molecular dynamics simulations have been used to investigate the forced unfolding pathway of FN-III$_{10}$ and of other similar modules and motifs.[28] Krammer *et al*. first used steered molecular dynamics (SMD) on FN-III$_{10}$ solvated in explicit water and found a large predominant force peak corresponding to separation of the G-strand.[17] Craig *et al*. used a similar system to find that the FN-III$_{10}$ module is mechanically one of the weakest FN-III modules and that β-strands progress from a twisted to an aligned state prior to unraveling.[29] Conversely, Paci & Karplus, using an implicit water model and biased molecular dynamics, found FN-III$_{10}$ to be considerably more stable than FN-III$_9$ and further reported an unfolding intermediate at an N terminus-to-C terminus distance of approximately 140 Å.[30] Finally, approaches using off-lattice models[31] have been able to reproduce a similar unfolding pathway as that reported in SMD simulations with explicit solvent models.

Previous SMD simulations have been limited due to limitations of computational resources. Simulations of explicit solvent models were restricted to earliest events up to or including separation of the first β-strand. Implicit water models were needed to simulate the whole unfolding process. These different approaches to mechanically stretching FN-III modules have yielded inconsistent results, such as the existence of intermediates in some studies but not others. Here, we have used multiple unfolding simulations of FN-III$_{10}$ solvated in a box of water large enough to permit stretching of FN-III$_{10}$ into its completely unfolded form. We investigate in detail the late stages of unfolding, probe for folding intermediates, and resolve discrepancies between the different molecular dynamics approaches. Our results show that β-strands separate from the main module by three possible pathways, A-strand separating first, G-strand separating first, or both A and G-strands separating simultaneously. The results also reveal an intermediate present only when the A-strand

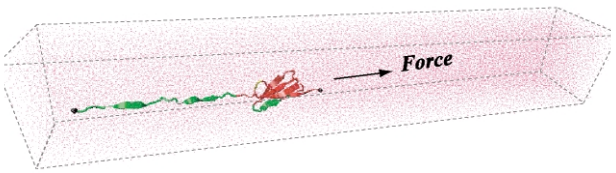

Figure 2. Setup of SMD simulations. An FN-III$_{10}$ module was solvated and stretched in a water box. In the simulations the C^{α} atom at the N terminus (Val1) was fixed, and C^{α} atom at the C terminus (Thr94) was stretched with constant force.

separates first. Below we first introduce the simulation methods, provide then a detailed comparison of trajectories, and finally discuss the physiological implications of our findings.

Results

Mechanical unfolding intermediates

SMD simulations with constant force protocol have been carried out to probe the unfolding intermediates of FN-III$_{10}$ solvated in a 367 Å long water box (see Figure 2). Figure 3 presents the extension *versus* time profile from a simulation SMD(500 pN)$_1$, in which an FN-III$_{10}$ module was stretched with 500 pN constant force from its equilibrated structure to its fully elongated configuration. A sequence of unfolding snapshots is shown in Figure 3(a)–(e). Initially, two β-sheets ABE and GFCD pack against each other in a twisted orientation (see Figure 3(a)). As the termini are straightened by an externally applied force, the protein extends to 11 Å at 50 ps, reaching the edge of a short 70 ps plateau (intermediate I_1). As discussed,[29] the plateau corresponds to a twisted state, for the angle at which two β-sheets pack against each other remains unchanged from that found in the equilibrium structure. Following the plateau, a pair of backbone hydrogen bonds between Arg6 on β-strand A and Asp23 on β-strand B break at 120 ps, permitting an additional extension of 7 Å. The protein subsequently enters an intermediate I_2, named "aligned state",[29] for two β-sheets aligned along the direction of the external force as shown in Figure 3(b). The aligned state is associated with a ~220 ps plateau, indicating a relatively stable intermediate. During the transition from I_1 to I_2,

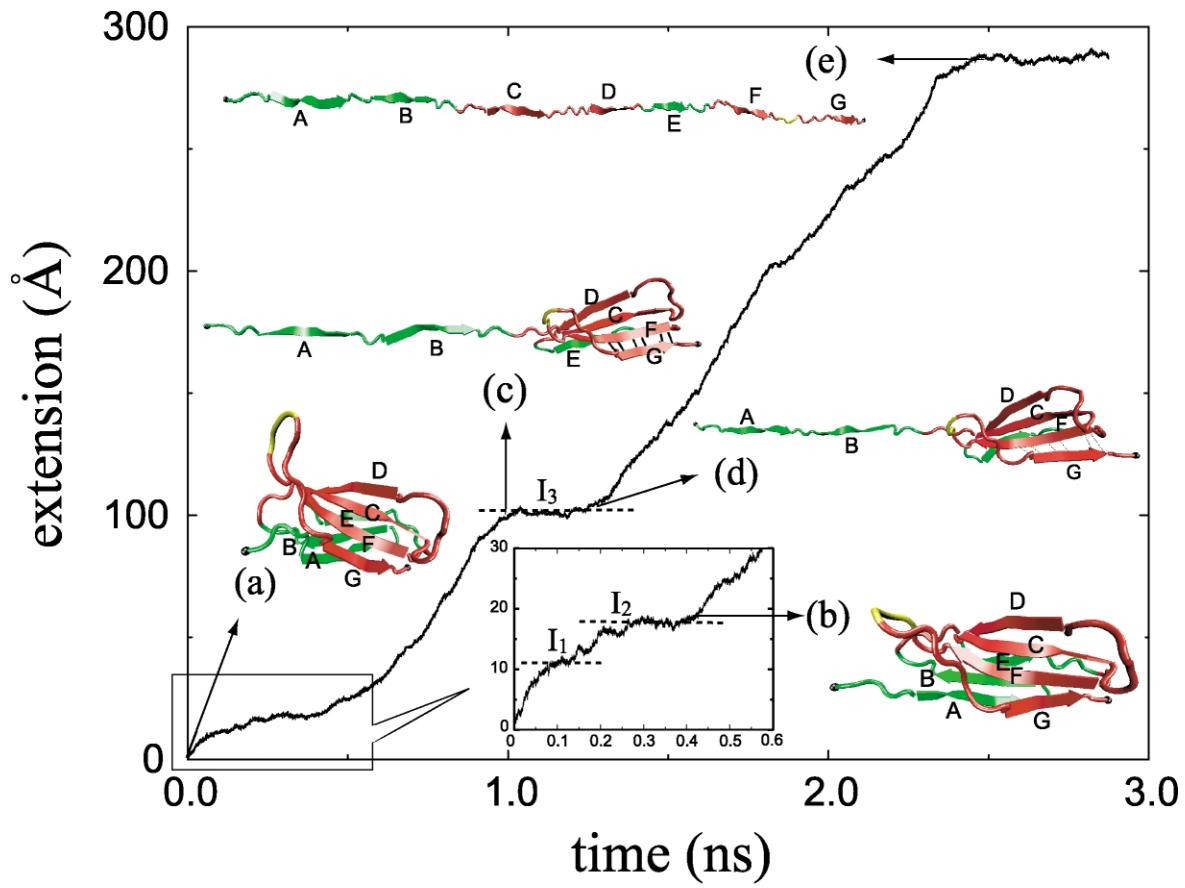

Figure 3. Extension-time profile from simulation SMD(500 pN)$_1$ and representative snapshots (a)–(e) illustrated during unfolding. (a) Native structure; (b) structure in aligned orientation at 0.42 ns; (c) structure of an intermediate at 100 Å in extension; (d) hydrogen bonds between β-strands G and F break at the end of the second plateau; (e) fully unfolded structure. Hydrogen bonds between β-strands F and G in (c) are shown as thick black lines; ruptured hydrogen bonds in (d) are shown as broken black lines.

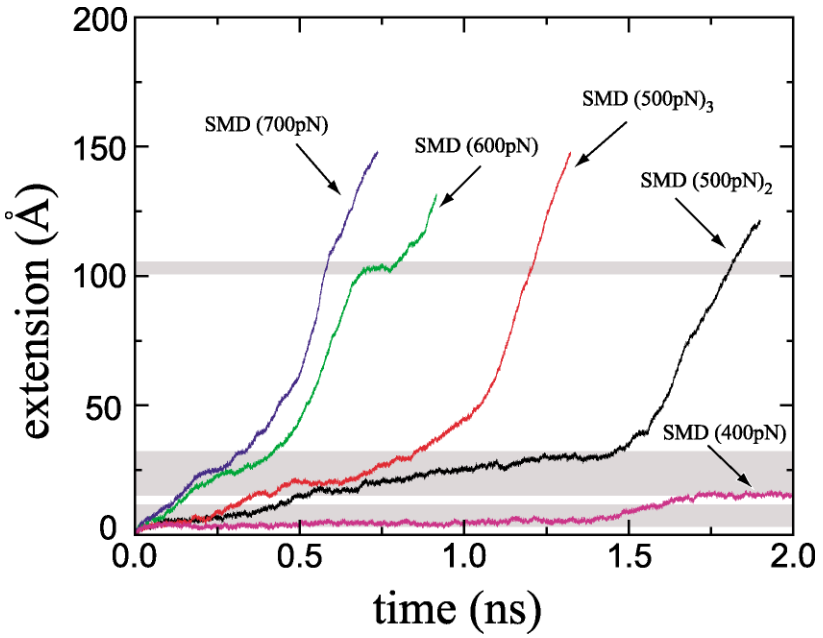

Figure 4. Representative extension-time curves from SMD simulations under constant force conditions. Simulations were performed in a 190 Å long water box. Two of three plateaus, indicated by shaded boxes, are evident in all simulations. A third plateau, at 100 Å, arises only for SMD(600 pN).

the hydrophobic core of FN-III$_{10}$ has been disrupted due to the rotation of the β-sheets. At the end of the second plateau the unfolding progresses along alternate pathways. In the simulation SMD(500 pN)$_1$, the remaining five intact hydrogen bonds between β-strands A and B rupture, leading to further extension. The protein unravels rapidly until it enters the third plateau at 100 Å extension (see Figure 3(c)), which corresponds to crossing of a barrier formed by six hydrogen bonds between β-strands G and F. This intermediate I_3 is characterized by having the A and B-strands separated from the remainder of the module. The passage time for crossing the barrier is ~250 ps, indicating the existence of a stable intermediate, named I_3, at an extension three times the length of native FN-III$_{10}$. After the concurrent breaking of six G–F backbone hydrogen bonds (see Figure 3(d)), the protein resumes unfolding until it is fully unfolded at 280 Å extension (see Figure 3(e)), which corresponds to the last plateau shown in the extension-time profile.

Figure 4 shows recordings of force-induced unfolding for FN-III$_{10}$ from five constant force SMD simulations performed in a 190 Å long water box. Two pronounced plateaus, common to all simulations, are found at extensions ranging from 5 Å to 11 Å and from 18 Å to 30 Å. A third plateau at 100 Å extension, similar to that observed in simulation SMD(500 pN)$_1$, appears only in simulation SMD(600 pN). The plateaus identified in SMD(500 pN)$_1$ can be associated with the three intermediates I_1, I_2, I_3 characterized in Table 1. Universal to all simulations is a jump by ~10 Å in extension between intermediate I_1 and I_2, corresponding to a transition from a twisted to aligned β-sheets, as illustrated in Figure 3. Although intermediate I_1 always precedes the alignment transition, intermediate I_2 can lead to three different unfolding pathways: in simulations SMD(600 pN) and SMD(700 pN), strand A detached from the remaining fold first; in simulation SMD(500 pN)$_2$ strand G detaches first; in simulation SMD(500 pN)$_3$ strands A and G simultaneously

Table 1. Location of stable intermediates and key events of hydrogen bond rupture observed in the unfolding studies of FN-III$_{10}$

	Location of stable intermediates (Å) and key events of hydrogen bond rupture								
Unfolding simulations	I_1 (twisted)	→	Rupture	→	I_2 (aligned)	→	Rupture	→	I_3	→	Rupture
SMD(400 pN)	5		R–D		≥16		–		–		–
SMD(500 pN)$_1$	11		R–D		18		A–B		100		G–F
SMD(500 pN)$_2$	9		R–D		25		G–F		–		–
SMD(500 pN)$_3$	9		R–D		20		A–B, G–F		–		–
SMD(600 pN)	11		R–D		25		A–B		100		G–F
SMD(700 pN)	11		R–D		25		A–B		–		–

The location of the intermediate is defined as the extension at the middle point of the corresponding plateau. A pair of backbone hydrogen bonds between Arg6 and Asp23 is designated as R–D. The hydrogen bonds between β-strands A and B and between G and F are abridged as A–B and G–F, respectively.

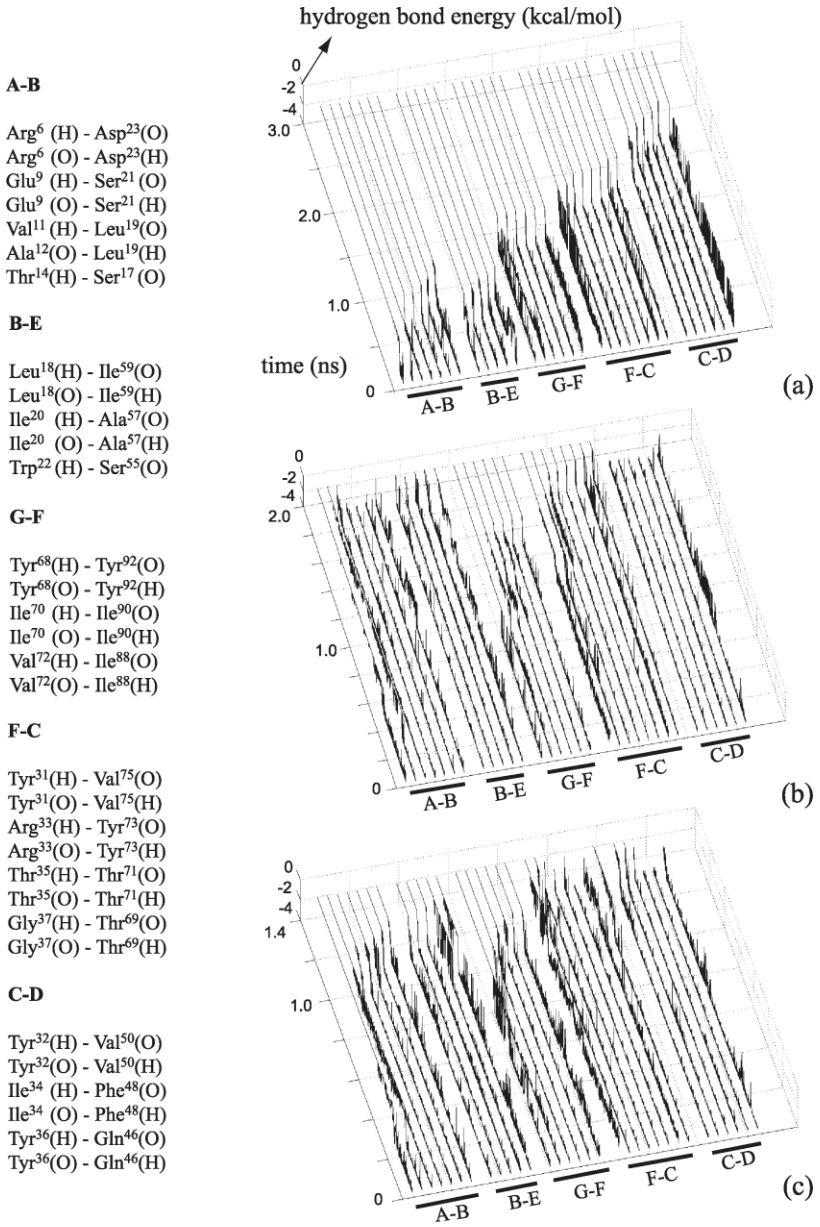

Figure 5. Inter-strand hydrogen bond *versus* time curves in three representative SMD trajectories from (a) SMD(500 pN)$_1$, (b) SMD(500 pN)$_2$, and (c) SMD(500 pN)$_3$. Energies are grouped by β-strand pairs, displayed from left to right according to the ordering in the list on the left side.

detach from their β-strand partners B and F. In simulation SMD(400 pN), none of the β-strands peeled away from the module within 2 ns.

Inter-strand hydrogen bonds

An analysis of the energies of all inter-strand hydrogen bonds was performed to characterize the three force-induced unfolding pathways of FN-III$_{10}$. Figure 5(a)–(c) shows the energies of 32 backbone hydrogen bonds of the module along three representative unfolding pathways from SMD(500 pN)$_1$, SMD(500 pN)$_2$ and SMD(500 pN)$_3$, respectively. As mentioned earlier, a pair of hydrogen bonds near the N terminus, connecting residues Arg6 and Asp23, break first in all simulations to allow the transition from intermediate I_1 to I_2 to occur. This pair of conserved hydrogen bonds was also found to be the first ruptured hydrogen bonds in SMD simulations for FN-III$_{7-9}$.[29] The aligned state I_1, therefore, is separated from the twisted state by a distinct energy barrier, barrier I. The hydrogen bond analysis shows that following the crossing of barrier I, backbone hydrogen bonds disassociate in different sequences. In Figure 5(a), the intact A–B and B–E hydrogen bonds break before the G–F bonds. This implies that β-strands A and B completely unravel before the β-sheet GFCD starts to unfold. The second plateau, corresponding to breaking of hydrogen bonds between A and B-strands, is named barrier II_{AB}. At 1.25 ns, the six hydrogen bonds between G and F were disrupted simultaneously, allowing further extension of the module as shown in Figure 3. The barrier formed by G–F hydrogen bonds can trap the protein in the meta-stable intermediate state I_3. The remaining intact hydrogen bonds between β-strands F and C and between C and D

have shown less resistance to external force as they were ruptured one by one, probably contributing to the short shoulders in the extension-time profile shown in Figure 3. Alternatively, as shown in Figure 5(b), G and F are the first strands to completely separate. A pair of hydrogen bonds near the C terminus, between residue Tyr68 and Tyr92, break at 0.95 ns, followed by the breaking of the remaining backbone hydrogen bonds between G and F at 1.5 ns. A similar scenario has been previously described in constant velocity stretching simulations of FN-III$_{10}$[17] and constant force stretching simulations.[29] We name the barrier associated with the hydrogen bond rupture between F and G, II_{GF}. Occasionally, we observed a third unfolding pathway as shown in Figure 5(c). In this pathway hydrogen bonds between G and F and between A and B break simultaneously at ~1.1 ns, except that the four hydrogen bonds near the termini break earlier. In this case, β-strand A and F detach from the protein simultaneously.

Forced unfolding pathways

According to the hydrogen bond analysis, the SMD simulations of FN-III$_{10}$ exhibit three different unfolding pathways, as shown in Figure 6. Simulations began with FN-III$_{10}$ equilibrated in a periodic box. Upon applying a constant force, the randomly coiled termini straightened slightly, while the overall tertiary structure still remained intact (see Figure 6(b)). The first significant event in unfolding was the rupture of two backbone hydrogen bonds, Arg6 and Asp23, between the A and B-strands, as shown in Figure 6(c). Immediately following this rupture, Pro5-Leu8 straightened, allowing solvation of the periphery of the hydrophobic core flanked by Trp22. Subsequently, the alignment of the β-strands between the upper and lower sheet was initiated by slipping of Ile88 of the G-strand past the A-strand (see Figure 6(d)). Unraveling of the β-strands began by either separation of the A-strand first (see Figure 6(e)), A and G-strands simultaneously (see Figure 6(f)), or G-strand first (see Figure 6(g)), leading to three different unfolding pathways. When unraveling of the β-strands began with the A-strand, an intermediate at 100 Å was observed (see Figure 6(i)).

RGD-loop conformations

The peeling of strand G away from β-sheet GFCD leads to a dramatic conformational change of the RGD-loop. This has been observed previously when strand G is the first strand to be stretched out.[17] Conformational analysis of the RGD loop in simulation SMD(500 pN)$_1$ shows a similar conformational transition when the G strand was stretched out after the rupture of β-sheet ABE, as shown in Figure 7. Prior to breaking of the G–F hydrogen bond, the RGD loop remains intact for 1.25 ns at the apex of a hairpin β-turn that connects strands G and F (see Figure 7(c) and (d)). The width of the RGD loop (see Figure 7(a) and the bending angle of the loop (see Figure 7(b)) exhibit a sudden increase after the hydrogen bond rupture. The width of the loop, defined as the distance between Arg78 and Pro82 (illustrated in Figure 7(c)), jumps from 7 Å to 14 Å, for the loop has been quickly straightened. The bending angle, defined as the angle formed by C^{α} atoms of Thr76, Arg78 and Asp80, displays a change from a bent (~110°) to a flatter (~170°) conformation. At 1.5 ns, the loop has been gradually dragged out and the apex of the loop has also been lowered to lie in the plane spanned by the remaining β-strands FCD (Figure 7(e)).

Discussion

The SMD simulations described here provide a detailed description of the entire force-induced unfolding pathway for FN-III$_{10}$ in a large periodic box. These simulations find that while there is a consensus unfolding pathway prior to the unraveling of β-strands for FN-III$_{10}$, the order in which β-strands separate from the module is variable. Consistent with earlier studies,[9,29] the simulations show that the upper and lower β-sheets progress from a twisted to aligned state prior to unraveling of the first β-strand. Invariably, we also find that the rupture of two inter-strand hydrogen bonds between Arg6 and Asp23 directly precedes extension to the aligned state. These hydrogen bonds play a stabilizing role in the FN-III$_{10}$ module: their rupture is initiated by solvation of the hydrophobic core. Stretching after the alignment of FN-III$_{10}$ β-sheets leads to either the separation of the G-strand, the A-strand, or both A and G-strands. An intermediate arises consistently at extension ~100 Å in case the A-strand separates first. Figure 6 provides a summary of the unfolding pathways for FN-III$_{10}$. Our simulations raise a number of questions that will now be discussed, including which of these pathways is preferable at physiological forces and what could be the function of an intermediate in the unfolding pathway.

Comparison with other models

SMD simulations of FN-III$_{10}$ have been conducted with explicit water sphere models,[17,29] with a small periodic water cell,[9] and with implicit solvation models.[30] In the case of water sphere models, FN modules were solvated in TIP3 water spheres of ~60 Å diameter, cut-offs being used for calculating electrostatic interactions. Intermediates I_1 and I_2 were previously identified for each model within an extension of <30 Å for proteins.[29] Although the sphere models provide a quick approach for studying mechanical stabilities and probing intermediates in the early stages of unfolding, they may exhibit certain artifacts at extension >60 Å when the modules extend beyond the water spheres. For example, one inherent problem

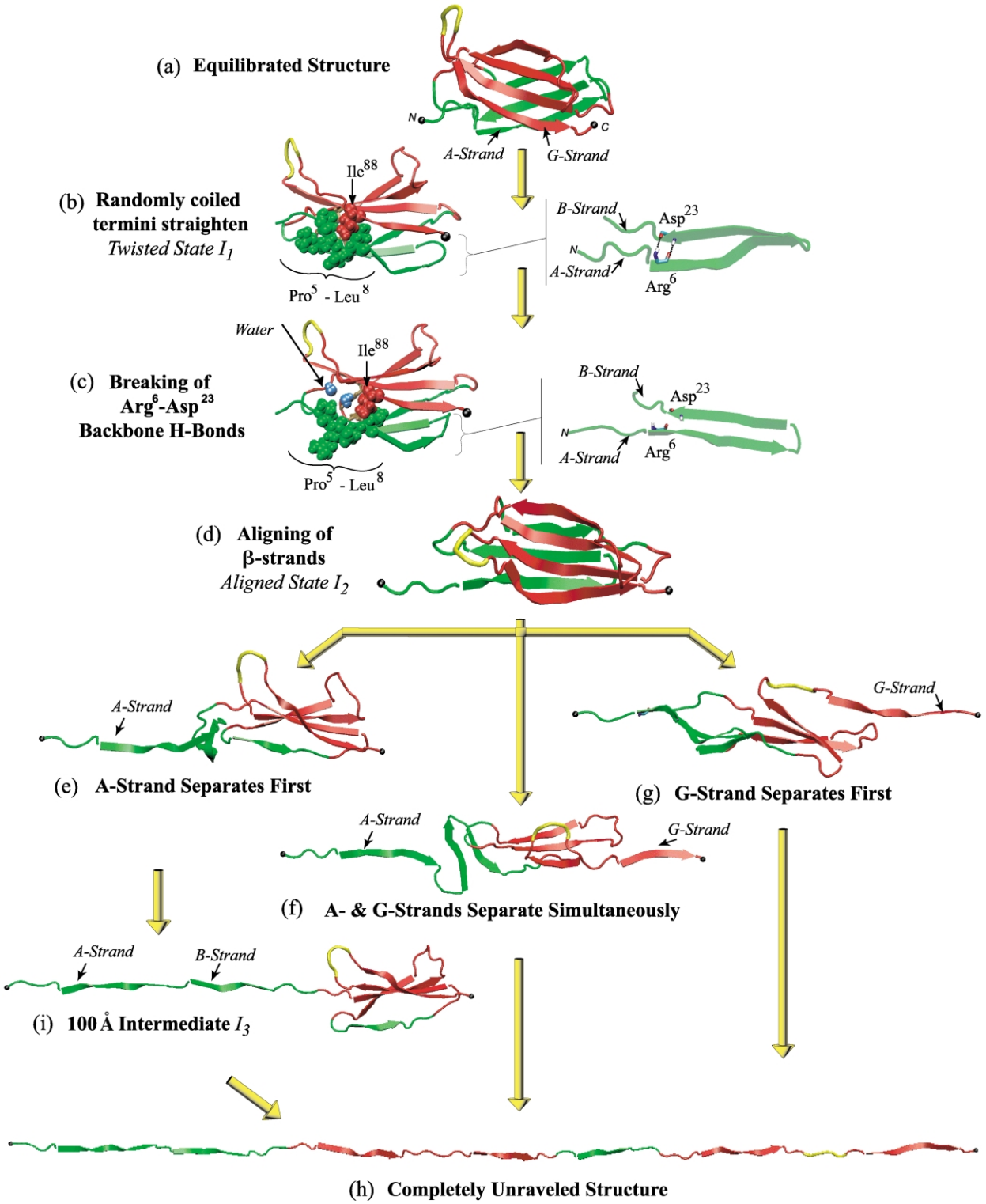

Figure 6. Unfolding pathways for FN-III$_{10}$ resulting from SMD simulations. (a) FN-III$_{10}$ equilibrated in a periodic box. (b) Upon applying a constant force the randomly coiled termini straightened slightly, while the overall tertiary structure remained intact. (c) The first significant event in unfolding, rupture of two backbone hydrogen bonds Arg6 and Asp23 between A and B-strands; immediately following the rupture the segment Pro5-Leu8 straightened, permitting solvation of the hydrophobic core periphery. (d) Alignment of β-strands between the two β-sheets is initiated by slipping of Ile88 of the G-strand past the A-strand. Unraveling of the β-strands begins by separation of (e) A-strand first, or (f) A and G-strands simultaneously, or (g) G-strand first. (h) Fully unraveled structure. (i) In the case that the A-strand separates first, an intermediate at 100 Å extension arises.

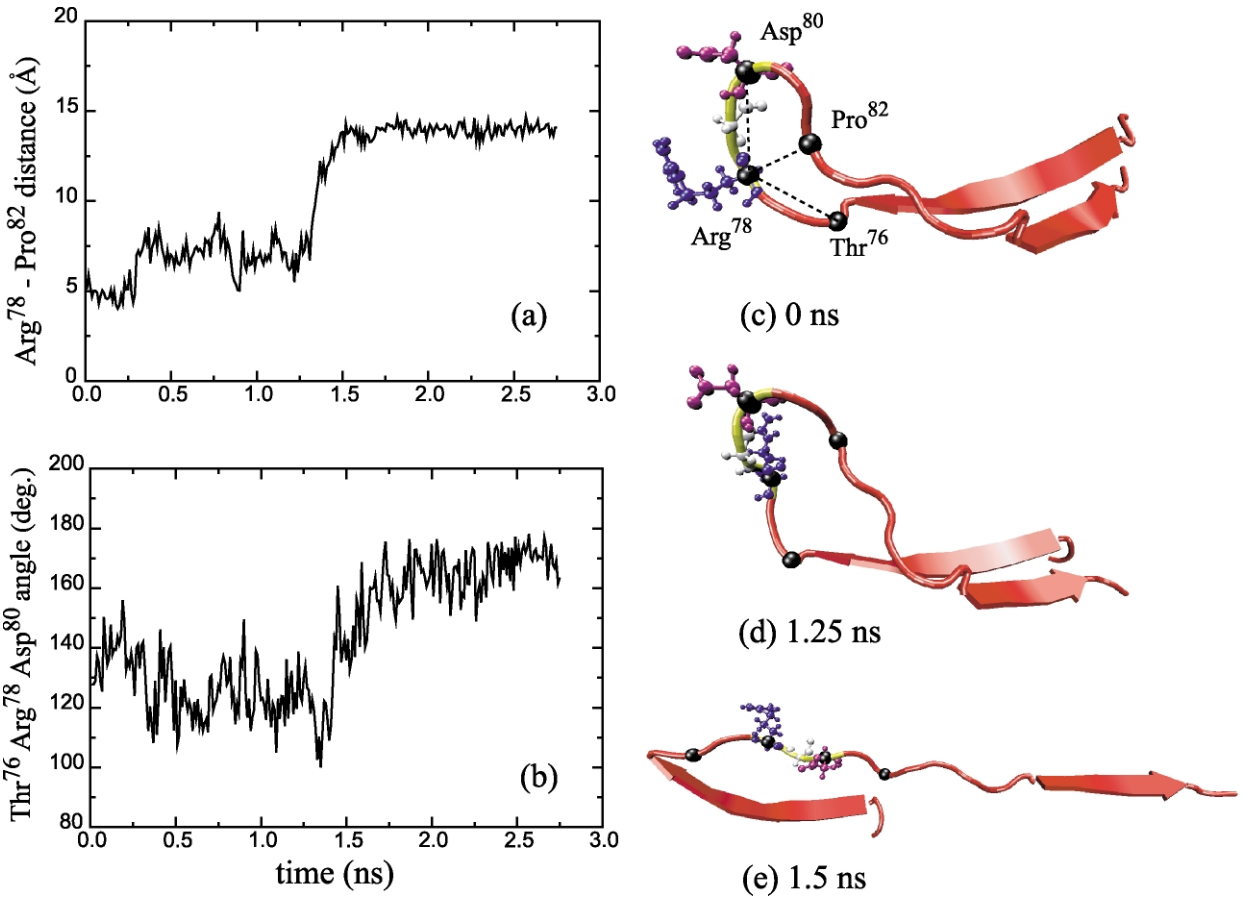

Figure 7. Conformational changes of the RGD-containing loop in simulation SMD(500 pN)$_1$. (a) The width of the RGD loop is measured by the distance between the C$^\alpha$ atoms of Arg78 and Pro82. (b) Angle formed by the C$^\alpha$ atoms of residues Thr76, Arg78, and Asp80 illustrates a change from a bent (~110°) to a more planar (~170°) conformation. (c) Snapshot of F and G-strands and the connecting loop in the native structure. (d) The RGD loop remains intact in the intermediate state at 100 Å. (e) The RGD loop is straightened out after the protein passes the intermediate state.

with using water spheres is the deformation of the water sphere, which requires additional, but artificial, forces. FN-III$_{9,10}$ have been also simulated in a periodic box with similar results to water sphere models, the simulations again being limited to only the early stages of unfolding because of the limited length of the periodic box used.[9] The implicit model, on the other hand, can be easily extended to large boxes, but it overlooks friction due to water molecules. Furthermore, it has intrinsic shortcomings because the effect of explicit water molecules on breaking inter-strand hydrogen bonds is not accounted for.[32] Implicit solvation models give an unphysical refolding rate of FN-III$_{9,10}$ modules of less than 50 ps.[30] The implicit models report FN-III$_{10}$ to be mechanically stronger than FN-III$_9$,[30] while employing an explicit model we found FN-III$_{10}$ is mechanically weaker than FN-III$_9$,[29] which is in agreement with experimental observations using AFM.[24] The water-bath models presented in this study address these problems by providing adequate explicit solvation for fully unfolding FN-III$_{10}$ modules and by faithfully describing long-range Coulomb forces.

Force-dependence of the unfolding pathways

Current computational resources restrict SMD simulations to a few nanoseconds, and forces higher than 400 pN need to be used to produce unfolding events that are fast enough to be observed in simulations. In AFM experiments unfolding proceeds on a ms time-scale and peak rupture forces of ~75 pN are measured.[24] Under physiological conditions forced unfolding of FN-III modules could take even longer and forces as low as several pN could be strong enough to unfold a single FN-III module.[8] Naturally the question arises: at low force, does FN-III$_{10}$ follow the three unfolding pathways revealed by SMD simulations? Among the three possible choices, the unfolding scenario in which β-strands A and G separate simultaneously seems the least likely under weak forces, for the protein needs to overcome two barriers simultaneously. To illustrate the preference of the other two unfolding pathways, which require crossing only one barrier, we compare the mean first passage times of the two processes. For this purpose we consider a very simple model for the unfolding process with a single

barrier, which has the form of a saw-tooth potential as introduced earlier:[29,33]

$$U(x) = \begin{cases} +\infty & x < a, \\ \Delta U(x-a)/(b-a) & a \le x \le b, \\ -\infty & x > b \end{cases}$$

ΔU is the height of the barrier and (a, b) defines the associated extension region. For this model, the mean first passage time to pass the barrier can be stated in the concise expression:

$$\tau = 2\tau_d \delta(F)^{-2}[e^{\delta(F)} - \delta(F) - 1]$$

where F is the external force and D is the effective diffusion coefficient. Other quantities are $\tau_d = (b-a)^2/2D$, $\delta(F) = \beta[\Delta U - F(b-a)$, $\beta = 1/k_B T$. The extension range b–a is ~4 Å, estimated from the fluctuation of the associated plateau (see Figure 3). According to the SMD simulations, crossing either barrier II_{AB} or II_{GF} leads to a different unfolding pathway. We assume that the heights of barriers II_{AB} and II_{GF}, designated as ΔU_{AB} and ΔU_{GF}, are determined by the number of intact hydrogen bonds between β-strands A and B and between G and F before barrier crossing. The height of the unfolding barrier ΔU probed in AFM experiments is 22 kcal mol^{-1}.[24] Divided by the maximum number of concurrently breaking hydrogen bonds (six), the experimental value gives an average energy contribution from each hydrogen bond of ~3.7 kcal mol^{-1}. From this we estimate that each hydrogen bond contributes approximatly 3.5 kcal mol^{-1} (note a deviation within ±1.5 kcal mol^{-1} from this number does not affect the conclusion) to the unfolding barrier. Hence, the five hydrogen bonds between β-strands A and B produce a barrier with $\Delta U_{AB} = 5 \times 3.5 \approx 18$ kcal/mol. To estimate ΔU_{GF}, we note that in some cases two hydrogen bonds close to the C terminus rupture before the concurrent breaking of the remaining four hydrogen bonds bridging F and G-strands, whereas in other cases all six hydrogen bonds break concurrently. We therefore assume for ΔU_{GF} two alternatives: $\Delta U_{GF} = 4 \times 3.5$ kcal mol^{-1} = 14 kcal mol^{-1}, and $\Delta U_{GF} = 6 \times 3.5$ kcal mol^{-1} = 21 kcal mol^{-1}. Denoting by τ_{AB} and τ_{GF} the mean first passage time for crossing barriers II_{AB} and II_{GF}, respectively, the ratio of τ_{GF} to τ_{AB} can be calculated as a function of the externally applied force F. The results are shown in Figure 8. One can recognize two regimes in Figure 8. In the regime of forces >400 pN, typical of SMD simulations, the mean first passage times for crossing both barriers are very close, i.e. two unfolding pathways have roughly the same chance to arise, as demonstrated in SMD simulations. In contrast, for forces <200 pN, a single unfolding pathway is more likely. Which pathway is followed depends on the barrier heights: for $\Delta U_{GF} > \Delta U_{AB}$, crossing barrier II_{AB} is more likely to happen first and *vice versa*. The analysis implies that in AFM experiments or under physiological conditions

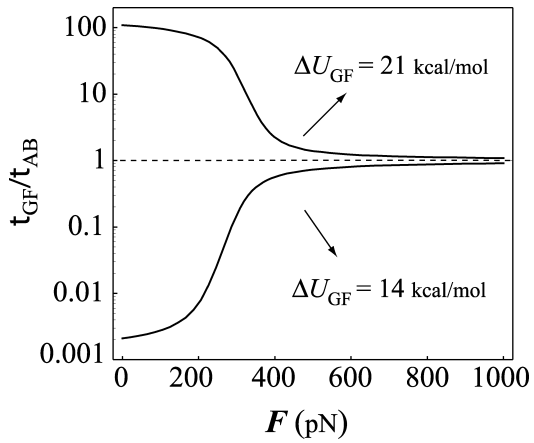

Figure 8. Ratio of mean first passage times τ_{AB} and τ_{FG} as a function of applied force. τ_{AB} and τ_{FG} are the first passage times for crossing the barriers II_{AB} and II_{FG}, respectively.

where small forces act, a single pathway should dominate unfolding. In agreement with this analysis, recent AFM unfolding experiments of FN-III$_{10}$ have shown that in most cases FN-III$_{10}$ prefers not to pass the unfolding intermediate I_3,[24] but in rare unfolding events the protein does demonstrate a transition state at extension close to intermediate I_3 (J. Fernandez, personal communications). Combining AFM experiments and the above analysis leads to the implication that the G-strand is likely to detach first from the protein upon a stretching force under physiological conditions.

Physiological implications of the intermediates in fibrillogenesis

The finding of multiple unfolding pathways and the finding that at least one unfolding pathway contains a quasi-stable intermediate for FN-III$_{10}$ has a number of implications. First, under physiological conditions, it is possible that the unfolding pathway may change as a result of binding to an integrin or another fibronectin protein. For example, binding of an integrin to the RGD loop between the F and G-strands of FN-III$_{10}$ may stabilize the interactions between these β-strands, thus favoring separation of the A-strand first. The existence of unfolding intermediates may be physiologically relevant because it suggests a mechanism for fibrillogenesis. It has been previously proposed that fibrillogenesis could occur through β-strand swapping.[13] The partially unfolded intermediate I_3 found in present SMD studies is able to undergo such swapping, since the two unraveled A and B-strands can self-assemble with other FN-III modules. This is consistent with experimental studies finding that FN-III$_{10}$ contains a buried cryptic site that binds to modules FN-III$_1$ and promotes formation of disulfide-bonded fibronectin fibers.[15] Thus, one hypothesis for fibronectin fibrillogenesis is that

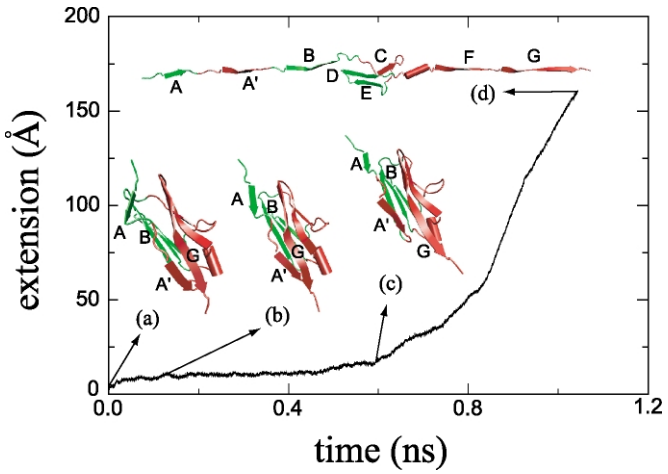

Figure 9. Extension-time profile from an SMD unfolding simulation of I27 with 750 pN constant force and representative unfolding snapshots ((a)–(d)). (a) Native structure; (b) strand A separates from strand B at 140 ps; (c) strand A′ separates from G at 16 Å extension, observed at the end of the plateau which corresponds to crossing of the barrier due to the breaking of inter-strand hydrogen bonds between β-strands A′ and G; (d) partially unfolded structure at 160 Å extension.

mechanical tension applied from an integrin would unravel the A and B-strands from FN-III$_{10}$ providing a nucleation site for other unraveled FN-III modules, such as FN-III$_1$. Conversely, FN-III$_{10}$ modules not bound to integrins would unravel by way of the G-strand separating first, thus making the RGD loop inaccessible. While further studies are needed to investigate such a model for fibrillogenesis, the simulations conducted here provide an explanation for previous findings and suggest a direction for future experimental studies.

Comparison with the muscle protein titin

It is of interest to compare the unfolding pathways of FN-III$_{10}$ to that of I27, a domain from muscle protein titin (see Figure 9(a)). Both FN-III$_{10}$ and I27 are built in a similar β-sandwich architecture. The forced unfolding of I27 has been studied extensively in AFM experiments[34] and in SMD simulations.[28,35] Unlike FN-III, the SMD simulations have shown a single unfolding pathway. However, so far unfolding I27 has never been simulated in water boxes permitting full extension and eliminating surface effects. Figure 9 presents the extension *versus* time profile from an SMD simulation of an I27 module in a 240 Å water box with a stretching force of 750 pN. β-Strand A separates from the B-strand at ~140 ps after the breaking of two backbone hydrogen bonds between A and B (see Figure 9(b)), corresponding to an intermediate identified earlier both in SMD simulations and AFM experiments.[36] At the end of the long plateau shown in the extension-time curve, strand A′ separates from G at the extension of 16 Å (see Figure 9(c)), initiating rapid unraveling. The separation of β-strand A′ from G, like the separations of β-strand A from B or G from F in SMD simulations of FN-III$_{10}$, involves concurrent breaking of a cluster of backbone hydrogen bonds, which constitute the main contribution to the mechanical stability of the proteins. However, there is a slight structural difference between I27 and FN-III$_{10}$ in the way that the N-terminal β-strand A/A′ connects with the C-terminal β-strand G. Strands A and G of FN-III$_{10}$ are separated, while I27 has β-strand A′ form six backbone hydrogen bonds with β-strand G. To unfold I27, all A–B and A′–G hydrogen bonds must be disrupted before it can extend further. The separation of strands A and A′ destabilized both β-sheets of I27, initiating simultaneous unraveling of both sheets.[37] In the case of FN-III$_{10}$, however, unfolding can be initiated by unraveling only one of the β-sheets through separating either the A-strand or the G-strand from the module. The other β-sheet maintains its tertiary structure until the module needs further extension. Disrupting the second β-sheet may result in an intermediate like I$_3$ in the middle of unfolding process, as observed in simulations using an explicit solvent model reported here and simulations using implicit models.[22]

Outlook

The intermediate I$_3$ of FN-III$_{10}$ is not easy to observe in AFM experiments. According to the mean first passage time-analysis described above, the reason is that overcoming barrier II$_{AB}$ is more difficult than crossing barrier II$_{GF}$. By engineering mutants which change the key mechanical elements, i.e. the hydrogen bonding structure of A–B or F–G strands, one can change the relative height of these two barriers and may thereby produce mutants that exhibit the intermediate I$_3$. The same reasoning for choice of mutants was previously applied to titin I27. For example, I27 mutants in which either A–B or A′–G hydrogen bonds were disrupted have been engineered and stretched with AFM.[36,38] Disrupting A–B hydrogen bonds of I27 eliminates the pre-burst intermediate.[36] Mutating residues involved in forming A′–G hydrogen bonds can mechanically weaken the module.[38]

Materials and Methods

An individual FN-III$_{10}$ module was adopted from the crystallographic structure of the tetramer FN-III$_{7-10}$

(Protein Data Bank entry code 1FNF[39]). After hydrogen atoms were added to the monomer with X-PLOR,[40] two TIP3[41] water boxes, one of size 60 Å × 60 Å × 370 Å and the other one of size 64 Å × 64 Å × 190 Å, were used for solvating the protein, resulting in systems of 126,082 and 72,856 atoms, respectively. Water molecules within 2.6 Å of the protein surface were deleted. A snapshot of the 126,082 atom simulation, with the water box framed, is shown in Figure 2.

The program NAMD[42] with CHARMm22 force fields[43] was subsequently used for all MD simulations described here. Initially, the system was minimized for 2000 conjugate gradient steps. A 20 ps equilibration was performed at 300 K, while keeping the protein fixed. Following the equilibration, the system was minimized for 400 steps. Then the constraints on the protein were released, except that the C^{α} atoms of the termini were harmonically restrained to prevent drifting. The whole system was equilibrated for additional 50 ps for the larger system and 150 ps for the smaller system. During equilibration, carried out under NpT conditions, the pressure was maintained at 1 atm using the Langevin piston method,[44] and the temperature was controlled by using Langevin dynamics at 300 K with a damping coefficient of 5 ps^{-1}. For the pressure control a piston period of 100 fs and a damping time-constant of 50 fs were employed. At the end of the second equilibration, the sizes of the simulation boxes were 55 Å × 60 Å × 367 Å and 61 Å × 63 Å × 182 Å, respectively. Finally, SMD simulations were carried out by fixing the C^{α} atom of the N terminus of FN-III$_{10}$, and applying constant forces to the C^{α} atom of the C terminus along the longest side of the boxes.

For all MD simulations, periodic boundary conditions were imposed. Full electrostatics was computed every 4 fs using the particle mesh Ewald (PME) method,[45] with grid spacing less than 1.0 Å. The van der Waals interactions were treated with cut-off by using a switching function between 10 Å and 12 Å. The integration time step was 1 fs and a uniform dielectric constant of 1 was assumed. A total of ten constant force SMD simulations encompassing 12 ns were completed. The simulations were carried out on the Origin2000 at the National Center for Supercomputing Applications (NCSA) and on a Linux cluster consisting of 32 1.33 GHz Athlon processors. A 1 ns simulation of the 126,082 atom system required ~170 hours running on the cluster. The SMD simulations presented here are referred to as SMD(force value), e.g. SMD(600 pN) for a 600 pN constant force SMD simulation. Multiple runs are denoted by a number subscript, e.g. SMD(500 pN)$_1$.

The trajectories of the SMD simulations were obtained by saving the atomic coordinates of the whole system every picosecond. The analysis of molecular structures and hydrogen bond energies were conducted using X-PLOR and VMD.[46] The extension of the protein is defined as the change of the end-to-end distance between the two terminal carbon atoms. An explicit hydrogen-bonding energy term was used in the trajectory analysis, with parameters adopted from param11.pro in X-PLOR.

Acknowledgements

This work was supported by the National Institutes of Health (NIH PHS5P41RR05969 KS, 1R01GM60946 KS, R01GM40963 VV, and 5T32GM08268 DC) and the National Science Foundation (NRAC MCA93S028 KS). We acknowledge Andre Krammer and Wendy Thomas for helpful discussions.

References

1. Mosher, D. R. (1989). *Fibronectin*, Academic Press, San Diego.
2. Hynes, R. O. (1990). *Fibronectins*, Springer, New York.
3. Geiger, B., Bershadsky, A., Pankov, R. & Yamada, K. (2001). Transmembrane crosstalk between the extracellular matrix and the cytoskeleton. *Nature Rev. Mol. Cell. Biol.* **2**, 793–805.
4. Vogel, V., Thomas, W. E., Craig, D. W., Krammer, A. & Baneyx, G. (2001). Structural insights into the mechanical regulation of molecular recognition sites. *Trends Biotechnol.* **19**, 416–423.
5. Ohashi, T., Kiehart, D. P. & Erickson, H. P. (1999). Dynamics and elasticity of the fibronectin matrix in living cell culture visualized by fibronectin-green fluorescent protein. *Proc. Natl Acad. Sci. USA*, **96**, 2153–2158.
6. Baneyx, G., Baugh, L. & Vogel, V. (2001). Coexisteing conformations of fibronectin imaged in cell culture using fluorescence resonace energy transfer. *Proc. Natl Acad. Sci. USA*, **98**, 14464–14468.
7. Baneyx, G., Baugh, L. & Vogel, V. (2002). Supramolecular chemistry and self-assembly special feature: fibronectin extension and unfolding within cell matrix fibrils controlled by cytoskeletal tension. *Proc. Natl Acad. Sci. USA*, **99**, 5139–5143.
8. Erickson, H. (1994). 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*, **91**, 10114–10118.
9. Krammer, A., Craig, D., Thomas, W. E., Schulten, K. & Vogel, V. (2002). A structural model for force regulated integrin binding to fibronectin's RGD-synergy site. *Matrix Biol.* **21**, 139–147.
10. Hocking, D. C., Sottile, J. & McKeown-Longo, P. J. (1994). Fibronectin's III-1 module contains a conformation-dependent binding site for the amino-terminal region of fibronectin. *J. Biol. Chem.* **269**, 19183–19187.
11. Sechler, J. L., Rao, H., Cumiskey, A., Vega-Colon, I., Smith, M., Murata, T. & Schwarzbauer, J. (2001). A novel fibronectin binding site required for fibronectin fibril growth during matrix assembly. *J. Cell Biol.* **154**, 1081–1088.
12. Ingham, K. C., Brew, S., Huff, S. & Litvinovich, S. (1997). Cryptic self-association sites in type III modules of fibronectin. *J. Cell. Biochem.* **272**, 1718–1724.
13. Litvinovich, S. V., Brew, S., Aota, S., Akiyama, S., Haudenschild, C. & Ingham, K. (1998). Formation of amyloid-like fibrils by self-association of a partially unfolded fibronectin type III module. *J. Mol. Biol.* **280**, 245–258.
14. Carnemolla, B., Leprini, A., Allemanni, G., Saginati, M. & Zardi, L. (1992). The inclusion of the type III repeat ED-B in the fibronectin molecule generates conformational modifications that unmask a cryptic sequence. *J. Biol. Chem.* **267**, 24689–24692.
15. Hocking, D. C., Smith, R. K. & McKeown-Longo, P. J. (1996). A novel role for the integrin-binding III-10 module in fibronectin matrix assembly. *J. Cell Biol.* **133**, 431–444.

16. Bultmann, H., Santas, A. & Peters, D. (1998). Fibronectin fibrillogenesis involves the heparin II binding domain of fibronectin. *J. Biol. Chem.* **273**, 2601–2609.
17. Krammer, A., Lu, H., Isralewitz, B., Schulten, K. & Vogel, V. (1999). Forced unfolding of the fibronectin type III module reveals a tensile molecular recognition switch. *Proc. Natl Acad. Sci. USA*, **96**, 1351–1356.
18. Litvinovich, S. V. & Ingham, K. C. (1995). Interactions between type III domains in the 110 kDa cell-binding fragment of fibronectin. *J. Mol. Biol.* **248**, 611–626.
19. Cota, E., Steward, A., Fowler, S. B. & Clarke, J. (2001). The folding nucleus of a fibronectin type III domain is composed of core residues of the immunoglobulin-like fold. *J. Mol. Biol.* **305**, 1185–1194.
20. Plaxco, K. W., Spitzfaden, C., Campbell, I. D. & Dobson, C. M. (1997). A comparison of the folding kinetics and thermodynamics of two homologous fibronectin type III modules. *J. Mol. Biol.* **270**, 763–770.
21. Carrion-Vazquez, M., Oberhauser, A., Fowler, S., Marszalek, P., Broedel, S., Clarke, J. & Fernandez, J. (1999). Mechanical and chemical unfolding of a single protein: a comparison. *Proc. Natl Acad. Sci. USA*, **96**, 3694–3699.
22. Paci, E. & Karplus, M. (2000). Unfolding proteins by external forces and temperature: the importance of topology and energetics. *Proc. Natl Acad. Sci. USA*, **97**, 6521–6526.
23. Best, R. B., Li, B., Steward, A., Daggett, V. & Clarke, J. (2001). Can non-mechanical proteins withstand force? Stretching barnase by atomic force microscopy and molecular dynamics simulation. *Biophys. J.* **81**, 2344–2356.
24. Oberhauser, A., Badilla-Fernandez, C., Carrion-Vazquez, M. & Fernandez, J. (2002). The mechanical hierarchies of fibronectin observed with single molecule AFM. *J. Mol. Biol.* **319**, 433–447.
25. Oberdörfer, Y., Fuchs, H. & Janshoff, A. (2000). Conformational analysis of native fibronectin by means of force spectroscopy. *Langmuir*, **16**, 9955–9958.
26. Rief, M., Gautel, M. & Gaub, H. E. (2000). Unfolding forces of titin and fibronectin domains directly measured by AFM. *Advan. Expt. Med. Biol.* **481**, 129–136.
27. Oberhauser, A. F., Marszalek, P. E., Erickson, H. & Fernandez, J. (1998). The molecular elasticity of tenascin, an extracellular matrix protein. *Nature*, **393**, 181–185.
28. Isralewitz, B., Gao, M. & Schulten, K. (2001). Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.* **11**, 224–230.
29. Craig, D., Krammer, A., Schulten, K. & Vogel, V. (2001). Comparison of the early stages of forced unfolding of fibronectin type III modules. *Proc. Natl Acad. Sci. USA*, **98**, 5590–5595.
30. Paci, E. & Karplus, M. (1999). Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. *J. Mol. Biol.* **288**, 441–459.
31. Klimov, D. K. & Thirumalai, D. (2000). Native topology determines force-induced unfolding pathways in globular proteins. *Proc. Natl Acad. Sci. USA*, **97**, 7254–7259.
32. Lu, H. & Schulten, K. (2000). The key event in force-induced unfolding of titin's immunoglobulin domains. *Biophys. J.* **79**, 51–65.
33. Lu, H. & Schulten, K. (1999). Steered molecular dynamics simulation of conformational changes of immunoglobulin domain I27 interpret atomic force microscopy observations. *Chem. Phys.* **247**, 141–153.
34. Fisher, T. E., Marszalek, P. E. & Fernandez, J. M. (2000). Stretching single molecules into novel conformations using the atomic force microscope. *Nature Struct. Biol.* **7**, 719–724.
35. Gao, M., Lu, H. & Schulten, K. (2001). Simulated refolding of stretched titin immunoglobulin domains. *Biophys. J.* **81**, 2268–2277.
36. Marszalek, P. E., Lu, H., Li, H., Carrion-Vazquez, M., Oberhauser, A. F., Schulten, K. & Fernandez, J. M. (1999). Mechanical unfolding intermediates in titin modules. *Nature*, **402**, 100–103.
37. Lu, H., Isralewitz, B., Krammer, A., Vogel, V. & Schulten, K. (1998). Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.* **75**, 662–671.
38. Li, H., Mariano, C. V., Oberhauser, A. F., Marszalek, P. E. & Fernandez, J. M. (2001). Point mutations alter the mechanical stability of immunoglobulin modules. *Nature Struct. Biol.* **7**, 1117–1120.
39. Leahy, D. J., Aukhil, I. & Erickson, H. P. (1996). 2.0 Å crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. *Cell*, **84**, 155–164.
40. Brünger, A. T. (1992). *X-PLOR, Version 3.1: A System for X-ray Crystallography and NMR*, The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University.
41. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
42. Kalé, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N. *et al.* (1999). NAMD2: greater scalability for parallel molecular dynamics. *J. Comput. Phys.* **151**, 283–312.
43. MacKerell, A. D., Jr, Bashford, D., Bellott, M., Dunbrack, R. L., Jr, Evanseck, J., Field, M. J. *et al.* (1998). All-hydrogen empirical potential for molecular modeling and dynamics studies of proteins using the CHARMM22 force field. *J. Phys. Chem. ser. B*, **102**, 3586–3616.
44. Feller, S. E., Zhang, Y. H., Pastor, R. W. & Brooks, B. R. (1995). Constant pressure molecular dynamics simulation—the Langevin piston method. *J. Chem. Phys.* **103**, 4613–4621.
45. Darden, T., York, D. & Pedersen, L. (1993). Particle mesh Ewald. An $N\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092.
46. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD—visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38.

Edited by M. Levitt

(Received 28 May 2002; accepted 5 September 2002)