MOLECULAR DYNAMICS STUDIES OF
MECHANICAL PROTEINS

BY

MU GAO

B.A., Zhejiang University, 1995
M.S., Peking University, 1998

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2003

Urbana, Illinois
To my parents, my sister
and my wife.
Acknowledgments

This work would not be possible without the inspiration, generous help and support from many people I work with. Especially I would like to thank my advisor, Klaus Schulten, who introduced me to the field, guided and encouraged me on every step along the way. His broad vision, keen insights, and endless energy inspired me and drived my progress.

It was a pleasure and privilege to collaborate with Viola Vogel, Julio Fernandez, Hongbin Li, Andres Oberhauser, Mariano Carrion-Vazquez, Olivier Lequin, Iain Campbell, Matthias Wilmanans, and especially David Craig, for many stimulating conversations and productive joint works.

Special Thanks to Hui Lu and Barry Isralewitz, who guided me to conduct my first molecular dynamics simulation, and inspired me in many illuminating discussions.

I am deeply grateful to my friends and colleagues in the Theoretical and Computational Biophysics group at Beckman Institute in University of Illinois at Urbana-Champaign, including, in no particular order, Emad Tajkhorshid, Justin Gullingsrud, Alexander Balaeff, Thorsten Ritz, Ana Damjanovic, Ioan Kosztin, Melih Sener, Rosemary Braun, Fangqiang Zhu, Sanghyun Park, Chalermpol Kanchanawarin, Markus Dittrich, Deyu Lu, Aleksei Aksimentiev, Ilya Balabin, Shige-hiko Hayashi, Jerome Baudry, Margit Mollhoff, Elizabeth Villa, Jordi Cohen, Edgar Larios, Marcos Sotomayor, Tim Isgro, Jin Yu, Jim Phillips, John Stone, Kirby Vandivort, Fatemeh Araghi, Kirby Vandivort, Robert Brunner, Sameer Kumar, Wei Wang, David Hardy, David Brandon, and Gengbin Zheng.

I want to express my appreciation to the administration team in the group, Gila Budescu, Tim Skirvin, Jo Miller, Marilyn Lynch, and Sara Lee, for technical support and maintaining a wonderful working environment.

I am deeply indebted to my parents and my sister, for their love, support and encouragement that I could not imagine to live without. I also want to thank members of my parents’ families and
my friends, for their love and encouragement.

I would like to thank my dear wife, Rui Zhao, for her love, understanding, and support that encouraged me to finish seventy percent of the work presented in this thesis.

This work was supported by the US National Institutes of Health Research grants PHS5P41RR05969 and 1R01GM60946, as well as the US National Science Foundation grant NRAC MCA93S028. Part of computation jobs were performed at National Center for Supercomputing Applications (NCSA) and Pittsburgh Supercomputing Center (PSC).
# Table of Contents

Chapter 1  **Introduction** .......................................................... 1  
1.1 Mechanical proteins ................................................................. 1  
   1.1.1 Muscle protein titin .......................................................... 1  
   1.1.2 ECM protein fibronectin and cell receptor integrin .................. 3  
1.2 Experiments on mechanical proteins ........................................... 4  
1.3 Steered molecular dynamics ....................................................... 7  
1.4 Outline ....................................................................................... 8  

Chapter 2  **Mechanical unfolding of titin modules** ......................... 9  
2.1 Introduction ................................................................................. 9  
2.2 Methods ...................................................................................... 12  
2.3 Results ....................................................................................... 14  
   2.3.1 Equilibration. ....................................................................... 15  
   2.3.2 Constant velocity unfolding. .................................................. 15  
   2.3.3 Constant force unfolding. ....................................................... 18  
2.4 Conclusions and Outlook ............................................................. 21  

Chapter 3  **Refolding of stretched titin I27** .................................... 25  
3.1 Introduction .................................................................................. 25  
3.2 Methods ....................................................................................... 26  
3.3 Results ....................................................................................... 29  
   3.3.1 Refolding of Intermediates I .................................................... 29  
   3.3.2 Refolding of Intermediate II ................................................... 33  
   3.3.3 Summary of Refolding Simulations ......................................... 35  
3.4 Discussions ................................................................................. 36  

Chapter 4  **Mechanical unfolding pathways of FN-III$_{10}$** .............. 39  
4.1 Introduction .................................................................................. 39  
4.2 Materials and Methods ............................................................... 42  
4.3 Results ....................................................................................... 44  
   4.3.1 Mechanical unfolding intermediates. ....................................... 44  
   4.3.2 Intra-strand hydrogen bonds. ............................................... 46  
   4.3.3 Forced unfolding pathways. .................................................... 48  
   4.3.4 RGD-loop Conformations. ..................................................... 51  
4.4 Discussion ................................................................................... 51  
   4.4.1 Comparison with other models. ............................................. 52  
   4.4.2 Force-dependence of the unfolding pathways. ....................... 52  


List of Tables

2.1 Extension of titin modules under constant forces ........................................ 19
3.1 List starting states for refolding simulations of titin I27 ............................ 27
3.2 Summary of refolding simulations of I27 ................................................... 35
4.1 Location of intermediates and key events during FN-III_{10} unfolding .......... 45
5.1 Statistics of FN-III_{1} NMR structure. ..................................................... 64
6.1 List of SMD simulations of FN-III modules ............................................. 75
# List of Figures

1.1 Structures of muscle sarcomere and titin molecule ........................................ 2  
1.2 Fibronectin structure, assembly and integrin ................................................. 4  
1.3 Schematic of AFM stretching experiments ..................................................... 5  
1.4 AFM experiments of titin Ig domains and FN-III modules .............................. 6  

2.1 Diagram and structures of titin modules ....................................................... 10  
2.2 A sample system for SMD simulation ............................................................ 13  
2.3 Equilibration of titin I1 .................................................................................. 14  
2.4 Constant velocity stretching of titin modules ................................................ 16  
2.5 Key events during titin I1 unfolding ............................................................... 17  
2.6 Snapshots of Ig modules under low constant forces ......................................... 20  
2.7 Constant force unfolding of I1 .......................................................................... 21  
2.8 Hydrogen bond analysis during I1 unfolding .................................................. 22  
2.9 Comparison of oxidized and reduced I1 modules .......................................... 23  

3.1 Starting configurations for refolding simulations of I27 .................................... 28  
3.2 Refolding of I27 AB intermediates .................................................................. 29  
3.3 Refolding and re-stretching I27 AB intermediates .......................................... 30  
3.4 Hydrogen bond analysis during AB intermediate refolding ............................ 31  
3.5 Refolding and re-stretching I27 A’G intermediates ......................................... 32  
3.6 Hydrogen bond analysis during A’G intermediate refolding ........................... 34  

4.1 Diagram of fibronectin modules ................................................................. 40  
4.2 FN-III_{10} solvated in a water box ............................................................... 42  
4.3 Snapshots of FN-III_{10} unfolding ............................................................... 44  
4.4 Constant force unfolding of FN-III_{10} ....................................................... 46  
4.5 Inter-strand hydrogen bond analysis during FN-III_{10} unfolding ................. 47  
4.6 Multiple unfolding pathways of FN-III_{10} .................................................. 49  
4.7 Conformational changes of the RGD loop ..................................................... 50  
4.8 First passage time analysis of FN-III_{10} unfolding pathways ....................... 53  
4.9 Comparison to unfolding of I27 ...................................................................... 56  

5.1 FN-III_{1} NMR Structures ............................................................................ 63  
5.2 Inter-strand hydrogen bonds of equilibrated FN-III_{1} ................................... 65  
5.3 Mechanical unfolding intermediates of FN-III_{1} ......................................... 66  
5.4 Energy analysis of unfolding FN-III_{1} ........................................................ 68  

6.1 Fibronectin Monomer and FN-III modules ................................................... 73
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2 Comparison of the key events for different FN-III modules</td>
<td>77</td>
</tr>
<tr>
<td>6.3 Mean first passage time for unfolding FN-III modules</td>
<td>78</td>
</tr>
<tr>
<td>6.4 Comparison of mechanical stability for FN modules</td>
<td>80</td>
</tr>
<tr>
<td>6.5 Comparison of different potential models</td>
<td>81</td>
</tr>
<tr>
<td>6.6 Unfolding of FN-III$_{10}$ at low pH.</td>
<td>86</td>
</tr>
<tr>
<td>7.1 Schematic view of integrin $\alpha_V\beta_3$</td>
<td>90</td>
</tr>
<tr>
<td>7.2 Setup of integrin-RGD ligand simulations</td>
<td>92</td>
</tr>
<tr>
<td>7.3 Equilibration of integrin-RGD ligand complex</td>
<td>94</td>
</tr>
<tr>
<td>7.4 Coordination of ions at integrin’s binding sites</td>
<td>95</td>
</tr>
<tr>
<td>7.5 Snapshots of RGD ligand unbinding from $\alpha_V\beta_3$</td>
<td>98</td>
</tr>
<tr>
<td>7.6 Ligand unbinding from mutated $\alpha_V\beta_3$-ligand systems</td>
<td>99</td>
</tr>
</tbody>
</table>
List of Abbreviations

SMD  steered molecular dynamics.

AFM  atomic force microscopy.

Ig  immunoglobulin

FN  fibronectin

FN-III  fibronectin type III modules

ECM  extracellular matrix

RGD  Arg-Gly-Asp

PRARI  Pro-Arg-Ala-Arg-Ile

ENB  extra domain B

vdW  van der Waals

H-bond  hydrogen bond
Chapter 1

Introduction

1.1 Mechanical proteins

Mechanical force is ubiquitous in biological molecular processes, such as in stretching cytoskeleton inside cells, in assembling fibrils connecting cells, and in regulating ligand-receptor association at the cell surface. Correspondingly, there are numerous proteins that participate in such processes involving mechanical forces. We term these proteins mechanical proteins. Although the folded structure of many mechanical proteins in equilibrium states has been resolved, little is known about the dynamic response of these proteins when subjected to mechanical forces in cell environment. The present study aims at bridging this gap by simulating the non-equilibrium transformations of mechanical proteins using molecular dynamics. For this purpose, we studied basic functional units of three example mechanical proteins: the muscle protein titin, the extracellular matrix (ECM) protein fibronectin, and the mechanical cell surface receptor, integrin.

1.1.1 Muscle protein titin

Striated vertebrate muscle fibers are made of myofibrils that are assembled bundles of elementary contractile units called sarcomeres, about 2.2 µm long. The sarcomere, a tiny machine developing forces and displacement in muscle, has a highly organized architecture (see Fig. 1.1a) built mainly with precisely assembled filaments of three large proteins: actin forming so called thin filaments,

myosin forming thick filaments, and titin attaching to both thick and thin filaments.

Figure 1.1: (a) The structure diagram of the sarcomere, the elementary functional unit of striated muscle (adapted from reference [1]). (b) The multi-modular structure diagram of titin I-band, which is responsible for the passive elasticity of muscle. Atomic structures of two Ig domains have become available, as shown in cartoon representations of titin I1 (left) and I27 (right).

Titin, also known as connectin, is a roughly 1 µm long macromolecule which spans half of the muscle sarcomere and plays important roles in generating passive elasticity of muscle [1]. It is the largest known polypeptide (up to $\sim 4$ MDa) coded by the longest gene identified in the human genome [2, 3]. Titin is a string-like molecule primarily composed of $\sim 300$ tandem repeats of two types of domains, immunoglobulin-like (Ig) domains and fibronectin type III (FnIII) domains, as well as unique sequences located both in the N2 region and the PEVK region [4, 5]. The latter is name so because of the abundance of proline (P), glutamate (E), valine (V) and lysine (K) residues found in this region.

One important function of titin is to develop a passive force which keeps sarcomere components uniformly organized during muscle extension and ensure the integrity of sarcomere. The I-band
part of titin is responsible for titin extensibility and passive elasticity. It consists of two tandem segments of Ig domains (proximal Ig segment and distal Ig segment), separated by the N2 region and the PEVK region. Each of the Ig domains possesses about 90 amino acids forming a ⍺-sandwich motif [6, 7] (see Fig. 1.1b). The N2 region contains both Ig domains and a unique sequence. The structure of the PEVK region is not clear, probably containing repeats of polyproline type II helices [8].

When stretched beyond its resting length, titin I-band first extends links between domains from a coiled confirmation. Further stress results in the unfolding of the I-band polypeptide. The PEVK region is mechanically unstable and unfolds first [4, 9], followed by the unravelling of the N2 region [10, 11]. Unfolding of individual Ig domains occurs if additional extension is required [12, 13, 14, 15].

1.1.2 ECM protein fibronectin and cell receptor integrin

In tissues cells are surrounded by the extracellular matrix (ECM), a flexible network containing several classes of proteins and polysaccharides secreted by cell themselves. Fibronectin (FN), a large glycoprotein (∼450 KDa) found in all vertebrates, is an important mechanical component of the ECM and acts as a specific adhesive, forming elastic FN fibrils that attach to cell surface receptors integrins (see Fig. 1.2). Fibronectin fibrils provide substrate for cells and guide cell migration during embryogenesis, tissue repairing and wound healing [16]. Crosslinked through a disulfide bond at the C-terminus, two identical FN subunits form a dimeric FN molecule consisting of ∼20 different modules in each subunit. About 15 modules are type III (FN-III) modules, structurally exhibiting a ⍺-sandwich motif similar to Ig domains found in titin. Since homologous FN-III modules have been found in 4% of mammalian proteins, the structural motif of FN-III represent a common design of proteins.

Integrin comprises a large family of important cell adhesion proteins, providing dynamic, bidirectional linkages between ECM and cytoskeleton. Mechanical stress has been found to regulate the assembly of FN fibrils, a process termed fibrillogenesis, through integrin receptors that mechanically couple intracellular actin filaments to extracellular FN molecules (see reviews [18, 19]). It has been hypothesized that the stretching of FN fibrils unfolds individual FN-III modules [13],
Figure 1.2: (a) Diagram for fibronectin assembly. Each fibronectin molecule consists of two identical subunits crosslinked by a couple of disulfide bonds. Mechanical force generated by actin filaments inside cells is transmitted through transmembrane receptor integrins to stretch fibronectins, exposing buried "cryptic sites" that allow fibronectin molecules to bind to each other and form fibrils. (b) Schematic structural view of a fibronectin monomer, including FN-I, FN-II, and FN-III modules, and highlighting key functional sites [17]. (c) The secondary structure for a typical FN-III module. The lower three β-strands are shown in green, the upper four β-strands in magenta, and loops in gray. Alignment of tertiary structures for several FN-III modules FN-III_{10,12,13,14,EDB} demonstrating similarity in tertiary structures.

providing the necessary extension to the FN fibril that can be stretched four times as long as its relaxed length [20]. In addition to providing elasticity, the unfolding of FN-III modules is functionally relevant to mediating fibrillogenesis by exposing otherwise buried cryptic sites [19, 21], thus promoting laterally association between individual fibronectin molecules and forming fibrils.

1.2 Experiments on mechanical proteins

During the past decade, the advent and rapid development of single biomolecule techniques enable researchers to study individual proteins, DNA or RNA molecules. There are two types of single molecule experiments: one is the visualization of the dynamics of single molecules, e.g., through fluorescence spectroscopy [22, 23]; the other is the direct manipulation of single molecules, e.g., by atomic force microscopy [24, 25, 26], laser optical tweezers [27, 28, 29], magnetic beads [30], and biomembrane force probe [31]. One advantage of single molecule experiments is to reveal
heterogeneous properties that are inundated in the ensemble average of a bulk system. The AFM experiments of titin Ig domains and FN-III modules, for example, have revealed the distinctly different mechanical properties of individual modules from a single molecule [26, 32].

Figure 1.3: Schematic of AFM stretching experiments (taken from [33]). (a) Diagram of AFM experiment setup. Typically the displacement of substrate $\Delta Z_p$ is changed at constant velocity. The force generated can be calculated by measuring the bending of the AFM cantilever. (b) Stretching and unfolding a series of linked proteins. (c) Force-extension profiles obtained from AFM experiments.

Due to the relative simplicity and high resolution, AFM has become one of the most popular instruments for probing single molecules [24, 25]. In a typical AFM experiment as shown in Fig. 1.3, a sample is anchored between the AFM cantilever and a moving substrate controlled by a piezoelectric positioner with Ångstrom resolution. The force developed can be calculated at pN resolution by measuring the bending of the cantilever.

The mechanical stability of individual Ig modules has been probed in force-unfolding experiments using AFM [25, 34, 35, 36, 37, 26] (Fig. 1.4(a – c)). The spacing between two consecutive force peaks matches the extended length of one Ig from its folded state, proving that, when these multi-domain proteins are stretched, their domains unfold one by one. The high magnitude of the force peaks, dependent on the pulling speed and type of protein, implies that the Ig domains are
designed to withstand significant stretching forces. At a pulling speed of 1 \( \mu \text{m/s} \), for example, AFM unfolding of titin Ig domains requires about 200 pN [25]. One remarkable property of these mechanical proteins is that they unfold reversibly. Refolding rates of around 1 s\(^{-1} \) were reported in AFM forced-unfolding experiments of titin I27 [25, 34].

![AFM experiments of titin Ig domains and fibronectin FN-III modules. Data and source for each type of modules are: (a) I27 ([35]); (b) I4 and I32 ([26]); (c–e) FN-III\(_{10}\), FN-III\(_{1}\), FN-III\(_{12}\), FN-III\(_{13}\) ([32]). The AFM force-extension data are fitted to the worm-like-chain model [38]. Note that distal Ig domains like I27 and I32 exhibit an 'hump' before the major force peak, whereas proximal Ig domains like I4 does not. With I27 serving as a fingerprint, stretching of FN-III\(_{1}\)-I27 repeats revealed that FN-III\(_{1}\) has at least one meta-stable intermediate. The force-unfolding experiments of FN-III exhibit a similar sawtooth pattern [39, 40, 32] (Fig. 1.4(d – e)). With genetically engineered identical repeats of FN-III modules or repeats of FN-III\(_n\)-I27 dimers (see Fig. 1.4(d)) where well-studied I27 serves as a mechanical fingerprint, the mechanical stability of individual modules can be quantified in terms of rupture forces. Suprisingly, although FN-III modules possess quite similar tertiary structures, they demonstrate distinctly different mechanical stability with peak rupture forces ranging from 75 – 200 pN at a pulling velocity of 0.6 \( \mu \text{m/s} \) [32]. More interestingly, FN-III modules such as FN-III\(_{1}\) have pronounced meta-stable
conformations, indicating the existence of unfolding intermediates. The mechanical variability and the intermediate states of FN-III modules can be closely related to the function of FN, such as in fibrillogenesis, the process of fibronectin fibril self-assembly initiated by the tensile forces transmitted through integrins.

AFM experiments have also been conducted on receptor-ligand systems. The first AFM pulling experiments studied the adhesion between biotin-avidin [24]. The strength of integrin-ligand systems have also been probed, including αVβ3/RGD-ligands [41], α5β1/fibronectin [42]. In addition to AFM experiments, laser optical tweezer has been exploited to study the binding affinity between integrin αIIbβ3/fibrinogen [43]. These experiments provides insights into the protein-protein association.

1.3 Steered molecular dynamics

It is amazing that individual Ig domains and FN-III modules can display distinct mechanical properties, even though they are built with seemingly identical structures. While single molecule experiments provide valuable dynamic force spectroscopy for mechanical proteins, the corresponding conformational changes of stretched proteins cannot be observed directly because conventional structure resolving methods such as X-ray and NMR are not applicable. On the other hand, the static 3D structures for two Ig domains [6, 7] and nearly half FN-III modules [44, 45] have been resolved at high resolution, but these structures cannot be directly related to the dynamic unfolding scenarios that one sees in AFM experiments. To link these static structures to their dynamically changing non-equilibrium states induced under external forces, steered molecular dynamics (SMD) simulations have been introduced to provide an atomic-level description of the unfolding processes for various mechanical proteins [46, 47] (reviewed in [48, 49, 50]).

Starting from an equilibrated x-ray or NMR structure, the dynamics of the protein are recorded during force-induced unfolding simulations, thus providing a better understanding of the structure-function relationship of the simulated protein. One advantage of SMD over conventional molecular dynamics is the ability to induce relatively large conformational changes in molecules on the nanosecond time scales accessible to computation. Two main pulling protocols have been used in simulating mechanical protein unfolding/unbinding: constant velocity pulling and constant force
pulling. Constant velocity pulling mimics the AFM experiment by applying a harmonic force to the system and, hence, allows one to compare SMD data with AFM recordings and to provide interpretation for experiments. Constant force pulling allows extensive studies of the barrier crossing event along the potential surface of the unfolding pathway. In combination with a statistical analysis, one can provide a quantitative description of the potential barrier separating the folded and unfolded domains and compare it to experimental observations.

1.4 Outline

The present dissertation describes molecular dynamics studies of three types of mechanical proteins. In Chapter 2 we will cover the unfolding of titin Ig domain I1, and compare to I27 that had been previously studied. Chapter 3 presents the refolding of stretched Ig domain I27. In Chapter 4 to 6 we discuss mechanical properties of fibronectin type III modules, focusing on the multiple unfolding pathway of FN-III_{10} (Chapter 4), the intermediate of FN-III_{1} (Chapter 5), and the mechanical stability of all structurally known FN-III modules (Chapter 6). Finally in Chapter 7 we will examine the ligand binding between integrin αVβ3 and an RGD containing ligand.
Chapter 2

Mechanical unfolding of titin modules

2.1 Introduction

Titin, \(\sim 1 \mu m\) long, is the longest covalently linked protein known in the human genome [3]. Spanning over half of the muscle sarcomere, a single titin molecule extends from the Z disc to the M line through both the A band and I band sections of sarcomere. Titin primarily consists of \(\sim 300\) modules in two motif types, immunoglobulin-like (Ig) and fibronectin type III (FnIII) domains. The titin A band is composed of regular arrangements of these domains that bind to myosin and, hence, cannot be extended upon tension. The titin I-band, however, is extensible and is thought to be responsible for the passive elasticity of muscle [51, 52, 53, 54, 55, 56]. In cardiac muscle the titin I-band contains four structural units: the proximal Ig region, the N2B or N2BA segment, the PEVK region, and the distal Ig region [5] (a diagram of titin I-band is shown in Fig. 2.1a). The PEVK region contains 163 or more amino acids, 70% of which are proline, glutamate, valine, and lysine. It is a mixture of unstable coiled conformations and polyproline type II helix, and easily elongates to develop passive tension under small forces of up to 20 pN [9, 57, 8, 58]. The structure of N2B or N2BA is still unknown. Studies have shown that the N2B segment is critical for reversible extensibility of cardiac myofibrils [59]. The proximal and distal Ig regions in human cardiac muscle have 15 and 22 tandem Ig domains, respectively. It has been suggested that some Ig domains unfold to provide extension for the over-stretched muscle [13, 60, 61, 15].

The reversible unfolding of titin Ig domains has been demonstrated in studies using atomic force microscopy (AFM) and optical tweezers [25, 34, 27, 28]. The AFM experiments have shown

\[^{3}\text{This chapter was adapted from the manuscript: M. Gao, M. Wilmanns, and K. Schulten, Steered molecular dynamics studies of titin II domain unfolding, (2002), Biophysical Journal. 83:3435–3445.}\]
Figure 2.1: (a) Modular structure of the I band section of cardiac titin. (b) Cartoon representations of the structures of titin I1 (left) and I27 (right) domains. Color scheme: sulfur atoms (yellow), A-strand (blue), A’-strand (red), B-, E-, C-strands (purple), G-, F-, C-, D-strands (cyan). Backbone hydrogen bonds between strands A and B and between A’ and G are represented as black dashed lines. (c) Sequence alignment of I1 and I27 domains following Mayans et al. [7]. The secondary structure is shown according to I1.

**a characteristic sawtooth pattern in force-extension profiles, which can be attributed to the subsequent unraveling of several Ig domains.** In order to exclude heterogeneous effects caused by different modules, polyproteins composed of identical I27 or I28 modules were genetically engineered and stretched in AFM experiments [62, 35, 63]. Analysis of the sawtooth force-extension profiles revealed that the unfolding of domains occurs in two steps within the ms timescale: forces of 50-150 pN extend the domains by \( \sim 7 \) Å; forces above 150 pN extend the domains further, inducing complete unfolding.

To interpret the AFM experiments at the atomic level, conformational changes of the Ig domains during the unfolding processes must be known. The AFM experiments motivated a series of Steered Molecular Dynamics (SMD) simulations of the unfolding/refolding pathways of the module [47, 64, 35, 65, 66] as reviewed in [48]. The SMD simulations of I27 revealed that the two-step
unfolding pathway observed in AFM experiments corresponds to two sequential events of inter-strand hydrogen bond rupture, in which two sets of hydrogen bonds connecting $\beta$-strands A and B, and $\beta$-strands A’ and G (see structure of I27 in Fig. 2.1b), are broken. At forces around 100 pN the first set of hydrogen bonds near the N-terminus breaks with a concomitant 4 to 7 Å extension, in agreement with the extension-force profile recorded in AFM experiments; the second set of hydrogen bonds breaks at forces of above 200 pN and initiates the complete unfolding [35]. The height of the kinetic barrier separating the folded and unfolded states has been probed in both AFM experiments [34] and SMD simulations [64]. Moreover, the scenarios of unfolding provided by SMD simulations using explicit solvent models revealed a key role of water molecules: the unfolding barrier is crossed with the help of water molecules that attack inter-strand hydrogen bonds [65]. The competition for hydrogen bond partners with water molecules is also important for the backbone oxygen and hydrogen atoms when they seek to reform hydrogen bonds in the spontaneous refolding process of I27: by driving water molecules away and reforming six A’-G backbone hydrogen bonds, a stretched I27 domain has been seen to spontaneously refold [66].

Alternative simulation approaches have also been carried out by other researchers. Paci and Karplus simulated the unfolding of I27 by employing implicit solvent models [67, 68]. Their simulations are computationally less expensive than the simulations based on explicit solvent described above. However, omitting water molecules yields lower force peaks than otherwise. Klimov and Thirumalai performed simulations using lattice models [69] and off-lattice models [70]. The latter generated unfolding peak forces of I27 in agreement with the measurements from AFM experiments. Although the correlation between unfolding and the rupture of intra-strand hydrogen bonds could not be established, simulations of off-lattice models apparently reproduced the same unfolding pathway of I27 as earlier molecular dynamics simulations of all-atom models.

Our previous SMD simulations of titin modules were focused on I27 from the distal Ig region of the titin I-band, until recently the only Ig domain with known structure. The now available crystallographic structure of titin I1 determined at 2.1 Å resolution [7] provides structural information of an Ig domain from the proximal Ig region of titin and a long awaited opportunity to compare modules from different Ig regions of the titin I-band. The secondary structures and sequences of I1 and I27 are compared in Fig. 2.1b,c. Both modules are built in a motif called $\beta$-sandwich, formed
by two $\beta$-sheets with four $\beta$-strands in each sheet. Compared to I27, however, I1 has two features that lead to different mechanical properties. First, I1 has a disulfide bond connecting its C- and E-strands, which restricts the relative movement of the two $\beta$-sheets in an oxidizing environment, i.e., when the bond is formed. Since 40\% of Ig domains in titin I-band have the potential to form a disulfide bond [7], the role of this bond in protecting the integrity of I1 has general implications to other homologous Ig domains. Second, I1 has more backbone hydrogen bonds between its A- and B-strands than between its A’ and G’ strands (six versus five), whereas I27 has fewer A-B hydrogen bonds than A’-G hydrogen bonds (two versus six). Previous SMD simulations have shown that the inter-strand hydrogen bonding structure of the A-B and the A’-G strand pairs are the key determinant for the mechanical response of I27; one would expect, therefore, to observe a difference in the mechanical function of I1 and I27.

In this chapter we present steered molecular dynamics simulations that compare the stretching and unfolding of I1 and I27. We will first describe the modeling and simulation procedure. An analysis of the pattern of inter-strand hydrogen bonds will be provided for latter classification of the unfolding pathways of these modules. The implications of the simulation results to upcoming AFM experiments will be discussed.

2.2 Methods

Initial atomic coordinates of the titin I1 domain were taken from the Protein Data Bank (entry code 1G1C [7]). Hydrogen atoms were added to the protein using X-PLOR [71]. Cysteines residues, Cys$^{36}$ and Cys$^{61}$, were patched for modeling the disulfide bridge of an oxidized I1 domain, while for modeling a reduced I1 domain the Cys residues were not bonded. A TIP3 water [72] sphere of 72 Å diameter was used for solvating the I1 domains, resulting in systems of 18,072 atoms for the oxidized I1 domain and 18,074 atoms for the reduced I1 domain. Fig. 2.2 shows the model of the oxidized I1 domain. All molecular dynamics simulations were performed using the program NAMD [73] with the CHARMm22 force field [74].

Simulations of oxidized and reduced I1 were carried out using the same protocol. First, an I1 system was minimized for 2000 conjugate gradient steps. Following the minimization, the system was heated from 0 K to 300 K in 10 ps, and was coupled to a 300 K heat bath for additional 10 ps.
The temperature control was released, and the whole system was subsequently equilibrated for 1 ns. Finally SMD simulations were carried out by fixing the C\textsubscript{\alpha} atom of the N-terminus of I1, and applying external forces to the C\textsubscript{\alpha} atom of the C-terminus. The forces were directed along the vector from the pulled atom to the fixed atom (see Fig. 2.2).

Both constant force and constant velocity protocols were used for the SMD simulations. In the latter case the pulling atom is harmonically constrained with a force $F = -k(x - vt)$, where $k$ is the spring constant, $x$ is the coordinate of the pulling atom, $v$ is the velocity of the atom, and $t$ is the time. The value of $k$ was set to 11.7 k\textsubscript{B}T/Å\textsuperscript{2}, corresponding to a thermal fluctuation of the pulling atom of $\sqrt{k_{\text{B}}T/k} = 0.29$ Å. An integration time step of 1 fs and a uniform dielectric constant of 1 were chosen. For calculating electrostatics and van der Waals interactions, a cut-off was employed, switching the interactions smoothly off between 10 Å and 13 Å.

Including simulations of I27 following the same protocol, fourteen SMD runs, altogether over 50 ns, were completed using a cluster of 32 1.33GHz Athlon processors, on which a 1 ns simulation requires $\sim$30 hours wall clock time. The fourteen simulations were carried out under different conditions, e.g., with different values of constant force or with different pulling velocities. These SMD simulations are referred to as cf-SMD(force value) for constant force stretching and cv-SMD(velocity
The analysis of molecular structures and hydrogen bond energies were conducted using X-PLOR and VMD. Atomic coordinates were saved every 1 ps. The coordinates for the pulling atom were saved every 10 fs for cv-SMD simulations, and were saved every 100 fs for cf-SMD simulations. The extension of the protein is defined as the change of the end-to-end distance between the two termini. An explicit hydrogen-bonding energy term was used in hydrogen bond energy calculations, with parameters adopted from param11.pro in X-PLOR.

2.3 Results

![Figure 2.3: Structure and stability of inter-strand hydrogen bonds in I1. Shown are hydrogen bonds between β-strands A, B, A’ and G of oxidized (a,b) and reduced I1 (d,e) at the end of the 1 ns equilibration, together with the hydrogen bond energy fluctuations of Gln18(H)-Gln96(O) during the equilibration of oxidized (c) and (d) reduced I1. The distances between oxygen (red) and hydrogen (white) of hydrogen bonds are given in Å. Other backbone atoms are shown in blue (A-, B-strands) and cyan (A’, G-strands). Water molecules are shown in green. Hydrogen bonds are represented as black dashed lines.](image-url)
2.3.1 Equilibration.

During the 1 ns free dynamics equilibration, both oxidized I1 and reduced I1 remained stable, exhibiting a C\_\alpha RMSD from the crystal structure of < 1.25 Å and an all-atom RMSD of < 2.0 Å. The backbone hydrogen bonding structures between A- and B-strands and between A\-'- and G-strands at the end of the equilibration are shown in Fig. 2.3. Except for the bond Q18(H)-Q96(O), all six A-B backbone hydrogen bonds, E3(O)-K31(H), E3(H)-K31(O), K6(O)-V29(H), K6(H)-V29(O), F8(H)-R27(O) and E9(O)-R27(H), and four A\-'-G bonds, Q14(H)-F92(O), Q14(O)-L94(H), V16(H)-L94(O), V16(O)-Q96(H), remained stable, i.e., these hydrogen bonds occasionally broke, but reformed quickly. Hydrogen bond Q18(H)-Q96(O), the A\-'-G hydrogen bond nearest to the C-terminus, appears to be weak. During the equilibration of oxidized I1, polar residues Q18 and Q96 continuously suffered from attacks by surrounding water molecules. As a result, the bond Q18(H)-Q96(O) was found to have been dissociated and reformed several times, reflected in the hydrogen bond energy fluctuations shown in Fig. 2.3c. At the end of the equilibration Q18(H) and Q96(O) form hydrogen bonds with solvent water (Fig. 2.3b). During the equilibration of reduced I1, Q18(H) broke up with its bond partner Q96(O) at 310 ps (Fig. 2.3f), forming a new hydrogen bond with A97(O) (Fig. 2.3e). The formation of this bond resulted in a more compact domain. The length of the reduced I1 module, defined as the distance between the two terminal C\_\alpha atoms, is 2 Å shorter than that of oxidized I1. However, the Q18(H)-A97(O) bond is easily broken under forces as small as 50 pN, leading to an additional extension of 2 Å as discussed below.

2.3.2 Constant velocity unfolding.

The results of forced unfolding of both oxidized and reduced I1 domains with constant velocities of 0.1 Å/ps and 0.5 Å/ps are compared in Fig. 2.4, together with results of unfolding I27. I1 domains exhibit a strong resistance against external forces in the extension range of 5-16 Å, a region broader than the major burst region of I27 of 12-15 Å [47]. Overcoming the initial resistance of I1 domains requires slightly weaker forces than I27 at the same pulling speed. To unfold Ig domains at 0.5 Å/ps, for example, oxidized I1 requires a peak force of 2397 pN recorded at 10 Å extension; reduced I1 requires a peak force of 2090 pN at 11 Å extension; in contrast, I27 requires a stronger peak force of 2479 pN. Unfolding these domains at a velocity of 0.1 Å/ps yields the same ordering of force...
peak values, but reduced by 20%-30%. The lower peak forces required for unfolding I1 implies that I1 is slightly less stable than I27. Unraveling the module to extensions beyond the main force peak, which corresponds to the unfolding barrier, requires weaker and weaker forces until the domain is fully extended when forces rise again. Since oxidized I1 contains a disulfide bond, the domain can only extend to $\sim 220$ Å as shown in Fig. 2.4, whereas reduced I1 can be stretched to $\sim 300$ Å, the length of the completely extended I1 domain.

What conformational changes of I1 domains can be related to the main peak forces? For both I1 domains, the peak force coincides with a burst of backbone hydrogen bonds between $\beta$-strands A and B and between $\beta$-strands $A'$ and G, as illustrated in Fig. 2.5 through the snapshots from
Figure 2.5: Representative force versus time curves (left) during early stages of cv-SMD(0.1 Å/ps) simulations and snapshots of key events (a-f). The results for oxidized and reduced I1 domains are shown at the top and bottom, respectively. Arrows mark the instances when the corresponding snapshots were taken. (a) At 100 ps and extension of 7 Å, a pair of backbone hydrogen bonds bridging Glu$^3$ and Lys$^{31}$ ruptures first. (b) The protein extends another 7 Å and the remaining backbone hydrogen bonds between A- and B-strands break at 170 ps. (c) 12 ps later four intact backbone hydrogen bonds between A' and G-strands rupture at 16 Å extension. The same order of hydrogen bond breaking occurs for reduced I1 at similar extensions (d-f). In all snapshots intact hydrogen bonds are represented as thick black lines while broken hydrogen bonds are shown as thin lines.

cv-SMD(0.1 Å/ps) simulations. In these simulations, the disruption of backbone hydrogen bonds started from a pair of bonds near the N-terminus, between Glu$^3$ on β-strand A and Lys$^{31}$ on β-strand B. For example, during the unfolding of oxidized I1, a peak force of 1600 pN was encountered at 100 ps when two E3-K31 hydrogen bonds were seen to break (see Fig. 2.5a). Following this a second force peak of 1677 pN, measured at 168 ps, proceeded the rupture of the remaining four A-B backbone hydrogen bonds at 170 ps and of four A'-G hydrogen bonds at 182 ps (see Fig. 2.5b-c). The extensions connected with the rupture of these bonds are 14 Å and 16 Å. Unfolding of reduced I1 exhibits the same sequence of ruptures of inter-strand A-B and A'-G hydrogen bonds. The peak force of 1655 pN was found to follow the disruption of four A-B hydrogen bonds at 159 ps and to precede the rupture of four A'-G hydrogen bonds. Reduced I1 has one more A'-G backbone hydrogen bond, Q18(H)-A97(O). This bond broke within the first 50 ps when the C-terminus was straightened and the bond did not contribute to the major force peaks. During the disruption
of A-B and A’-G hydrogen bonds, the secondary structure of the remaining part of the module were maintained. After the burst of A-B and A’-G hydrogen bonds, the module gradually lost its secondary structure by separating β-strands. The force peaks beyond the main reaction region from 5 Å to 16 Å extension are due to disruption of packing interactions and zipper-like unraveling of individual backbone hydrogen bonds.

2.3.3 Constant force unfolding.

Constant forces of 50 pN, 200 pN, 650 pN and 750 pN have been applied to the I1 domain. Fig. 2.6 and Table 2.1 compare the extension of reduced I1 and I27 at forces of 50 pN and 200 pN for up to 10 ns. Under 50 pN the extension of I1 fluctuated between 1 Å to 3 Å, corresponding to the disruption (Fig. 2.6a) and re-formation of the hydrogen bond Q18(H)-A97(O) near the C-terminus. Applying a stronger force of 200 pN prohibited the reformation of this bond, and broke additionally two A-B hydrogen bonds between Glu^3 and Lys^31 at 5.7 ns (Fig. 2.6b). However, the average extension of I1 increased only ∼2.0 Å from 2.0 Å at 50 pN to 4.3 Å at 200 pN, because the preserved other four hydrogen bonds between A- and B-strands prevented further extension. In comparison, titin I27 responded differently to the stretching forces. For a constant force of 50 pN the module appeared rigid with no rupture of any inter-strand hydrogen bond observed (Fig. 2.6c). As a result, the module experienced only a small extension of less than 1 Å on average. Being Stretched with a constant force of 200 pN, however, I27 extended up to 8 Å, mainly due to the disruption of a pair of hydrogen bonds between A- and B-strands (Fig. 2.6d). Re-formation of these two bonds at 9.18 ns resulted in an extension drop from over 7 Å to less than 4 Å. On average, the extension of I27 increased from 0.6 Å at 50 pN to 6.3 Å at 200 pN. Since this ∼6 Å elongation characterizes an I27 intermediate [35], which is associated with the unraveling of A-strand from B-strand, I1 seems unlikely to have a similar intermediate as I27 because the separation of its A-strand is prevented before crossing the main unfolding barrier and the 2 Å change of extension prior to the barrier crossing is very small.

As the forces were increased, the breaking of all A-B and A’-G hydrogen bonds and separation of A- and A’-strands from the remaining fold were observed. Fig. 2.7 demonstrates that unfolding oxidized II domain happens in three key steps, discernible as three plateaus in the extension versus
Table 2.1: Extension of titin domains from their equilibrated structures at constant forces of 50pN and 200pN. The extensions were calculated as the average of the extension during the last 4 ns of simulations.

<table>
<thead>
<tr>
<th>titin domains</th>
<th>extension (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>reduced II</td>
<td>2.0 ± 2.0</td>
</tr>
<tr>
<td>127</td>
<td>0.6 ± 2.0</td>
</tr>
</tbody>
</table>

An analysis of the A-B inter-strand hydrogen bond energy during the cf-SMD(750pN) simulation of oxidized I1 is provided in Fig. 2.8. Prior to the stretching, the equilibrated I1 domain has six inter-strand hydrogen bonds between its A- and B-strands, as shown in Fig. 2.8. Upon stretching, two hydrogen bonds between E3 and K31 broke first. This pair of bonds, functioning like a 'lock' to protect the integrity of the I1 domains, was also observed to be the first hydrogen bonds broken in all other cf-SMD simulations with forces higher than 650 pN. As shown in Fig. 2.8, bonds E3(H)-K31(O) and E3(O)-K31(H) were destabilized at 100 ps and 70 ps, respectively, with an energy jump from -3.5 kcal/mol to -1 kcal/mol. Fluctuating with an energy around -1.0 kcal/mol for
Figure 2.6: Extension-time profiles of reduced I1 (left) and I27 (right) for constant force of 50 pN and 200 pN, together with snapshots (a-d) of these two modules. (a) The extension fluctuation of I1 from 1 Å to 3 Å under 50 pN can be related to the rupture and re-formation of the hydrogen bond between Q18(H) and A97(O). (b) For a constant force of 200 pN two hydrogen bonds between Glu3 and Lys31 of I1 broke at 5.7 ns, but this rupture only led to an additional extension of less than 2 Å. (c) Extension of I27 fluctuated around 1 Å for a constant force of 50 pN. (d) For a constant force of 200 pN I27 elongated 6 Å further by disrupting the two A-B hydrogen bonds.

about 350 ps, both bonds were completely broken at 420 ps. The remaining four A-B hydrogen bonds, K6(O)-V29(H), K6(H)-V29(O), F8(H)-R27(O) and E9(O)-R27(H), maintained intact until 540 ps and were completely ruptured at ~600 ps as shown in Fig. 2.7b.

The reduced I1 domain exhibits the same sequence of hydrogen bond ruptures during SMD simulations as oxidized I1. The absence of the disulfide bond between two β-sheets, however, reduces the mechanical stability of the module. Fig. 2.9 shows a comparison of the results from SMD simulations of oxidized and reduced I1 domains for forces of 650 pN and 750 pN. Three transition states can be identified in Fig. 2.9 as plateaus or shoulders in extension-time profiles from SMD simulations. Similar to unfolding oxidized I1 analyzed above, the first plateau at ~5 Å, shown in Fig. 2.9, corresponds to disrupting the hydrophobic core and overcoming the resistance imposed by A-B backbone hydrogen bonds, especially between residue E3 and K31. The second
Figure 2.7: Extension-time profile and snapshots (a-d) of key events from a cf-SMD(750pN) simulation of the oxidized I1 domain. (a) Two backbone hydrogen bonds between A- and B-strands broke at 410 ps. (b) The remaining four A-B hydrogen bonds broke at 585 ps. (c) Two inter-strand hydrogen bonds between A’- and G-strands broke at 585 ps, followed (d) by the rupture of the remaining three bonds at 1110 ps.

plateau/shoulder, corresponding to the separation of A-strand from B-strand, are generally short and were found at an extension of $\sim 9$ Å for oxidized I1, and of $\sim 11$ Å for reduced I1. Oxidized I1 turns out to be more stable than reduced I1 when one compares the third plateau/shoulder, which corresponds to disturbing the four A’-G backbone hydrogen bonds. A constant force of 650 pN could not separate A’-strand from G-strand of oxidized I1 within 2.5 ns, whereas for reduced I1 the A’-strand was peeled away from the G-strand by the same force after the rupture of the A’-G hydrogen bonds at 2.1 ns. The results suggest that the disulfide bond enhances the stability of oxidized I1 by restricting the disruption of backbone hydrogen bonds between A’ and G-strands.

2.4 Conclusions and Outlook

I1 and I27 are homologous modules with the same $\beta$-sandwich architecture. SMD simulations of I1 show that the main mechanical resistance to an external force occurs within the initial 16 Å
Figure 2.8: Energy analysis of six backbone hydrogen bonds between A- and B-strands during a cf-SMD(750pN) simulation of oxidized I1. The equilibrated structure of β-strands A and B are shown top left. The energy versus time profiles for individual bonds reveal that the inter-strand hydrogen bonds rupture in two steps. Two of them, between Glu³ and Lys³¹ (top right), broke earlier than the other four bonds (bottom). The latter four bonds broke concurrently at ∼580 ps.

extension and arises mainly from its inter-strand hydrogen bonding between A-B and A’-G strands. The force peaks observed in constant velocity pulling simulations (see Fig. 2.5) or the plateaus observed in constant force pulling simulations (see Fig. 2.7), corresponding to crossing the barrier that separates folded and unfolded states, coincide with the breaking of A-B and A’-G hydrogen bonds. Since I1 modules encounter the unfolding barrier very early, namely at extension of ∼5 Å, it is likely that the transitional state of unfolding I1 is close to its native state, similar to I27 which has a transition state of only a 3 Å extension from the native state [34]. After the rupture of the two clusters of hydrogen bonds and separation of the A- and A’-strands, the remaining inter-strand backbone hydrogen bonds are easily ruptured by “unzipping”. These observations are similar to what has been reported from SMD simulations of I27 [47, 64, 35, 65].

Although the general unfolding pathways of I1 and I27 are similar, the modules exhibit a different mechanical design in terms of A-B and A’-G backbone bonding structure and the presence of a disulfide bridge, leading to different mechanical responses upon stress. First, the mechanical
Figure 2.9: Representative extension-time profiles from cf-SMD simulations of oxidized (top) and reduced I1 (bottom). The snapshots of I1 domains at the end of the simulations are shown on the right. The shaded areas in profiles correspond to key transition states for crossing barriers formed by hydrogen bonds: Glu$^3$-Lys$^{31}$ hydrogen bond rupture (cyan); rupture of the remaining bonds bridging A- and B-strands (light purple); rupture of the bonds between A’- and G-strands (purple).

The stability of I1 is largely due to the six backbone hydrogen bonds between its A- and B-strands, not the A’-G hydrogen bonds as for I27. This is because I1 has a larger number of hydrogen bonds between A- and B-strands than of hydrogen bonds between A’- and G-strands. Second, the mechanical stability of I1 domains is slightly less than that of I27, as reflected in the maximum unfolding force shown in Fig. 2.4. The main reason is that in I1 domains no more than four backbone hydrogen bonds breaking concurrently during stretching (six A-B bonds break in two steps: two bonds break first, followed by the break of the remaining four bonds), which contribute to the force peaks recorded in simulations; in contrast, I27 has six A’-G backbone hydrogen bonds that break simultaneously [47]. Comparison of the extension of I1 and I27 at forces of 50 and 200 pN (see Table 2.1) leads to a third difference. I27 exhibits an $\sim$6 Å extension 'hump' revealed in force-extension curves [35]. I1 domains, however, should not exhibit such hump at forces up
to 200 pN because of the large number of A-B inter-strand hydrogen bonds. A fourth difference between I1 and I27 is due to the disulfide bond. Our simulations of oxidized and reduced I1 revealed that the disulfide bridge between Cys$^{36}$ and Cys$^{31}$ increases the mechanical stability and limits the extension of I1 within 220 Å. Indeed, reduced spacing between force peaks has been observed in unfolding oxidized I1 (J. Fernandez, private communication).

Limited by current available computational resources, the timescale accessible to SMD simulations, i.e., nanosecond, is six orders of magnitude shorter than the millisecond timescale over which titin modules are stretched and unfolded in AFM experiments. This timescale gap requires a pulling velocity used in SMD simulations about six order of magnitude faster than in experiments, leading to a discrepancy in the unfolding forces as discussed previously [76, 64]. This problem may be solved in the future with simulations applying reduced pulling speeds close to experimental values, requiring, however, vastly improved computational resources.

Nevertheless the hypothesis suggested by our SMD simulations that hydrogen bonds protect Ig domains is worth experimental examinations. For example, I27 mutants in which either A-B or A'-G hydrogen bonds were disrupted have been engineered and stretched with AFM after suggestions derived from SMD simulations [35, 36]. Disrupting A-B hydrogen bonds of I27 through mutation eliminates the pre-burst intermediate [35]. Mutating residues involving A'-G hydrogen bonds produced proteins that require weaker unfolding forces than wild type I27 [36]. A similar approach may be applied to I1. For example, by mutating residue K6 to proline reduces the number of A-B hydrogen bonds, and, correspondingly, separation of the A-strand should occur at weaker forces. AFM stretching of this mutant may even produce a pre-burst 'hump', corresponding to A-strand separation before the burst of A'-G backbone hydrogen bonds.

New structures of Ig domains will likely be solved in the near future, providing further opportunities to compare their mechanical responses to AFM generated force and simulated force, as well as to develop an understanding of the evolutionary design and function of the remarkable protein titin.
Chapter 3

Refolding of stretched titin I27

3.1 Introduction

If proteins need to unfold to fulfill their physiological role, they must be able to do it reversibly. The reversible unfolding is indeed demonstrated in both AFM and optical tweezer experiments of titin Ig domains [25, 34, 27, 28]. In AFM experiments, it takes no less than 1 s for an I27 module to completely refold [34]. A faster refolding rate of ∼20 s\(^{-1}\) of human fibronectin FnIII domains was reported in chemical experiments [77, 78], and a refolding rate of ∼42 s\(^{-1}\) of FnIII domains of tenascin was estimated in AFM experiments [40]. These refolding rates are, however, much slower than unfolding rates, which typically range from 10\(^{-3}\) to 10\(^{-5}\) s\(^{-1}\) [79, 80, 25, 34]. Due to the limited resolution of current experimental instruments, we know little about the characteristic of the refolding pathway of Ig-like domains at the atomic level. With MD tools that complement the experimental techniques, one can characterize the refolding pathway, especially the reformation of interstrand hydrogen bonds.

Starting from partially unfolded structures of the FnIII\(_9\) domain of titin at different stages of unfolding, Paci and Karplus have simulated the refolding by employing an implicit solvent model [67, 68]. Their simulations, however, showed an extremely fast refolding process that lasted less than 50 ps, which is even shorter than the time that they spent on simulated stretching; refoldings also ended in non-native secondary structures.

In this chapter we will simulate the refolding process starting from different stages of partially unfolded I27 intermediates by employing an explicit solvent model. The energy of two sets of

\[^3\]This chapter was adapted from the manuscript: M. Gao, H. Lu, and K. Schulten, Simulated refolding of stretched titin immunoglobulin domains, (2001), Biophysical Journal, 81:2268–2277.
hydrogen bonds, which involve the minor and major unfolding events characterized above, will be analyzed to characterize the refolding process. The simulations correspond to the final stages of the I27 folding process and the formation of the key force bearing element of the domain which secures the integrity of the protein under external forces. We will demonstrate that water-protein interactions play an important role during refolding. To test the mechanical elasticity of the refolded proteins, external forces are applied to the resulting structures from two of the refolding simulations.

3.2 Methods

We started our simulations on model immunoglobulin I27 domains that had been partially unfolded in previous studies through application of 750 pico-Newton (pN) constant forces \[65\]. The pre-stretched proteins were allowed to relax spontaneously for two nanoseconds, and were then stretched by applying constant force SMD protocols to them. The simulations carried out are listed in Table 3.1.

Eight I27 unfolding intermediates \(s_k\) \((k = 1, 2, ..., 8)\) were selected from two independent SMD unfolding simulation trajectories SMD(nat)_1 and SMD(nat)_2 \[65\], as shown in Fig 3.1. These intermediates can be classified as two sets: AB intermediates \(s_k\) \((k = 1, 2, 3)\), named after the disruption of the interstrand hydrogen bonds only between \(\beta\)-strands A and B during forced unfolding, and A’G intermediate \(s_k\) \((k = 4\) to \(8)\) where all the six hydrogen bonds between \(\beta\)-strand A’ and G were ruptured. The AB intermediates \(s_1\) and \(s_2\) were selected at 200 ps and 299 ps from trajectories of SMD(nat)_1 and SMD(nat)_2 respectively. Intermediate \(s_3\) was selected after the hydrogen bonds between \(\beta\)-strands A and B had been broken for about 110 ps in SMD(nat)_1. The intermediates \(s_4\) to \(s_8\) were selected at timesteps no less than 1000 ps from SMD(nat)_1 and SMD(nat)_2, after all six hydrogen bonds between \(\beta\)-strand A’ and G were broken. Each of the simulations describing the refolding of the intermediates \(s_k\), referred below as \(R(s_k)\), was carried out for 2 ns. In these simulations no external forces were applied, permitting the domains to refold. To start these simulations, we have used restart files to initiate the simulations for \(R(s_3), R(s_4)\) and \(R(s_8)\) with actual atomic velocities, whereas in other five refolding simulations the velocities of atoms were randomly assigned according to Maxwellian distributions for a temperature of 300 K.
Table 3.1: Sequences of simulations utilized in this study. Two independent SMD unfolding simulations reported
in [65], named SMD(nat)1 and SMD(nat)2, were started from the native I27 structure by applying forces of 750 pN.
At various time steps the unfolding simulations were stopped and the intermediate protein structures, named s_k
(k = 1 – 8), were selected. These structures were allowed to relax for 2 ns without application of external forces,
resulting in structures named s'_k (k = 1 – 8). The corresponding simulations described in this paper are denoted
as R(s_k) (k = 1 – 8). Finally, two SMD simulations, named SMD(s'_3) and SMD(s'_4), were carried out on partially
refolded structures s'_3 and s'_4, by applying forces of 100 pN and 750 pN, respectively, to test the mechanical elasticity
of s'_3 and s'_4.

Simulations R(s_k) lead to resulting structures denoted as s'_k. For s'_3 and s'_4, a second simulation
was carried out, referred to as SMD(s'_3) and SMD(s'_4), in which constant forces of 100 pN and
750 pN were applied to the Cα atom of the C-terminus residue of the I27 modules, the Cα atom
of the N-terminus being fixed, and the force being directed along the vector connecting the initial
positions of the N-terminus to the C-terminus.

Before the original SMD simulations were initiated, the native protein was solvated by a sphere
of TIP3 water molecules [72] of 31 Å radius and equilibrated for 1 ns. The molecular surface of the
domain was covered at this point by at least four shells of water molecules.

All molecular dynamics simulations were carried out using the software packages NAMD [73]
and X-PLOR [71] with the CHARMM22 force field [74]. An integration time step of 1 fs and
a uniform dielectric constant of 1 was chosen. A switching function for calculating non-bonded
Figure 3.1: Extension-time profiles from simulations SMD(nat)\textsubscript{1} and SMD(nat)\textsubscript{2}, adapted from [65].

Eight frames $s_k$ ($k = 1 - 8$) were selected from two trajectories as the initial structures for the refolding simulations $R(s_k)$ ($k = 1 - 8$), as indicated by circles and arrows. Three of structures, $s_1$ (200ps), $s_2$ (299ps), $s_3$ (311ps), have hydrogen bonds between $\beta$-strands A and B broken, while the other five structures, $s_4$ (1000ps), $s_5$ (1010ps), $s_6$ (1067ps), $s_7$ (1072ps), and $s_8$ (1085ps), have all the hydrogen bonds between $\beta$-strands A’ and G and between $\beta$-strands A and B broken.

Coulomb and van der Waals interactions was employed, switching interactions off between 10 Å and 13 Å. The atomic coordinates of the entire system were recorded every picosecond.

The analysis of molecular structures and hydrogen bond energies were conducted using X-PLOR and VMD [75]. The extension of the protein is defined as the change of the end-to-end distance between the two terminal carbon atoms (L1(C\textsubscript{\alpha}) and L89(C\textsubscript{\alpha})) at the end of the 1 ns equilibration, i.e., prior to the first round of SMD simulations. An explicit hydrogen-bonding energy term was used in the trajectory analysis, with parameters adopted from param11.pro in XPLOR. The minimum energy of the nitrogen-hydrogen-oxygen (N-H-O) hydrogen bond is -3.5 kcal/mol, occurring when the O-H distance is 1.9 Å and the N-H-O angle is 180°. An hydrogen bond was considered broken when the O-H distance is larger than 3 Å, or when the N-H-O angle is less than 90°, situations that correspond to an H-bond energy to above -1 kcal/mol.
3.3 Results

Eight refolding simulations \( R(s_1) - R(s_8) \) have been performed, and two constant force SMD simulations \( \text{SMD}(s'_3) \) and \( \text{SMD}(s'_4) \) with stretching forces of 100 pN and 750 pN have been carried out, as described in Methods.

3.3.1 Refolding of Intermediates I

![Figure 3.2: Extension-time profiles from (a) simulation \( R(s_1) \) and (b) simulation \( R(s_2) \), together with snapshots of \( \beta \)-strands A and B at the initial and final timesteps. Both curves show a fast decrease in extension, but slow reformation of hydrogen bonds between \( \beta \)-strands A and B. In all snapshots only the backbone atoms and the hydrogen atoms involved in the backbone hydrogen bonds are shown. Oxygen atoms, hydrogen atoms, water molecules and other backbone atoms are colored in purple, black, green, and blue, respectively.](image)

Figure 3.2 presents the extension-time profiles resulting from refolding simulations \( R(s_1) \) and \( R(s_2) \), together with snapshots of backbone hydrogen bonds between \( \beta \)-strands A and B at the initial and final time steps of the simulations. In Fig. 3.2a, the extension decreases within 1 ns from 11 Å to less than 1 Å, which is very close to the native fully-folded case, then fluctuates around 2 Å in the remaining time of the simulation. Although the extension of the protein shrinks dramatically, water molecules remain intercalated between \( \beta \)-strands A and B, preventing backbone hydrogen bonds from reforming. All three hydrogen bonds are still broken and backbone atoms
form hydrogen bonds with nearby water molecules at the end of the simulations. The simulation results of \( R(s_2) \) shown in Fig. 3.2b are similar to those shown in Fig. 3.2a. The extension of the I27 domain is seen to diminish from 13 Å to 1 Å in 500 ps, then to fluctuate around 4 Å for the remaining simulation period. Nevertheless, simulations \( R(s_2) \) did not lead to a stable reformation of backbone hydrogen bonds between \( \beta \)-strands A and B in two nanoseconds.

Figure 3.3: Time development of extension of a pre-stretched I27 domain in simulations \( R(s_3) \) (black) and SMD(\( s'_3 \)) (red), together with snapshots of backbone hydrogen bonds between \( \beta \)-strand A and B at various stages of the simulations. The left snapshot shows the \( \beta \)-strands A and B of the initial structure of refolding \( s_3 \). After two nanosecond relaxation (the center snapshot), the extension of the protein has decreased from about 12.5 Å to about 3 Å. The protein was then stretched by applying a constant force of 100 pN, and reached an extension of ∼5 Å (right snapshot). Oxygen atoms, hydrogen atoms, water molecules and other backbone atoms are colored in purple, black, green, and blue, respectively.

To further investigate the unfolding properties of the partially refolded domains resulting from \( R(s_3) \), the \( s'_3 \) structure was subjected to a 100 pN constant force in SMD simulation SMD(\( s'_3 \)). The resulting extension vs. time curves are shown in Fig. 3.3, covering both simulations \( R(s_3) \) and SMD(\( s'_3 \)). The refolding resulting in \( R(s_3) \) resembles the curve shown in Fig. 3.2 for \( R(s_1) \) and \( R(s_2) \). At the end of simulation \( R(s_3) \), the extension of the protein decreased by about 8 Å and fluctuated around a value of 4 Å. Although the backbone hydrogen and oxygen atoms came closer in a tendency toward reformation of the interstrand hydrogen bonds (see center snapshot),
water molecules still remain intercalated between strands A and B prevented hydrogen bonds from reforming. Since those hydrogen bonds did not reform, the protein in the state \( s_3' \) was expected to be more easily extendable than in the case of the native structure. In fact, during simulation SMD(\( s_3' \)) in which a force of 100 pN was applied, the extension reached a peak of 7 Å in 300 ps and later fluctuated around 5 Å. The \( \beta \)-strands A and B separated further (as seen in the right snapshot) to provide a transitional extension before the main unfolding event.

Figure 3.4: Backbone hydrogen bond energies versus time for individual hydrogen bonds between \( \beta \)-strands A and B. Top row: simulation \( R(s_1) \); middle row: simulation \( R(s_2) \); bottom row: simulation \( R(s_3) \).

Figure 3.4 shows the hydrogen bond energy versus time for all hydrogen bonds between \( \beta \)-strands A and B. As one can see, the energy of the hydrogen-bond between E3(O) and S26(H) is always higher than -1 kcal/mol, i.e., this bond remains in a ruptured state. Hydrogen bonds K6(H)-E24(O) and K6(O)-E24(H) show energy fluctuations around -1 kcal/mol, i.e., they are stronger than the E3(O)-S26(H) bond, yet these bonds cannot be considered to have stably reformed.
Figure 3.5: Extension versus time profiles from simulations R(s₄) (black) and SMD(s'₄) (red), together with snapshots of β-strands A-B and β-strands A’-G at various stages during the simulations. The left snapshots show β-strands A-B (bottom left) and β-strands A’-G (upper left) of the initial structure s₄. After two nanosecond relaxation, the extension of the domain decreased from over 15 Å, to a value of around 0 Å. One hydrogen bond between A and B (bottom center) and five hydrogen bonds between A’ and G (upper right) have stably reformed. The protein was then stretched by applying a constant force of 750 pN. The bottom right snapshot of β-strands A’ and G was captured when the hydrogen bonds between β-strand A’ and G were broken again and water molecules formed hydrogen bonds with backbone oxygens and hydrogens. Oxygen atoms, hydrogen atoms, and water molecules are colored in purple, black, and green. Other backbone atoms in snapshots of A-B and A’-G are colored in blue and red, respectively.
3.3.2 Refolding of Intermediate II

Figure 3.5 presents the extension versus time profile from the contiguous simulations R(s₄) and SMD(s'₄), together with snapshots of backbone hydrogen bonds between β-strands A-B and β-strands A'-G. During simulation R(s₄), the domain gradually became more compact with its extension decreasing from over 15 Å to 0 Å, the native folded length. At the end of the 2 ns refolding simulation, one hydrogen bond between A-B, K6(O)-E24(H) (see bottom right snapshot), stably reformed; five of the six hydrogen bonds connecting A' and G stably reformed, except the bond between Y9(O) and N83(H) (see upper right snapshot). There is no any water molecule observed within 3 Å of the hydrogen and oxygen atoms that pairs these stably reformed hydrogen bonds. In simulation SMD(s'₄), the protein, under a force of 750 pN, extends by 10 Å in less than 10 ps, then fluctuates in the middle plateau region for about 500 ps. After this, the protein crosses the main unfolding energy barrier (at ~2.5 ns). The right snapshot in Fig. 3.5 was made when the hydrogen bonds between β-strands A’ and G were ruptured again, and water molecules attacked backbone oxygens and hydrogens to form protein-water hydrogen bonds.

The energy of individual interstrand hydrogen bonds was monitored during simulation R(s₄), as shown in Fig. 3.6. Hydrogen bonds between β-strands A and B displayed a tendency to be stably reformed, except for the bond between E3(O) and S26(H). The bond K6(H)-E24(O) reforms (the energy is less than -1 kcal/mol) at around 200 ps, but breaks after 1 ps. It reaches a more favorable low energy configuration at around 1600 ps and stays there for about 200 ps. The bond then breaks again and is found to fluctuate between formed and broken states for the remaining time of the simulation. The other bond between K6 and E24, K6(O)-E24(H), is adjusted to a configuration with a bond energy lower than than -3 kcal/mol in less than 200 ps; it then fluctuates between formed and broken states until it is stabilized. No water molecule forms a hydrogen bond with either K6(O) or E24(H) at the end of simulation R(s₄).

Hydrogen bonds between β-strand A’ and G show a stronger tendency to reform, except for the one between hydrogen bond partners Y9(O) and N83(H), located at the end of the A’-G strands. The four hydrogen bonds, V11(H)-N83(O), V11(O)-K85(H), V13(H)-K85(O), V13(O)-K87(H) reform in less than 200 ps and are stable after the reformation. It takes hydrogen bond V15(H)-K87(O) 400 ps to reform. After the 2 ns simulation R(s₄) five of the six hydrogen bonds
Figure 3.6: Fluctuation of interstrand hydrogen bonds between \( \beta \)-strands A and B, and \( \beta \)-strands A' and G during simulation R(s\text{4}). Hydrogen bonds between A' and G reform in 500 ps, except the one between Y9 and N83, which never reforms during 2 ns simulation. Hydrogen bonds between \( \beta \)-strands A and B have a tendency to reform, except the one between E3 and S26, but remain unstable and are easily disrupted during the simulations.

between A' and G are reformed.

While none of three hydrogen bonds between \( \beta \)-strands A and B stably forms in simulations R(s\text{5}), R(s\text{6}), and R(s\text{7}), results of hydrogen bond formation between \( \beta \)-strands A' and G from these simulations are similar to those from R(s\text{4}). The three hydrogen bonds, V11(H)-N83(O), V11(O)-N83(H) and V13(H)-K85(O), located in the middle of the hydrogen bond cluster between A'-G strands, reform faster than the three H-bonds, V13(O)-K87(H), V15(H)-K87(O) and Y9(O)-N83(H), located at the edges of the hydrogen bond cluster. However, in simulation R(s\text{8}) none of the hydrogen bonds have been found to reform over a 2 ns time period, though the extension of the protein decreased from 23 Å to around 9 Å. Five of the six broken A'-G backbone hydrogen bonds in R(s\text{8}), however, were observed to shorten their O-H distance from 10 - 13 Å to 7 - 10 Å, demonstrating a tendency for reformation.
3.3.3 Summary of Refolding Simulations

Simulations R(s₁) to R(s₃) were performed to simulate the refolding of A-B β-strands. The simulations R(s₄) to R(s₈) simulated the reformation of hydrogen bonds between both A-B strands and A'-G strands. In Table 3.2, we have summarized the reformation of hydrogen bonds and overall extension from all eight refolding simulations. We consider a hydrogen bond stably reformed if the bond energy has been less than -1 kcal/mol for over 500 ps during the second half of the simulation period, i.e., the period from 1 ns to 2 ns. The three hydrogen bonds between β-strands A and B have seldom been stably reformed, except for one case in R(s₄), as shown in Fig. 3.6. On the other hand, in four out of five simulations of A'-G reformation, most of the backbone hydrogen bonds characterizing this pair of β-strands do reform quickly in the first half of the 2 ns simulation period, and then stay in an energetically favorable configuration (less than -1 kcal/mol) for over 500 ps for the rest of time, especially those located in the middle of the hydrogen bond cluster.

<table>
<thead>
<tr>
<th>Refolding Simulations</th>
<th>AB H-bonds</th>
<th>A’G H-bonds</th>
<th>Extension (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reformed</td>
<td>Stabilized</td>
<td>Reformed</td>
</tr>
<tr>
<td>R(s₁)</td>
<td>3</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>R(s₂)</td>
<td>2</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>R(s₃)</td>
<td>3</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>R(s₄)</td>
<td>2</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>R(s₅)</td>
<td>2</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>R(s₆)</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>R(s₇)</td>
<td>2</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>R(s₈)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3.2: A summary of the number of reformed hydrogen bonds between A-B strands and A'-G strands and the extension of the I27 domain in eight refolding simulations R(sₖ) (k = 1 – 8), respectively. A hydrogen bond is defined as reformed if the value of its hydrogen bond energy has been ever less than -1 kcal/mol during the 2ns refolding simulation; and it is counted as a stably reformed hydrogen bond if the energy of the bond has been less than -1 kcal/mol for over 500 ps during the second half of the simulation period from 1 ns to 2 ns. The final extension is calculated by averaging the extension of the protein over the last 500 ps in a simulation.

In all simulations, except R(s₈), the domain length shrinks to a near-native level. In R(s₄), where β-strands A'-G reform stably and one hydrogen bond in β-strands A-B reforms, the domain refolds and fluctuates around its native folded length. Six simulations for which β-strands A-B become not stably bonded reveal an overall extension that is longer by 2 Å to 5 Å than the native fold. In R(s₈), where both β-strands A'-G and β-strands A-B do not reform, the extension is 9 Å longer than for native fold. The observation indicates that the extension of the protein can serve well as the reaction coordinate in the unfolding process.
3.4 Discussions

Starting from the partially stretched I27 domain, the refolding simulations showed that the reformation of \( \beta \)-sheet AB and \( \beta \)-sheet A’G exhibit different characteristics. Comparing the current study with the 1 ns equilibration described in [65], it is interesting to notice that the more stable a hydrogen bond appeared during equilibration, the easier this hydrogen bond reforms stably during refolding. In both equilibration and refolding simulations, the hydrogen bonds between the same pair of \( \beta \)-strands showed a certain cooperativeness. The larger number of hydrogen bonds involved in \( \beta \)-sheet A’G than those in \( \beta \)-sheet AB may determine why \( \beta \)-sheet A’G reforms much faster than \( \beta \)-sheet AB.

AFM experiments revealed a small transitional extension of I27 tandem repeats before the main unfolding event [35]. This hump-like extension of 6.6 Å per module was suggested to correspond to the unfolding of hydrogen bonds between \( \beta \)-strands A and B, and further analysis implied that an average time period of 40 ms must pass before I27 can fully refold from this intermediate. Our refolding simulations of intermediates with broken interstrand hydrogen bonds between \( \beta \)-strands A and B demonstrated a decrease in the extension of I27, from a configuration of early forced unfolding to close to its native fully-folded length, in less than 1 ns, as shown in Figs. 3.2 and 3.3. The energy analysis of hydrogen bonds shown in Figs. 3.4 and 3.6 seldom exhibit a stable reformation of the bonds between \( \beta \)-strands A and B, except that the K6(O)-E24(H) in R(s_4) stably reforms. Yet the tendency for refolding in terms of overall domain length was clearly demonstrated within the nanosecond time scale of the simulations. The rare reformation of hydrogen bonds between \( \beta \)-strands A and B may explain the low probability of ‘hump’ events in the post-initial force peaks of AFM experiments, because hydrogen bonds bridging \( \beta \)-strands A and B of unfolded domains might remain ruptured during the unfolding of the first module and the following unfolding events. In SMD(s’3) we applied a constant force of 100 pN to the partially refolded intermediate, showing that the domain elongates from 4 Å to 5 Å in less than 200 ps. As suggested in [35], the disruption of hydrogen bonds between A and B corresponds to the transitional extension of \( \sim 6 \) Å. The extension of the I27 domain in the re-stretching SMD simulation shows that the separation between \( \beta \)-strands A and B increases the extension again easily, which is in good agreement with the observations in AFM experiments. The current work provides new evidence to support our previously proposed
mechanism that hydrogen bonds between β-strands A and B control the unfolding intermediate of I27.

The refolding simulations in this study extend previous studies [65] showing that water plays a major role in the reformation of backbone hydrogen bonds. As one can see in snapshots shown in Figs. 3.2 and 3.3 that characterize refolding of the AB intermediates, water molecules compete for hydrogen bonds with backbone atoms on β-strands A and B, hindering the reformation of hydrogen bonds between backbone atoms themselves. This behavior, however, cannot be accounted for by the implicit solvent models that Paci et al. [67, 68] employed.

In contrast to the refolding of β-sheet AB, the refolding of β-sheet A’G occurs very fast, reforming hydrogen bonds between β-strands A’ and G shortly after their concurrent disruption. As an example, one can see from Fig. 3.5 of simulation R(s4), water molecules in the initial snapshot of A’-G have been driven out after 2 ns relaxation and the extension is reduced to the native folded length of the domain. Further details of hydrogen bond energy analysis in Fig. 3.6 show that, except for the bond between Y9(O) and N83(H), the five other interstrand hydrogen bonds between β-strands A’ and G reform in less than 400 ps and remain in an energetically favorable configuration thereafter. Since the length of hydrogen bond Y9(O)-N83(H) had been demonstrated to be more flexible (fluctuating between 2 Å and 5 Å during 1 ns equilibration reported in [65]) than that of the other five hydrogen bonds between β-strand A’ and G, it is not surprising that the reformation of hydrogen bond Y9(O)-N83(H) requires longer time. The fluctuation of the energies of the other five hydrogen bonds shown in Fig. 3.6 resembles closely the fluctuation observed during the equilibration of the native I27 domain [65] and reflects stable bonds. When the I27 domain is stretched further (beyond 17 Å), it is more difficult for A’G hydrogen bonds to reform during the 2 ns simulation period. As shown in Table 3.2, simulation R(s8) was started at an extension of 23 Å, and none of the hydrogen bonds between β-strands A’ and G are reformed subsequently, though the distance between hydrogen bond partners decrease about 3 Å.

The relatively fast refolding between β-strands A’ and G may be a mechanical property that leads to the characteristic saw-tooth pattern generated in AFM experiments [25, 34]. With this property, even when the patch of hydrogen bonds between β-strands A’ and G of two domains is ruptured concurrently by accident, if the domain remained in near-native extension the hydrogen
bonds reform quickly and contribute to a second force peak in a later unraveling event. In SMD(s’4) we applied a constant force of 750 pN to the partially refolded structure, showing a 500 ps long plateau where the domain is trapped in a transitional state. This passage time of 500 ps, spent on crossing the energy barrier, is comparable to what we observed earlier in stretching native I27 domains (about 800 ps as shown in Fig. 3.1). The close agreement of the first mean passage times indicates good reformation of hydrogen bonds between \( \beta \)-strands A’ and G, reformed at the end of simulation R(s4), and restoration of the I27’s resistance against stretching.

One must note that the fast reformation of \( \beta \)-sheet A’G does not correspond to the refolding of I27 from a random coil or completely extended conformation which would take milliseconds to seconds. The simulation of the complete folding process in these cases is still impossible due to the limitation of computational power. According to the current understanding of protein folding, a major barrier lies in the formation of native topology. Our starting structures are only partially unfolded I27 domains which already have the correct topology. Thus the refolding simulations carried out in this study correspond to the final step of the protein folding process. Our emphasis here is the atomic level description of the formation of the force bearing hydrogen bonds between \( \beta \)-strands A’ and G, which control I27’s mechanical behavior.
Chapter 4

Mechanical unfolding pathways of FN-III_{10}

4.1 Introduction

The extracellular matrix protein fibronectin not only provides a substrate for cell anchorage, but also regulates cell adhesion by transmitting environmental signals to cells [81, 16]. 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 [19, 21].

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 Fig. 4.1). Cells assemble fibronectin dimers into fibrillar networks that provide mechanical stability to the extracellular matrix and connective tissue [81, 16]. Integrin binding to fibronectin’s RGD-loop on FN-III_{10} results in the formation of a cytoskeletal apparatus that mechanically couples actin contractility to extracellular fibronectin fibers [19]. Fluorescent microscopy studies using chimeric fibronectin with green fluorescent protein have demonstrated that cells stretch fibronectin fibers up to 4-fold their equilibrium length [20]. Recent studies using fluorescent resonance energy transfer to measure fibronectin fibrils stretched into non-equilibrium, found that cells not only...
integrate fibronectin in an extended confirmation into fibrils, but also hyperextend many fibrils so that FN-III modules become unfolded [82, 22]. 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 buried so-called cryptic sites [13, 19, 21, 83]. Functionally relevant cryptic sites have been specifically proposed to exist on a number of FN-III modules including FN-III$_1$ [84], FN-III$_2$ [85], FN-III$_7$ [86], FN-III$_9$ [87], FN-III$_{EDB}$ [88], FN-III$_{10}$ [89], FN-III$_{13-14}$ [90], and FN-III$_{15}$ [86].

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 [91]. By swapping β-strands of FN modules, intermediates in the unfolding pathway may serve as nucleation sites [87] 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\textsubscript{10} is one of the most thermostable FN-III modules [92]. These studies suggest that the A- and G-strands are the last to form during folding of FN-III\textsubscript{10} [93, 78]. 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 exists a correlation between mechanical stability and thermodynamic stability [34] and that the respective unfolding pathways may differ [68, 94, 32]. 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 [95, 96, 40, 32].

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\textsubscript{10} and of other similar modules and motifs [48]. Krammer et. al. first used steered molecular dynamics on FN-III\textsubscript{10} solvated in explicit water and found a large predominant force peak corresponding to separation of the G-strand [91]. Craig et. al. used a similar system to find that the FN-III\textsubscript{10} module is mechanically one of weakest FN-III modules and that β-strands progress from a twisted to an aligned state prior to unraveling [97]. Conversely, Paci and Karplus using an implicit water model and biased molecular dynamics, found FN-III\textsubscript{10} to be considerably more stable than FN-III\textsubscript{9} and further reported an unfolding intermediate at an N-terminus-to-C-terminus distance of approximately 140 Å [67]. Finally, approaches using off-lattice models [70] have been able to reproduce a similar unfolding pathway as 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\textsubscript{10} solvated in a box of water large enough to permit stretching of FN-III\textsubscript{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
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.

4.2 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 [44]). After hydrogen atoms were added to the
monomer with X-PLOR [71], two TIP3 [72] water boxes, one of size 60×60×370 Å³ and the other
one of size 64×64×190 Å³, 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 Fig. 4.2.

![Figure 4.2: Setup of SMD simulations. An FN-III_{10} module was solvated and stretched in a water box. In
the simulations the C\textsubscript{\alpha} atom at the N-terminus (Val1) was fixed, and C\textsubscript{\alpha} atom at the C-terminus (Thr94)
was stretched with constant force.](image)

The program NAMD [73] with CHARMM22 force fields [74] was subsequently used for all MD
simulations described in this paper. 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\textsubscript{\alpha} 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 [98], 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 $55 \times 60 \times 367$ Å$^3$ and $61 \times 63 \times 182$ Å$^3$, respectively. Finally SMD simulations were carried out by fixing the C$_\alpha$ atom of the N terminus of FN-III$_{10}$, and applying constant forces to the C$_\alpha$ 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 [99], 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 10 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 1.33GHz Athlon processors. A 1 ns simulation of the 126,082 atom system required $\sim$170 hours running on the cluster. The SMD simulations presented in this paper are referred to as SMD(force value), e.g., SMD(600pN) for a 600pN constant force SMD simulation. Multiple runs are denoted by a number subscript, e.g., SMD(500pN)$_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 [75]. 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.
Figure 4.3: Extension-time profile from simulation SMD(500pN) 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 dashed black lines.

4.3 Results

4.3.1 Mechanical unfolding intermediates.

SMD simulations with constant force protocol have been carried out to probe the unfolding intermediates of FN-III\textsubscript{10} solvated in a 367 Å long water box (see Fig. 4.2). Fig. 4.3 presents the extension versus time profile from simulation SMD(500pN), in which an FN-III\textsubscript{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 Fig. 4.3(a-e). Initially, two β-sheets ABE and GFCD pack against each other in a twisted orientation (see Fig. 4.3(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-long plateau (intermediate I\textsubscript{1}). As discussed previously [97], 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 I₂, named “aligned state” previously [97], for two β sheets aligned along the direction of the external force as shown in Fig. 4.3(b). The aligned state is associated with a ~220 ps plateau, indicating a relatively stable intermediate. During the transition from I₁ to I₂, the hydrophobic core of FN-III₁₀ has been disrupted due to the rotation of the β-sheets. At the end of the second plateau the unfolding progress along alternate pathways. In the simulation SMD(500pN)₁ shown in Fig. 4, 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 Fig. 4.3(c)), which corresponds to crossing of a barrier formed by six hydrogen bonds between β-strands G and F. This intermediate I₃ 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 I₃, at an extension three times that of the native FN-III₁₀. After the concurrent breaking of six G-F backbone hydrogen bonds (see Fig. 4.3(d)), the protein resumes unfolding until it is fully unfolded at 280 Å extension (see Fig. 4.3(e)), which corresponds to the last plateau shown in the extension-time profile.

<table>
<thead>
<tr>
<th>Unfolding Simulations</th>
<th>Location of Stable Intermediates (Å) and Key Events of Hydrogen Bond Rupture</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMD(400pN)</td>
<td>I₁ (twisted) → rupture → I₂ (aligned) → rupture → I₃ → rupture</td>
</tr>
<tr>
<td>SMD(500pN)₁</td>
<td>5 R-D ≥16 I₂ R-D 18 A-B 100 G-F</td>
</tr>
<tr>
<td>SMD(500pN)₂</td>
<td>9 R-D 25 G-F — —</td>
</tr>
<tr>
<td>SMD(500pN)₃</td>
<td>9 R-D 20 A-B, G-F — —</td>
</tr>
<tr>
<td>SMD(600pN)</td>
<td>11 R-D 25 A-B 100 G-F</td>
</tr>
<tr>
<td>SMD(700pN)</td>
<td>11 R-D 25 A-B — —</td>
</tr>
</tbody>
</table>

Table 4.1: Location of stable intermediates and key events of hydrogen bond rupture observed in the unfolding studies of FN-III₁₀. The location of the intermediate are 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.

Fig. 4.4 shows recordings of force-induced unfolding for FN-III₁₀ 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(500pN)\(_1\), appears only in simulation SMD(600pN). The plateaus identified in SMD(500pN)\(_1\) can be associated with the three intermediates I\(_1\), I\(_2\) and I\(_3\) characterized in Table 4.1. Universal to all simulations is a jump by \(\sim10\) Å in extension between intermediate I\(_1\) and I\(_2\), corresponding to a transition from a twisted to aligned \(\beta\)-sheets, as illustrated in Fig. 4.3. Although intermediate I\(_1\) always precedes the alignment transition, intermediate I\(_2\) can lead to three different unfolding pathways: in simulations SMD(600pN) and SMD(700pN), strand A detached from the remaining fold first; in simulation SMD(500pN)\(_2\) strand F detaches first; in simulation SMD(500pN)\(_3\) strands A and G simultaneously detach from their \(\beta\)-strand partners B and F. In simulation SMD(400pN), none of the \(\beta\)-strands peeled away from the module within 2 ns.

### 4.3.2 Intra-strand hydrogen bonds.

An analysis of the energies of all intra-strand hydrogen bonds was performed to characterize the three force-induced unfolding pathways of FN-III\(_{10}\). Fig. 4.5(a-c) shows the energies of 32 backbone
Figure 4.5: Inter-strand hydrogen bond versus time curves in three representative SMD trajectories from (a) SMD(500pN)$_1$, (b) SMD(500pN)$_2$, and (c) SMD(500pN)$_3$. Energies are grouped by $\beta$-strand pairs, displayed from left to right according to the ordering in the list on the left side.
hydrogen bonds of the module along three representative unfolding pathways from SMD(500pN)\textsubscript{1}, SMD(500pN)\textsubscript{2} and SMD(500pN)\textsubscript{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\textsubscript{1} to I\textsubscript{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\textsubscript{7–9} [97]. The aligned state I\textsubscript{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 Fig. 4.5(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\textsubscript{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 Fig. 4.3. The barrier formed by G-F hydrogen bonds can trap the protein in the meta-stable intermediate state I\textsubscript{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 Fig 4.3. Alternatively, as shown in Fig. 4.5(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\textsubscript{10} [91] and constant force stretching simulations [97]. We name the barrier associated with the hydrogen bond rupture between F and G, II\textsubscript{GF}. Occasionally, we observed a third unfolding pathway as shown in Fig. 4.5(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.

4.3.3 Forced unfolding pathways.

According to the hydrogen bond analysis, the SMD simulations of FN-III\textsubscript{10} exhibit three different unfolding pathways, as shown in Fig 4.6. Simulations began with FN-III\textsubscript{10} equilibrated in a periodic
Figure 4.6: Unfolding pathways for FN-III\textsubscript{10} resulting from SMD simulations. (a) FN-III\textsubscript{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 Arg\textsubscript{6} and Asp\textsubscript{23} between A- and B-strands; immediately following the rupture the segment Pro\textsubscript{5}-Leu\textsubscript{8} straightened, permitting solvation of the hydrophobic core periphery. (d) Alignment of β-strands between the two β-sheets is initiated by slipping of Ile\textsubscript{88} of the G-strand past the A-strand. Unraveling of the β-strands begins by either separation of (e) A-strand first, (f) A- and G-strands simultaneously, or (g) G-strand first. (h) Fully unraveled structure. (i) In case that the A-strand separates first, an intermediate at 100 Å extension arises.
box. Upon applying a constant force, the randomly coiled termini straightened slightly, while
the overall tertiary structure still remained intact (see Fig. 4.6(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 Fig. 4.6(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 Fig. 4.6(d)). Unraveling of the β-strands began by either
separation of the A-strand first (see Fig. 4.6(e)), A- and G-strands simultaneously (see Fig. 4.6(f)),
or G-strand first (see Fig. 4.6(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 Fig. 4.6(i)).

Figure 4.7: Conformational changes of the RGD-containing loop in simulation SMD(500pN)_1. (a) The
width of the RGD loop is measured by the distance between the Cα atoms of Arg78 and Pro82. (b) Angle
formed by the Cα 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.
4.3.4 RGD-loop Conformations.

The peeling of strand G away from $\beta$-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 [91]. Conformational analysis of the RGD loop in simulation SMD(500pN) shows a similar conformational transition when the G strand was stretched out after the rupture of $\beta$-sheet ABE, as shown in Fig. 4.7. Prior to breaking of the G-F hydrogen bond, the RGD loop remains intact for 1.25 ns at the apex of a hairpin $\beta$-turn that connects strands G and F (see Fig. 4.7(c,d)). The width of the RGD loop (see Fig. 4.7(a)) and the bending angle of the loop (see Fig. 4.7(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 Fig. 4.7(c)), jumps from 7 Å to 14 Å, for the loop has been quickly straightened. The bending angle, defined as the angle formed by $C_\alpha$ atoms of Thr76, Arg78 and Asp80, displays a change from a bent ($\sim110^\circ$) to a more flat ($\sim170^\circ$) 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 $\beta$-strands FCD (Fig. 4.7(e)).

4.4 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 $\beta$-strands for FN-III$_{10}$, the order in which $\beta$-strands separate from the module is variable. Consistent with earlier studies [97, 83], the simulations show that the upper and lower $\beta$-sheets progress from a twisted to aligned state prior to unraveling of the first $\beta$-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$_{10}$’s $\beta$-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 $\sim100$ Å in case the A-strand separates first. Fig. 4.6 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.

### 4.4.1 Comparison with other models.

SMD simulations of FN-III$_{10}$ have been conducted with explicit water sphere models [91, 97], with a small periodic water cell [83], and with implicit solvation models [67]. In case of water sphere models, FN modules were solvated in TIP3 water spheres of $\sim$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 [97]. 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 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 [83]. 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 [65]. Implicit solvation models give an unphysical refolding rate of FN-III$_{9,10}$ modules of less than 50 ps [67]. The implicit models report FN-III$_{10}$ to be mechanically stronger than FN-III$_9$ [67], while employing an explicit model we found FN-III$_{10}$ is mechanically weaker than FN-III$_9$ [97], which is in agreement with experimental observations using AFM [32]. 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.

### 4.4.2 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. In AFM experiments unfolding proceeds on a ms time scale and peak rupture forces of $\sim$75 pN are measured [32]. Under
physiological conditions forced unfolding of FN-III modules could take even longer and forces as low as several pN could be strong enough to unfold a single FN-III module [13]. Naturally the question arises: at low force regime, does FN-III$_{10}$ follow the three unfolding pathways revealed by SMD simulations? Among the three possible choices, the unfolding scenario in which $\beta$-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 previously [64, 97],

$$U(x) = \begin{cases} +\infty & x < a, \\ \Delta U(x-a)/(b-a) & a \leq x \leq b, \\ -\infty & x > b. \end{cases} \quad (4.1)$$

$\Delta U$ 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

$$\tau = 2\tau_d \delta(F)^{-2}[e^{\delta(F)} - \delta(F) - 1] \quad (4.2)$$

Figure 4.8: Ratio of mean first passage times $\tau_{AB}$ and $\tau_{FG}$ as a function of applied force. $\tau_{AB}$ and $\tau_{FG}$ are the first passage times for crossing the barriers $\Pi_{AB}$ and $\Pi_{FG}$, respectively.
where $F$ is the external force and $D$ is the effective diffusion coefficient. Other quantities are $\tau_d = (b-a)^2/2D$, $\delta(F) = \beta[\Delta U - F(b-a)]$, $\beta = 1/k_BT$. The extension range $b-a$ is $\sim 4\,\AA$, estimated from the fluctuation of the associated plateau (see Fig. 4.3). According to the SMD simulations, crossing either barrier $\Pi_{AB}$ or $\Pi_{GF}$ leads to a different unfolding pathway. We assume that the heights of barriers $\Pi_{AB}$ and $\Pi_{GF}$, designated as $\Delta U_{AB}$ and $\Delta U_{GF}$, are determined by the number of intact hydrogen bonds between $\beta$-strands A and B and between G and F before barrier crossing. The height of the unfolding barrier $\Delta U$ probed in AFM experiments is $22\,\text{kcal/mol}$ [32]. Divided by the maximum number of concurrently breaking hydrogen bonds (six), the experimental value gives an average energy contribution from each hydrogen bond of $\sim 3.7\,\text{kcal/mol}$. From this we estimate that each hydrogen bond contributes approximate $3.5\,\text{kcal/mol}$ (note a deviation within $\pm 1.5\,\text{kcal/mol}$ from this number does not affect the conclusion) to the unfolding barrier. Hence, the five hydrogen bonds between $\beta$-strands A and B produce a barrier with $\Delta U_{AB} = 5 \times 3.5 \approx 18\,\text{kcal/mol}$. To estimate $\Delta U_{GF}$, 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 $\Delta U_{GF}$ two alternatives: $\Delta U_{GF} = 4 \times 3.5 = 14\,\text{kcal/mol}$, and $\Delta U_{GF} = 6 \times 3.5 = 21\,\text{kcal/mol}$. Denoting by $\tau_{AB}$ and $\tau_{GF}$ the mean first passage time for crossing barriers $\Pi_{AB}$ and $\Pi_{GF}$, respectively, the ratio of $\tau_{GF}$ to $\tau_{AB}$ can be calculated as a function of the externally applied force $F$. The results are shown in Fig. 4.8. One can recognize two regimes in Fig. 4.8. In the regime of forces $> 400\,\text{pN}$, 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 $< 200\,\text{pN}$, 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 $\Pi_{AB}$ is more likely to happen first and vice versa. The analysis implies that in AFM experiments or under physiological conditions 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 passing the unfolding intermediate I$_3$ [32], but in rare unfolding events the protein does demonstrate a transition state at extension close to intermediate I$_3$ (J. Fernandez, private 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.

4.4.3 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 $\beta$-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 $\beta$-strand swapping [87]. The partially unfolded intermediate I$_3$ found in present SMD studies is able to undergo such swapping since the two unravelled 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 [89]. Thus, one hypothesis for fibronectin fibrillogenesis is that 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.

4.4.4 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 Fig 4.9(a)). Both FN-III$_{10}$ and I27 are built in a similar $\beta$-sandwich architecture. The forced unfolding of I27 has been studied extensively in AFM experiments [100] and in SMD simulations [48, 66]. 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
Figure 4.9: 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 intra-strand hydrogen bonds between \( \beta \)-strands A and G; (d) partially unfolded structure at 160 Å extension.

eliminating surface effects. Fig. 4.9 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. \( \beta \)-strand A separates from the B-strand at \( \sim 140 \) ps after the bursting of two backbone hydrogen bonds between A and B (see Fig 4.9(b)), corresponding to an intermediate identified earlier both in SMD simulations and AFM experiments [35]. 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 Fig 4.9(c)), initiating rapid unraveling. The separation of \( \beta \)-strand A' from G, like the separations of \( \beta \)-strand A from B or G from F in SMD simulations of FN-III\textsubscript{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\textsubscript{10} in the way that the N-terminal \( \beta \)-strand A/A' connects with the C-terminal \( \beta \)-strand G. Strands A and G of FN-III\textsubscript{10} are separated, while I27 has \( \beta \)-strand A' form six backbone hydrogen bonds with \( \beta \)-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 simultaneously unraveling of both sheets [47]. In case of FN-III₁₀, 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 I₃ in the middle of unfolding process, as observed in simulations using an explicit solvent model reported here and simulations using implicit models [68].

4.4.5 Outlook.

The intermediate I₃ of FN-III₁₀ is not easy to observe in AFM experiments. According to the mean first passage time analysis described above, the reason is that overcoming barrier IIₐₕ is more difficult than crossing barrier II₉₉. 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 I₃. 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 [35, 36]. Disrupting A-B hydrogen bonds of I27 eliminates the pre-burst intermediate [35]. Mutating residues involving in forming A’-G hydrogen bonds can mechanically weaken the module [36].
Chapter 5
Mechanical unfolding of FN-III₁

5.1 Introduction

Fibronectin is a key extracellular matrix protein that not only provides a substrate for cell anchorage, but also serves as a regulatory protein in processes such as cell adhesion, motility, differentiation, and proliferation [19]. Structurally, fibronectin is a ∼540 kDa multimodular protein existing both as a soluble dimer and as an insoluble fibrillar component that incorporates into the ECM. Cells bind fibronectin through transmembrane proteins of the integrin family, which mechanically couple the actin cytoskeleton to the extracellular matrix. The mechanical responses of fibronectin are defined by its multimodular structure composed predominantly of three different repeats termed FN-I, FN-II, and FN-III. Individual FN-III modules have been proposed to unfold upon mechanical stretching of fibronectin, providing for the elasticity of fibronectin fibrils [13]. Consistent with this hypothesis, cells have been observed to stretch fibronectin up to four-fold their relaxed length [20, 82, 22].

In addition to providing necessary elasticity for accommodating cell movements, stretching of FN-III modules can expose buried binding sites that, for example, serve as nucleation sites for the assembly of fibronectin into its fibrillar form. These buried binding sites, termed cryptic sites, presumably exist either within the FN-III core or buried between the hinge regions of two neighboring FN-III modules. However, the role of mechanical force in this process of transforming fibronectin to a fibrillar form, fibrillogenesis, is only partially understood. It is known that cell-derived me-

---

3This chapter was adapted from the manuscript: M. Gao, D. Craig, O. Lequin, I. Campbell, V. Vogel, and K. Schulten. The structural basis for the mechanical unfolding intermediates of fibronectin type III₁ modules, (2003). Proceedings of the National Academy of Sciences, USA. In press. The NMR structure of FN-III₁ was resolved by Olivier Lequin and colleagues of the Campbell group.

58
Mechanical force is a prerequisite for FN polymerization \textit{in vitro}. Fibrillogenesis is initiated by both artificial [101] and cell-derived tensile forces, and by addition of partially or completely denatured FN-III modules (see review [19]). In contrast, fibrillogenesis does not occur if the mechanical linkage between cells and fibronectin has been disrupted by Rho inhibitors, myosin inhibitors, or actin-disrupting agents [102, 103, 104].

Cryptic sites for fibrillogenesis have been proposed to exist on FN-III$^1$ [84, 105], FN-III$^2$ [85], FN-III$^7$ [86], FN-III$^9$ [87], FN-III$^{10}$ [89], and FN-III$^{13-15}$ [86, 90]. Thermal or chemical unfolding of these modules is associated with increased binding by either fibronectin or a 70-kDa NH$_2$-terminal fibronectin fragment. In particular, partially folded FN-III$^1$ has been found to induce fibrillogenesis.

Studies have shown that heat-denatured FN-III$^1$ was able to bind FN and its 70 kDa amino terminal fragment [84]. Furthermore, a 76 amino-acid fragment from FN-III$^1$, termed anastellin and obtained by cutting A- and B-strand off the N-terminus region, is structurally stable and promotes assembly of so-called superfibronectin macromolecular structure [84, 86, 106, 107, 108], which was found to inhibit the growth of malignant cells [109].

While FN-III$^1$ and other FN-III modules have been probed for biologically relevant cryptic sites, these experiments were typically conducted using thermal and chemical denaturing or proteolytic cleavage. Consequently, it remained unclear whether partial denaturing can occur under physiological conditions and whether these denatured fragments are thus physiologically relevant. Recently, single molecule experiments using atomic force microscopy (AFM) have revealed that the mechanical unfolding pathway of FN-III$^1$ is markedly different from that of other FN-III modules such as FN-III$^{10}$, FN-III$^{12}$, and FN-III$^{13}$ [32]. FN-III$^1$ exhibits pronounced intermediate states during forced-unfolding. Since no structural information of the FN-III$^1$ module or its forced-unfolding intermediate was available, it was not known if these intermediates were functionally relevant or if these intermediates correlated to fragments promoting fibrillogenesis. To address this issue, we determined the structure of FN-III$^1$ using NMR spectroscopy. We then employed steered molecular dynamics (SMD) [48], which has been successfully applied to identify the mechanical unfolding intermediate of titin Ig domains previously [47, 35], to simulate the forced-unfolding of FN-III$^1$ and probe its mechanical intermediates. We find that the structure of one stable mechanical intermediate is closely related to the adhesive FN-III$^1$ fragment previously implicated to initiate
formation of fibronectin. Taken together with known experimental findings, these results suggest how mechanical stretching induces partial unfolding and exposure of cryptic sites.

5.2 Materials and Methods

5.2.1 Sample preparation

Human FN cDNA encoding FN-III₁ repeat (residues S609-S701 in Swiss-Prot accession No. P02751) was subcloned into a pGEX-6P-2 expression vector (Pharmacia). The recombinant module was expressed in *E. coli* BL21 strain, as a GST fusion protein. Uniformly $^{15}$N-labelled FN-III₁ was expressed in M9 minimal media by using 0.1% (w/v) $^{15}$NH₄Cl, as the sole nitrogen source. Recombinant FN-III₁ was purified using the protocol previously used for FN-III₁₃ preparation [110]. The identity and purity of the protein was confirmed by electrospray mass spectrometry. NMR samples typically comprised 1-2 mM protein in either 90 % H₂O / 10 % D₂O or 100 % D₂O containing 50 mM sodium phosphate buffer at pH 6. Sample used for residual dipolar coupling measurement contained 5% (w/v) of a 30/10/1 molar ratio of ditetradecyl-phosphatidylcholine/dihexyl-phosphatidylcholine/cetyltrimethylammonium bromide (CTAB).

5.2.2 NMR spectroscopy

NMR experiments were acquired at $^1$H frequencies of 500.1, 600.1 and 750.1 MHz on spectrometers built in-house at the Oxford Centre for Molecular Sciences, incorporating Oxford Instruments magnets, and on a Bruker DMX500 spectrometer. 3D $^1$H-$^{15}$N experiments were recorded at 500.1 MHz at 20 and 40 °C. $^1$H and $^{15}$N resonance assignments were obtained using 3D NOESY-HSQC (75 and 125 ms mixing times) and 3D TOCSY-HSQC (34 and 61 ms isotropic mixing times). Vicinal $^{3}$J$_{HNH\alpha}$ coupling constants were measured from a 3D HNHA experiment. Homonuclear 2D NOESY and TOCSY spectra were recorded at 750 MHz and 40 °C with mixing times of 125 ms and 55 ms, respectively. 2D NOESY, TOCSY and DQF-COSY spectra were also recorded at 500 MHz with a sample dissolved in D₂O. Slow-exchanging amide protons were identified on a series of 2D $^1$H-$^{15}$N HSQC spectra recorded after dissolving the sample in D₂O. Residual dipolar couplings were measured from IPAP $^1$H-$^{15}$N HSQC spectra [111]. Data were processed using the FELIX 2.3
software package (Accelrys, Inc., San Diego, CA) and Bruker XWINNMR program, running on Sun and SGI workstations. Spectra were analyzed with the aid of the XEASY program [112].

5.2.3 Structure calculation

Structures were calculated by torsion angle dynamics using the DYANA program [113]. A set of 100 structures was calculated, starting with randomized conformers and using a standard annealing protocol. The input for the structure calculation consisted of 1113 distance restraints and 81 dihedral angle restraints. 23 hydrogen bonds were constrained corresponding to slow-exchanging amide protons. The 25 conformers with the lowest residual target function (1.33 ± 0.31 Å²) were subsequently refined using the XPLOR-NIH program [114] and topallhdg forcefield. The structures were submitted to a simulated annealing protocol during which 57 dipolar couplings restraints were introduced. Finally the structures were energy-minimized using the CHARMM22 force field [74]. One structure was eliminated from the final set owing to large residual violations.

5.2.4 Molecular dynamics simulations

The MD simulations of FN-III₁ were carried out following the method described previously [47, 91]. The program NAMD [73] was used with the CHARMM22 force field and TIP3 water parameters [72]. Simulations were performed with a time step of 1 fs, a uniform dielectric constant of 1, and a cutoff of nonbonded forces with a switching function starting at 10 Å and reaching zero at 13 Å. FN-III₁ was solvated in a 64 Å diameter water sphere, resulting in a system of 12,532 atoms. The whole system was minimized for 2000 steps, followed by gradually heating from 0 K to 300 K in 10 ps. The temperature was maintained at 300 K for another 10 ps and the system was then equilibrated for 1.8 ns. Starting configurations for SMD simulations were obtained every 100 ps after the protein had been equilibrated for 1.5 ns. During SMD simulations, external harmonic force was applied to the C-terminus of the protein while fixing the N-terminus, with the stretching direction chosen along the vector pointing from N-terminus to C-terminus and a spring constant of 480 pN/Å. Trajectories were saved every picosecond. The end-to-end distance $R_{NC}$ is defined as the distance between the $C_{\alpha}$ atoms of the two termini. Force and $R_{NC}$ presented in Figure 3 were averaged every 4 ps and 20 ps for simulations at pulling velocities 0.05 Å/ps and 0.01 Å/ps, re-
spectively. The hydrogen bond energy presented in Figure 4 was calculated using explicit hydrogen bond parameters in the CHARMM19 force field. The total simulation time was 82 ns.

5.3 Results

5.3.1 Structure of FN-III

The structure of the recombinant FN-III module was determined by $^1$H, $^{15}$N NMR spectroscopy. The NMR structures exhibit few residual violations and low energies, indicating good agreement with experimental restraints and good geometry (Table 1). The structure of the FN-III module consists of two antiparallel $\beta$-sheets arranged as a $\beta$-sandwich with a long axis of $\sim$35 Å (Fig. 5.1). One $\beta$-sheet is formed by three strands (A, B, E) and the other $\beta$-sheet contains four strands (G,F,C,D) as in other FN-III modules resolved previously [115, 44, 45]. The $\beta$-strands are well-defined with a backbone RMSD of 0.37 Å (Tab. 5.1). Loops DE and EF show good precision, while loops AB, BC, CD and FG are more disordered, especially loop AB (backbone RMSD of 1.95 Å), which is longer than in other structurally known FN-III modules.

Equilibration of the explicitly solvated protein for 1.8 ns yields a stabilized structure with backbone RMSD from the NMR structure of less than 2.5 Å (Fig. 5.2a). The A-strand in the triple-stranded $\beta$-sheet ABE is unusually short with only four backbone hydrogen bonds between Glu5, Phe7 and Asn23, Gln21 of the B-strand in the NMR structure. This hydrogen bonding network was marginally extended during the equilibration through the formation of an additional hydrogen bond between Thr9 and Pro19 (Fig. 5.2b). Furthermore, two prolines, Pro12 and Pro19, twist the A- and B-strands away from each other and act as $\beta$-breakers, disrupting further hydrogen bonding between these two strands and resulting in an unusually weak $\beta$-sheet among FN-III modules (Fig. 5.2a). In contrast, in the four stranded $\beta$-sheet GFCD, the G- and F-strands form ten inter-strand hydrogen bonds (Fig. 5.2c), more than the F-G hydrogen bonds of any other structurally resolved FN-III modules. This structural difference, i.e., strengthening of the GFCD sheet, can be attributed to the mutation of a proline located in the FG loop, well conserved among other FN-III modules (Fig. 5.1e).
Figure 5.1: Structures and FN-III module sequence alignment. (a) Stereo views of the backbone heavy atom superposition of 25 minimized NMR structures. The two $\beta$-sheets and loops connecting $\beta$-strands are colored in green, red, and grey, respectively. The same color code was used for all other representations of FN-III$_1$. (b) Stereo views of the C$_\alpha$ atom trace and (c) cartoon drawings of a single FN-III$_1$ structure in two different orientations. (d) Schematic diagram of FN-III$_1$. (e) Sequence alignment of fourteen FN-III modules. The secondary structures of solved modules are highlighted. A key conserved proline among all other modules, but not in FN-III$_1$ and FN-III$_2$, is highlighted in purple. Protein snapshots were generated with VMD [75].
### Table 5.1: Statistics of FN-III NMR structure.

<table>
<thead>
<tr>
<th>Category</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residual distance constraint violations a</td>
<td></td>
</tr>
<tr>
<td>number &gt; 0.1 Å</td>
<td>2.4 ± 1.4</td>
</tr>
<tr>
<td>RMS deviation (Å)</td>
<td>0.014 ± 0.001</td>
</tr>
<tr>
<td>Residual dihedral constraint violations</td>
<td></td>
</tr>
<tr>
<td>number &gt; 5˚</td>
<td>3.0 ± 1.0</td>
</tr>
<tr>
<td>RMS deviation (˚)</td>
<td>2.1 ± 0.3</td>
</tr>
<tr>
<td>RMS deviation of dipolar couplings (Hz)</td>
<td>0.11 ± 0.003</td>
</tr>
<tr>
<td>RMS deviation from idealized covalent geometry</td>
<td></td>
</tr>
<tr>
<td>bond lengths (pm)</td>
<td>0.94 ± 0.01</td>
</tr>
<tr>
<td>bond angles (˚)</td>
<td>1.97 ± 0.02</td>
</tr>
<tr>
<td>improper torsion angles (˚)</td>
<td>2.35 ± 0.11</td>
</tr>
<tr>
<td>Van der Waal’s energy (kcal.mol⁻¹)</td>
<td>-427 ± 8</td>
</tr>
<tr>
<td>Ramachandran diagram (% residues)</td>
<td></td>
</tr>
<tr>
<td>most favored regions</td>
<td>80.7</td>
</tr>
<tr>
<td>additional allowed regions</td>
<td>17.6</td>
</tr>
<tr>
<td>generous allowed regions</td>
<td>1.6</td>
</tr>
<tr>
<td>disallowed regions</td>
<td>0.2</td>
</tr>
<tr>
<td>RMS deviation to mean structure b (Å)</td>
<td></td>
</tr>
<tr>
<td>N, Cα, C’ (1-93)</td>
<td>0.88 ± 0.19</td>
</tr>
<tr>
<td>N, Cα, C’ of β-strands</td>
<td>0.37 ± 0.08</td>
</tr>
<tr>
<td>Heavy atoms of β-strands</td>
<td>0.88 ± 0.09</td>
</tr>
<tr>
<td>N, Cα, C’ in loops c</td>
<td>1.26 ± 0.28</td>
</tr>
</tbody>
</table>

a The final input for the structure calculation consisted of 1113 distance restraints (206 intraresidual, 362 sequential, 93 medium-range and 452 long-range), 81 dihedral angle restraints (71 φ, 10 χ1) and 57 1H-15N dipolar couplings restraints. b Average coordinates for the 24 conformers after superposition for best fit of N, Cα, and C’ atoms of residues in β-strands (residues 5-8, 17-23, 30-39, 47-51, 57-61, 67-77 and 82-91). c RMS deviations of N, Cα, C’ atoms for individual loops are (Å): loop AB 1.95 ± 0.73, loop BC 1.00 ± 0.30, loop CD 1.08 ± 0.39, loop DE 0.27 ± 0.13, loop EF 0.54 ± 0.22, loop FG 1.10 ± 0.38
Figure 5.2: Structures of equilibrated FN-III\textsubscript{1}. (a) Alignment of NMR (cyan) and equilibrated (pink) structures shown in two different orientations. Two prolines that prevent A- and B-strands forming more inter-strand hydrogen bonds are colored in purple. (b) Inter-strand hydrogen bond networks of the ABE and (c) of the GFCD \( \beta \)-sheets, the latter in different orientations. Hydrogen and oxygen atoms are colored in white and orange. Hydrogen bonds are represented by black dashed lines.

### 5.3.2 Forced-unfolding Intermediates

Since inter-strand hydrogen bonds have been shown to play important roles in the mechanical stability of proteins [35, 97], the unusual hydrogen bonding network of FN-III\textsubscript{1} suggests a unique mechanical design. Indeed, mechanical unfolding experiments using AFM have shown that FN-III\textsubscript{1} has distinct intermediate states [32], which are not typical for other FN-III modules such as FN-III\textsubscript{12} and FN-III\textsubscript{13}. To further examine the mechanical response of FN-III\textsubscript{1}, we conducted five SMD simulations at constant stretching velocities of 0.05 Å/ps and 0.01 Å/ps for a total time of 82 ns.
Figure 5.3: Mechanical unfolding intermediates of FN-III$_1$. Curves shown at top left are force versus distance $R_{NC}$ profiles from constant velocity SMD unfolding simulations of FN-III$_1$ at speeds of 0.05 Å/ps and 0.01 Å/ps. $R_{NC}$ is defined as the end-to-end distance between two termini. Multiple runs at the same velocity are differentiated by a numeric suffix. Two force peak regions separated about 95 Å are highlighted. Four unfolding intermediates (I$_1$ - I$_4$) characterizing the unfolding pathway of FN-III$_1$ are shown as snapshots in cartoon representation. Transition from aligned state I$_2$ to I$_3$, and from I$_3$ to I$_4$ requires disrupting certain inter-strand hydrogen bonds between two β-sheets (a) ABE and (b, c) GFCD. The disruptions need assistance from surrounding water molecules (purple) that attack hydrogen bonds. Formed hydrogen bonds between backbone oxygen (orange) and hydrogen (white) atoms are represented as thick black lines.
Two pronounced peaks were observed at end-to-end distances $R_{NC}$ of $\sim 55 \, \text{Å}$ and $\sim 150 \, \text{Å}$ in all five unfolding simulations, and four intermediates have been identified along the mechanical unfolding pathway (Fig. 5.3). Upon stretching, the module first elongates the coiled termini to $\sim 40 \, \text{Å}$ with no obvious disruption of the tertiary structure, reaching state I$_1$. This state, also called “twisted state”, is similar to the first transitional state observed previously in studies of FN-III$_{7-10}$ modules [97]. Following I$_1$, the two $\beta$-sheets rotate to align themselves in the direction of the force vector, leading the protein to the second transitional state I$_2$ at $\sim 55 \, \text{Å}$, also termed aligned state.

The aligned state occurs prior to separation of the A-strand and corresponds to disrupting the hydrophobic core of the module, requiring a force of $\sim 500 \, \text{pN}$ in 0.01 Å/ps stretches. Entering the first peak area the force subsequently remains high, around 500 pN to 800pN, until the A-strand separates from the fold.

While the unfolding pathway is similar to that of other FN-III modules before the first $\beta$-strand detaches from the remaining fold [97], a significance divergence in the unfolding pathway for FN-III$_1$ was observed following the detachment. The A- and B-strands gradually unravel first under a lower force until they are completely separated from the folding core, extending the protein to $\sim 150 \, \text{Å}$. At this point, the protein reaches its most stable intermediate, termed I$_3$, or $C^{-G}$FN-III$_1$, because it contains C- to G-strands in the remaining folded core. Transiting further from I$_3$ requires a second peak force of up to $\sim 1400 \, \text{pN}$ to disrupt the remaining GFCD $\beta$-sheet. After this most prominent force peak the module gradually extends under a much lower force. Occasionally a fourth intermediate, I$_4$, was seen before a third force peak; the intermediate, also termed $C^{-E}$FN-III$_1$, contains only C-, D- and E-strands folded. Finally the protein fully extended at $\sim 330 \, \text{Å}$.

Transition from one intermediate to the next requires forces to overcome energy barriers contributed from inter-strand hydrogen bonds, and from van der Waals (vdW) and electrostatic packing interactions of the protein (Fig. 5.4). Transition from I$_2$ is connected to the first force peak between 40 to 70 Å. The force peak is correlated with the breaking of five intra-strand hydrogen bonds between A- and B-strands, the breaking of two to four hydrogen bonds between G- and F-strands near the C-terminus, and a severe perturbation of the hydrophobic core. Reaching I$_3$ and stretching the protein away from the intermediate, the second force peak at $\sim 150 \, \text{Å}$ arises, correlating with the burst of six to eight F-C hydrogen bonds and two to four G-F hydrogen bonds, as well as a
5.4 Discussion

5.4.1 Comparison with AFM experiments

Unfolding experiments of FN-III$_1$ repeats using AFM technique have revealed several mechanically stable conformations, separated by 90 Å or 200 Å and characterized through rupture forces of 90 pN and 120 pN, respectively [32]. This suggests the existence of intermediates which are consistent with
our SMD results, despite the fact that rupture forces measured in SMD are about one magnitude higher than in AFM experiments due to the different pulling velocity regimes [64]. According to SMD simulations, the extensions of 90 Å and 200 Å observed in AFM experiments are caused by the transition from I₂ to I₃ and by completely unraveling I₃. The transition from I₂ to I₃ is characterized by breaking the A-B inter-strand hydrogen bonds at distance $R_{NC}$ of 40 to 70 Å (55 ± 15 Å), and unraveling I₃ is characterized by breaking the G-F hydrogen bonds at $R_{NC}$ of ~150 Å. Accordingly, our simulations give an extension of ~95 Å from I₂ to I₃, and an extension of ~180 Å for unraveling I₃, in close agreement with AFM experiments. Furthermore, the SMD rupture force corresponding to disrupting I₃ is higher than the force needed to extend I₂, in qualitatively agreement with AFM experiments.

### 5.4.2 Comparison with anastellin NMR structure

Ruoslahti and colleagues have found that anastellin, a 76 residue FN-III₁ fragment beginning with residues NAPQ of the BC loop and spanning the remaining C-to-G-strands, is able to stimulate fibrillogenesis of fibronectin to form so-called superfibronectin [107]. This fragment has shown anti-metastatic effects on tumors grown in mice [109]. Unless the 76 residue fragment assumes a new conformation, structurally it would correspond to the folded core of the I₃ intermediate C-G-FN-III₁ found in our SMD studies, containing only the C-G-strands. Indeed, the predicted structure of intermediate I₃ is in excellent agreement with a recently published anastellin NMR structure [108], which exhibits a similar single antiparallel β-sheet consisting of C-, D-, F- and G-strands, as well as a flexible E-strand covering partially the hydrophobic core. Formed inter-strand hydrogen bonds shown in Fig. 5.3b match well with those observed in the NMR structure (Fig. 3 in [108]). A difference is found at the G-F inter-strand hydrogen bonding. Two bonds at residues Glu83 and Thr85 are missing in case of anastellin but are present in the simulated I₃, whereas two bonds at Phe89 near C-terminus were observed in NMR structure but are broken in I₃ due to the stretching.

### 5.4.3 Tuning mechanical stability of intermediates

Assuming the prominent unfolding intermediate I₃ has functional relevance, it is of interest to understand the origin of this intermediate, based on the NMR structure reported here. In this
respect, we note that the NMR structure shows that FN-III$_1$ contains four more hydrogen bonds between the F- and G-strands than FN-III$_{10}$, which exhibits an intermediate similar to I$_3$, but only rarely [116]. In fact, the intermediate is found only in the seldom case that stretching of FN-III$_{10}$ unravels the A-strand first. However, as a consequence of the strong F-G inter-strand hydrogen bonding, unfolding of FN-III$_1$ always begins with unraveling the A-strand and, accordingly, always leads to the I$_3$ intermediate. The structural reason for the strong F-G hydrogen bonding is the absence of a proline residue that is otherwise highly conserved and found at the beginning of the G-strand of other FN-III modules (Fig. 5.1). Based on other FN-III modules with known structure (FN-III$_7$–10,12–14), this proline creates a β-bulge, preventing formation of more backbone hydrogen bonds and increasing solvent accessibility to existing backbone hydrogen bonds between the F- and G-strands. In contrast, FN-III$_1$ and FN-III$_2$ belong to the few FN-III modules that do not include this proline and, consequently, do not contain the β-bulge. The absence of this proline thus enhances interactions between the F- and G-strands and stabilizes FN-III$_1$ and FN-III$_2$ over other FN-III modules. This hypothesis is consistent with AFM experiments in which FN-III$_1$ and FN-III$_2$ modules were found to be the mechanically most stable FN-III modules [32]. The absence of the stated proline in FN-III$_2$ implies that FnIII$_2$ is capable of forming a intermediate similar to I$_3$ described above. Interestingly, FN-III$_2$ has also been proposed to contain buried cryptic sites necessary for fibrillogenesis [85]. However, further research is needed to determine under what physiological conditions the I$_3$ intermediate may be exposed by mechanical stretching.

5.4.4 Significance of the intermediates

The intermediates of FN-III$_1$ can have profound physiological relevance. The states I$_1$ and I$_2$ may expose cryptic sites buried between domains or doubling the length of modules without inducing the unraveling of the entire FN-III tertiary structure. The presence of two late stage intermediates following the unraveling of the first β-strands, I$_3$ and I$_4$, can expose further cryptic sites buried deep in the hydrophobic core. Without providing a structural basis, it was previously suggested that fibrillogenesis involves the swapping of β-strands with other partially unfolded FN-III modules [87]. Our findings of a stable intermediate I$_3$ in the mechanical unfolding pathway support this suggestion. They provide for the first time a structural model revealing how cells exploit me-
chanical forces to expose cryptic sites on the FN-III module. Once an FN-III module is stretched into this intermediate, the straightened A- and B-strands can potentially be swapped with adjacent partially unfolded FN-III modules, thereby, promoting the binding between FN molecules.
Chapter 6

Mechanical Stability of FN-III modules

6.1 Introduction

Fibronectin type III modules (FN-III) are structurally ubiquitous, found in 2% of mammalian proteins, and essential to the function of many multimodular proteins. For the extracellular matrix (ECM) protein fibronectin containing more than fifteen repeating FN-III modules (see Fig. 6.1), it is remarkable that despite high similarity in tertiary structures, the sequence homology for these FN-III modules is conspicuously low, typically less than 20% between modules. By comparison, the sequence homology for the same module across multiple species is uniformly high, approximately 80% to 90% between modules, suggesting that the high variability in sequence is essential to function [16]. To a certain extent, some variability in sequence is expected for exposing different FN-III surface chemistries that specify protein-protein interactions. In fact, many of fibronectin’s major functions are linked to specific FN-III modules, such as cell binding through the RGD-loop on FN-III\textsubscript{10} [19]. However, potential interactions with other proteins do not completely explain the high degree of conserved sequence variability between FN-III modules nor do they provide an explanation as to why conserved variations occur in amino acids that are buried within modules while in equilibrium.

Recent evidence suggests that non-equilibrium states of FN-III modules may also play a critical role in fibronectin function. Cell stretching of extracellular fibronectin affects fibronectin’s func-

\footnote{This chapter was adapted from the manuscript: D. Craig, M. Gao, K. Schulten, and V. Vogel, Tuning the mechanical stability of fibronectin type III modules through sequence variation, (2003). *Structure.* In press.}
Figure 6.1: Fibronectin Monomer and FN-III modules. (a) Schematic of fibronectin emphasizing structural motifs, including FnI, FnII, and FN-III modules, and highlighting key functional sites [17]. The secondary structure for a typical FN-III module. The lower three $\beta$-strands are shown in green, the upper four $\beta$-strands in magenta, and loops in gray. (c) Alignment of tertiary structures for modules FN-III$_{10}$, FN-III$_{12}$, FN-III$_{13}$, FN-III$_{14}$, EDB and T-FN-III$_3$ demonstrating similarity in tertiary structures (coloring is the same as in b). (d) Sequences of FN-III modules found in human fibronectin and of T-FN-III$_3$ of tenasin. Coloring is green for lower $\beta$-strands and magenta for upper $\beta$-strands. Amino acids referred to in the text and participating in key hydrogen bonds are highlighted either yellow for backbone bonds or brown for side chain bonds. A key proline substitution is shown in blue.

Figure 6.1: Fibronectin Monomer and FN-III modules. (a) Schematic of fibronectin emphasizing structural motifs, including FnI, FnII, and FN-III modules, and highlighting key functional sites [17]. The secondary structure for a typical FN-III module. The lower three $\beta$-strands are shown in green, the upper four $\beta$-strands in magenta, and loops in gray. (c) Alignment of tertiary structures for modules FN-III$_{10}$, FN-III$_{12}$, FN-III$_{13}$, FN-III$_{14}$, EDB and T-FN-III$_3$ demonstrating similarity in tertiary structures (coloring is the same as in b). (d) Sequences of FN-III modules found in human fibronectin and of T-FN-III$_3$ of tenasin. Coloring is green for lower $\beta$-strands and magenta for upper $\beta$-strands. Amino acids referred to in the text and participating in key hydrogen bonds are highlighted either yellow for backbone bonds or brown for side chain bonds. A key proline substitution is shown in blue.

Figures 6.1: Fibronectin Monomer and FN-III modules. (a) Schematic of fibronectin emphasizing structural motifs, including FnI, FnII, and FN-III modules, and highlighting key functional sites [17]. The secondary structure for a typical FN-III module. The lower three $\beta$-strands are shown in green, the upper four $\beta$-strands in magenta, and loops in gray. (c) Alignment of tertiary structures for modules FN-III$_{10}$, FN-III$_{12}$, FN-III$_{13}$, FN-III$_{14}$, EDB and T-FN-III$_3$ demonstrating similarity in tertiary structures (coloring is the same as in b). (d) Sequences of FN-III modules found in human fibronectin and of T-FN-III$_3$ of tenasin. Coloring is green for lower $\beta$-strands and magenta for upper $\beta$-strands. Amino acids referred to in the text and participating in key hydrogen bonds are highlighted either yellow for backbone bonds or brown for side chain bonds. A key proline substitution is shown in blue.

Figures 6.1: Fibronectin Monomer and FN-III modules. (a) Schematic of fibronectin emphasizing structural motifs, including FnI, FnII, and FN-III modules, and highlighting key functional sites [17]. The secondary structure for a typical FN-III module. The lower three $\beta$-strands are shown in green, the upper four $\beta$-strands in magenta, and loops in gray. (c) Alignment of tertiary structures for modules FN-III$_{10}$, FN-III$_{12}$, FN-III$_{13}$, FN-III$_{14}$, EDB and T-FN-III$_3$ demonstrating similarity in tertiary structures (coloring is the same as in b). (d) Sequences of FN-III modules found in human fibronectin and of T-FN-III$_3$ of tenasin. Coloring is green for lower $\beta$-strands and magenta for upper $\beta$-strands. Amino acids referred to in the text and participating in key hydrogen bonds are highlighted either yellow for backbone bonds or brown for side chain bonds. A key proline substitution is shown in blue.

Figures 6.1: Fibronectin Monomer and FN-III modules. (a) Schematic of fibronectin emphasizing structural motifs, including FnI, FnII, and FN-III modules, and highlighting key functional sites [17]. The secondary structure for a typical FN-III module. The lower three $\beta$-strands are shown in green, the upper four $\beta$-strands in magenta, and loops in gray. (c) Alignment of tertiary structures for modules FN-III$_{10}$, FN-III$_{12}$, FN-III$_{13}$, FN-III$_{14}$, EDB and T-FN-III$_3$ demonstrating similarity in tertiary structures (coloring is the same as in b). (d) Sequences of FN-III modules found in human fibronectin and of T-FN-III$_3$ of tenasin. Coloring is green for lower $\beta$-strands and magenta for upper $\beta$-strands. Amino acids referred to in the text and participating in key hydrogen bonds are highlighted either yellow for backbone bonds or brown for side chain bonds. A key proline substitution is shown in blue.
enhance or weaken the mechanical stability of FN-III modules and to explain how these variations
are connected to fibronectin’s unique functional properties.

Steered molecular dynamics (SMD) both probe mechanical stability and provide the structural
data necessary for elucidating key interactions [48]. Previous SMD predictions agree well with
atomic force microscopy (AFM) results showing that FN-III10, containing the cell binding RGD-
loop, is mechanically one of the least stable FN-III modules [97]. These SMD results describe
the structural events immediately prior to separation of the first $\beta$-strand from FN-III modules
beginning with the breaking of one to two backbone hydrogen bonds connecting the A- and B-
strands. This event is followed by straightening in the A-strand, water entering the periphery of
the hydrophobic core, and subsequent slipping between the A- and G-strands such that upper and
lower $\beta$-sheets transition to a state where their $\beta$-strands are aligned with the external force vector,
the so-called aligned state. Finally, breaking of clusters of backbone hydrogen bonds results in the
separation of either the A- or G-strand followed by rapid unraveling of the entire module [97, 118,
91].

Below we compare the mechanical stability of FN-III modules with known high resolution
structures and provide structural insights into how the variability of single amino acids relates to
the mechanical stability of FN-III. SMD was used to investigate the early unfolding events for
modules FN-III10, FN-III12, FN-III13, FN-III14, FN-III$\text{EDB}$, and the third FN-III module from
the protein tenascin (T-FN-III$\text{3}$). We found that SMD protocols adopted prior to recent AFM
studies correctly predict the mechanical hierarchy for FN-III modules. We further find that while
mechanical unfolding ultimately begins with the breaking of a few key backbone hydrogen bonds,
amino acid side chains modulate mechanical stability by shielding key hydrogen bonds that break
early in the unfolding pathway from attack by water molecules. Our results also indicate that
the mechanical hierarchy for FN-III modules can be changed by interaction with the environment.
For example, FN-III10 becomes mechanically more stable as a result of protonation of carboxyl
side chains at lower pH, and FN-III13 becomes more stable as a result of heparin binding. These
results suggest a mechanism by which cells can tune the sequence in which FN-III modules undergo
structural and thus functional changes during mechanical stretching of fibronectin.
### 6.2 Method

<table>
<thead>
<tr>
<th># SMD runs:</th>
<th>barrier height (kcal/mol)</th>
<th>rupture forces (pN) at 0.6 μm/ps</th>
<th>hydrogen bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td>FN-III$_{10}$</td>
<td>15:15ns</td>
<td>15</td>
<td>50</td>
</tr>
<tr>
<td>FN-III$_{10}^{H/4.7}$</td>
<td>6:10ns</td>
<td>20</td>
<td>150</td>
</tr>
<tr>
<td>FN-III$_{12}$</td>
<td>11:20ns</td>
<td>24</td>
<td>230</td>
</tr>
<tr>
<td>FN-III$_{13}$</td>
<td>10:10ns</td>
<td>19</td>
<td>130</td>
</tr>
<tr>
<td>FN-III$_{14}$</td>
<td>13:17ns</td>
<td>18</td>
<td>100</td>
</tr>
<tr>
<td>FN-III$_{EDB}$</td>
<td>8:10ns</td>
<td>20</td>
<td>140</td>
</tr>
<tr>
<td>T-FN-III$_{3(1–90)}$</td>
<td>8:10ns</td>
<td>23</td>
<td>210</td>
</tr>
<tr>
<td>I-27</td>
<td>12:24ns</td>
<td>31</td>
<td>390</td>
</tr>
</tbody>
</table>

Table 6.1: Barrier heights and rupture forces determined by first passage time analysis for each of the modules investigated. Key hydrogen bonds corresponding with barrier crossing are highlighted in Fig. 6.1

SMD simulations were used to probe the mechanical properties of FN-III modules under force. Using the program NAMAD [73], molecular dynamic simulations were carried out on an all-atom model of individual FN-III modules with the CHARMM22 force field [74] and including explicit solvent described by TIP3P [72] water parameters. Forced unfolding was accomplished by adding a constant force to the C$_\alpha$ atom of the carboxy terminus along the direction of the vector connecting the two termini while the amino terminus was held fixed. Modules were solvated in a sphere of explicit water molecules with a radius of 32 Å, resulting in approximately 13,000 atoms so as to be consistent with previous studies and allow for good sampling per module [97, 91]. Single FN-III modules were adopted from known X-ray crystallographic or NMR structures. FN-III$_{10}$ was adopted from the FN-III$_{7–10}$ fragment (PDB code 1FNF) [44]. FN-III$_{12}$, FN-III$_{13}$, and FN-III$_{14}$ were adopted from the FN-III$_{12–14}$ fragment (PDB code 1FNH) [45]. An extended version of T-FN-III$_3$ was modeled by adding two amino acids to the carboxy terminus using published NMR data [119]. FN-III$_{EDB}$ was adopted from PDB code 2FNB [120] and T-FN-III$_3$ from PDB code 1TEN [121]. Simulations were also conducted on I27 to provide further comparison to AFM studies [47]. A total of ~85 SMD simulations encompassing together ~175 ns were conducted (see Tab. 6.1). Each nanosecond of simulation utilized ~200 CPU hours on Beowulf clusters of either 12 or 36 Athlon 1.2 GHz Processors.

Forces were applied to single FN-III modules ranging from 300pN to 1000pN. Extension is defined as the difference in end-to-end distance between equilibrated and stretched structures.
Extension vs. time plots clearly showed plateaus early in the unfolding pathway. The unfolding barrier height, $\Delta U$, is determined from Eq. 6.4 using a weighted least squared fit solving for $\Delta U$ and $D$, with a barrier width of 3.0 Å and using plateau duration as a measure of mean first passage (barrier crossing) time $\tau$. In some cases, barriers appeared less as distinct plateaus but more as regions of gradual extension due to straightening of the termini. In these cases, we analyzed changes in hydrogen bonding patterns, side-chain contacts, van der Waals (VDW) energies, $\Phi/\psi$ angles and solvent accessibility to identify a distinct event correlating with barrier crossing consistent with the other simulations. The most probably rupture force was determined by iteratively solving Eq. 6.3.

6.3 Results and Discussion

6.3.1 The mechanical hierarchy of FN-III modules

The relative order of mechanical stability for FN-III modules, defined as mechanical hierarchy, indicates the sequence in which different FN-III modules undergo functional changes when fibronectin is stretched. To determine this hierarchy, modules were stretched under constant force (cf-SMD). Plateaus in extension vs. time plots obtained from simulations correspond to overcoming unfolding barriers of FN-III modules. The duration of plateaus characterize the strength of barriers and, as expected, the passage time spent in a plateau region increases as force is decreased (see in Fig. 6.2) [76, 122]. Extension through plateaus correlates with the breaking of distinct hydrogen bonds [65, 97], shown in Fig. 6.2 and discussed in more detail in the next subsection.

SMD simulations provide an approach to predict the mechanical hierarchy of FN-III modules. As shown in Fig. 6.3, distinct differences in mechanical stability are visible when first passage times are plotted versus the applied pulling force. Each data point represents an independent cf-SMD simulation. The mechanically more stable modules are shifted to the right, reflecting longer first passage times to cross the barriers from the twisted to the aligned state (see Fig. 6.2). Overall, even without further quantitative analysis one can immediately observe the following relative mechanical stability: $I_{27} > FN-III_{12} \approx T-FN-III_{3} > FN-III_{13} \approx FN-III_{EBD} > FN-III_{14} > FN-III_{10}$.

Although in principle it is possible, in practice it is difficult to obtain from MD simulations the detailed shape of the mechanical unfolding barriers [123, 124]. Recently the Jarzynski equality has
Figure 6.2: Comparison of the key events that correlate with crossing of the first energy barrier for different FN-III modules. (a) The early unfolding pathway of FN-III modules typically exhibits two dominant barriers that are separated by plateaus in extension-time plots prior to unraveling the first β-strand. (b) The first barrier was typically highest for all simulations and its crossing correlated with the breaking of key force bearing hydrogen bonds as a result of increased attack from neighboring water molecules. For modules FN-III₁₀, FN-III₁₃, FN-III₁₄, FN-III₁₂, and T-FN-III₃, the following is shown: a time-extension plot in the upper left, an energy-time plot for key hydrogen bond(s) in the middle left, a snapshot of the FN-III module at the time indicated by the yellow arrow in the lower left, a surface representation of the yellow box illustrating how water approaches the buried hydrogen bonds is shown in the upper right, and a licorice representation of key amino acids cut out of the above surface representation is shown in the lower right. FN-III modules are ordered and labeled based on similarities in barrier crossing events as discussed in the text. Key amino acid side chains are colored by atom (red oxygen, white hydrogen, cyan carbon, and blue nitrogen), key buried hydrogen bonds are red dotted lines, water molecules are orange, the upper β-sheet is colored magenta, and the lower β-sheet is green.
been successfully applied to reconstruct a potential of mean force from SMD simulations [125] and quantitative analysis has been provided [126]. However, the construction poses extreme computational demands due to the need of sampling over many trajectories. A systematic reconstruction of the unfolding potential for FN-III modules is not possible at present due to the need of computing many trajectories, but the present results captures the key characteristics of the unfolding barrier along the spatial and the energy scales. For this purpose we use the mean first passage time theory [76, 127] together with dynamic force spectroscopy theory [128] that permits one to extrapolate from the SMD timescale the mechanical stability of FN-III modules for pulling velocities typical to AFM experiments.

![Figure 6.3](image)

Figure 6.3: Mean first passage time vs. force for FN-III_{EDB}, FN-III_{10}, FN-III_{12}, FN-III_{13}, FN-III_{14}, T-FN-III, and for comparison I-27. Data points represent time to cross the first barrier and fitted curves assume a 3Å barrier width. The stability decreases as follows: I27 > FN-III_{12} >≈ T-FN-III > FN-III_{13} ≈ FN-III_{EDB} >≈ FN-III_{14} > FN-III_{10}.

As suggested previously [76, 127], the mean first passage time theory relates the potential barrier to the duration of plateaus probed at different forces. By solving the Smoluchowski equation one can determine the mean first passage time \( \bar{\tau}(x_0; F) \) for extending the module from a position \( x_0 \) to \( b \) across the barrier \( U(x) \) under a constant force \( F \)

\[
\bar{\tau}(x_0; F) = \frac{1}{D} \int_{x_0}^{b} e^{\beta(U(x)-Fx)} dx \int_{a}^{x} e^{-\beta(U(y)-Fy)} dy. \tag{6.1}
\]

In this equation \( \beta = 1/k_B T \), \( x, y \) denote the reaction coordinate in the interval \([a, b]\), and \( D \) is the effective diffusion coefficient. For the present purpose of a simple characterization of the unfolding
barrier we assume the saw-tooth potential

\[
U(x) = \begin{cases} 
+\infty & x < a, \\
\Delta U (x - a)/(b - a) & a \leq x \leq b, \\
-\infty & x > b.
\end{cases}
\]

Here, \(\Delta U\) is the height of the potential barrier, and \((a - b)\) corresponds to the width of the barrier. The mean first passage time of interest is \(\tau(F) \equiv \bar{\tau}(x_0 = a; F)\). The saw-tooth potential inserted into 6.1 yields the relationship

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

where \(F\) is the externally applied force, \(\tau_d = (b - a)^2 / 2D\), \(\delta(F) = \beta[\Delta U - F(b - a)]\). This analytic solution permits one to estimate the barrier height, \(\Delta U\), and \(D\) using plateau duration as a measure of mean first passage time \(\tau\). The barrier width \(\sim 3.0\) Å is estimated from the extension value during the plateau (Fig. 6.2), a value that corresponds well to results from AFM experiments, for FN-III\(_{13}\) 3.4 Å [32], for FN-III\(_{10}\) 3.8 Å [32], and for I\(_2\) 2.5 Å [34]. A more detailed specification of the unfolding barrier would be inadequate given the fact that the input data taken from single SMD trajectories are extension and duration corresponding to observed plateaus, i.e., solely two values for each module. Fig. 6.3 matches \(\tau\) to the data points for the most prominent plateau of each module sampled. To provide upper and lower bounds for each module’s barrier height, the curves were also fitted with a barrier width of 2.5 Å and 3.5 Å, shown as the upper and lower bounds of the boxes in Fig. 6.4 below.

The most probable rupture force \(f\) for crossing an unfolding barrier is velocity dependent and can be approximated by iteratively solving [128]

\[
f \approx f_\beta \ln(v_s/v_\beta) + f_\beta \ln(f/f_\beta - 3/2) + (1/2) \ln(f/f_\beta)
\]

where \(v_s\) is the pulling speed and \(f_\beta \equiv k_B T/x_\beta\); the quantity \(v_\beta \approx L_p/4\tau_0\sqrt{x_\beta/b}\) represents the characteristic velocity for pulling a worm-like polymer taut, \(L_p\) is the contour length, \(b\) is the persistence length, \(\tau_0\) is the passage time extrapolated to zero force, and \(x_\beta\) is the barrier width. The
contour length and persistence length is assumed to be 28 nm and 0.4 nm, respectively, consistent with those used in AFM experiments. The variable $\tau_0$ is determined from Eq. 6.3 with a zero force using the previously determined barrier height.

A quantitative comparison between rupture forces delineated from SMD using Eq. 6.3 and Eq. 6.4, and those obtained from AFM data are shown in Fig. 6.4 assuming a pulling velocity of 0.6 $\mu$m/s. This velocity corresponds to those used in AFM experiments and is similar to the velocities associated with many cell processes, such as movement of kinesin along a microtubule [129]. Table 6.1 summarizes the calculated barrier heights at zero force and the calculated rupture forces at 0.6 $\mu$m/s for all the modules simulated here by SMD. It furthermore specifies the key hydrogen bonds that break at the onset of the barrier crossing.

We note that the relative order of mechanical stability for modules investigated by both AFM and SMD agree well. Both AFM and SMD find the relationship: I27 > FN-III12 > FN-III13 > FN-III10. Partial qualitative agreement is also found between AFM studies of T-FN-IIIAll and
our SMD simulations of T-FN-III₃. Here, we find that the mechanical stability of T-FN-III₃ is approximately equal or slightly less than the mechanical stability of FN-III₁₂. In comparison AFM finds an average rupture force for all FN-III modules in tenascin of 137±12 pN similar to the rupture force of 124±18 pN found in separate studies of FN-III₁₂ [40, 32].

Predicting rupture forces at AFM pulling speeds from SMD requires extrapolating pulling velocities over several orders of magnitude. One cannot expect a uniform scaling behavior over the broad range of velocity linking AFM and SMD [49, 130, 122]. It is thus noteworthy that the predicted rupture forces for FN-III modules are within less than an order of magnitude of the rupture forces found in AFM experiments, even if precise quantitative agreement is not yet accomplished. One reason for the lack of quantitative agreement could be that a “saw-tooth” potential that has been assumed to solve Eq. 6.3 is an oversimplification of the actual, but yet unknown barrier shape.

\[ U(x) = \begin{cases} \Delta U (x - a) / (a - b), & U_1(x) \\ \exp[-(x-b)^2/2], & U_2(x) \\ \exp[-2(x-b/2)^2], & U_3(x) \end{cases} \]

with \( a = 0, b = 3 \text{ Å}, \) and \( \Delta U = 22 \text{ kcal/mol}. \) \( U_1 \) corresponds to the saw-tooth model. The mean first passage time \( \tau \) for crossing the corresponding barrier \( U_i \) from \( a \) to \( b \) is calculated as a function of (b) externally applied force, (c) of barrier height \( \Delta U, \) and (d) of barrier width \( (b-a) \) for \( U_1, \) all using 6.1 with a diffusion constant \( D = 1 \text{ Å}^2/\text{ns}, \) and boundary conditions \( \bar{\tau}(a; F) = 0, \bar{\tau}'(b; F) = 0. \)

Figure 6.5: Comparison of the mean first passage time for three different barrier models demonstrating the effect of the different barrier shapes on mean first passage time, \( \tau. \) (a) \( U_1(x) = \Delta U (x - a) / (a - b), \)
\( U_2(x) = \exp[-(x-b)^2/2], \) and \( U_3(x) = \Delta U \exp[-2(x-b/2)^2], \) with \( a = 0, b = 3 \text{ Å}, \) and \( \Delta U = 22 \text{ kcal/mol}. \) \( U_1 \) corresponds to the saw-tooth model. The mean first passage time \( \tau \) for crossing the corresponding barrier \( U_i \) from \( a \) to \( b \) is calculated as a function of (b) externally applied force, (c) of barrier height \( \Delta U, \) and (d) of barrier width \( (b-a) \) for \( U_1, \) all using 6.1 with a diffusion constant \( D = 1 \text{ Å}^2/\text{ns}, \) and boundary conditions \( \bar{\tau}(a; F) = 0, \bar{\tau}'(b; F) = 0. \)

Since the mean first passage time \( \tau \) is dependent on the detailed shape of the barrier potential, it is of interest to compare impact of different barrier shapes. Fig. 6.5 shows three potentials and corresponding passage time as a function of external force, barrier height or barrier width. Clearly \( \tau \) is dependent on the barrier shape (Fig. 6.5a and b). Furthermore, steeper and/or narrower
barriers require much longer passage time as shown in Figure 5c and d. Nevertheless, we note that for a particular potential the passage time is monotonous as one changes the applied force. It has been shown that this monotonousness property is not dependent on the shape of the barrier [76]. The duration of passage time, hence, can be used to characterize the relative strength of FN-III modules as long as the same type of barrier shape is assumed. The facts that FN-III modules are structurally highly homologous and have been found to have similar barrier width is consistent with the hypothesis that FN-III modules extend across similar shaped barriers during the earliest stages of unfolding. Indeed, the good agreement between the mechanical hierarchy of FN-III modules predicted by SMD and obtained from AFM studies argues that the largest energy barrier probed by AFM is the same as the one investigated by SMD, i.e., that our analysis of SMD trajectories should yield the same ordering of the mechanical stability (rupture forces) as AFM experiments. In the future, improved computer resources will permit sampling of many trajectories per module, in which case one can accurately reconstruct the barrier shape using the approach in [125, 126] and hopefully achieve quantitative agreement with the AFM experiments.

6.3.2 Structural Basis for Mechanical Stability

SMD simulations furnish insight into the sequence of structural events that lead to mechanical unfolding. For each FN-III module, the events correlating with plateaus are detailed below and shown in Fig. 6.2. While some simulations were continued until each module was fully unfolded, we will only discuss noteworthy events around the separation of the first $\beta$-strand.

For all modules, the first significant structural event was the solvation and breaking of one to two conserved backbone hydrogen bonds connecting the A- and the B-strands. These backbone hydrogen bonds occurred between the 6th and 23rd amino acids and are next to a conserved proline as the 5th amino acid. Breaking of these hydrogen bonds was followed by one or two water molecules entering the periphery of the hydrophobic core, slipping between $\beta$-sheets such that the $\beta$-strands align with the external force vector, and finally led to separation of the first $\beta$-strand. Variability in these events could be correlated to amino acid substitutions and, likewise, provided an explanation for observed differences in FN-III mechanical stability.
**FN-III**<sub>10</sub>. Crossing the first plateau correlated to the solvation and breaking of the backbone hydrogen bond between Arg<sup>6</sup><sup>+</sup> of the A-strand and Asp<sup>23</sup><sup>−</sup> of the B-strand, resulting in a slight straightening of a highly conserved bulge between Pro<sup>5</sup> through Asp<sup>7</sup><sup>−</sup>. Immediately afterwards, the periphery of the hydrophobic core flanked by Trp<sup>23</sup> was solvated, followed by movement of the A- and G-strands until the β-strands of the two sheets appeared approximately aligned. Typically, a second plateau or a region of slow extension (3-7 Å) was observed that ended with the breaking of backbone hydrogen bonds and separation of either the A- or G-strand [97, 116].

**FN-III**<sub>13</sub>. Similar to FN-III<sub>10</sub>, the unfolding pathway of FN-III<sub>13</sub> contained two significant plateaus in time-extension plots prior to separation of the first β-strand, with the first plateau being longest in duration. Crossing the first plateau correlated with the solvation and breaking of the backbone hydrogen bonds between Arg<sup>6</sup><sup>+</sup> and Arg<sup>23</sup><sup>+</sup> and/or between side chain Tyr<sup>73</sup>(O<sup>γ</sup>) and Ser<sup>3</sup>(O). Passage across the second barrier again correlated with separation of either the A- or G-strand.

**FN-III**<sub>14</sub>. Crossing of the first plateau correlated with the breaking of backbone hydrogen bonds between two polar amino acids, Ser<sup>6</sup> and Gln<sup>23</sup>. Crossing of the shorter second plateau correlated with the breaking of several hydrogen bonds connecting the A-strand to the rest of the module. If the simulations were continued to full extension, a pronounced intermediate was found later in the unfolding pathway whereby the A- and B-strands were completely separated while the remainder of the FN-III module remained structurally intact. Similar late-stage intermediates have been previously described for other FN-III modules, although they are typically evident only when the FN-III modules were solvated in a large periodic box [116] or when an implicit water model was used [67, 68]. One functional importance of this intermediate might be that separation of the A- and B-strands exposes the Pro-Arg-Ala-Arg-Ile (PRARI) sequence at the BC loop which has been implicated in heparin mediated cell binding [131, 45]. The physiological importance of the PRARI sequence has been disputed previously since heparin binding was only apparent using peptide fragments and not observed in the fully folded FN-III<sub>14</sub> module [132]. These simulations suggest a mechanism by which mechanical force can expose this sequence under physiological conditions.
FN-III\textsubscript{12}. Unlike the FN-III modules described above, breaking of the Thr6-Thr23 backbone hydrogen bonds and partial solvation of the hydrophobic core did not result in significant extension. Rather, slipping of individual water molecules between the $\beta$-strands was prevented by an additional hydrogen bond between Met63 and Tyr68. The added stability likely resulted from substitution of a highly conserved proline to Val64 adding flexibility in the E-F loop and allowing for formation of a second hydrogen bond between Met63 and Tyr68. A proline is found at this position for all other fibronectin FN-III modules with the exceptions of FN-III\textsubscript{12} and FN-III\textsubscript{15} (highlighted in blue in Figure 1).

T-FN-III\textsubscript{3}. Similar to FN-III\textsubscript{12}, breaking of the Ser6 and Phe23 hydrogen bonds resulted in solvation of the periphery of the hydrophobic core but did not result in significant extension. Also similar to FN-III\textsubscript{12}, the $\beta$-strand alignment was prevented by a hydrogen bond between Tyr68 and Lys63\textsuperscript{+} that linked the upper and lower $\beta$-sheets. Extension beyond the longest plateau correlated with the solvation and breakage of either Gln8(HN) - Thr20(O) or Ala18(HN) - Iso59(O). Previous experimental studies have found that extending the 90 amino acid T-FN-III fragment used for the crystal structure by two residues, referred to as T-FN-III (1-92), resulted in thermodynamic stabilization of the module [133, 119]. Using the extended T-FN-III\textsubscript{3} (1-92) modeled from previously published NMR data [119], these additional residues did not appear to substantially affect the duration of the first plateau in SMD. For example, for T-FN-III\textsubscript{3}(1-92) the first plateau extended for 3.5 ns at 500pN and 0.7 ns at 550pN, slightly longer than 2.5 ns at 500pN and 0.6 ns at 550pN for T-FN-III\textsubscript{3}(1-90). This result is consistent with other observations that mechanical and thermodynamic stability are not necessarily correlated since the unfolding might occur along different trajectories [94, 34].

FN-III\textsubscript{EDB}. Unlike other FN-III modules investigated, FN-III\textsubscript{EDB} does not contain a proline at the fifth amino acid and, as predicted from the observations made above, the hydrogen bond between Thr6 and Thr23 broke without a major plateau. Noteworthy, absence of the proline at the fifth position in FN-III modules has been linked to FN-III aggregation though the detailed mechanism is unclear [134]. Thus, one possible role for the proline at the fifth amino acid is to stabilize the 6-23 hydrogen bonds. For FN-III\textsubscript{EDB}, further extension was prevented by a single side
chain hydrogen bond connecting the upper and lower $\beta$-sheets between Tyr33 and Leu26. A similar hydrogen bond involving a tyrosine side chain had also been found to substantially increase the mechanical stability of FN-III$_7$ [97] as it locks the $\beta$-sheets in place thus preventing the alignment of their $\beta$-strands with the external force vector. In FN-III$_{EDB}$, however, the stabilizing contribution of this hydrogen bond is less than in FN-III$_7$ due to the proline substitution and easier solvent access to Tyr33 and Leu26.

**Comparison.** Common themes emerge when comparing the structural origins of mechanical stability in FN-III modules. The mechanical stability of FN-III modules is related to how well key hydrogen bonds that break early in the unfolding pathway are shielded from attack by water molecules. In FN-III$_7$ [97], FN-III$_{12}$, FN-III$_{EDB}$, T-FN-III$_3$ hydrogen bonds buried in the hydrophobic core between the O$^\gamma$ atom of a tyrosine in the upper $\beta$-sheets and a HN backbone atom in the lower $\beta$-sheets prevented alignment of the $\beta$-strands with the external force vector even when parts of the hydrophobic core had been solvated. In particular for FN-III$_{12}$, this tyrosine hydrogen bond appeared to be even further stabilized by a lower solvent accessibility, apparently resulting from the absence of a proline at position 64. Second, it was found that amino acid side chains, particularly those containing charged side groups contacting the 6$^{th}$ and 23$^{rd}$ amino acid in the A- and B-strand respectively, lowered mechanical stability by shielding key backbone hydrogen bonds from attack by water molecules. For example, the mechanical stability of FN-III$_{10}$ containing charged amino acids between the Arg6$^+$ and Asp23$^-$ amino acids was considerably lower than that of other FN-III modules not containing similarly charged amino acids. This finding is investigated in more detail in the next section.

### 6.3.3 The mechanical stability of FN-III$_{10}$ is pH dependent.

Previous experimental and SMD studies agree in that FN-III$_{10}$ is one of the mechanically weakest modules despite having an essential role in cell binding and the formation of a mechanically stable complex [97, 32]. In contrast to mechanical weakness, FN-III$_{10}$ has high thermodynamic stability [135]. Recently Koide et. al. found that FN-III$_{10}$ becomes even more thermodynamically stable at $\sim$pH 4.7 as a result of protonation of three negative amino acids: Asp7$^-$, Asp23$^-$, and Glu9$^-$ [136]. While mechanical stability does not necessarily correlate with thermodynamic stabil-
ity [94, 34], this finding was intriguing since these amino acids also play a key role in protecting hydrogen bonds between R6\(^{+}\) and Asp23\(^{-}\) from attack by water molecules.

We conducted additional SMD simulations of FN-III\(_{10}\) with Asp7\(^{-}\), Asp23\(^{-}\), and Glu9\(^{-}\) neutralized by protonation to determine the effect of a lower pH on mechanical stability. It was immediately found during thermalization and equilibration that a new side chain hydrogen bond formed between the protonated carboxyl groups of Asp23 and Asp7. During a 1ns equilibration of FN-III\(_{10}\) at a pH of approximately 4.7, this new hydrogen bond would stochastically break and the side chains of Asp7 and Glu9 would temporarily come closer (see Fig. 6.6). This finding is particularly interesting because the side chains of these amino acids are partially separated during equilibration of neutral pH FN-III\(_{10}\) due to the repulsive interactions of these three negative side groups. In SMD simulations, the first passage times for FN-III\(_{10}\) became significantly longer at lower pH and the barrier height increased by roughly 33\%, from 15 kcal/mol to 20 kcal/mol. As a result, the mechanical stability of FN-III\(_{10}\) rises from being the least stable of the FN-III modules tested.
here to being even more stable than FN-III$_{13}$. Further stabilization may result from conformational changes occurring beyond the timescale of molecular dynamics equilibration.

The finding that FN-III$_{10}$ is mechanically more stable at lower pH may have physiological relevance. Lower extracellular pH is associated with cells in stressed environments, such as in wound and cancer tissues where measurements of the bulk pH are approximately 6-7 and 5-6, respectively [137, 138]. Additionally, fibronectin is known to play a key role in bone resorption which is also associated with a reduction in pH to the surrounding environment [139]. The pH of the microenvironment immediately neighboring FN-III$_{10}$ will be lowered further due to the overall negative charge found on modules FN-III$_{7-10}$ and as a result of local pH gradients. Finally, the three negative residues involved in the thermal [136] and mechanical stabilization of FN-III$_{10}$ are highly conserved and unique to FN-III$_{10}$ across multiple organisms, consistent with their predicted functional importance in regulating the mechanical stability of FN-III$_{10}$.

6.3.4 Heparin binding stabilizes FN-III$_{13}$

FN-III$_{13}$ contains a heparin-binding site, Hep-II, that is required for inducing stress fibers and formation of focal contacts [132]. Recent studies mapping the Hep-II binding site find that heparin binding to FN-III$_{13}$ involves interactions with the first 29 amino acids including six positive amino acids that form a cationic cradle [132, 140, 110, 45].

Notably, these same amino acids involved in heparin binding also play key roles in the early stages of unfolding. Unfolding of FN-III$_{13}$ begins with separation of the A- and B-strands containing 25 of the first 29 previously mentioned amino acids. More specifically, the amino acids involved in the formation of key force bearing hydrogen bonds between Ser3 - Thr73 and Arg6$^+$ - Arg23$^+$ as identified by SMD were also found to undergo significant NMR chemical shifts upon heparin binding (Ser3, Arg6$^+$, Arg23$^+$), indicating a decrease in their solvent exposure. This is consistent with measurements indicating that heparin binding to FN-III$_{13}$ is entropically driven with a high energy of solvation. Further mutagenesis studies also identified Arg6$^+$ and Arg7$^+$ to be necessary for heparin binding [110]. Taken together, it is likely that binding of heparin to this Hep-II site prevents water molecules from attacking the force bearing hydrogen bonds between Arg6$^+$ - Arg23$^+$. If complexation limits the access of water molecules to these backbone hydrogen bonds,
it is expected that heparin binding to FN-III$_{13}$ stabilizes the FN-III$_{13}$ module.

6.4 Conclusions

Determination of the mechanical hierarchy in which FN-III modules unfold if mechanically stretched is essential to learn how cells use force to regulate the exposure of molecular recognition sites by stretching extracellular matrix proteins. Our SMD predictions regarding the hierarchy in which FN-III modules unfold agree well with experimental findings performed by atomic force microscopy. Here we explain how single amino acid changes, or even the protonation of single carboxyl groups in FN-III$_{10}$, alter the accessibility of water to key backbone or backbone-side chain hydrogen bonds that break early in the unfolding pathway. Identification of amino acids that regulate the mechanical stability of FN-III modules provides novel functional explanations into why the sequence homology of FN-III modules is low despite their remarkable structural similarities. Our data further suggest that the mechanical hierarchy of unfolding is not strictly given by the amino acid sequence but can be regulated by the local environment and by complex formation with other molecules. For example, for FN-III$_{10}$ we found that protonation of the carboxyl groups of Asp7, Glu9, Asp23, corresponding to a lowered pH, can substantially increase the first major energy barrier contributing to its mechanical stability. Furthermore, insights into the role played by key residues in the early unfolding events suggest that heparin binding to FN-III$_{13}$ will increase its mechanical stability. Consequently, our studies imply that the mechanical stability of FN-III modules involved in cell binding can be altered in response to environmental changes. Our results also uncover novel structural principles that regulate mechanical stability: variations in the mechanical stability of different proteins depend not simply on the pattern of backbone hydrogen bonds, but more critically on how well key hydrogen bonds are shielded from attack by water molecules. Taken together with the known biological functions of fibronectin, these results suggest how subtle variations in amino acid sequence can have a profound effect on protein function and ultimately cellular behavior.
Chapter 7

Ligand Unbinding from Integrin $\alpha_V\beta_3$

7.1 Introduction

Cells generate mechanical tension in their actin cytoskeleton and exert forces on the extracellular matrix (ECM) through a type I transmembrane class of proteins termed integrins. The shifting balance of forces from different integrins on the ECM regulates essential cell processes including migration, proliferation, and differentiation [19, 141]. Consequentially, strong binding of integrins to the ECM is required to withstand considerable mechanical forces. Thus, it is amazing that despite the large size of integrins and the large size of many ECM proteins, integrin-ECM adhesion is typically mediated through the binding of integrins to short ligand peptides, such as the famous RGD peptide. This observation is highlighted in the recent crystal structure of integrin $\alpha_V\beta_3$ bound to a cyclic RGD peptide whereby only the Arg and Asp side chains are found to be significantly buried within the integrin [142]. Considering the importance of integrin binding to the ECM for cell function, it is of interest to determine how the rather small RGD peptide sequence remains bound to an integrin under considerable mechanical force.

The first X-ray crystallographic structures of the extracellular portion of integrin $\alpha_V\beta_3$ [145, 142] revealed that, as predicted, the ligand binding region consists of a $\beta$-propeller domain and an $\beta$A (or I-like) domain homologous to the $\alpha$A-domain (or I-domain) found in other integrins. Also expected was the observed coordination of a divalent cation at MIDAS (Metal Ion-Dependent Adhesion Site) that contacts the aspartate of the RGD peptide. However, the structures showed also unexpected features. For example, both the liganded and unliganded $\alpha_V\beta_3$ exhibit a 45 degree “V” shape bend

\footnote{This chapter was adapted from the manuscript: D. Craig, M. Gao, K. Schulten, and V. Vogel. Molecular dynamics simulations reveal a critical role for divalent manganese cations in integrin binding of an RGD ligand., (2003).}
structure (see Fig. 7.1). Unexpected were also the observed role of two additional divalent cations found near the MIDAS. The MIDAS is occupied by a divalent ion in the liganded state but such ion cannot be identified in the unliganded state, unlike the cases for integrins containing an αA domain where both the liganded and unliganded αA domain can capture a divalent ion (see, for example, [146]). In addition to the MIDAS, two other divalent cation sites, termed ADMIDAS (Adjacent MIDAS) and LIMBS (Ligand-associated Metal Binding Site), were found in αVβ3 integrin near the MIDAS. Located 8 Å from the MIDAS, the ADMIDAS can be occupied by either a Mn^{2+} or a Ca^{2+} ion, no matter whether the ligand is present or not. The LIMBS, located 6 Å from the MIDAS, acquires a Mn^{2+} ion when the RGD peptide is present, but does not favor binding a Ca^{2+} ion. From the structure alone the roles for these unexpected ions are unclear, although it has been
suggested earlier that the ions coordinate the MIDAS ion [142].

We conducted molecular dynamics simulations of the headpiece of integrin \( \alpha V/\beta 3 \), consisting of the \( \beta A \) domain, the hybrid domain, and the \( \beta \)-propeller domain bound to a cyclic RGD peptide (shown in Fig. 7.1 and Fig. 7.2). Equilibrium MD simulations were first used to investigate the contacts between the RGD peptide and the binding sites. Non-equilibrium SMD simulations were then used to study the structural mechanism by which integrins remain bound to their RGD ligand under significant mechanical forces that arise between contractile cells and the extracellular matrix. Our results reveal a novel role for the LIMBS ion and provide insight into how divalent cations facilitate non-covalent binding that can withstand significant mechanical force.

7.2 Methods

The extracellular portion of integrin \( \alpha V/\beta 3 \), together with a cyclo-RGDfV ligand bound to the integrin, was adopted from the crystallographic structure determined at 3.2 Å resolution (Protein Data Bank entry code 1L5G). To reduce the size of the system, domains other than \( \beta A \), hybrid domain and \( \beta \)-propeller were removed. The remaining parts were subsequently solvated in a TIP3 [72] water cell of size \( 130 \times 130 \times 120 \) Å\(^3\), resulting in a system of 146,291 atoms. The divalent cations found in the original crystal structure were included in the solvated structure. Seven additional divalent ions were introduced to neutralize the system. MD simulations were then carried out with the program NAMD [73]. An MSI version of CHARMM22 force fields [74] with additional parameters for Mn\(^{2+}\) cation was utilized [147]. The system was first minimized for 2000 conjugate gradient steps. After minimization it was gradually heated from 0 K to 300 K in 30 ps. Subsequently the system was equilibrated for 1 ns under NPT conditions. The pressure was maintained at 1 atm using the Langevin piston method [98], and the temperature was controlled by using Langevin dynamics at 300 K with a damping coefficient of 5 ps\(^{-1}\). A piston period of 100 fs and a damping time constant of 50 fs were employed for the pressure control. To prevent drifting, the backbone C\( _\alpha \) atoms of both C-termini residues Arg\(^{438}\) on the \( \alpha \) subunit and Asp\(^{432}\) on the \( \beta \) subunit were fixed. During 1ns equilibration the system exhibited an all-atom RMSD of less than 1.5 Å for the \( \alpha \) subunit and of less than 2.5 Å for the \( \beta \) subunit.

External forces were applied to stretch out the RGD ligand using constant velocity SMD simu-
lations (cv-SMD). Pulling velocities of 0.05 Å/ps and 0.02 Å/ps were used. The center of mass $R_c$ of the ligand cyclo-RGDfV was harmonically constrained with a force $F = k(vt - x)$, where $k$ is the spring constant, $v$ is the velocity, and $t$ is the time. The spring constant was set to $10 k_B T/\AA^2$. The direction of the force was chosen by the sum of two vectors, pointing from the two fixed C$_\alpha$ atoms on the $\alpha/\beta$ subunits to $R_c$ respectively. The variable $x$ is the coordinate along the chosen direction of the force.

![Diagram](image)

**Figure 7.2**: Regions of integrin $\alpha_V/\beta_3$ investigated by molecular dynamics. (a) Schematic of the integrin binding headpiece, including the $\beta$-propeller, $\beta_A$, and hybrid domains. (b) The integrin binding headpiece was solvated in a periodic box containing 44,300 water molecules (blue semi-transparent lines) and 140,000 atoms overall. (c) Cartoon representation emphasizing secondary structure of $\beta$-propeller, $\beta_A$, and hybrid domain. (d) Surface representation of $\beta$-propeller, $\beta_A$, and hybrid domain shown from the same point of view as in (c). For all pictures, the $\beta_3$ integrin subunit is colored magenta, the $\alpha_V$ subunit is colored light blue, and the cyclic RGDFV peptide is colored yellow.

Periodic boundary conditions were imposed for all MD simulations. Full electrostatics was computed every 4 fs using the Particle Mesh Ewald (PME) method [99], with a grid spacing of $\sim$1.0 Å. A cut-off smoothly switching off between 10 Å to 12 Å was employed for van der Waals interactions. An integration time step of 1 fs and a uniform dielectric constant of 1 were
chosen. A total of 3 ns simulation time was completed. The simulations were carried out at the Pittsburgh Supercomputing center (PSC) and local Linux clusters. A 1 ns simulation required 40 hours running on 32 Compaq Alphaserver ES45 nodes consisting of 128 1GHz processors. During simulations atomic coordinates of the whole system were saved every picosecond. Visualization and analysis of these trajectories were conducted using VMD [75].

7.3 Results

Below we first describe the structural changes observed during the equilibration. Then we applied harmonic forces at constant velocities to probe both wild type and mutated αV/β3-RGD systems.

7.3.1 Structural changes during the equilibration

Integrin modules participating in RGD binding, including the βA domain, the hybrid domain from the αV subunit, and the β-propeller from the β3 subunit, were equilibrated in an NPT ensemble simulation of approximately 140,000 atoms over a one nanosecond (see Fig. 7.3), yielding a stable system with the backbone heavy atom RMSD of <1.5 Å for the β-propeller domain and of <2.5 Å for the βA domain.

While only minor changes were observed in the integrin complex as a whole, several important structural changes were observed at the RGD binding site. The most significant change is that one of the carboxylic oxygen from Asp<sub>RGD</sub> formed a new contact with the LIMBS ion, while the other carboxylic oxygen of Asp<sub>RGD</sub> maintained its original contact with the MIDAS ion, as shown in Fig. 7.3. Thus, following the equilibration Asp<sub>RGD</sub> contacted both the MIDAS and LIMBS ions, whereas the residue contacted only the MIDAS ion in the crystal structure. As expected water molecules not present in the original crystal structure were found to coordinate the solvent accessible divalent cations, i.e., the MIDAS and ADMIDAS ions. Two water molecules surround the ADMIDAS ion and one water molecule coordinate with the MIDAS ion (see Fig. 7.4).

Changes are also apparent between the αV subunit and Arg<sub>RGD</sub> of the cyclo-RGDiV ligand (Fig. 7.3a). Arg<sub>RGD</sub> maintains its bidentate salt bridge to Asp<sup>218</sup> of αV. However, a new bidentate salt bridge forms between the Arg<sub>RGD</sub> and Asp<sup>150</sup> of αV, whereas only one salt bridge was observed between these amino acids in the crystal structure. New contacts are also formed between Arg<sub>RGD</sub>
Figure 7.3: Snapshots of RGD binding site of integrin $\alpha_\text{V}/\beta_3$ from (a) the original crystal structure and (b) after a one nanosecond equilibration. During the 1 ns equilibration, several new contacts are made to the cyclo-RGDF ligand. The aspartic acid (Asp$^{RGD}$) amino acid makes a new contact with the LIMBS ion such that after the equilibration one carboxylate oxygen is coordinated by the MIDAS ion and the other carboxylate oxygen is coordinated by the LIMBS ion. In addition, both Asp218 and Phe177 make new contacts with the arganine of the cyclo-RGDfV ligand (Arg$^{RGD}$). Representations are chosen to show as much of the binding site as possible. Both the $\beta_3$ integrin subunit (magenta) and $\alpha_\text{V}$ the subunit (light blue) are shown in surface representation. The cyclo-RGDFV ligand is in a CPK representation. Amino acids from integrin $\alpha_\text{V}/\beta_3$ that form salt bridges with the cyclo-RGDFV ligand are shown in van der Waals representation (carbon atoms are cyan, nitrogen ions blue, oxygen atoms are red, polar hydrogens are white, and non-polar hydrogen are not visible). The ADMIDAS (gray), MIDAS (green) and LIMBS (purple) ions are shown in VDW representations. Salt bridges and hydrogen bonds are shown as dotted white lines, and a backbone trace of the RGD peptide is shown as a yellow dotted line.
Figure 7.4: Coordination spheres of the LIMBS, ADMIDAS, and MIDAS ions. Atoms are shown using a CPK representation (carbon atoms are cyan, nitrogen ions blue, oxygen atoms are red, and hydrogen are not visible) and a portion of the integrin backbone connecting these amino acids is shown as a gray ribbon. The ADMIDAS (gray), MIDAS (green) and LIMBS (purple) ions are shown in VDW representation. Contacting atoms are shown as red dotted lines if no change occurred during equilibration, whereas yellow dotted lines indicate a contact was broken during equilibration, blue lines indicate a contact with a water molecule, and green dotted lines indicate a contact was made during equilibration.
and backbone oxygen of Gln$^{180}$ and Phe$^{177}$. Overall, most of the original contacts between the cyclo-
RGDfV ligand and $\alpha V/\beta 3$ were well maintained throughout the course of the equilibration. Emerging
new contacts suggest that the cyclo-RGDfV ligand is in a slightly more stable conformation after
the equilibration.

It is highly surprising to observe that Asp$^{RGD}$ reoriented such that it coordinated with both the
MIDAS and LIMBS ions. This event occurs within a few picoseconds of the equilibration and does
not require substantial changes in the coordination spheres of these ions. As shown in Fig 7.4, the
MIDAS coordination sphere is nearly identical when comparing the crystallized and equilibrated
structures. The only change is that one of the empty coordination positions are filled by a water
molecule, as would be expected. Only minor changes were observed in the coordination sphere of
the LIMBS ion. Here, the backbone oxygen of Asp$^{217}$ (of $\beta 3$) was replaced by the two side chain
carboxylic oxygen of the same amino acid (Asp$^{217}$). The backbone oxygen of Glu$^{156}$ filled an empty
coordination position at approximately 100 ps and through the rest of the 1 ns equilibration. Upon
co-coordinating Asp$^{RGD}$, the distance between the MIDAS and LIMBS ions decreased from 5.9Å
to 4.9Å and Glu$^{220}$ moved slightly such as to not reside directly between the LIMBS and MIDAS
ions. Thus when considered together, the co-coordination of Asp$^{RGD}$ required only very minor
changes at the binding sites of the LIMBS and MIDAS ions.

Motivated by the observation that Asp$^{RGD}$ can be coordinated to both the LIMBS ion and to
the MIDAS ion, additional equilibration was conducted. It has been found in the crystallography
studies that ADMIDAS could coordinate either Mn$^{2+}$ or Ca$^{2+}$, whereas the MIDAS and LIMBS
coordinated only Mn$^{2+}$ ions [145, 142]. To investigate the role of different divalent cations, we sim-
ulated the integrin complex as described replacing the Mn$^{2+}$ ions with Ca$^{2+}$ ions. Because of the
relatively short (nanosecond) simulation time, the Ca$^{2+}$ ions would not disassociate. Interestingly,
in two of three equilibration simulations involving Ca$^{2+}$ ions, Asp$^{RGD}$ coordinated only with the
MIDAS ion, consistent with the crystal structure, whereas in only one simulation Asp$^{RGD}$ coor-
dinated with both the LIMBS and MIDAS ions, in a fashion similar to the structure assumed in
the equilibration involving Mn$^{2+}$ ions. These simulations provide some insight into the differential
effects of Mn$^{2+}$ and Ca$^{2+}$ ions on the binding of RGD ligands.
7.3.2 Forced ligand unbinding

To induce the unbinding of the cyclo-RGDfV ligand from the integrin binding pocket, a harmonic constraint was applied to the ligand moving at different constant velocities. Fig. 7.5 presents the key steps of the unbinding process identified in a cv-SMD simulation at velocities of 0.05 Å/ps. Identical events revealed in Fig. 5 have also been observed when the ligand was pulled at the slower velocity of 0.02 Å/ps.

Upon pulling the cyclo-RGDF ligand, the first major structural event observed is breaking of the contact between Asp$^{RGD}$ and the LIMBS ion, and a slight reorientation of Asp$^{RGD}$ so that both of its carboxylic oxygen contact only the MIDAS ion (Fig. 7.5b). This event typically revealed itself as a shoulder in the major force peak, in force vs. extension curves. The major force peaks correlates with the breaking of the contact between Asp$^{RGD}$ and the MIDAS ion (Fig. 7.5d). It is important to note that breaking of this contact is accomplished by attacks from surrounding water molecules, which appear to compete for the carboxylic oxygen of Asp$^{RGD}$ that normally interact only with the MIDAS ion. This competition appears to be very analogous to the breaking of backbone hydrogen bonds in the forced unfolding of FnIII/I27 modules where water molecules attack hydrogen bonds, thereby weakening the electrostatic interaction between the oxygen and hydrogen atoms [65, 97, 118, 116, 148]. Similarly it has been demonstrated that water is critical in other ligand detachment studies, such as unbinding of biotin from streptavidin [149, 150].

Following detachment of Asp$^{RGD}$ from the MIDAS and LIMBS ions, the cyclo-RGDF ligand continued to pull away from the integrin binding pocket until a series of significantly smaller force peaks arise in the force-extension curves. The first smaller force peak corresponds to the detachment of Arg$^{RGD}$ from Asp$^{150}$ and Phe$^{177}$. A third and even smaller force peak was observed as Arg$^{RGD}$ broke the final salt bridge with Asp$^{218}$ (Fig. 7.5e). In some simulations, the minor force peaks were observed as transient salt bridges formed as the cyclo-RGDF ligand would make a temporary salt bridge. Similar to breaking of the Asp$^{RGD}$-MIDAS contacts, water molecules were found to compete with and attack the salt bridges connecting Arg$^{RGD}$ to the integrin, facilitating its detachment.
Figure 7.5: Stretching the cyclo-RGDF ligand out of the binding sites of $\alpha V\beta 3$ at a constant velocity of 0.05Å/ps ($a$–$f$). The corresponding force-extension curves is shown below. One major force peak is observed, corresponding to the detachment of $\text{Asp}^{\text{RGD}}$ from the coordination sphere of the MIDAS ion, followed by smaller peaks corresponding to the breaking of salt bridges between $\text{Arg}^{\text{RGD}}$ and the $\alpha v$ subunit. Note that multiple water molecules are required to facilitate detachment of $\text{Asp}^{\text{RGD}}$ from the MIDAS coordination sphere and that the water molecule that was originally coordinated at MIDAS from the equilibration does not facilitate detachment without the contribution of other water molecules. Coloring is the same as in other figures, with the exception that all water molecules within 3Å of the MIDAS ion, LIMBS ion, and $\text{Asp}^{\text{RGD}}$ carboxylic oxygens ($\text{Asp}^{\text{RGD}}:\text{OD1}$, $\text{Asp}^{\text{RGD}}:\text{OD2}$) are shown as red (oxygen) and white (hydrogens) spheres.
Figure 7.6:  

(a) Force-extension curves at pulling velocities of 0.05 Å/ps and 0.02 Å/ps. Typically, one major force peak was observed followed by smaller sub-peaks. The major force peak corresponds to the breaking of the contact between the MIDAS ion and the carboxylate atom(s) of Asp$_{RGD}$. A shoulder on the main force peak was typically observed and corresponded to the breaking of the contact between one (but not both) of the carboxylate oxygens of Asp$_{RGD}$ with a divalent cation. Variations were observed in the smaller force peaks, which generally corresponded to the breaking of individual salt bridges between amino acids of αV and Arg$_{RGD}$. 

(b) Force-extension curves at 0.05 Å/ps for the electrostatic component of the MIDAS and/or LIMBS turned off (ΔMIDAS or ΔLIMBS). Notably, the largest force is reduced only slightly when the MIDAS ion is turned off, as the carboxylate oxygens of Asp$_{RGD}$ reorient to coordinate exclusively with the LIMBS ion. However, when both the MIDAS and LIMBS ions are neutralized, the primary force peak is substantially decreased, and only force peaks associated with breaking of salt bridges to Arg$_{RGD}$ are observed.

7.3.3 Unbinding of mutants

The SMD simulations revealed that the largest barrier to ligand detachment corresponds to disrupting the interaction between the MIDAS ion and the carboxylic oxygen of Asp$_{RGD}$. Despite forming up to five salt bridges only minor force peaks were observed at these pulling velocities, suggesting that the interactions between Asp$_{RGD}$ and divalent cations within the binding pocket are essential for the adhesion between RGD peptides and the integrin. To examine this hypothesis, additional SMD simulations were conducted where we turned off the electrostatic component of either the MIDAS ion (ΔMIDAS simulations), the LIMBS ion (ΔLIMBS simulations), or both the MIDAS and LIMBS ions (ΔMIDAS-ΔLIMBS simulations), as shown in Fig 7.6. In ΔMIDAS
simulations, the primary force peak was reduced slightly, since \( \text{Asp}^{RGD} \) reoriented such that its carboxylic oxygen only contacted the LIMBS ion. A similar small reduction in the force peak was observed for \( \Delta \text{LIMBS} \) simulations, where the cyclo-RGDF ligand reoriented to only bind the MIDAS ion. The major force peak at an approximate extension of 6\AA{} was almost non-existent in \( \Delta \text{MIDAS}-\Delta \text{LIMBS} \) simulations. These simulations demonstrate that the electrostatic interaction between the divalent cations and the carboxylate oxygen of \( \text{Asp}^{RGD} \) are essential for forming a non-covalent bond between RGD and integrin \( \alpha_V/\beta_3 \). Interestingly, whereas more than one water molecule was needed to be within 3\AA{} of the \( \text{Asp}^{RGD} \) carboxylate oxygen in order for it to detach when both MIDAS and/or LIMBS were active, no additional water molecules were needed to detach \( \text{Asp}^{RGD} \) in \( \Delta \text{MIDAS}-\Delta \text{LIMBS} \) simulations.

### 7.3.4 Coordination with water molecules

During equilibration of integrin-ligand complex a water molecule is observed to be coordinated to the MIDAS ion and just 3-4\AA{} from the carboxylate oxygen of \( \text{Asp}^{RGD} \). This water molecule comes within 3\AA{} of the oxygens of \( \text{Asp}^{RGD} \) several times during constant velocity pulling without leading to the breaking of this contact. Instead, this water molecule remains preferentially coordinated to the MIDAS ion. It is not until both an additional water molecule comes within 3\AA{} of \( \text{Asp}^{RGD} \)‘s oxygen and the cyclo-RGDF has stretched 5-6\AA{} out of it’s binding pocket, that breaking of the \( \text{Asp}^{RGD} \)-MIDAS contact occurs. The water appears to block the path of other water molecules that could facilitate detachment of the cyclo-RGD peptide. Based on these observations, it seems likely that the highly positive charge on the divalent cation keeps the water molecule from facilitating the detachment of \( \text{Asp}^{RGD} \).

### 7.4 Discussion

The results demonstrate divalent cations incorporated by integrin \( \alpha_V/\beta_3 \) at its binding sites have several critical functions that are not apparent from the static crystal structure alone. Molecular dynamics simulations reveal that the \( \text{Asp}^{RGD} \) of the cyclic RGD peptide is coordinated not only by the MIDAS ion but also to the LIMBS ion. Additionally, SMD simulations show how these divalent cations add mechanical stability to the binding between integrin \( \alpha_V/\beta_3 \) and an RGD peptide.
by preferentially coordinating water molecules such that they do not facilitate detachment of the cyclo-RGDFV peptide.

One of the principal questions regarding integrin biology concerns the structural basis for integrin binding [151, 152]. The integrin $\alpha_V\beta_3$ crystal structures led to many models regarding the changes in the overall quarternary and tertiary structure during integrin activation. It was not clear whether the bend in integrin $\alpha_V\beta_3$ was an artifact of crystallization, represented an inactivated form, or only indicated generally flexibility of the molecule. While our results do not provide definite answers to these questions, the fact that significant force was required to detach the cyclo-RGDFV peptide integrin $\alpha_V\beta_3$ does suggest that the hybrid domain, and the $\beta$-propeller domains are in a conformation that is capable of binding an RGD peptide with a high barrier to disassociation.

The contact between the LIMBS ion and the Asp$^{RGD}$ was not anticipated from the crystal structure. In fact, the function of the LIMBS ion was previously unclear. Here, we show that the LIMBS ion plays a critical role in stabilizing the cyclo-RGDFV peptide. With this finding, new questions emerge, such as why two divalent cations are found to mediate ligand binding for integrin $\alpha_V\beta_3$? Neither cation was observed in the unliganded structure, suggesting that their occupation is required for activating the integrin. Perhaps, two divalent cations allows for further modulation of $\alpha_V\beta_3$’s activation state or increasing the tolerance of the integrin on occupation of ion binding sites. As shown in our simulations, deleting either MIDAS or LIMBS ions only marginally affect the unbinding peak force.

Ligand binding involving divalent metal ions is a recurrent theme in cell adhesion. Our results provide evidence that, in particular, the electrostatic attraction between the metal ions and a ligand is competent at withstanding high mechanical loads. The notion that electrostatic interactions provide for mechanical stability has been previously found for other systems. For example, the interstrand hydrogen bonds and electrostatic packing between $\beta$-strands are known to play a critical role in the mechanical stability of immunoglobulin domains, such as titin’s Ig domains or fibronectin’s FnIII modules [35, 65, 97, 118, 116, 148, 153]. This mechanical stability critically depends on the accessibility of hydrogen bonds to water molecules [65, 97]. Key force-bearing hydrogen bonds have been shown to assume greater mechanical stability when buried within the
hydrophobic core. Thus, it is interesting to see that immediately following pulling the cyclo-RGDfV peptide from its binding site, a water molecule is found to be coordinated to the MIDAS ion, only ~2.5Å from the Aspartate carboxylate group. Instead of facilitating detachment of the Asp$^{RGD}$ from the MIDAS ion, this water molecule remained coordinated to the MIDAS ion, even as the ligand was pulled with force exceeding 1500 pN. Detachment eventually did occur, but only after several additional water molecules had inserted into the binding site.

The incorporation of metal ions by integrin receptors establishes a mechanically stable binding to ligands through minor conformational changes. Precise coordination of molecules surrounding the metal ion provides a mechanism whereby a mechanically stable linkage can be created without requiring the ligand to be substantially buried within the integrin. Based on this insight, we propose that the reason why ligand-integrin binding involves ions is that it strengthens binding and permits the integrin-RGD contact withstand substantial mechanical force, without requiring the ligand to be significantly buried within the binding site, such as in biotin-streptavidin.
References


Vita

Mu Gao was born in Neijiang, Sichuan Province, China on April 20, 1974. In 1995 he graduated from Zhejiang University in Hangzhou, China, with a bachelor degree in Physics. From 1995 to 1998 he studied in Peking (Beijing) University, Beijing, China, where he received Master of Science degree in Physics. He then came to University of Illinois at Urbana-Champaign and continued his graduate study in Physics. In 1999 he joined Klaus Schulten’s group to study theoretical biophysics.

Publications:


