Size-Dependent Mechanical Properties of Beta-Structures in Protein Materials

by

Sinan Keten

B.S., Civil Engineering
Boğaziçi University (2005)
M. Eng., Civil and Environmental Engineering
Massachusetts Institute of Technology (2006)

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2010

Author .....
Department of Civil and Environmental Engineering
May 15, 2010

Certified by .....
Markus J. Buehler
Associate Professor of Civil and Environmental Engineering
Thesis Supervisor

Accepted by .....
Daniele Veneziano
Chairman, Departmental Committee for Graduate Students

Size-Dependent Mechanical Properties of Beta-Structures in Protein Materials

by

Sinan Keten

Submitted to the Department of Civil and Environmental Engineering
on May 15, 2010, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in the field of Civil and Environmental Engineering

Abstract

Protein materials such as spider silk can be exceptionally strong, and they can stretch tremendously before failure. Notably, silks are made entirely of proteins, which owe their structure and stability to weak molecular interactions, in particular, hydrogen bonds (H-bonds). Beta-structures, a class of protein folds that employ dense arrays of H-bonds, are universal in strong protein materials such as silks, amyloids, muscle fibers and virulence factors. The biological recipe for creating strong, tough materials from weak bonds, however, has so far remained a secret. In this dissertation, size, geometry and deformation rate dependent properties of beta-structures are investigated, in order to provide a link between the nanostructure and mechanics of protein materials at multiple length scales. Large-scale molecular dynamics (MD) simulations show that beta-structures reinforce protein materials such as silk by forming H-bonded crystalline regions that cross-link polypeptide chains. A key finding is that superior strength and toughness can only be achieved if the size of the beta-sheet crystals is reduced to a few nanometers. Upon confinement into orderly nanocrystals, H-bond arrays achieve a strong character through cooperation under uniform shear deformation. Moreover, the size-dependent emergence of a molecular stick-slip failure mechanism enhances toughness of the material. Based on replica-exchange MD simulations, the first representative atomistic model for spider silk is proposed. The computational, bottom-up approach predicts a multi-phase material with beta-sheet nanocrystals dispersed within semi-amorphous domains, where the large-deformation and failure of silk is governed by the beta-structures. These findings explain a wide range of observations from single molecule experiments on proteins, as well as characterization studies on silks. Results illustrate how nano-scale confinement of weak bond clusters may lead to strong, tough polymer materials that self-assemble from common, simple building blocks.

Thesis Supervisor: Markus J. Buehler
Title: Associate Professor of Civil and Environmental Engineering

Acknowledgments

I would like to express my gratitude for many who have supported me during my studies at MIT. First and foremost, I would like to thank my advisor Professor Markus J. Buehler, who has introduced me to the nano-scale realm, for his excellent mentorship and support during my Ph.D. He has been a great source of motivation and inspiration for me. I would like to thank Professor Kausel, Professor Connor, Professor Kamm, Professor Lang and Professor Grossman for fruitful discussions, and for their guidance. I would also like to acknowledge Professor Michael Feig at MSU and Ross Walker at SDSC for their help with supercomputing applications. For their friendship and excellent discussions, I am also thankful to my peers at LAMM. Over the past few years, I have had to opportunity to co-advise several undergraduate and visiting students: Andrea Nova, Britni Ihle, Maya Srinivasan, Xuefeng Chen, J. Fernando Alvarado and Jessica Duran have made great contributions to this work. Finally, I am grateful to my family and friends for their never-ending support, encouragement and faith in my academic endeavors.

This research was funded by an MIT Presidential Fellowship, and grants by the Office of Naval Research (N00014-08-1-00844), National Science Foundation (CMMI-0642545 and MRSEC DMR-0819762), the Army Research Office (W911NF-06-1-0291), DARPA (HR0011-08-1-0067), as well as the UROP office at MIT. Their support is greatly appreciated. I would also like to acknowledge an allocation of advanced computing resources supported by the National Science Foundation (TeraGrid, grant no. TG-MSS080030).

Dedicated to my parents, Nilüfer and Okan Keten.

Contents

1 Background **17**

1.1 Physical concepts of failure at different length scales 19

1.2 Chemistry and hierarchical structure of protein materials 25

1.3 Outline 30

2 Methodology **33**

2.1 Atomistic and molecular modeling 33

2.1.1 Classical molecular dynamics formulation 33

2.1.2 Force fields 37

2.1.3 Coarse-grained modeling approaches for protein structures . . 40

2.1.4 Treatment of solvent: Implicit vs. Explicit 43

2.2 Sampling and force manipulation methods 45

2.2.1 Steered molecular dynamics 45

2.2.2 Quasi-static constant force method 46

2.2.3 Improving sampling: Replica exchange MD 48

2.3 Data analysis and visualization 50

2.3.1 Analysis methods for simulation results 50

2.3.2 Visualization schemes 50

2.4 Experimental techniques 51

2.5 Theoretical bond strength models at the atomic level 52

3 Generic fracture strength model for beta-structures in proteins **57**

3.1 Mechanical significance of beta-sheet proteins 57

3.2 Molecular structure of beta-sheets and atomistic modeling results . . 60
3.2.1 MD Simulations in tear and shear modes for a simple beta-sheet model 63
3.3 A generic fracture strength model for proteins 65
3.3.1 Griffith-Irwin energy balance concept and the protein strength model 67
3.3.2 Prediction of maximum strength for H-bond clusters 74
3.3.3 Application to double strand shear in beta-sheets 74
3.3.4 Influence of H-bond cluster size 77
3.4 Experimental validation 80
3.4.1 Direct comparison with AFM experiments 80
3.4.2 Comparison with proteomics data 84
3.5 Discussion and conclusions 87

4 Nanomechanics of beta-solenoid structures 91
4.1 Background on beta-solenoids 93
4.2 Single beta-helices in tension and compression 95
4.3 Triple beta-helices and the cell-puncture needle application 105
4.4 Solenoid cross-sectional shape its mechanical implications 115
4.5 Discussion and conclusion 119

5 Size dependence of the strength and toughness of beta-sheet nanocrystals in silk 125
5.1 Background on spider silk and significance of nanocrystals 126
5.2 Simulation setup and theoretical background 127
5.2.1 MD simulation setup 129
5.2.2 Continuum theory formulations 131
5.3 Size-effect of beta-sheet nanocrystals: Predictions and experimental validation 133
5.4 Discussion and implications for materials design 140

6 Nanostructure and molecular mechanics of spider dragline silk protein assemblies **145**

6.1 Background on spider silk ultrastructure and mechanics 146

6.2 Simulation approach 148

6.2.1 Structure identification 150

6.2.2 Mechanical analysis 152

6.3 REMD and mechanical stretching simulation results 153

6.3.1 Predicted representative structures and experimental validation 153

6.3.2 Molecular mechanics of dragline silk protein assemblies 156

6.4 Development of a constitutive law 161

6.5 Implications of nanostructure on micromechanics 168

6.6 Discussion and conclusion 172

7 Conclusion **177**

7.1 Summary of key findings and significance 177

7.2 Opportunities for future research 181

A Script library **185**

A.1 .tcl scripts 185

A.1.1 Script to calculate H-bonds in VMD 185

A.1.2 Script to calculate phi-psi angles in VMD 185

A.1.3 Script to calculate secondary structure ratios in VMD 186

A.2 MATLAB scripts 187

A.2.1 Code for averaging force-extension curves from NAMD 187

A.2.2 Serial spring model for spider silk 189

A.3 Shell scripts 192

A.3.1 Basic shell script for setting up REMD with EEF1 on Teragrid Abe cluster 192

B Extension of the beta-sheet strength model **195**

B.1 Application to alpha-helices 195

List of journal publications

1. **S. Keten**, M. J. Buehler, "Nanostructure and molecular mechanics of dragline spider silk protein assemblies", in review.

2. **S. Keten**, Z. Xu, M. J. Buehler, "Triangular core as a universal strategy for stiff nanostructures", in review.

3. M. J. Buehler, **S. Keten**, "Colloquium: Failure of molecules, bones, and the Earth itself", *Reviews of Modern Physics*, 2010, 82(2). Highlighted by M. Buchanan, "Learning from Failure", *Nature Physics*, 2009, 5(705).

4. **S. Keten**, Z. Xu, B. Ihle, M. J. Buehler, "Nanoconfinement controls stiffness, strength and mechanical toughness of β-sheet crystals in silk", *Nature Materials*, 2010, 9, p. 359-367.

5. **S. Keten**, M. J. Buehler, "Atomistic model of the spider silk nanostructure", *Applied Physics Letters*, 2010, 96: p. 153701. (Cover article)

6. R. Paparcone, **S. Keten**, M.J. Buehler, "Nanomechanical properties of Alzheimers Aβ(1-40) amyloid fibrils under compressive loading", *Journal of Biomechanics*, 2010, 43 (6): p. 1196-1201.

7. M. Srinivasan, S.G.M. Uzel, A. Gautieri, **S. Keten**, M. J. Buehler, "Alport Syndrome mutations in type IV tropocollagen alter molecular structure and nanomechanical properties". *Journal of Structural Biology*, 2010, 168(3): p. 503-510.

8. **S. Keten**, J. F. R. Alvarado, S. Muftu and M. J. Buehler, "Nanomechanical characterization of the triple β -Helix domain in the cell puncture needle of bacteriophage T4 Virus", *Cellular and Molecular Bioengineering*, 2009, 2(1): p. 66-74.

9. T. Ackbarow, **S. Keten**, and M.J. Buehler, "A multi-timescale strength model of alpha-helical protein domains". *Journal of Physics: Condensed Matter*,

2009,(3): p. 035111.

10. M. J. Buehler, **S. Keten**, and T. Ackbarow, "Theoretical and computational hierarchical nanomechanics of protein materials: Deformation and fracture". *Progress in Materials Science*, 2008, 53(8): p. 1101-1241.

11. **S. Keten** and M.J. Buehler, "Geometric confinement governs the rupture strength of H-bond assemblies at a critical length scale". *Nano Letters*, 2008, 8(2): p. 743-748.

12. **S. Keten** and M.J. Buehler, "Asymptotic strength limit of hydrogen bond assemblies in proteins at vanishing pulling rates". *Physical Review Letters*, 2008, 100: p. 198301.

13. **S. Keten** and M.J. Buehler, "Large deformation and fracture mechanics of a beta-helical protein nanotube: Atomistic and continuum modeling". *Computer Methods in Applied Mechanics and Engineering*, 2008, 197(41-42): p. 3203-3214.

14. **S. Keten** and M.J. Buehler, "Strength limit of entropic elasticity in beta-sheet protein domains". *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 2008, 78(6): p. 061913.

15. M. J. Buehler and **S. Keten**, "Elasticity, strength and resilience: A comparative study on mechanical signatures of α-helix, β-sheet and tropocollagen domains". *Nano Research*, 2008, 1(1): p. 63-71.

16. T. Ackbarow, X. Chen, **S. Keten**, M. J. Buehler, "Hierarchies, multiple energy barriers, and robustness govern the fracture mechanics of alpha-helical and beta-sheet protein domains". *Proceedings of the National Academy of Sciences*, 2007, 104(42), p.16410-16415. (Cover article)

List of Figures

1-1 Multi-scale mechanisms of failure. 24

1-2 Schematic describing protein synthesis. 26

1-3 Hierarchical features in biological materials. 28

1-4 Multi-scale modeling. 29

2-1 Molecular dynamics simulation approach. 34

2-2 Historical development of computational power. 37

2-3 Coarse-grained modeling of proteins with bead models. 41

2-4 AFM experiments and SMD simulation. 47

2-5 REMD simulation method. 49

2-6 Bell model and the energy barrier concept 55

3-1 Shear topology of beta-sheets in proteins. 59

3-2 Molecular structure of beta-sheets and simulation setup. 61

3-3 Molecular dynamics simulation results from tear and shear mode. . . 64

3-4 Pulling rate dependence of proteins. 66

3-5 Schematic figures for explaining theory and simulations. 68

3-6 Strength prediction for a double-strand slip case. 76

3-7 AFM experiments on beta-sandwich I27 domain in titin. 80

3-8 Application of BSSM to I27 domain in titin. 83

3-9 Size dependent strength and length distribution of beta-strands. . . . 85

3-10 Characteristic dimensions of common protein secondary structures. . 86

3-11 Sensitivity analysis to BSSM parameters. 88

4-1 Atomistic model of a single-beta-helix. 92
4-2 Tension and compression of a single-beta-helix. 96
4-3 Rupture events in a single-beta-helix under tension. 98
4-4 Relationship between rupture events and mechanical behavior. 99
4-5 Compression simulation setup on the single-beta-helix. 102
4-6 Compressive deformation and failure map of single-beta-helices. . . . 103
4-7 Compression simulation setup for the triple-beta-helix. 108
4-8 Snapshots of deformation for the triple-beta-helix. 109
4-9 Rate dependence of force and stiffness. 111
4-10 SPS simulation results for the triple-beta-helix. 112
4-11 Continuum formulations for triple-beta-helix. 113
4-12 Triangular core of amyloids and beta-solenoids. 116

5-1 Simulation setup and theoretical considerations for nanocrystals. . . . 128
5-2 Deformation and failure of beta-sheet nanocrystals. 134
5-3 Size-dependent stiffness of nanocrystals. 136
5-4 Size scaling of strength and toughness in nanocrystals. 138
5-5 Hierarchical effects in the architecture of nanocrystals. 141

6-1 REMD simulation protocol and results. 149
6-2 Comparing MaSp1 and MaSp2 predictions with experiments. 154
6-3 Secondary structure of representative replicas. 157
6-4 Force-displacement behavior of selected structures. 159
6-5 Stretching structural transformation in MaSp1 and MaSp2 models. . 160
6-6 Mechanics of high-turn content silk structures. 162
6-7 Constitutive elements of spider dragline silk. 165
6-8 Stress-strain response of the silk fibril model. 169

B-1 Schematic for alpha-helix unfolding. 196

List of Tables

1.1 Strength of chemical bonds in biology. 27

4.1 Diameter and persistence length of biological filaments. 121

5.1 Elastic properties of beta-sheet nanocrystals. 137

6.1 Parameters for the semi-amorphous region in the mesoscale model. . . 165

6.2 Parameters for the beta-sheet nanocrystal in the mesoscale model. . . 167

Chapter 1

Background

Natural and biological phenomena are among the key driving forces of engineering practice, as they not only inspire the design of man-made systems, but also continuously test their function during service. For example, flight, inspired by bird locomotion, enabled mankind to defy gravity and explore far reaches of space, only through designs that can stand the the worst natural hazards, including weather storms, radiation and vast temperature as well as pressure variations. This is perhaps the most well-known example of biomimetics, the study of nature and biology for transferring ideas to engineering. In this regard, the practice of making and using natural materials for improving the human condition dates back to early ages, starting with the use of readily available resources such as stone, wood and bone. Most strikingly, the rate with which we develop new materials has grown tremendously over centuries. Just within the past few decades, many key technological advancements have been pioneered using new materials and material design strategies. Composites have revolutionized aviation and space exploration, polymers have prolonged life expectancy through their use in armor technology and medicine, and silicon has transformed our world by facilitating the information technology revolution. The race to develop materials that can surpass those readily found in nature, however, has lead to massive use of resources and unsustainable practices. New developments in materials science and engineering will have to acknowledge this need to return to environmentally friendly practices [7].

Following Feynman's famous address [79] in 1959 entitled "There is plenty of room at the bottom", referring to intriguing physical phenomena at the nano-scale, current trends in developing new materials rely heavily on fabricating systems at exceedingly smaller length scales through nanotechnology, reaching the limit of what can even be detected or observed with the most sophisticated characterization tools available. Controlling matter at the nano-scale in a bottom-up scheme has many advantages compared with classical top-down approaches. Key phenomena that cannot be anticipated through continuum predictions can be investigated and exploited at the atomic scale with nano-scale engineering. Most importantly, biological materials form in this bottom-up fashion, for instance through self-assembly, and have superior properties compared with common engineered materials [7, 42, 84, 245]. A key issue that needs to be resolved for truly mastering biomimetics is understanding the role of hierarchies that extend from nano to macro in structural biological materials, and how this translates into enhanced physical properties that we seek and want to improve.

Most of the structural materials in biology employ proteins as a key building block, and most proteins are made from only a handful of the 20 common amino acids that define all life forms on our planet. The genius in biology is, therefore, not in the diversity of the building blocks, but rather in the complexity of their structural arrangement. Examples for such structural ingenuity include protein materials such as spider silk, hair, connective tissue (skin, cartilage, tendon) and cells [245]. All of these materials have appealing properties such as very high extensibility and fracture toughness, making them suitable for their functions requiring flexibility and mutability [232, 93]. Proteins also serve as templates for biomineralized composites employing inorganic materials, and thereby lead to a whole different class of materials such as abalone shell, lobster cuticle and bone [231, 27]. This different material design strategy leads to lower deformation capacity as required by their function but very high strength and toughness. The fundamental understanding of self-assembly (*association*) and failure (*dissociation*) of protein materials at the level of chemical bonds is very important for transferring key biological material design concepts into engineering solutions [41, 42].

In order to address the need for understanding mechanics of protein materials, this thesis focuses on one of the most common topological features in structural protein materials called beta-structures. These structures are common to silks, muscle fibers, as well as tough pathological fibrils known as amyloids. The aim is to investigate how the size and geometry of beta-structures in protein materials, for instance in spider silk, influence their mechanical properties such as strength and toughness. The work focuses on nano-scale behavior, but also discusses the impact of variations in the nanostructure on macro-scale behavior for these materials. The results presented in this thesis are based entirely on computational and theoretical approaches, and all findings are extensively compared with experimental data available to provide validation and support to the hypotheses presented in this work.

1.1 Physical concepts of failure at different length scales

Failure and deformation of engineering materials has been studied extensively and has changed our world by enabling the design of complex structures and advanced devices. Compared with the current understanding of the behavior of engineered materials, the fundamental physics of many biological phenomena continue to pose substantial challenges with respect to model building, experimental studies, and simulation. Because of our lacking ability to engineer biological materials, we remain hindered in our capacity to mass produce and utilize these materials for daily life applications, through consumer products, medical devices and large-scale systems in aerospace, defense and building technologies. The hierarchical bottom-up design approach in biology, from the level of genes (DNA), to proteins, to tissues, organs and organisms, originates at the molecular scale and requires a bottom-up description from a fundamental perspective. For this reason, approaches rooted in physics that consider the structure-process-property paradigm of materials science are a powerful means to investigate the properties of biological materials.

This thesis focuses on discussing the origin and mechanisms of materials failure, in the context of biological materials consisting of protein building blocks. The starting point for discussing failure in materials is coming up with a rigorous definition for failure. Simply put, failure occurs when an engineered or natural component suddenly loses its capacity to provide the service it was originally designed for, rendering it either impossible or risky to use. The key factor here is that this loss is often sudden yet significant, and that it occurs during the expected lifetime of the component. With regards to this simple explanation, failure in structural materials and structures occurs when the load bearing capacity of the designed system is significantly reduced or completely lost due to a sudden, generally unforeseen development. In the case of natural or biological systems, the definition remains the same, and is characterized by a sudden loss of function. This could be or instance the sudden rupture and slipping of the tectonic plates in an earthquake, which affects the ground's ability to provide stable foundation for the built environment. An excerpt from Darwin's The Voyage of the Beagle, describing an earthquake he experienced in Chile, illustrates how we perceive the failure of the earth's crust:

> "A bad earthquake at once destroys our oldest associations: the earth, the very emblem of solidity, has moved beneath our feet like a thin crust over a fluid; – one second of time has created in the mind a strange idea of insecurity, which hours of reflection would not have produced".

In a simple view of failure, there are typically two aspects to the problem, the designed material system and service conditions (for example, mechanical loads). Materials deform when they are subjected to loads, this may or may not be observable by the naked eye but is definitely observable in the microscopic world, as the molecular bonds stretch, rotate and shear, which provides the basis for a material's ability to change its shape. When the loads exceed a certain limit, bonds begin to rupture, initiating the atomistic mechanism for failure. Depending on the properties of interatomic bonds and the structure of the material at the nanoscale, failure will occur through a variety of atomistic mechanisms, leading to for instance brittle or

ductile failure, or very slow failure as observed in creep and fatigue. Once the governing unit processes such as cracks, dislocations, diffusional mass transport, molecular unwinding or sliding propagate through the material, they become observable at the macro-scale and lead to failure of a larger component in the system, for instance a beam in the case of a building collapse, bone in case of an injury, or the breakdown of cells in genetic disease.

It is quite interesting from a historical perspective to consider how the field of fracture and failure evolved since the earliest scientific works in the field, particularly in the context of size-effects. While the foundation of the field is attributed to the work of Griffith [95] and Irwin [114] in developing analytical methods for studying fracture of solids, many other historical notables have shown interest in the field, such as Leonardo da Vinci, who studied scaling of the failure strength of iron wires as a function of their length and flaw presence [252]. Although his study was not definitive due to the making and quality of the wires at that time, he was way ahead of his time in his insight to hypothesize an inverse proportionality of length and strength, such that shorter wires are stronger for a given thickness [10]. Galileo Galilei also studied the strength of wires as a function of thickness, and applied the same concept to testing of marble columns to conclude that the strength depends on the cross-sectional area of the column yet not on the length, thereby providing the intellectual basis for the concept of stress, defined as force per unit area. Mariotte, a court engineer at the time of Louis XIV of France developed the concept of failure strain to describe fracture strength of pressurized vessels, and also realized that larger structures are likely to fail more easily due to the increased probability of having a weakened zone. Some of these ideas were extended after the Industrial Revolution, but no significant scientific development was achieved until Griffith proposed that the physical basis for strength limit of materials is governed by flaws in the materials, such as voids and cracks and other structural imperfections [95].

Following this breakthrough, the 20th century marked the rapid development of the field of fracture mechanics, where the analytical treatment of how glass, ceramics, metals, polymers, thin films and most recently, biological materials and tissues was

developed. The most recent expansion of the concepts of fracture models towards biological materials and biological systems still bases on the fundamental concept that flaws in the material ultimately control their overall strength; and the question of how biological systems are capable of tolerating and healing such flaws has received particular interest from the physics community [86]. A failure of a biological organism to function is often related to a catastrophic response of a system to existing or newly emerging flaws, such as genetic mutations, protein misfolding or the production of foreign material in tissues.

As pointed out before, identifying properties of materials is only half of the task; predicting service conditions is an equally demanding undertaking. Many of the colossal failures in engineering practice or in medicine are rooted in extreme loading conditions or a combination of factors (where each of which alone would not be catastrophic) that were not anticipated in the design process or under typical evolutionary constraints. Examples of such failures are many, and they have shaped our understanding of materials design for increasingly safer practices and have driven our scientific curiosity to elucidate the physical principles of life. The wind induced collapse of the Tacoma Narrows Bridge, or massive seismic activities such as the Northridge earthquake in California provided us with clues about how dynamic nature of loading can lead to unforeseen failures in large structures. Brittle fracture of the Liberty Ships during World War II illustrated how low temperatures in cold climates can literally cause ships to snap like matchsticks. Fatigue induced failure of the Comet airplanes, and later the Aloha Airlines Boeing 737 jets illustrated the importance of corrosion and cyclic loading due to pressure changes. Failure of tissues and organs in genetic or infectious disease are other vivid examples that illustrate the great significance of failure in the context of life sciences, with severe impacts on our very human existence. The central modern day challenges involve understanding failure across a vast range length and time scales; encompassing materials that will last for years in the harsh, unearthly conditions of the far reaches of space, or on the quite contrary, within the smallest scales of human physiology as part of an effort to develop "invisible" implants that will monitor, regulate and repair biological

processes at molecular precision.

The framework of understanding failure provides us with the foundation to ask fundamental questions about the multi-scale behavior of materials under extreme loading conditions and varying external constraints. One of the long-term goals of this research field is to develop a new engineering paradigm that encompasses the seamless analysis and design of structures and materials, starting from the molecular level. The work that roots in first addressing fundamental concepts of materials and structures may lead to the development of a new set of tools that can be applied, together with advanced synthesis methods, to select, design, and produce a new class of materials, similar to the approaches used today in computer aided design of buildings, cars and machines, but now applied to engineer the fundamental molecular makeup of materials.

Now that we have established a basic layman's definition of failure, the next step is to come up with a rigorous physical explanation for how materials break. The key challenge here is that clearly, not all materials are the same; glass breaks differently than a metallic paper clip, and that is different than how a muscle tear in an injury takes place. A technical definition of materials failure requires understanding different failure modes, which may be activated under a variety of different boundary conditions, and most importantly, by the multi-scale makeup of the material that controls the most fundamental unit mechanisms of failure. For all these phenomena, a consideration of physical processes at multiple time- and length-scales is essential in order to develop rigorous models of failure.

The most fundamental source of the difference in materials behavior lies at the atomistic scale, essentially controlled by the atomic interactions. Typically, materials feature different types of chemical bonds, which lead to significantly variant nanostructures that influence macroscale properties. In the case of glass, we observe that fracture occurs suddenly and propagates through the specimen at extremely high speeds (close to the order of sound speeds on the order of several km/sec). However, it is extremely tough to break a metallic paper clip by trying to pull it apart, and certainly, the same type of rapid fracture as observed in glass is not found. Yet, if the

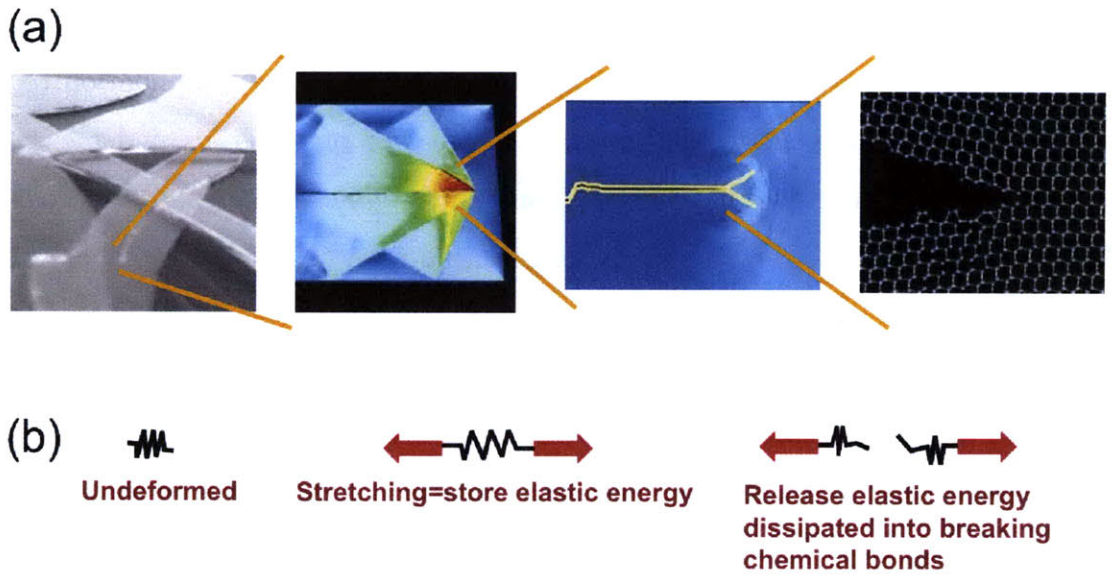

Figure 1-1: Multi-scale mechanisms of failure. (a) Multiscale view of failure of glass, from macro to nano. (b) Fracture can be envisioned as dissipation of elastic (reversible) energy. This basic view of fracture holds for a very broad range of failure phenomena, from failure of the Earth during earthquakes, failure of engineering materials, to failure of proteins. Figure adapted from Ref. [28, 41].

material microstructure is altered by for instance bending a paper clip it repetitively, it can eventually be broken with less effort. Muscle fibers on the other hand are extremely efficient in carrying loads repeatedly, but stretching them beyond their limits may lead to sudden tearing of fibers, resulting in injury. Mechanical deformation of biological tissues (e.g. blood vessels) is a natural cue that initiates the formation of this very tissue through a process called angiogenesis (growth of new blood vessels) [256]; however, changes in the material structure due to the buildup of calcium deposits and a heightened blood pressure might lead to catastrophic failure, causing heart attack and stroke. So what leads to these rather distinct material phenomena, and how can we formulate a fundamental physical model to predict the onset of materials failure?

At a fundamental level, fracture of a material due to mechanical deformation can be understood as dissipation of elastic energy into breaking of chemical bonds and heat. This concept can be exemplified by envisioning an elastic material such as a rubber band; by stretching it, elastic energy is stored inside the material. At

the moment of fracture, this elastic energy is dissipated, where most of the energy goes into breaking or tearing of molecules and atomic bonds and into heating up the sample. Whereas the storing of elastic energy is a process associated with the length scales of a macroscopic specimen, the tearing of molecular bonds typically happens at molecular and submolecular levels. This intimate connection of small and large is a universal hallmark of fracture, and the development of appropriate models provides the basis for exciting intellectual challenges and opportunities. Figure 1-1 illustrates the basic process of fracture, including a schematic multi-scale view of failure of glass (for which crack extension via repeated breaking of interatomic bonds is a unit mechanism of fracture), as well as the mechanism of dissipation of energy during the basic unit event of fracture.

1.2 Chemistry and hierarchical structure of protein materials

Proteins are the essential building blocks of living matter on Earth. Protein synthesis occurs in the cell based on segments of a DNA code which defines the primary structure of a protein given as a sequence of amino acids. A schematic describing this process is shown in Figure 1-2. Amino acids typically have a polar backbone and sidechains that can be charged, polar or hydrophobic. The linear primary structure of a protein occurs through covalent peptide bonds that link amino acids. Through weak interactions such as hydrophobic effects, van der Waals interactions and hydrogen bonding, proteins in the physiology create folds and filamentous structures that carry out physiological functions. Tightly folded structures called globular proteins have a wide range of functions, for instance as enzymes that control reactions or receptors that control cell signaling and communication. Distinct from globular proteins, fibrous proteins are large assemblies of polypeptide chains play mostly structural and mechanical functions [120]. Many biological structural materials, including cell cytoskeleton, tendon, hair, nail, hoof and spider silk employ largely fibrous proteins that

have unique thermal and mechanical properties. Structure and mechanics of these materials are largely controlled by the weak interactions at nano-scale, in particular, hydrogen bonds.

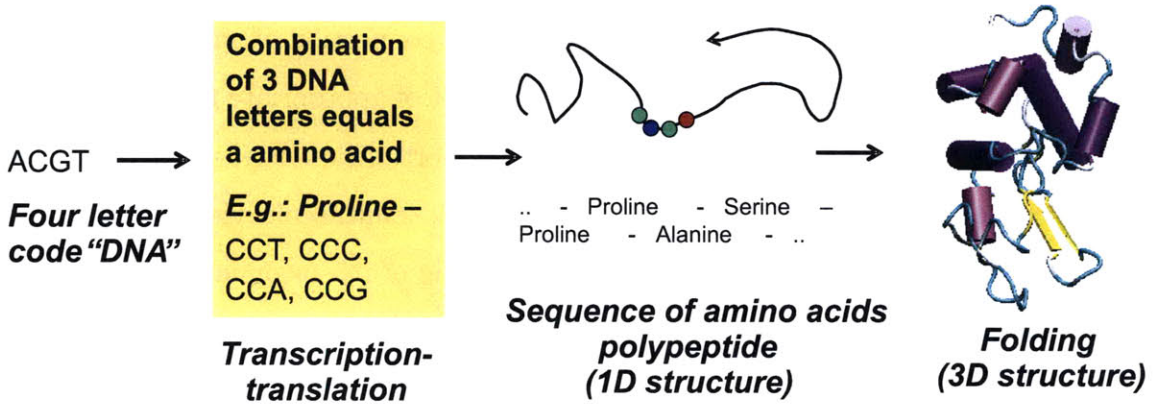

Figure 1-2: Figure illustrates how proteins are synthesized in biological systems based on sequence information encoded in the DNA. A series of letter codes (nucleotides) is translated into the polypeptide sequence, which folds into complex three-dimensional geometries. Three DNA letters correspond to a specific amino acid building block (further information, see e.g. Ref. [4]). Figure reprinted from Ref. [42].

The behavior of materials, in particular their mechanical properties, are intimately linked to the atomic microstructure of the material. Whereas crystalline materials show mechanisms such as dislocation spreading or crack extension [22, 107, 57], biological materials feature molecular unfolding or sliding, with a particular significance of rupture of chemical bonds such as hydrogen bonds, covalent cross-links or intermolecular entanglement. Other mechanisms operate at larger length scales, where the interaction of extracellular materials with cells and of cells with one another, different tissue types and the influence of tissue remodeling become more evident. The dominating mechanism emerges as a result of geometrical parameters, the chemical nature of the molecular interactions, as well as the structural arrangement of the protein elementary building blocks, across many hierarchical scales, from nano to macro. Table 1.1 provides examples of the strength of key chemical bonds found in biological systems.

Overall, a major trait of virtually all biological materials is the occurrence of hierarchies (Figure 1-3) and, at the molecular scale, the abundance of weak interactions.

Bond Type	Bond Strength (kcal/mol)
C ≡ C (Covalent)	195
C = C (Covalent)	146
C – C (Covalent)	83
S – S (Covalent)	51
$-COO^- \cdots {}^+H_3N-$ (Ionic)	21
$H_2O \cdots H_2O$ (H-bond)	5
$-CH_3 \cdots H_3C-$ (London dispersion forces)	0.07

Table 1.1: Strength of chemical bonds in biology (Data summarized from Ref. [152, 246]).

The presence of hierarchies in biological materials may be vital to take advantage of molecular and sub-molecular features, often characterized by weak interactions, and multiply their properties so that they become visible at larger scales, in order to provide a link between structural organization and function [82]. Utilization of weak interactions to produce strong materials at moderate temperatures, and thus with limited energy use, is very appealing for sustainable engineering practice, however, the means to achieve this has so far remained elusive. Another important distinction between traditional and biological materials is the geometrical occurrence of defects. While defects are often distributed randomly over the volume in crystalline materials, biological materials consist of an ordered arrangement of structure that reaches down to the nano-scale. In many biological materials, defects are controlled with atomistic or molecular precision, and their arrangement may play a major role in the material behavior observed at larger scales. These features have been observed in bone, nacre, collagenous tissue or cellular protein networks, among others [84, 30, 29].

The mechanical properties of biological materials have wide ranging implications for biology. In cells for instance, mechanical sensation is used to transmit signals from the environment to the cell nucleus or to control tissue formation and regeneration [71]. The structural integrity and shape of cells is controlled by the cell's cytoskeleton, which resembles an interplay of complex protein structures and signaling cascades arranged in a hierarchical fashion [4, 109]. Bone and collagen, providing structure to our body, or spider silk, used for prey procurement, are examples of materials that have incredible elasticity, strength and robustness unmatched by many synthetic

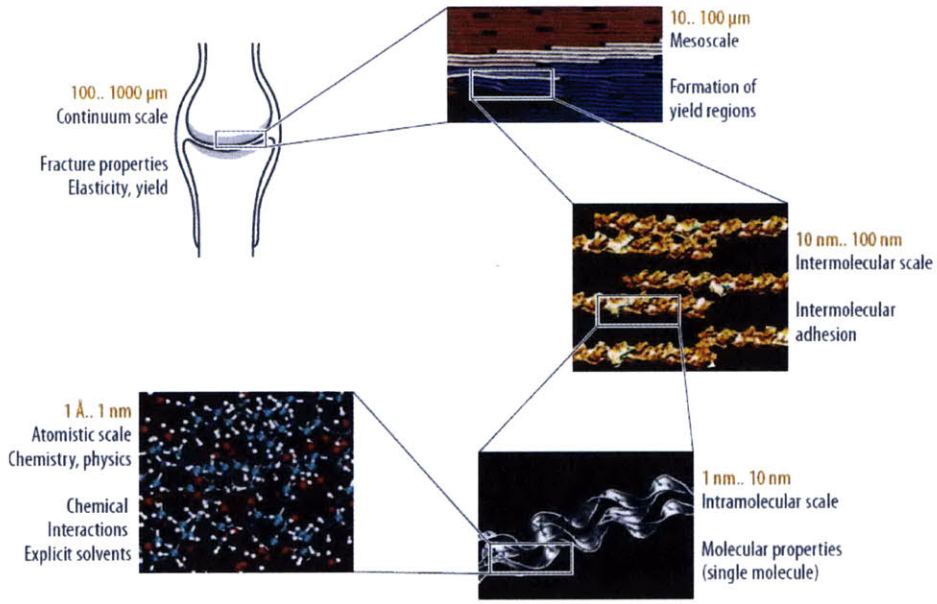

Figure 1-3: Overview of different material scales, from nano to macro, here exemplified for collagenous tissue [30, 39, 42, 84, 89]. Biological protein materials such as collagen, skin, bone, spider silk or cytoskeletal networks in cells feature complex, hierarchical structures. The macroscopic mechanical material behavior is controlled by the interplay of properties throughout various scales. In order to understand deformation and fracture mechanisms, it is crucial to elucidate atomistic and molecular mechanisms at each scale and how they interact. Figure reprinted from Ref. [42].

materials [187, 138, 139, 36, 39, 229, 67, 83]. Molecular precision in the bottom-up hierarchical synthesis is believed to be the source of the exceptional properties of these materials.

Promising strategies for probing biological materials at the nano-scale can be currently developed by combining experiment and simulation concurrently. State-of-the-art of materials modeling at multiple length scales has reached an unprecedented level of sophistication both in computation and instrumentation, as evident from the arsenal of techniques shown in Figure 1-4. Development and utilization of Atomic Force Microscope (AFM)[186], optical tweezers [62, 229, 46, 16] or nanoindentation [8, 70, 231] to analyze biological materials, in conjunction with advanced quantum mechanical (QM), molecular dynamics (MD) and finite element method (FEM) methods will lead to the development of superior materials through an improved understanding of the influence of nano-scale on the macro-scale. It is therefore reasonable to say that modeling and simulation has evolved from explaining experimental observations

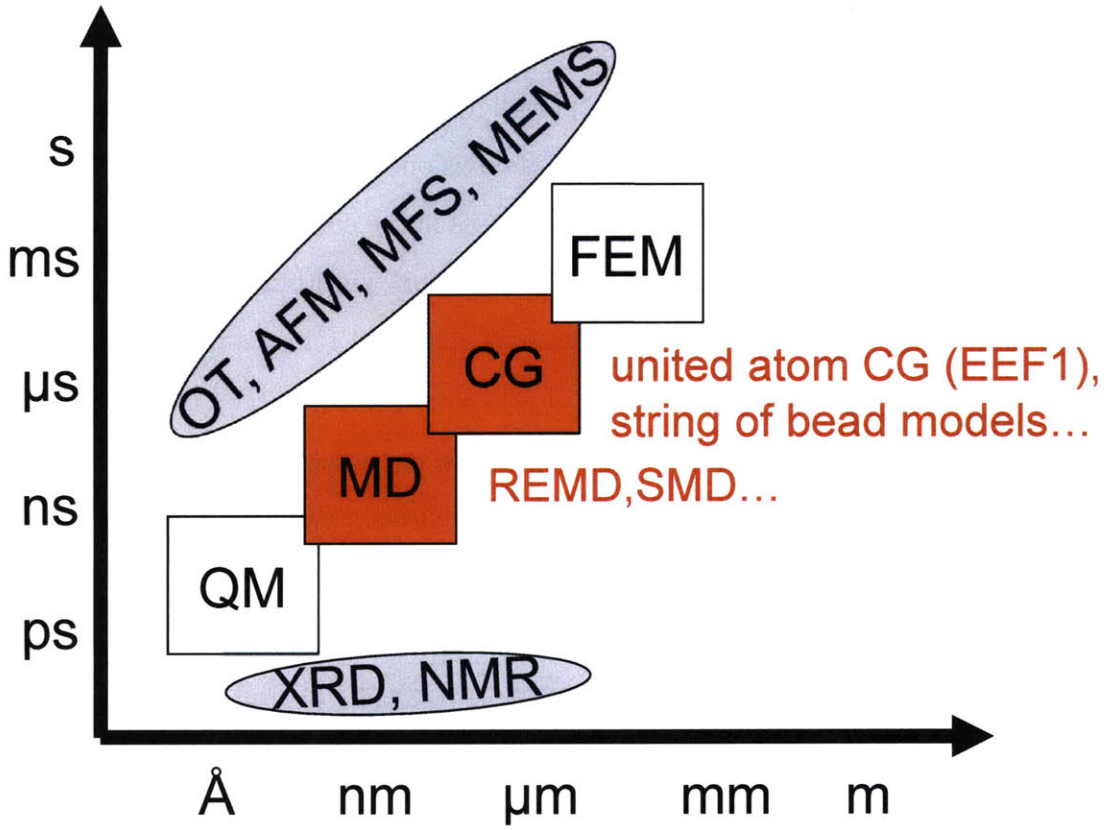

Figure 1-4: Coupling of different computational and experimental tools can be used to traverse throughout a wide range of length- and time-scales. Such methods enable to provide a fundamental insight into physical phenomena, across various time- and length-scales. Handshaking between different methods enables one to transport information from one scale to another, either hierarchically or concurrently. Eventually, results of atomistic, molecular or mesoscale simulation may feed into constitutive equations or continuum models. Experimental techniques such as atomic force microscope (AFM), molecular force spectroscopy (MFS), nanoindentation or optical tweezers (OT) now overlap into atomistic and molecular approaches, enabling the direct comparison of experiment and simulation. High-resolution structure prediction methods such as nuclear magnetic resonance (NMR) and X-ray diffraction (XRD) provide atomistic models which can be investigated and refined further with molecular dynamics (MD), coarse-grained (CG) models and the finite element method (FEM). Methods used in this thesis (MD,CG) are highlighted.

into predictive tools that complement experimental analyses (see Figure 1-4).

The characterization of material properties for biological protein materials may play a crucial role in better understanding diseases. Injuries and genetic diseases are often caused by structural changes in protein materials, resulting in failure of the material's intended function. In this sense, material properties play an essential role in biological systems, and the current paradigm of focusing on biochemistry alone as the cause of diseases is insufficient. It is envisioned that the long-term potential impact of this work can be used to predict diseases in the context of diagnostic tools by measuring material properties rather than focusing on symptomatic chemical readings alone. Such approaches have been successfully explored for example for cancer and malaria [60, 230]. A recent development of potentially great impact is the application of nano-scale characterization and simulation tools to investigate amyloids, which are tough protein beta-structures associated with neurodegenerative diseases [58, 112, 120, 124, 140, 176, 177, 210, 250]. Insight gained from these studies can aid the development of novel detection methods and treatment strategies for some of the related diseases.

1.3 Outline

The scope of this thesis is the investigation of the size-dependent mechanical properties of beta-structures in protein materials, focusing on applications in silks, amyloids and beta-solenoid structures observed in virulence factors. All of these protein materials feature exceptional mechanical properties that have been linked to the existence of beta-structures in the material, however, no systematic study has focused on assessing geometry, size and rate dependence of the beta-topologies in a variety of systems. Addressing this issue by laying out a theoretical and computational framework for studying beta-structures is the fundamental goal of this thesis.

The content is arranged as follows: Chapter 2 outlines briefly methods used for studying biological materials, focusing mostly on the approaches used in this thesis; Chapter 3 summarizes the development of a generic fracture strength model for

protein materials, with the scope of assessing the size-dependent rupture strength of hydrogen bond assemblies; Chapter 4 reports one of the first studies on the compressive strength of proteins, particularly focusing on the mechanics of beta-solenoid protein nanotubes such as the cell-puncture needle of bacteriophage T4 virus; Chapters 5 summarizes findings on the size-dependent strength and toughness of beta-sheet nanocrystals in silks and other protein materials; Chapter 6 reviews modeling efforts in understanding the molecular structure and mechanics of spider dragline silk, the toughest material known. Chapter 7 provides a outlook for anticipated applications of protein based materials, lays out recommendations for areas that require future research, and concludes with a discussion of the importance of size-effects in protein materials.

Chapter 2

Methodology

In this Chapter, a brief review of the methods used in analyzing the nanomechanics of biological molecules is provided. Along with theoretical methods, atomistic and molecular simulation approaches are reviewed, focusing on molecular dynamics simulation, a selection of force fields, coarse-graining methods, treatment of solvent and enhanced sampling techniques. Experimental techniques that complement and validate this work are briefly summarized for the reader's convenience.

2.1 Atomistic and molecular modeling

This section overviews atomistic and molecular modeling approaches used in this thesis. Particular emphasis is given to molecular dynamics method and its variants. A discussion of force fields, solvation models as well as enhanced sampling techniques relevant to biomolecular simulations is provided.

2.1.1 Classical molecular dynamics formulation

Molecular dynamics (MD) is a tool for elucidating motion of atoms and molecules at the nano-scale. The MD method is capable of describing atomistic mechanisms that control many physical phenomena, in particular those related to the mechanics of materials. An interesting example is fracture, where simulations can reveal, both

quantitatively and visually, the atomistic mechanisms that control deformation and rupture of chemical bonds at the onset of failure at the nano-scale [243, 199, 42, 31, 40, 38, 34, 35, 28].

The basic idea behind atomistic simulation using MD is to compute the dynamical trajectory of each atom in the material, by considering their atomic interaction potentials, by solving each atom's equation of motion according to $F = ma$, leading to positions $r_i(t)$, velocities $v_i(t)$ and accelerations $a_i(t)$. The numerical integration of Newton's law by considering proper interatomic potentials to obtain interatomic forces enables one to simulate a large ensemble of atoms that represents a larger material volume, albeit typically limited to several nanoseconds of time scale. The basic concept of molecular dynamics is shown in Figure 2-1(a,b).

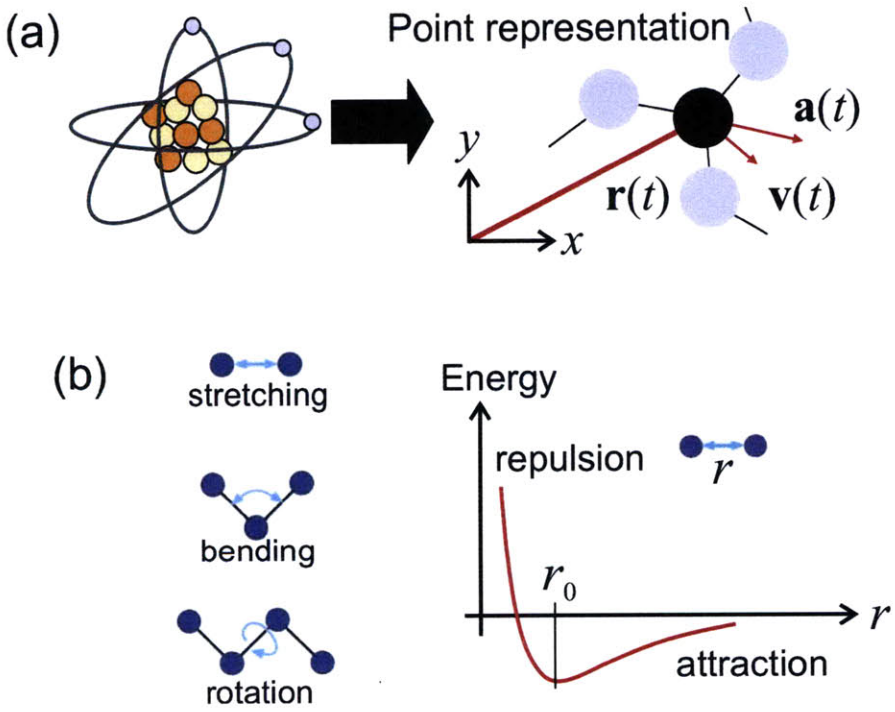

Figure 2-1: Model of the individual energy contributions due to bond stretching, bond bending, bond rotation as well as electrostatic and vdW interactions. The combination of these terms constitutes the entire energy landscape of interatomic and intermolecular interactions. Figure reprinted from Ref. [41].

Classical molecular dynamics generates the trajectories of a large number of particles, interacting with a specific interatomic potential. Thereby, the complex 3D structure of an atom (composed of electrons and a core of neutrons and protons) is

approximated by a point particle, as shown in Figure 2-1(b)). Molecular dynamics is an alternative approach to methods like Monte-Carlo, with the distinction that MD provides full dynamical information and deterministic trajectories. It should be emphasized that Monte-Carlo schemes provide certain advantages as well; however, this point will not be discussed further here as all simulation studies presented here are carried out with a MD approach. The total energy of the system is written as the sum of kinetic energy (K) and potential energy (U),

$$E = K + U \tag{2.1}$$

where the kinetic energy is

$$K = \frac{1}{2} m \sum_{j=1}^{N} v_j^2, \tag{2.2}$$

and the potential energy is a function of the atomic coordinates r_j,

$$U = U(r_j), \tag{2.3}$$

with a properly defined potential energy surface $U(r_j)$. The numerical problem to be solved is a system of coupled second order nonlinear differential equations:

$$m \frac{d^2 r_j}{dt^2} = -\nabla_{r_j} U(r_j) \qquad j = 1..N, \tag{2.4}$$

which can only be solved numerically for more than two particles, $N > 2$. Typically, MD is based on updating schemes that yield new positions from the old positions, velocities and the current accelerations of particles. In the commonly used Verlet scheme, this can be mathematically formulated as

$$r_i(t_0 + \Delta t) = -r_i(t_0 - \Delta t) + 2r_i(t_0)\Delta t + a_i(t_0)\,(\Delta t)^2 + O((\Delta t)^4) \tag{2.5}$$

The forces and accelerations are related by $a_i = f_i/m$. The forces are obtained from the potential energy surface – sometimes also called force field – as

$$F = m\frac{d^2 r_j}{dt^2} = -\nabla_{r_j} U(r_j) \qquad j = 1..N. \tag{2.6}$$

This technique can also be used for not only single atoms but also groups of atoms as in the case of coarse-grained meso-scale approaches. The basis of simulations is simply Newton's law and a definition of how atoms interact with each other. The availability of interatomic potentials for a specific material (based on the characteristic type of chemical bonding) is often a limiting factor for the applicability of the MD method. Provided interatomic potentials are available, MD is capable of directly simulating a variety of very complex physical phenomena, including self-assembly of polymers, diffusion, fracture, as well as protein folding, unfolding and aggregation.

Unlike many continuum mechanics approaches, atomistic techniques require no *a priori* assumption on the defect dynamics or elastic properties, such as isotropy. Once the atomic interactions are chosen, the complete material behavior is determined. Choosing appropriate models for interatomic interactions provides a rather challenging and crucial step that remains subject of a very active discussion in the scientific community. A variety of different interatomic potentials are used in the studies of biological materials at different scales, and different types of protein structures require the use of different atomistic models. A drawback of atomistic simulations is the difficulty of analyzing results and the large computational resources necessary to perform the simulations. Due to computational limitations, MD simulations are restricted with respect to the time scales that can be reached, limiting overall time spans in such studies to tens of nanoseconds, or in very long simulation studies to fractions of microseconds. Therefore, many MD simulations on stretching proteins and crystalline materials have been carried out at large deformation rates, exceeding several m/sec.

Recent advances in computational power now enables the simulation of billions of particles in MD simulations, reaching dimensions on the order of micrometers. Figure 2-2 depicts the historical development of computational power.

The goal of Sections 2.1 and 2.2 is to provide a brief overview of popular in-

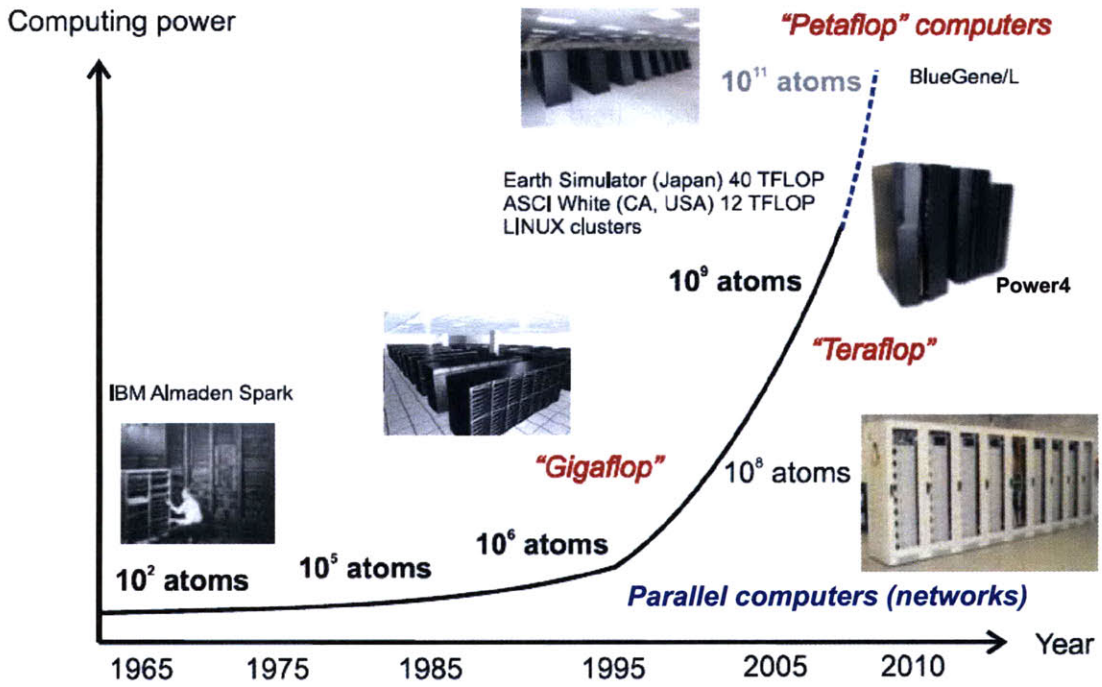

Figure 2-2: Development of computing power over the past decades. The development illustrates the emergence of petaflop computers in the next few years. The plot also summarizes the number of atoms that can be treated with these computing systems; these numbers are developed for simple interatomic potentials with short cutoffs. For CHARMM or ReaxFF potentials, the number of atoms is significantly smaller. Figure reprinted from "Atomistic Modeling of Materials Failure", Copyright 2008, with permission from Springer [28].

teratomic force fields and modeling approaches suitable for simulating the behavior of protein structures. For additional information, the reader may refer to extensive review articles, in particular regarding force field models [249, 159, 206, 63].

2.1.2 Force fields

All-atom force fields are used in molecular dynamics simulations of biological materials at the nanoscale. Among all force field approaches, all-atom potentials are considered to be highly accurate and computationally efficient way of studying dynamics of macromolecules. A wide range of force fields and simulation programs are currently available, most notably AMBER [179], CHARMM force fields and programs [25, 160], OPLS force field [118], GROMOS/GROMACS packages are commonly used in

all-atom molecular dynamics [242], NAMD program [169, 182] which can carry out computations using CHARMM and other force fields. For the sake of brevity, the main aspects of the CHARMM force field and its implementation in NAMD will be discussed here; the basic concepts of the MD technique and force field formulations are common to all packages used in the field (for a general review, see for instance [159, 183].

The CHARMM force field is widely used in the protein and biophysics community, and provides a reasonable description of the behavior of proteins. The parameters in force fields are often determined from more accurate, quantum chemical simulation models by using the concept of force field training [92]. Parameters for the CHARMM force field have been meticulously optimized and revised over the years taking into consideration a wide variety of input including *ab initio* results, experimental crystal structures and geometries, as well as vibrational spectra [160].

The potential includes bonding and non-bonding (interaction) terms to describe short and long-range forces between particles. In the CHARMM model, the mathematical formulation for the empirical energy function that contains terms for both internal and external interactions has the form:

$$\begin{aligned} U(\vec{R}) = & \textstyle\sum_{\text{bonds}} K_b(b-b_0)^2 + \sum_{\text{UB}} K_{UB}(S-S_0)^2 + \sum_{\text{angle}} K_\theta(\theta-\theta_0)^2 + \\ & \textstyle\sum_{\text{dihedrals}} \mathrm{K}_\chi(1+\cos(n\chi-\delta)) + \sum_{\text{impropers}} \mathrm{K}_{\text{imp}}(\phi-\phi_0)^2 + \\ & \textstyle\sum_{\text{nonbond}} \varepsilon \left[\left(\frac{R_{\min(i,j)}}{r_{ij}} \right)^{12} - \left(\frac{R_{\min(i,j)}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\varepsilon_1 r_{ij}} \end{aligned} \tag{2.7}$$

where the constants after summation are bond, Urey-Bradley, angle, dihedral angle, and improper dihedral angle force constants, respectively; and are terms in parantheses are the bond length, Urey-Bradley 1,3-distance, bond angle, dihedral angle, and improper torsion angle, respectively, with the subscript zero representing the equilibrium positions for the individual terms. Figure 2-1 shows a schematic of the individual energy contributions listed in eq. (2.7).

The Coulomb and Lennard-Jones 6-12 terms constitute the external or nonbonded

interactions; characterized by the Lennard-Jones well depth and the distance at the Lennard-Jones minimum, the partial atomic charge, the effective dielectric constant, and the distance between atoms i and j. In CHARMM, no additional terms are used for hydrogen bonds, the charge and LJ components were verified to be adequate for describing protein, solvent and interface hydrogen bonding. In all-atom force fields, water molecules are generally also treated explicitly. Parameters of the force field generally are specified considering a specific water model (e.g. TIP3P dimer model for CHARMM) [160, 159].

The CHARMM force field belongs to a class of models with similar descriptions of the interatomic forces; other models include the DREIDING force field [164], the UFF force field [190] or the AMBER model [249, 179]. In CHARMM and other classical force fields, bonded terms are modeled with harmonic springs or its variations, and therefore can't be modified or broken once defined by the connectivity input obtain from the topology of the molecule. Further, the atom charges are fixed and cannot change during simulation. These simplifications improve the simulation speed drastically and are not a major issue for most simulations studying conformation of proteins under ambient physiological conditions. On the other hand, simulations in extreme conditions such as mechanical perturbations (e.g. protein unfolding studies) or harsh chemical environments require reactive force fields that can take into account changes in fixed charges of the molecules, formation and breaking of bonds, as well as variations in bond order.

Several flavors of reactive potentials have been proposed in recent years [21, 223, 69]. Reactive potentials can overcome the limitations of empirical force fields and enable large-scale simulations of thousands of atoms with quantum mechanics accuracy. The reactive potentials, originally only developed for hydrocarbons, have been extended recently to cover a wide range of materials, including metals, semiconductors and organic chemistry in biological systems such as proteins [33, 35, 50, 51, 52, 69, 222, 221].

While the reactive force fields show promise for capturing intricate chemical details of materials at extreme conditions [51], they are also more expensive computation-

ally and require time to perfect for each application. The parameters developed for some the most common elements such as C,H,O,N,S have been tested extensively in many different benchmark studies, however, applications to material systems require training a new set of parameters through extensive quantum simulations. As with other methods bridging the quantum to nano-scale systems, the ideal solution depends critically on the application. Typically, after careful consideration of the size of the system and phenomena of interest, most QM/MM strategies developed are a compromise between fully quantum and fully classical treatment, providing adequate sampling time and chemical information.

2.1.3 Coarse-grained modeling approaches for protein structures

Although a very accurate description of macromolecules, all-atom modeling approaches have historically been prohibitively extensive when large systems and long simulation times must be considered. This lead to the development of coarse-grained models [237] (Figure 2-3), which provide simplified representations of macromolecules employing less degrees of freedom and simple bonded and non-bonded interactions that can be rapidly calculated in each time step. Coarse-grained models have so far been successfully applied to a wide range of problems including protein folding, allostery, aggregation, molecular biomechanics as well as multi-scale description of complex materials such as bone. The various approaches are briefly reviewed here.

Single bead models are perhaps the earliest approach taken for studying macromolecules. The term single-bead derives from the idea of using single beads (masses) for describing each amino acid in a protein structure. Elastic Network Model (ENM) [236], Gaussian Network Model (GNM) [99] and Go-model [104] are well-known examples of this simplistic approach.

Simple models such as ENM and Go-like models treat each amino acid as a single bead located at the Cα position, with mass equal to the mass of the amino acid. The beads are interconnected by harmonic or nonlinear springs representing the covalently

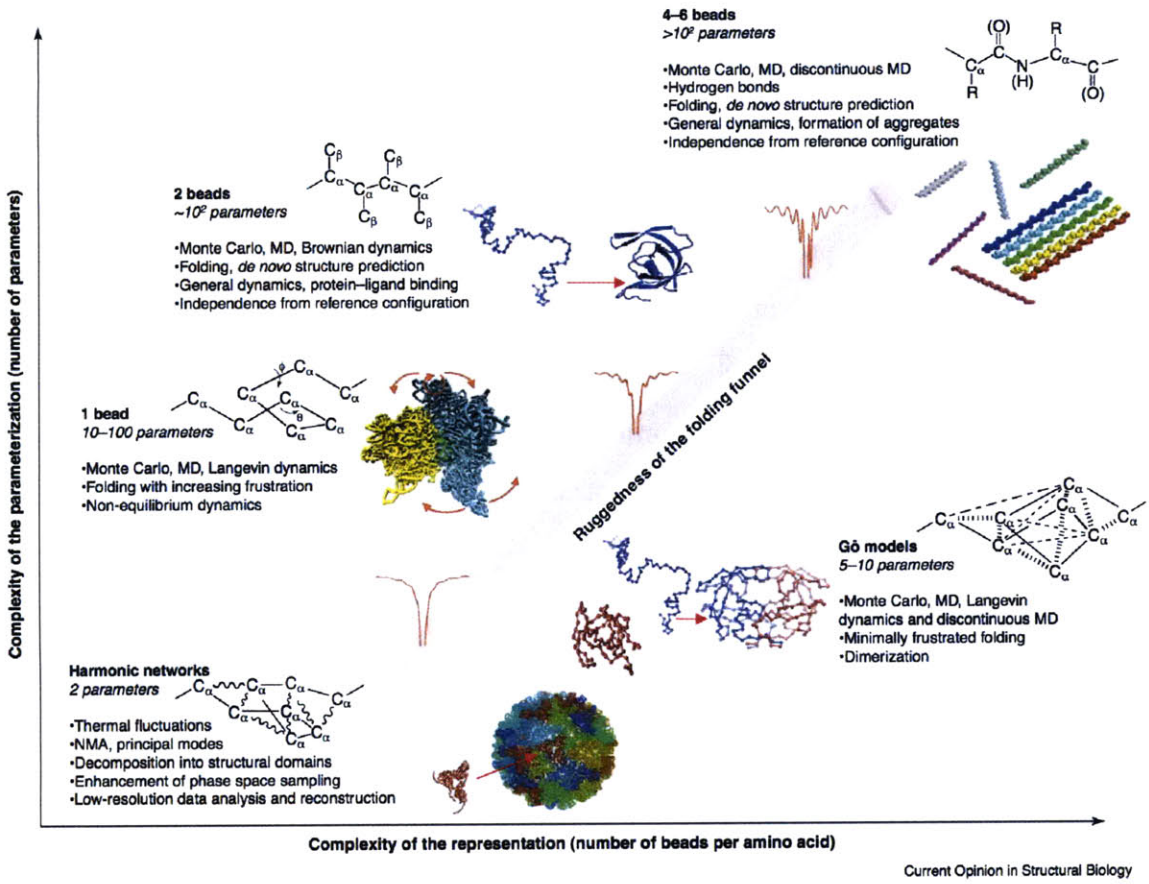

Figure 2-3: Overview over various coarse-graining techniques [237] . The figure illustrates commonly used coarse-graining applications in biomolecular simulations. A schematic representation of the model, indicative number of parameters, methods of solution, main characteristics and applications are shown. Axes indicate increasing complexity of the models in parameterization and molecular representations. Figure reprinted from "Current Opinions in Structural Biology", "Coarse-grained models for proteins", Vol. 15, pp. 144-150, Copyright 2005, with permission from Elsevier [237].

bonded protein backbone. In the Go-like models, an additional Lennard-Jones term is included in the potential to describe short-range non-bonded native interactions between atoms within a cutoff distance. Despite their simplicity, these models have been extremely successful in explaining thermal fluctuations of proteins [237] and have also been implemented to the unfolding problem to elucidate experimental results [251, 226, 65]. A more recent direction is coupling of ENM models with finite element frameworks for mechanistic studies on protein structures and assemblies [9].

Due their simplicity, single-bead models have several shortcomings. With classic ENM, only harmonic deviations from the initial configuration are possible. In

the Go-model, native interaction definitions lead to a minimally frustrated landscape which his highly biased towards the input configuration of the molecule; such models therefore cannot predict folding/unfolding intermediates and meta-stable states. Explicit treatment of protein-solvent interactions, non-native interactions and hydrogen bonds is also not possible with single-bead models. For unfolding, results are only qualitative at best, although revealing important aspects of topology dependent mechanical resistance [251, 226].

Using more than one bead per amino acid can lead to more detailed descriptions of macromolecules. In the simplest case, addition of another bead can be used to describe specific side-chain interactions [6]. Four to six bead models capture even higher amount of detail by explicit or united atom description for backbone carbon atoms, sidechains, carboxyl and amino groups of amino acids. A great example of this approach is the coarse-grained models developed for studying folding and aggregation in proteins using discontinuous molecular dynamics [171, 170].

Although multi-bead models have superior qualities compared to single-bead descriptions, dozens of additional energetic terms involving pseudobonds and other means to avoid complex dihedral/improper potentials that stabilize the conformation of the polypeptide chain have to be introduced for generic models (see Figure 2-3). Even with the introduction of these terms, some of which are physically not as intuitive, the models offer limited applicability, as specific side chain interactions are only valid for simple residues such as glycine and alanine. More complex yet fast to compute potentials that intrinsically take into account sequence specificity are extremely challenging to develop, making readily available all-atom descriptions and simulation packages more favorable for most applications.

More recently, coarser-level modeling approaches have been applied to model biomolecular systems at larger time and length scales. These models typically employ superatom descriptions that treat clusters of amino acids as beads. The elasticity of the polypeptide chain is captured by simple harmonic bond and terms. This viewpoint of hierarchical multi-scale modeling can be computationally very efficient, and capture shape dependent mechanical phenomena in large biomolecular structures [5]

and can also be applied to collagen fibrils in connective tissue as well as mineralized composites such as nascent bone [27]. An alternative approach to hierarchical multi-scale approaches is the use of concurrent molecular dynamics simulation methods, where a different level of detail can be employed in different regions of the system studied and computations are carried out simultaneously, with information exchange between the coarser and finer descriptions[153, 32].

2.1.4 Treatment of solvent: Implicit vs. Explicit

Historically, there has been two main directions of thought in modeling solvent in proteins and polymers. Most biomolecular simulations have either an explicit or implicit treatment of solvent around the molecule, as performing simulations in vacuum generally leads to inaccurate results, since the molecular friction, hydrophobic effects and dielectric screening properties of the solvent are not captured. In the explicit solvent approach, each water molecule is treated individually in the simulation. The molecular forces and solvent dynamics are computed on each of these molecules as they are done on the solute. As a result of including most of the details of water in the simulation, explicit solvent approaches are considered to be the most truthful to experimental data. In the implicit solvent approach, the overall effect of water with respect to the conformation of the molecule is taken into consideration. Instead of extensive dynamics calculations on the solute, these models carry out theoretical calculations on the behavior of solvent around the solute as a continuum, and compute forces and energies accordingly. Since water and pH = 7 is the most common physiological solvent conditions, the focus here is on the models developed for this general purpose.

Several molecular models have been developed for explicit water. Most approaches take the molecule as rigid, and consider only non-bonded interactions such as electrostatics (Coulomb's law), dispersion and repulsion forces, typically described with a 12-6 Lennard jones potential (see eq. (2.7)). Three-site models such as SPC and TIP3P, with interactions on the 2 H and O atoms of the molecule, are the most commonly used approaches. Models such as TIP3P have been implemented with small

modifications for their use with empirical protein force fields such as CHARMM [160]. Higher accuracy models, based on a larger number of interactions sites (e.g. TIP4P, TIP5P) have also been proposed [118, 161]. While these are still in the testing stages, TIP4P has gained acceptance in applications with CHARMM as it provides an improvement over TIP3P and is only approximately 7% slower [15]. The extent of the implementation of these water models varies greatly based on the MD engine and protein force field used, and transferring water among different force fields requires extensive testing and careful consideration. All explicit solvent simulations in this thesis were carried out using TIP3P solvent model with CHARMM22 all-atom force field [160].

Despite the increase in accuracy, explicit treatment of water brings with itself significant challenges. Explicit treatment of water is a severe obstacle against scaling up in biological simulations, since a large water box or water sphere needs to be used to keep the protein within a periodic boundary. In many cases, majority of the computational effort goes into simulating motion of the solvent, whereas typically the solute is of more interest. This makes the explicit calculations very prohibitively expensive for large-scale simulations. Another consideration is the difficulty of the calculation of the solvation energy and free energies of folding (ΔG) from explicit solvent MD simulations, whereas this is readily available with implicit solvent methods.

This has lead to the development of implicit solvent models, which treat water as a continuum and calculate its effective influence on the solute. In certain cases, implicit solvent techniques can be as fast as 50% slower than simulations done in vacuum, which are generally hundreds of times faster than explicit solvent runs. EEF1 effective energy model [147] is one of the fastest solvent models; it uses a Gaussian based solvent exclusion model in combination with a modified CHARMM19 polar hydrogen energy function. This model used in Chapter 6 to sample amorphous structures taken by spider silk protein segments. Other methods take into account free energy of solvation by calculating the accessible surface area of the protein, which can capture more accurate solvent physics, but are computationally much more expensive. Such methods can be used in combination with the generalized Born (GB) formulation

(numerical approximation to the exact Poisson-Boltzmann equation that is solved in a continuum electrostatic model) to accurately capture the electrostatic and non-polar effects of water on folding mechanisms on proteins [49]. Viscosity of water molecules can also be implemented in simulation by using Langevin dynamics with a friction coefficient. Recent developments in analytical treatment of solvent have rendered implicit solvation a reasonable alternative to explicit treatment of water. For more information on mathematical basis and details of different solvent models, the reader may refer to the more comprehensive reviews in this field [48, 183, 202].

2.2 Sampling and force manipulation methods

Due to the short time scale of MD simulations and the challenges in observing rare events such as folding, unfolding and structural transitions, many different methods have been developed to speed up the sampling rate of MD simulations. Here we discuss methods to assess how protein structures react to external mechanical perturbations, focusing on steered molecular dynamics and constant force methods. This is followed by a brief overview of advanced sampling methods, focusing on replica exchange molecular dynamics approach implemented here and discussed in more detail in Chapter 6.

2.2.1 Steered molecular dynamics

Application of mechanical forces to biomolecules is of interest to scientists and engineers for several reasons. First, mechanical characterization of individual proteins is crucial for understanding larger-scale behavior of many fibrous materials, such as collagen, titin, spider silk or intermediate filaments in the cell. Second, rare events such as unfolding or structural transitions are often cued by mechanical forces in the cell or extra-cellular matrix (ECM), therefore single-molecule experiments and simulation allow understanding key these biophysical processes. Another key reason for force manipulation is to probe the stability of proteins, where by applying systematic forces, one can probe the energy landscape of the protein and thereby quantify the

thermodynamical stability of the system in a wide range of physiological conditions. To apply forces to the molecule that induce deformation, steered molecular dynamics (SMD) has evolved into a useful tool [154]. Steered MD is based on the concept of adding a harmonic moving restraint to the center of mass of a group of atoms. This leads to the addition of the following potential to the Hamiltonian of the system:

$$U(r_1, r_2, ..., t) = \frac{1}{2}k[vt - (\vec{R}(t) - \vec{R}_0) \cdot \vec{n}]^2. \tag{2.8}$$

where $\vec{R}(t)$ is the position of restrained atoms at time t, $\vec{R}_0$ denotes original coordinates and v and $\vec{n}$ denote pulling velocity and pulling direction respectively. The net force applied on the pulled atoms is $F(r_1, r_2, ..., t) = k(vt - \vec{R}(t) - \vec{R}_0) \cdot \vec{n}$. By monitoring the applied force (F) and the position of the atoms that are pulled over the simulation time, it is possible to obtain force-versus-displacement data that can be used to derive the mechanical properties such as bending stiffness or the Young's modulus (or other mechanical properties). SMD studies are typically carried out with a spring constant $k = 10$ kcal/mol/ Å^2. The SMD method mimics an AFM nanomechanics experiment, as illustrated in Figure 2-4. Due to the time-scale limitations of MD to several nanoseconds, there is typically a large difference in the pulling rates and transducer stiffnesses in simulation and experiment. This requires additional consideration in order to interpret MD results in light of experimental findings.

2.2.2 Quasi-static constant force method

An alternative approach to SMD is applying constant forces to a molecular system and observing the fluctuations in the system as a function of time. This method is useful when the loading conditions are complex, for instance, when the response of a molecular assembly to a combination of shear, tensile and bending loads are of interest.

While this is a favorable approach, there are certain differences of this technique from SMD. First, the method is load controlled, in other words, a loading scenario where load is increased incrementally throughout the simulation by updating forces

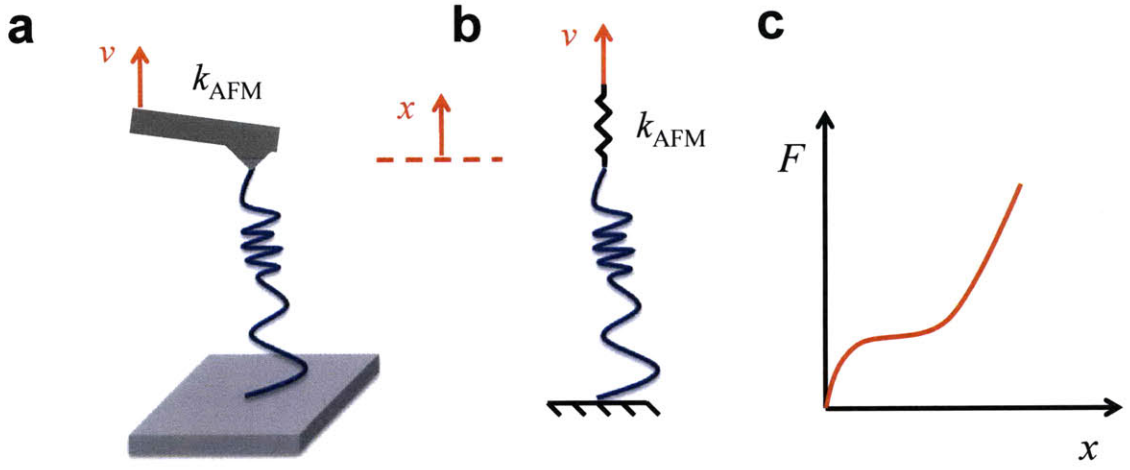

Figure 2-4: Single molecule pulling experiments. Panel (a) depicts an experimental setup based on AFM, and panel (b) depicts a steered molecular dynamics (SMD) analogue. In the SMD approach, the end of the molecule is slowly pulled. This leads to a slowly increasing force over the displacement, as schematically shown in subplot (c). Both approaches, AFM and SMD lead to force-displacement information. In addition to the *F(x)* curve, SMD provides detailed information about associated atomistic deformation mechanisms.

can lead to very large deformation on local bonds if the forces selected are relatively large. Using very small forces (picoNewtons or lower if possible) and allowing the system to equilibrate under the load in a quasi-static manner can alleviate this issue. A second issue is that very large accelerations can be observed upon rupture of bonds and unfolding in load controlled simulations of molecular assemblies where the failure force is high. This may lead to unrealistic scenarios where the rebinding of molecules or interaction with water may not be sampled due to the fast motion of the molecule. For these scenarios, a displacement controlled approach such as SMD, or more precise control of the force at the onset of rupture may be preferred. In scenarios where the influence of a specific force value is of interest, for instance in computing distance of stable transition state activated mechanically, constant force is again a very useful method.

2.2.3 Improving sampling: Replica exchange MD

As discussed earlier, time-scale limitations of MD arise from the necessity to use femtosecond time-steps required for capturing the vibrations of light atoms correctly. Since the simulation time is limited to nanoseconds, and some of the phenomena of interest happen in the order of microseconds or higher, improving the sampling of the equilibrium states of a molecular system is of great interest. This is particularly important for polymers and biological systems, where the native state of the protein requires traversing through a rough energy landscape with many local minima (Figure 2-5(a).

Many different enhanced sampling methods have been developed in the past [3, 13, 146]. The focus here is on the replica exchange molecular dynamics (REMD) method, which is used in Chapter 6 to investigate the molecular structure of spider dragline silk proteins. The REMD method combines the best aspects of Monte-Carlo (MC) simulation [3] with MD to improve sampling. The basis of replica exchange (parallel tempering) is that instead of a single simulation, M replicas are simulated, each in the canonical ensemble, and each at a different temperature, T_i. In general $T_1 < T_2 < ...T_M$, where T_1 is normally the temperature of the system of interest. Figure 2-5(b-c) illustrates this concept. The replicas do not interact energetically. If the probability of performing a swap move is equal for all conditions, exchanges between ensembles i and j are accepted with the MC type probability given as

$$A = \min(1, \exp[+(\beta_i - \beta_j)(U(\vec{r}_i^N) - U(\vec{r}_j^N)]), \tag{2.9}$$

where $U(\vec{r}_i^N)$ denotes the potential energy of the system as a function of atomic coordinates, and $\beta_i = 1/(k_B T_i)$. Exchanges are typically attempted between systems with adjacent temperatures, $j = i + 1$. In MD, one must also take into account the momenta of all the particles in the system according to temperature. According to the algorithm by Sugita and Okamoto, the velocities of the system are rescaled after the exchange as

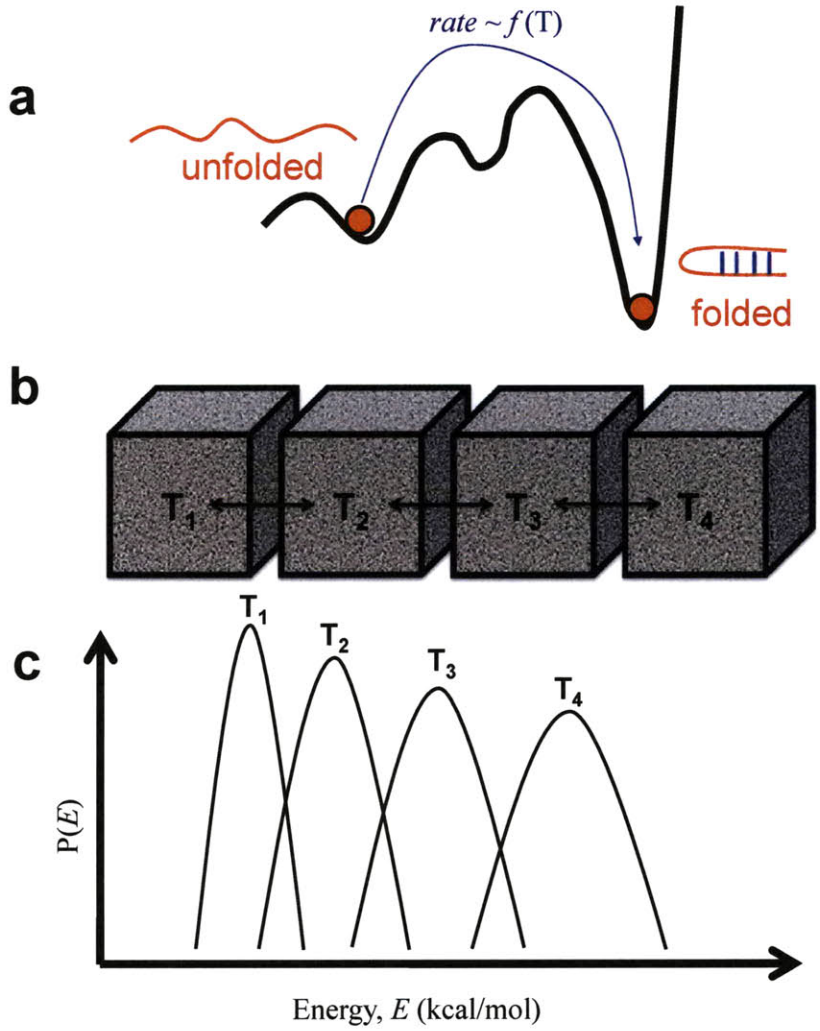

Figure 2-5: Replica exchange molecular dynamics (REMD) method. Complex systems such as proteins have a very rugged energy landscape that cannot be sampled efficiently with classical MD, as shown in panel (a). Higher temperatures increase the diffusion rate of a protein and allow overcoming larger energy barriers that cause kinetic trapping. Extensive runs at high temperature however also introduce artifactual, non-native states, which are eliminated in REMD through more detailed investigation at low temperatures. Panel (b) shows the concept of replicas, where identical systems are modeled at different temperatures (possibly starting from different initial conditions). Higher temperatures typically sample an much greater variety of conformations, leading to a broader probability distribution of energies as shown in panel (c). Favorable conformations that have a lower energy can be exchanged with the adjacent lower temperature replica.

$$\vec{p}(i)_{\text{new}} = \vec{p}(i)_{\text{old}}\sqrt{T_{new}/T_{old}} \tag{2.10}$$

to satisfy balance criterion of statistical mechanics [224].

REMD is considered to be an effective tool for investigating folding and aggregation of proteins, as it reduces the likelihood of kinetic trapping at non-native states [203]. Through a fast search of the conformation space at high temperatures and more detailed investigation at low temperatures, it allows the system to overcome energy barriers and local minima corresponding to non-native structures [78, 189, 192, 166] of proteins and allows identifying native protein structures from the amino acid sequence, with atomistic resolution. An application of this method to investigate spider silk nanostructures can be found in Chapter 6.

2.3 Data analysis and visualization

A wide variety of scripts are commonly used for analyzing data from MD simulations. Basic python and shell scripts are used to generate assemblies of structures. For equilibration simulations, root mean square deviation (RMSD), hydrogen bond dynamics, trajectory smoothing and distribution of strains can all be calculated by means of .tcl scripts integrated that can be used in conjunction in VMD [111]. A wide range of functions available in MATLAB have also been used in this thesis to use raw data from simulations to extract material properties, displacement profiles and strain distributions.

All scripts used in this thesis are provided in Appendix A with appropriate descriptions within the script.

2.3.1 Analysis methods for simulation results

2.3.2 Visualization schemes

Visualization plays a crucial role in the analysis of MD simulation results, as the raw data represents merely a collection of positions, velocities and accelerations as a

function of time. In particular, structural features and patterns of proteins are difficult to analyze. To address this point, many visualization tools have been developed for displaying protein molecules and assemblies. A rather versatile, powerful and widely used visualization tool is the Visual Molecular Dynamics (VMD) program [111]. VMD enables one to render complex molecular geometries using particular coloring schemes. It also facilitates describing important structural features of proteins by using a simple graphical representation, such as alpha-helices, or the protein's backbone. The simple graphical representation is often referred to as cartoon model. These visualizations are often the key to understand complex dynamical processes and mechanisms in analyzing the motion of protein structures and protein domains, and they represent a filter to make useful information visible and accessible for interpretation.

2.4 Experimental techniques

Recent advances in experimental techniques further facilitate analyses of ultra-small scale material behavior. For instance, techniques such as nanoindentation, optical tweezers, or atomic force microscopy (AFM) can provide valuable insight to analyze the molecular mechanisms in a variety of materials, including metals, ceramics and proteins. Mechanical signature of proteins and other single biomolecules can be obtained by AFM where the biomolecule is attached to a surface and manipulated by a cantilever that pulls the molecule at constant force. Figure 2-4 depicts AFM experiments on protein structures. A saw-tooth shaped force-displacement profile is commonly observed and linked to sequential unfolding of certain domains in the protein. Worm-like chain model (WLC) [45, 162] is frequently used to describe the entropic elasticity of these domains. The reader may refer to articles regarding details of these experimental approaches (see, e.g. [186, 255, 227, 229, 167, 98, 62, 18]). A selection of experimental techniques is summarized in Figure 1-4, illustrating the overlap with multi-scale simulation methods.

An important experimental technique in conjunction with atomistic modeling of protein materials is x-ray diffraction; results of such experiments provide the initial

atomistic and molecular structure, the starting point for all atomistic simulations. The structure of many proteins, elucidated using such experiments, has been deposited in the Protein Data Bank [14].

Since recent advances in experimental methods now enable one to probe time- and length-scales that are also directly accessible to large-scale atomistic based simulation, the combination of experiment and simulation might lead to a particularly fruitful interaction. This is particularly promising since the kind of information obtained from experiment and simulation might be complementary.

2.5 Theoretical bond strength models at the atomic level

The strength of a material is ultimately governed by the cohesive strength of the various bonds that it employs at the atomic scale. Evaluation of bond strength is therefore a fundamental question for materials science applications. Furthermore, most biological systems consist of weak bonds that can break and reform at room temperature and small forces, which plays a functional role in the structural transformations of proteins, cell adhesion and motility [76, 142]. Developing quantitative models that can predict the strength of individual bonds under mechanical, chemical and thermal perturbations is therefore an issue of fundamental importance to biophysicists, chemist and materials scientists.

One of the first atomic level models for assessing bond strength was Bell's model for cell adhesion [12]. Bell's model is a simple and popular phenomenological model that describes the the frequency of failure of reversible bonds. The concept of reversibility relates to the idea that the bond is weak, and therefore can reform under thermal fluctuations. For instance, electrostatic, van der Waals, or H-bond interactions can be considered to be reversible bonds that can break and reform at high frequency. The frequency of dissociation is characterized by an off-rate, scaling inversely with a bond lifetime [126].

Bell's theory explains the force dependence of the off rate, and thus provides an understanding of how molecular forces drive bond breaking at the nano-scale. Bell's model builds upon the fundamental concepts of the Arrhenius equation and transition state theory of reactions; it was also inspired by Zhurkov's work on the kinetic fracture theory of the strength of solids [259] [100]. Bell predicted that the off rate of a reversible bond, which is the inverse of the bond lifetime, increases when subjected to an external force f. Indeed, the rupture of bonds occurs via thermally assisted crossing of an activation barrier E_b which is reduced by $f \cdot x_b$ as the applied force f increases, x_b being the distance between the bound state and the transition state. Thus the Bell off rate expression is

$$k = \omega_0 \exp\left(-\frac{E_b - f \cdot x_b}{k_B \cdot T}\right), \tag{2.11}$$

where ω_0 is a natural vibration frequency and $k_B \cdot T$ the thermal energy. The force $f_0 = E_b/x_b$ represents the force to vanish completely the energy barrier and gives a very rough value of the rupture force.

This conjecture was established long before single molecule experiments were performed. Later, it became very successful especially to describe forced unfolding of biological molecules. The simplicity of the model and its capacity to predict energy barriers and rupture strength of bonds has been useful for interpreting experimental observations. The model has also been the basis of more advanced theories regarding the strength of single bonds and bond clusters.

Despite its success, Bell's model has several shortcomings. For instance, when Bell's model is applied to complex problems such as protein unfolding, the multidimensional nature of the energy landscape of biomolecules has to be reduced to a single relevant reaction coordinate. The loss of information from this simplification may be important in systems that have multiple unfolding pathways and transition states. Another limitation is the fact that the distance to the transition state, x_b is assumed to be constant with different rates. An extension to Bell's model based on Kramer's theory was provided by Evans *et al.*, in order to provide a more general expression

for the off-rate by taking into account by taking into account force dependent terms, in particular $x_b(f)$ [75, 165].

An important contribution of the Bell model is the prediction that the strength of bonds depends crucially on the loading rate. A key prediction is that above a critical loading rate, the force of rupture increases logarithmically with the loading rate, a phenomenon observed in many experiments. The bond strength f^* is defined as the most probable rupture force and corresponds to the peak of the force density distribution. Given that x_b, E_b and the loading rate $r_f = \Delta f/\Delta t$ remain constant, Bell model predicts the logarithmic dependence of strength on the loading rate:

$$f^* = \frac{k_B \cdot T}{x_b} \ln\left(\frac{r_f \cdot x_b}{k_B \cdot T \cdot k_0}\right) \tag{2.12}$$

where k_0 is the off rate in the absence of force. This expression reported in [55, 255] is very powerful as it can directly quantify energy landscape and distance to the transition state by using a series of rate-dependence simulations or experiment. An extension of the Bell model has been formulated for stiff transducer systems (e.g. MD simulations), where loading rate is characterized by the pulling velocity of the transducer and time required to reach the transition state [1]. In this case, pulling velocity can be written in terms of the off-rate, k_0 and x_b as $v = k_0 \cdot x_b$ and expression in eq. (2.12) becomes

$$f^* = \frac{k_B \cdot T}{x_b} \ln(v) - \frac{k_B \cdot T}{x_b} \ln(v_0) \tag{2.13}$$

where $v_0 = \omega_0 \cdot x_b \cdot \exp\left(-E_b/(k_B \cdot T)\right)$

Similar to the eq. (2.12), eq. (2.13) also predicts a linear relationship of the failure force with $\ln(v)$. Figure 2-6 summarizes the concept of the effect of external force on energy landscape and logarithmic dependence of strength.

In force probe techniques, a transducer is usually moved at constant speed relative to a substrate in which the bond is anchored. The pulling speed v is linked to the loading rate r_f through the stiffness K of the transducer $r_f = K \cdot v$. Thus, this description of rate dependence is very important because it enables to get the

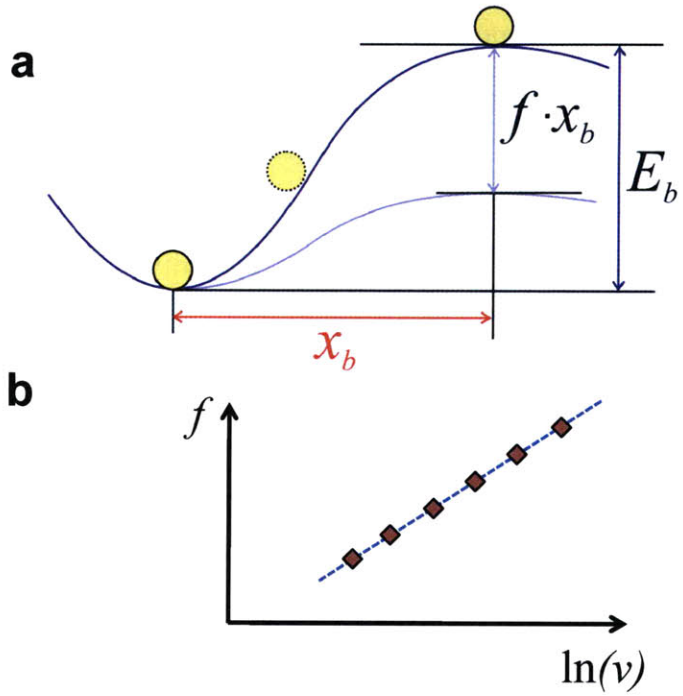

Figure 2-6: Panel (a) shows the evolution of the energy landscape of a bond subjected to a force. The minimum corresponds to the bound state. The transition state is the peak of the potential and corresponds to the bond rupture. The parameter x_b is the distance between the bound state and the transition state. According Bell's theory, the rupture of a bond occurs via thermally assisted crossing of an activation barrier E_b which is reduced with an applied force f. Panel (b) shows the logarithmic dependence of the force on pulling velocity. Linear fit to the f vs. $\ln(v)$ allows prediction of phenomenological parameters for E_b and x_b from experiment or simulation.

bond constants k_0 and x_b from a simple linear regression on the f vs. $\ln(r_f)$ curve. This rationalizes the variation among rupture force values obtained from different experimental and simulation techniques, where different loading rates were used. It is typically observed however that the logarithmic dependence breaks down when probing across many time-scales [1, 2]. An explanation can be that x_b and E_b do not remain constant (change of mechanism, multiple energy barriers) or the elastic behavior changes with the loading rate.

Several attempts have been made to extend the simple Bell model. For instance, it has been shown that rebinding can have a great impact on strength [209, 208, 75]. Similar to a force dependent x_b term, the existence of a rebinding rate can alter the logarithmic rate dependence of rupture force. In non-equilibrium pulling regime, other models attempt also to explain non-logarithmic rate dependence [110] Moreover,

other extensions try to implement the influence of the transducer stiffness in order to explain the disparities in measured unbinding force among different methods [248] . Another important concept to be included in these extensions is the energy landscape roughness of bonds [113].

According to Bell's theory, the lifetime of an individual weak bond such as an H-bond is very low. However, in biology, weak bonds can provide more significant strength by forming arrangement of multiple bonds. Several attempts have been made to study the strength of multiple parallel bonds [208, 74, 73]. These models aim to describe the influence of various parameters on the strength. Key parameters include the shape of the single bond potential profile, bond cluster size, rebinding rate, the constant external force, the loading rate and the stiffness of the transducers. Given the complexity of the structure, elasticity and energy landscape of biological materials, perhaps hybrid approaches combining concepts from fracture mechanics, statistical mechanics and chemistry will provide the most insight into generalized theories of protein strength and elasticity. One good example of this approach is a model developed for biopolymer extensibility, which combines the Worm Like Chain (WLC) elasticity with a thermodynamic two-state model extended from Bell's theory [193]. Similarly, combination of elastic network models with statistical theories may also provide to be a fruitful approach [65]. A challenge that remains, however, is to provide generic analytical theories that can explain both the elastic behavior and rupture strength of biopolymers. Chapter 3 will focus on a recent model developed based on fracture theory, chain elasticity, and Bell model to predict the mechanical strength of H-bond assemblies in protein materials.

Chapter 3

Generic fracture strength model for beta-structures in proteins

At a fundamental level, material failure is linked to how atoms and molecules detach from each other through nanoscale processes. For instance, in order to understand the failure mechanics of proteins and biomolecular assemblies from a physical science perspective, we need to identify the universal and diverse features of polypeptides, and the relative importance of these features on mechanical properties of these materials. This calls for a simplified framework that is focused on the general rather than the specific properties of proteins. In line with this goal, this chapter presents a generic approach to describe failure of protein materials at the molecular level [128, 131, 129].

3.1 Mechanical significance of beta-sheet proteins

Natural fibers can match the strength of steel, as in the case of spider silk [247], yet the chemical bonds that give rise to these properties, namely H-bonds, are almost one thousand times weaker than the interatomic metallic bonding. The source of the macroscopic strength remains a mystery: The fact such strong materials can be created using weak H-bonds is counter-intuitive and has been discussed controversially. For instance, mechanical resistance of proteins has been linked to their orientation [198, 251, 23, 94]. However, how the topology, the number of parallel

strands and their lengths influence the strength remains divisive. Most theoretical concepts that attempt to explain the strength of spider silk and similar materials are based on phenomenological continuum-type theories that are not directly linked to specific atomistic and molecular processes. As a consequence, nano-scale deformation and fracture mechanisms in silk and other beta-sheet dominated protein materials remain unknown.

Steered molecular dynamics (SMD) simulations [155, 115, 149] have confirmed that the mechanical resistance of beta-sheet rich proteins derive from hydrogen bonding between beta-strands of beta-sheet structures. Mechanical resistance is highly dependent on pulling geometry; the key beta-sheets must be loaded in shear, parallel to the strand direction [198, 251, 23, 94] such that parallel hydrogen bond assemblies in between the beta-strands cooperatively resist deformation. This loading geometry also corresponds to physiological conditions in for instance the I27 domains in the muscle tissue [163], and it justifies the orientation of beta-sheets predominantly along the fiber direction in spider silk [173] (see also Figure 3-1).

The shear loading leads to stronger structures since it induces concurrent as opposed to sequential breaking of H-bonds. Based on this view, longer strands would have higher mechanical strength, since more H-bonds contribute to the mechanical resistance. In contrast, it is observed in experimental analyses of protein structures that strand lengths do not exceed several residues (one residue equals roughly one H-bond), or equivalently a few nanometers in virtually all beta-sheet structures currently known [180]. These short strands are typically arranged between short random coil segments, forming a composite of H-bond networks and chains exhibiting entropic elasticity. These structural features appear to be a universal phenomenon.

Moreover, this particular geometry is astonishing since other protein structures such as those found in collagenous tissue [39] or cytoskeleton networks [4] display much larger structural features, reaching 300 nm and more. However, in contrast to beta-sheets, the intermolecular interactions in these protein materials are not primarily based on H-bonds, but include much stronger electrostatic and covalent bonding. This suggests that the ultra small features in beta-sheets may be related to the

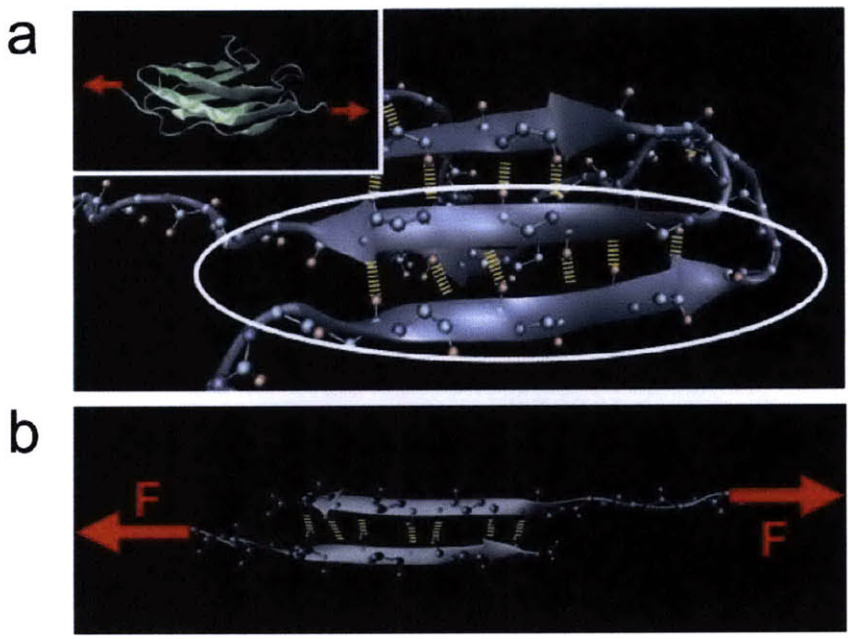

Figure 3-1: Beta-sandwich structures are a characteristic feature of mechanical proteins that employ networks of parallel H-bonds that work cooperatively in shear. Interstrand hydrogen bonds (thicker yellow lines) act as mechanical clamps that resist unfolding, as shown in panel (a) (the inset shows the characteristic shear topology of I27 domain from muscle protein titin). The single double-strand system panel (b) is taken as a model to study the strength of beta-sheet strands.

specific properties of the H-bond networks.

Despite significant advancements in our understanding of the nanomechanics of beta-structures in materials [195, 194, 172, 196, 198, 251, 23, 94], several key fundamental questions remains unanswered. Using theoretical and computationals tools, this chapter aims to address the following points:

- Estimating the strength limit of H-bond assemblies in protein materials
- Improving current understanding of the structure-mechanics relationship of strong beta-sheet proteins
- Understanding how the length of beta-strands in beta-sheets influences their mechanical properties

Addressing these key issues through studies on complex protein structures is extremely difficult, due to the challenges in isolating rupture events in the unfolding

process. For this purpose, the next section discusses atomistic results based on a simple beta-sheet model system. The structure represents a mechanical clamp as found in beta-structured proteins. An isolated nano-scale model captures intricate key details that govern force-induced unfolding of such systems, in particular entropic elasticity of polypeptide chains at low force levels, and the energetics of hydrogen bonding. These two aspects of the protein unfolding problem had been commonly addressed separately in experimental and computational studies. The simplistic theoretical and computational results presented here aim to capture the coupling between the elasticity of protein backbone and H-bond energy release due to bond rupture during the unfolding process.

3.2 Molecular structure of beta-sheets and atomistic modeling results

The beta-sheet, along with the alpha-helix, is one of the two most common types of secondary structures that make up protein materials [246]. Typically a sheet consists of several parallel (or anti-parallel) peptide chain segments with multiple interstrand H-bonds that stabilize the structure. Regular structure of the sheet is extended onto less regular loop sections and random coils that exhibit entropic elasticity under low external force levels. Unlike alpha helices, which consist generally of a single polypeptide chain folded onto itself in helix form by localized H-bond interactions, beta-sheets can be formed not only by intramolecular attraction between two regions of the same polypeptide chain but also by intermolecular forces between individual polypeptides, serving as Nature's glue for assembling very complex nanostructures.

The complex nature of full protein domains makes it extremely difficult to find generic, quantitative predictions of the strength of beta-sheet structures. The goal here is to overcome this challenge by selecting a simple structural model and well-defined loading conditions to provide crucial physical insight into the common observations made by both computational and experimental approaches in the earlier

works on beta-structures.

As a simple model representation of larger beta-sheet rich protein structures (Figure 3-2(a)), a small, full atomistic protein model embedded in explicit solvent (TIP3P water), as shown in Figure 3-2(b) is considered here. The model contains three beta-strands with inter-strand H-bonds, representing the beta-sheet protein motif that forms larger protein structures as shown in Figure 3-2(a). The strands in this model system have free chains at their extremities, representing already unfolded domains of a macromolecule.

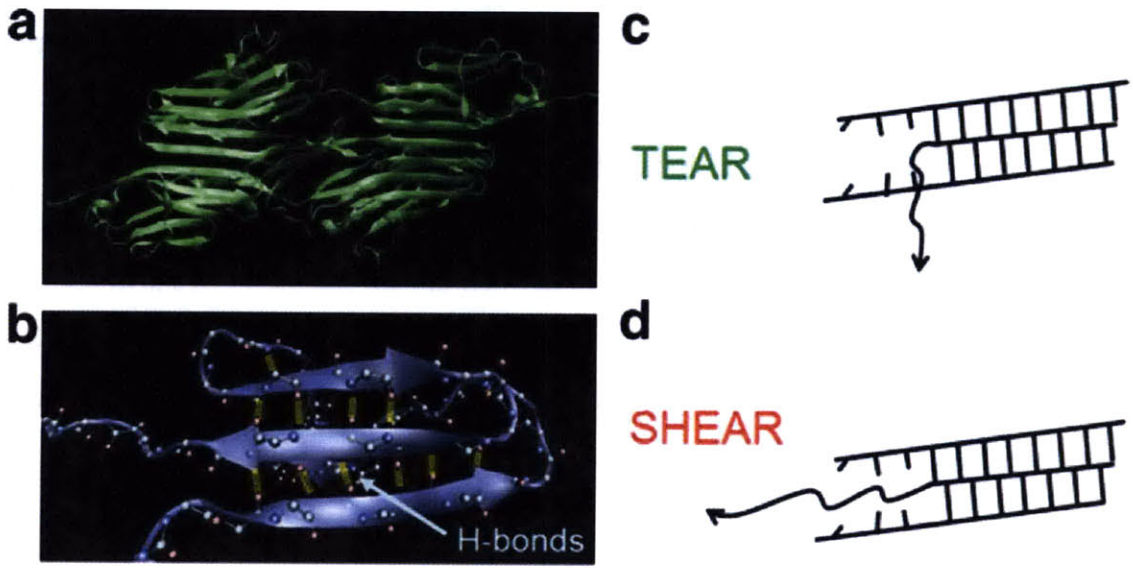

Figure 3-2: Molecular structure and loading conditions, used for molecular dynamics simulations. Beta-sandwich structures are a characteristic feature of mechanical proteins, which employ networks of parallel H-bonds that work cooperatively in shear, like the Z1Z2-telethonin complex in titin (shown in panel (a)). Inter-strand H-bonds (thick yellow lines) act as mechanical clamps that resist unfolding, as shown in panel (b) for the I27 domain of titin. A a three-strand system with free chains at the ends, as shown in panel (b) is taken as a model to study the strength of beta- strands. Two deformation modes are studied, in-plane shear and out-of-plane shear, which are referred to here as SHEAR and TEAR modes, respectively (panels (c) and (d)). In the TEAR mode, the middle strand is pulled out in the direction orthogonal to the strand direction. In the SHEAR mode, the middle strand is pulled in the direction of the beta-strand, leading to uniform loading of H-bonds.

All-atom MD simulations were performed using segments from the crystal structure of a beta-barrel protein obtained from Protein Data Bank [14] (PDB ID 1i78). The model three-strand beta-sheet system studied corresponds to residues 134-144

(lower strand), 161-176 (middle strand and free chain) and 203-211 (upper strand) in the PDB file 1i78.

In MD simulations, energy minimization is carried out for 20,000 steps, which allows for a favorable conformation to be achieved. This is followed by an equilibration procedure, during which the molecular assembly is heated up to 300 K with a rate of 25 K every 25 steps. An NVT ensemble is employed to hold the temperature constant at the final temperature of 300 K. Using a time step of 1 fs, each molecular assembly is subjected to equilibration for 2 ns. The system is equilibrated while all $C\alpha$ atoms on top and bottom strands were fixed. This constitutes the boundary conditions adopted for uniformly shearing two strands with minimal translational and rotational motion of the molecular assembly.

To simulate forced rupture of H-bonds between the strands, the Steered Molecular Dynamics procedure (SMD) with a constant velocity pulling scheme is used, in line with the general description in Chapter 2. An SMD spring constant of 10 kcal/mol/Å^2 is used. Pulling rates ranging from 20 m/s to 0.1 m/s are employed in this study to estimate the energy barriers and location of the transition state for each deformation regime. These pulling velocities are reasonable in light of previous studies [155, 87, 149] on beta-sheets. To subject the assembly to shear, the terminal $C\alpha$ atom of the middle strand free chain is pulled, while the other strands are subject to fixed boundary conditions described in the equilibration process.

Bond dynamics analysis is done based on NAMD results and visualization schemes using the VMD program [111]. H-bond rupture is defined to occur at an extension of 4 Å. A large sampling of trajectories (10 bond extension data points per ps) are used to estimate rupture dynamics accurately in each deformation mode.

Maximum force for rupture is obtained from MD simulations. A moving average is employed such that force-displacement plots employ 1 point per 1 Å extension. In tear mode, since the bond rupture events occur sequentially and a series of peaks are observed, the mean force over all rupture events is taken as the representative maximum force, as it is statistically more significant than the absolute maximum for this deformation mechanism.

3.2.1 MD Simulations in tear and shear modes for a simple beta-sheet model

As a simple model representation of larger beta-sheet rich protein structures (Figure 3-2(a)), a small, full atomistic protein domain embedded in explicit water is considered, as shown in Figure 3-2(b). The model contains three beta-strands with inter-strand H-bonds, representing the beta-sheet protein motif that forms larger protein structures as shown in Figure 3-2(a). The strands in this model system have free chains at their extremities, representing unraveled protein domains. The behavior of this protein assembly is studied by large-scale molecular dynamics (see Chapter 2 for details regarding the numerical procedure). In the analyses of the mechanical response of this system, the center strand is pulled in different directions of loading, while the outermost strands remain fixed.

The goal of the computational experiments is to probe the strength and rupture mechanism of the model system in two extreme modes of deformation. In the out-of-plane shear mode (TEAR, Figure 3-2(c)), the middle strand is pulled perpendicular to the plane of the sheet. The computational experiment in the TEAR mode is designed so that the H-bonds break sequentially, one-by-one. This case enables assessment of the rupture behavior and energy landscape of individual H-bonds. The second case considered is the the in-plane shear mode (SHEAR, Figure 3-2(d)). In this case, the middle strand is pulled out in the axial direction of the three strands. The SHEAR case is designed to asses the rupture behavior and the energy landscape of a large number of H-bonds under uniform shear loading. For both deformation modes, a systematic study of the rupture force F against pulling rate v is carried out, and the results are plotted on a logarithmic scale over three orders of magnitude (Figure 3-3(a)). The curves obtained for each deformation mode fall on two distinct lines in the F vs. $\ln(v)$ plane, which indicates according to Bell Model (see Chapter 2) [1] that the unfolding event can be characterized by a two state system with a discrete and unique energy barrier, E_b and location of the transition state, x_b [12] .

The energy barrier in the TEAR deformation mode is $E_b^{TEAR} = 2.83$ kcal/mol. An

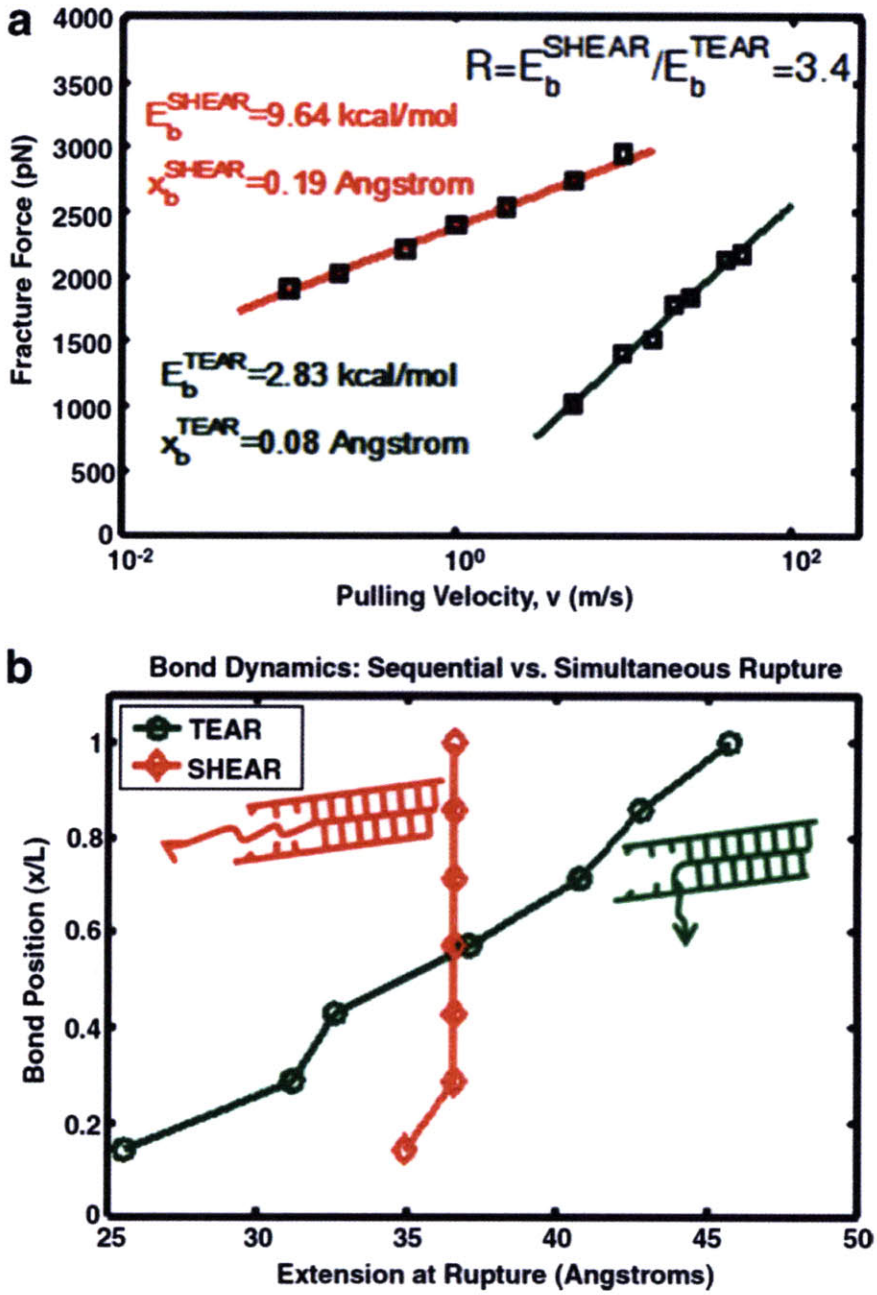

Figure 3-3: Unfolding force vs. deformation speed, TEAR vs. SHEAR deformation modes, and associated H-bond rupture mechanism (results of full atomistic molecular dynamics simulations in explicit water solvent) [129]. Panel (a): Logarithmic relationship between the pulling speed and rupture force, for both deformation modes. The parameters of the linear fit for each regime are used in conjunction with the Bell Model to determine the energy landscape for each deformation mode. Panel (b) displays the H-bond rupture dynamics, which reveals that the localization of strain in the TEAR mode leads to sequential rupture of H-bonds. In contrast, several H-bonds rupture concurrently in the SHEAR mode.

analysis of the bond rupture dynamics reveals that the H-bonds break sequentially (Figure 3-3(b)), indicating that the energy barrier to break a single H-bond in this protein is $E_{HB}^{0} = 2.83$ kcal/mol. The energy barrier in the SHEAR deformation mode is $E_b^{SHEAR} = 9.64$ kcal/mol, significantly larger than in the TEAR mode. The analysis of the rupture mechanism in the SHEAR mode reveals that H-bonds break in clusters of multiple bonds (Figure 3-3(b)).

The comparison of the energy barriers in the TEAR and SHEAR mode can reveal how many H-bonds participate in the rupture mechanism in SHEAR loading. This is possible because the energy barrier in the TEAR mode corresponds to rupture of individual H-bonds. The number of H-bonds that break concertedly in the SHEAR mode is given by the ratio of the E_b values for SHEAR and TEAR; it is $E_b^{SHEAR}/E_b^{TEAR} \approx 3.4$. This suggests that H-bonds break in clusters of 3 to 4, but not more, even when the loading conditions (uniform shear) allow them to respond cooperatively.

The key finding from this study is that despite the presence of more than a dozen H-bonds in the system, all loaded in uniform shear, H-bond rupture occurs in clusters of 3-4. In terms of the energy barrier, it appears to be limited to a finite value, despite the possibility for a manifold increase of the resulting energy barrier due to the large number of H-bonds. An explanation to this phenomenon is provided in the next section, where a theoretical framework of the asymptotical strength limit prediction of hydrogen bond assemblies is presented.

3.3 A generic fracture strength model for proteins

Non-covalent bonds such as H-bonds are indispensable to biological function as they play a key role in cell adhesion and motility, formation and stability of protein structures and nucleic acids, and receptor-ligand interactions. The weak nature of these bonds makes it a necessity for them to work cooperatively and function in assemblies. Strength of multiple parallel bonds has received notable interest and has been studied both theoretically, using statistical mechanics approaches as well as experimentally with the atomic force microscope (AFM) [75, 165, 105, 208, 73, 101, 225].

The fundamental question of strength is also linked to the protein unfolding problem, where the rupture of parallel inter-strand H-bonds controls the mechanical unfolding pathway. Atomistic simulation [156, 149] and single-molecule force microscopy studies [194, 195, 47] have shown that beta-sheet rich proteins exhibit higher rupture forces, since they employ parallel strands with numerous H-bonds that act as mechanical clamps under shear loading [198, 251, 23, 94]. In both experiment and simulation of mechanically resistant proteins, maximum force peak observed is linked to an individual event corresponding to the unraveling of a single beta-sheet in the protein. The dependence on pulling velocity has also been widely discussed in the literature [1, 219] and explains the discrepancy of force values between MD simulations and experimental methods, the former employing fast and the latter near equilibrium pulling rates. It is astonishing to see that regardless of the variation in topology and size of structures studied, almost all so called 'mechanical' proteins (e.g. fibronectin and titin domains) examined so far exhibit a rupture force of a few hundred piconewtons (pN) at experimental pulling rates [219, 226].

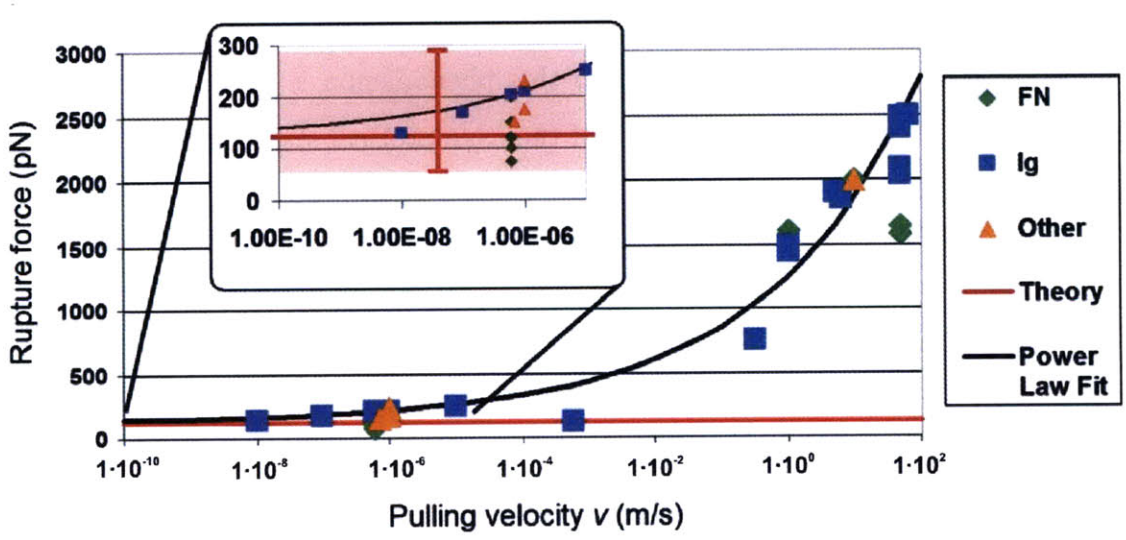

Figure 3-4: Summary of earlier studies on rupture force of beta-domains as a function of pulling rate as reported in [128]. This figure summarizes findings (adapted from [226]) of the strength of beta-sheets rich proteins (focus on fibronectin, immunoglobulin domains in ECM and titin). The overall behavior suggests that the rupture force asymptotically approaches a limiting value for vanishing pulling rates (continuous line is a power law fit to data for I27 [219]).

Earlier experimental and simulation results are presented in Figure 3-4 where rupture force is plotted against pulling velocity on log (v) scale. The asymptotical limit at vanishingly slow (quasi-static) rates can be inferred from the overall behavior as well as from the power law fit to data based on unfolding force data corresponding to the I27 domain in titin (data obtained from references [219, 226]). This observation suggests that without the presence of salt bridges or covalent links, the strength of individual protein domains asymptotically approaches a limiting strength. Thus far, no theoretical basis or prediction has been proposed for such an intrinsic strength limit; most earlier analysis have been focused on the rate dependent behavior in order to explain the increase in unfolding forces with increasing pulling speed (Figure 3-4). An explanation to this phenomenon can be provided by applying basic concepts from fracture mechanics to the protein unfolding problem. To simplify the complex unfolding problem, let's focus on a model system that consists of a single beta-sheet strand stabilized by H-bonds, which has been identified as the mechanical clamp that governs strength during unraveling of the protein [198, 251, 23, 94] (see Figure 3-2). The next section focuses on developing a theoretical framework for this problem.

3.3.1 Griffith-Irwin energy balance concept and the protein strength model

This section reviews a theoretical model based on the structure presented in Figure 3-5, consisting of a single strand with multiple peptide H-bonds and a free chain representing unraveled protein domains. The idealized model is shown in Figure 3-5(a), where a polypeptide chain of arbitrary length is attached to a substrate and strained in one end with force F. This setup serves as the model to develop the theoretical framework. As a maximum strength value that is not hindered by too few H-bonds is sought after, let's consider an infinitely long beta-strand (that is, the length of the strand, $L_x \rightarrow \infty$).

Here, Griffith's theory of fracture mechanics ([95] is adopted to predict the critical rupture force required to break the bonds in the theoretical model system (see Figure

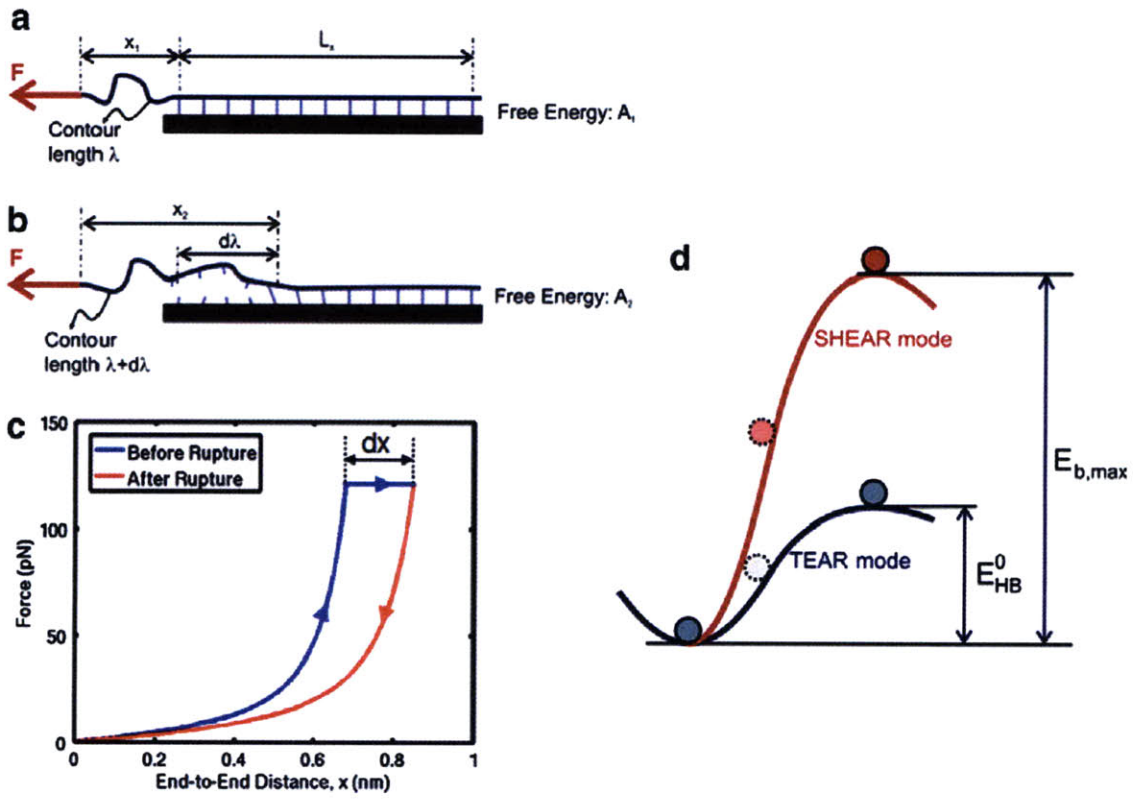

Figure 3-5: Schematics to explain the main concepts of the WLC-based fracture theory, as well as the implications on the energy landscape [128, 129]. Panel (a): A single double-strand system as a polypeptide chain of infinite length ($L_x \to \infty$), with a free end of length a is considered. The polypeptide chain is stabilized by an array of parallel H-bonds, and strained due to the external force F. At the onset of rupture (panel (b)), several H-bonds break, and the contour length increases due to the detachment of a piece of the chain of length $d\lambda$. Panel (c): The change in the contour length and the end-to-end distance at the onset of rupture yields two distinct WLC curves, relating to the states before and after rupture; the area enclosed between the WLC force extension curves is equivalent to the change in free energy before and after rupture. This dissipated (released) energy must equal the adhesion energy per unit length at the onset of rupture. Panel (d) shows a schematic of how the energy barrier changes from rupture of a single H-bond up until the limiting energy barrier $E_{b,max}$, when 3-4 H-bonds rupture simultaneously. This energy barrier is the highest energy barrier of an assembly of H-bonds under uniform loading.

3-5(a) and (b)). The concept is based on the idea that the onset of fracture is characterized by the condition that the change in potential energy W_P of a system due to extension of a crack balances the energy necessary to create new surfaces.

In the Griffith theory, the negative of the expression for the potential energy change with respect to a crack advance of one unit distance δa is called the energy release rate or crack driving force given as:

$$G = -\frac{\delta W_P}{\delta a}. \tag{3.1}$$

The total potential energy in the system can be expressed as the sum of the elastic energy stored in the material, denoted by Φ, and the work done by the external load, denoted by W_F:

$$W_P = \Phi + W_F \tag{3.2}$$

The onset of fracture is characterized by the condition that the potential energy released due to a unit crack advance δa must equal the energy required per unit length to create a new fracture surface. The idea that fracture surface energy, γ_s must equal the energy release rate at the onset of fracture leads to

$$G = -\frac{\delta W_P}{\delta a} = \gamma_s. \tag{3.3}$$

Despite being originally proposed as a continuum fracture theory, the energy balance notion relies on purely thermodynamical concepts and has been used in the past to describe fracture phenomena at various length scales, including atomistic phenomena in collagenous tissue [39].

The model system can be broken down into three components, the external force, the peptide H-bonds, and the elasticity of the free chain that represents the backbone of unfolded protein domains. Since the strength at vanishing pulling rates is of interest here (quasi-static deformation),the system is assumed to be in equilibrium and the force is constant over infinitesimal time scales. The strength of H-bonds are characterized by their dissociation energy, E_b.

To understand the fracture behavior of polypeptides bonded by H-bonds, one can extend the theoretical framework laid out by Griffith and Irwin by taking into consideration that at the force levels of interest, the elasticity of the protein backbone is primarily due to entropic rather than energetic effects. The Marko-Siggia worm-like chain (WLC) model [45, 162] is one of the most widely used expressions to predict the entropic elasticity of polypeptide chains, and has been adopted widely as the elastic description of the backbone (a wider range of earlier studies provide substantial evidence that this model is an excellent model for the behavior of individual, unconstrained polypeptide chains [194, 195, 80, 163, 172, 46]). According to the WLC model, the force-extension behavior of a polypeptide chain can be given as:

$$F(x) = \frac{k_B T}{4\,\xi_p} \left[\left(1 - \frac{x}{\lambda}\right)^{-2} + 4\frac{x}{\lambda} - 1 \right], \tag{3.4}$$

where ξ_P is the persistence length, λ is the contour length (length when fully-extended), and x is the end-to-end length of the chain. While the Griffith theory considers the change in internal energy, U of the system, in the case of entropic elasticity, the expression sought after is related to how the free energy – rather than the internal energy, changes as a function of crack extension. It is emphasized here that this generalization from the internal energy to the free energy does not change the concept behind the Griffith model and thus the Griffith model is applicable without any further modifications.

The approach can be summarized as follows: (i) free energy is released per unit polypeptide rupture distance is calculated (free energy release rate, it increases with larger applied force); (ii) the critical free energy release rate is determined, such that it is equal to the cohesion energy of a H-bond system, and then (iii) the critical force required to initiate rupture based on this critical free energy release rate is computed.

Integration of the force-extension equation of the WLC model yields the free energy $A = U - TS$ at a specific deformation state. This is because the WLC equation describes forces that arise due to changes in the entropy of the system and is derived by calculating the change of free energy as a function of the deformation

variable, x. The potential free energy stored in the molecule of end-to-end distance $\hat{x}$ and contour length $\hat{\lambda}$, subject to force F is given by the integral of the force F, integrated from the reference point taken at zero to the end-to-end distance $\hat{x}$ (this is a general expression valid for any contour length $\hat{\lambda}$ and any end-to-end distance $\hat{x}$):

$$A(\hat{x},\hat{\lambda}) = \int_0^{\hat{x}} F(\tilde{x})\, d\tilde{x} = \frac{k_B T}{\xi_p}\left[\frac{1}{4}\left(1-\frac{\tilde{x}}{\hat{\lambda}}\right)^{-1}\hat{\lambda} - \frac{\tilde{x}}{4} + \frac{\tilde{x}^2}{2\hat{\lambda}}\right]_0^{\hat{x}}. \tag{3.5}$$

Before the onset of rupture, the molecule has a free end-to-end distance of length x_1 and an initial contour length of λ_1 and is strained by an external force, F which remains constant at the onset of rupture. When H-bonds rupture the end-to-end distance as well as the contour length both increase. Let's assume that a piece of the molecule of length $\delta\lambda$ detaches at the moment of rupture. The contour length then changes from λ_1 (before rupture) to $\lambda_2 = \lambda_1 + \delta\lambda$ (after rupture). The change of the end-to-end distance from x_1 (before rupture) to x_2 (after rupture) can be calculated by considering that the force before and after rupture is identical. Before fracture,

$$F_1 = \frac{k_B T}{4\xi_p}\left[\left(1-\frac{x_1}{\lambda_1}\right)^{-2} + 4\frac{x_1}{\lambda_1} - 1\right], \tag{3.6}$$

and after rupture,

$$F_2 = \frac{k_B T}{4\xi_p}\left[\left(1-\frac{x_2}{\lambda_1+\delta\lambda}\right)^{-2} + 4\frac{x_2}{\lambda_1+\delta\lambda} - 1\right], \tag{3.7}$$

Assuming that the system is equilibrated at all times, the force remains constant before and after fracture ($F = F_1 = F_2$). Based on this assumption, one can equate eqns. (3.6) and (3.7) and obtain an expression for end-to-end distance after rupture, x_2 as a function of the initial end-to-end distance, x_1:

$$x_2 = \frac{\lambda_1 + \delta\lambda}{\lambda_1} x_1. \tag{3.8}$$

The change in end-to-end distance during rupture is

$$\delta x = \left(\frac{\lambda_1 + \delta\lambda}{\lambda} - 1\right) x_1. \tag{3.9}$$

Knowledge of these parameters now enables us now to calculate the change in free energy ($\delta W_P = \delta\Phi + \delta W_F$) due to rupture of a piece of length $\delta\lambda$. Recall that the quantity $G = -\delta W_P/\delta\lambda$ is the energy release rate. Since

$$\delta\Phi = A_2 - A_1, \tag{3.10}$$

and

$$\delta W_F = -F\delta x, \tag{3.11}$$

the net change in free energy is given by

$$\delta W_P = A_2 - A_1 - F\delta x. \tag{3.12}$$

To greatly simplify these expressions, one can introduce a deformation variable $\alpha = x/\lambda$, which denotes the stretch level of the chain. The physical meaning of α is the ratio of end-to-end length of the free chain to its contour length, and is defined in the range from 0 to 1 in context of the WLC theory. The WLC expression for force then becomes:

$$F_{WLC} = \frac{k_B T}{4\xi_P}\left((1-\alpha)^{-2} + 4\alpha - 1\right) \tag{3.13}$$

Similarly, $A_{WLC} = \int_0^\alpha F_{WLC}(\alpha)d\alpha$ becomes the free energy of the chain as a function of the stretch state, based on the worm-like chain (WLC) model. Eq. (3.13) represents a continuum approximation for the elasticity of polypeptide chains; any other appropriate elasticity model could also be used in context of the approach presented here. Noting that the deformation variable α remains constant as 'crack' propagates, the free energy state function of the system can then be written as:

$$W_P(x,\lambda) = \lambda A_{WLC}(\alpha) - Fx - \gamma_s(L - \lambda). \tag{3.14}$$

The condition for rupture can be found by taking the derivative of eq. (3.14) with respect to $d\lambda$; $dA/d\lambda = 0$ yields the critical α value (denoted as α_{cr}) that will enable propagation of bond rupture under constant force. The critical value, α_{cr}, can then be substituted into eq. (3.13) to obtain the failure force F_{cr}. The force extension behavior showing energy dissipated in a unit cycle of bond rupture is given in Figure 3-5(a-c). The fracture condition can simply be given as:

$$G(\alpha) = \frac{k_B T}{4\,\xi_p}\left[\alpha\,(1-\alpha)^{-2} - (1-\alpha)^{-1} + 2\alpha^2 + 1\right] \tag{3.15}$$

The value of G depends only on the ratio α but not on the individual values of x_1 and λ_1.This equation can immediately be used to calculate a critical condition for initiation of failure by assuming $G = \gamma_s$ and obtaining the value α_{cr} that satisfies this condition. The parameter γ_s is the 1D analogy to cohesive energy (twice the fracture surface energy), and is defined as:

$$\gamma_s = \frac{E^0_{HB}}{L_{x,0}}, \tag{3.16}$$

where E^0_{HB} is the energy released by a single bond and $L_{x,0}$ is the average distance between two bonds. Once the critical value, α_{cr} is solved for, the rupture force can be obtained as:

$$F_{\text{break}}(\alpha_{cr}) = \frac{k_B\,T}{4\,\xi_p}\left[(1-\alpha_{cr})^{-2} + 4\,\alpha_{cr} - 1\right]. \tag{3.17}$$

Figure 3-5 (c) displays the process of loading the structure, the rupture event, and the return to the initial configuration. The area between the two curves is the dissipated free energy that is employed in the breaking of the H-bonds.

The only input parameters in this model are the persistence length ξ_p of the polypeptide chain, as well as the dissociation energy of a H-bond, E^0_{HB} (γ_s is calculated directly from E^0_{HB}). E^0_{HB} for peptide H-bonds is known to generally lie in the range of 2-9 kcal/mol, whereas the persistence length has been found to be approximately 0.4 nm for many different polypeptides and also for DNA strands [194, 172, 46, 173, 213].

3.3.2 Prediction of maximum strength for H-bond clusters

To obtain a numerical prediction for N_{cr}, the following parameters are chosen: T= 300 K and $\gamma_s = 0.94$ kcal/(mol ·Å) (for $E^0_{HB} = 2.83$ kcal/mole from the TEAR mode MD simulation), and ξ_p= 0.4 nm from previous experimental studies ([194, 172, 46, 173, 213]. The equation can be solved numerically to find the parameter $\alpha = x/\lambda$, the ratio of the end-to-end distance to contour length. It turns out that α =0.85 for the particular value of γ_s selected. This means that rupture occurs when the free chain is stretched to 85 % of its contour length. The predicted asymptotic force limit is 127 pN. It should be noted that this value does not depend on the number of H-bonds that participate; rather, it is an intrinsic maximum shear strength limitation.

This is the most important result of this analysis: The force F_{break} is the absolute maximum strength that can be attained by any parallel bond arrangement, at near equilibrium pulling rates. The rupture force F_{break} is independent of the contour length λ_1, since the solution to eq. (3.15) only depends on α.

The significance of this result is evident by the comparison with experimental and computational results of the unfolding force as a function of pulling speed, as shown in Figure 3-4. It is apparent that the pulling speed dependence suggests that a limiting force of approximately 100 to 300 pN is reached for vanishing pulling rates. This observation has been noted previously in the literature; however, it has never been explained. This work closes this gap and provides the first rigorous explanation.

These findings indicate that the key to understand this strength limitation is the interplay between entropic elasticity and the energy of H-bond adhesion; since both components are 'material intrinsic properties' of protein structures, the force limit represents a universally valid value.

3.3.3 Application to double strand shear in beta-sheets

One of the underlying assumptions in the beta-sheet strength model is that the bonded chain is fully relaxed in the context of the WLC model leading to the condition $A(x,\lambda) = \lambda A_{WLC}(x/\lambda) - F(x) - \gamma_s(L-\lambda)$. This assumption is a result of the fact

that in the first term, free energies are consistently calculated with respect to the fully relaxed state, such that $A_{WLC} = \int_0^{\alpha} F_{WLC}(\alpha)d\alpha$. This represents a particular case of the theory that which leads to the least number of parameters and yields the simplest analytical formulation possible.

To extend the theory to take into account other cases, additional terms have to be considered. In general, the attached segment of the chain would have a finite average end-to-end distance given as sL, where $0 < s \leq 1$ and L is the contour length of the attached segment (see inset, Figure B-1). In addition to the terms already present, one should also take into account the initial entropic free energy of the bonded segment of the chain due to its partially extended configuration. Then, the free energy state function becomes $A(x,\lambda) = \lambda A_{WLC}(x/\lambda) - F(x - s\lambda) - \gamma_s(L - \lambda) + (L-\lambda)A_{FIX}(s)$ where the last term $A_{FIX} = \int_0^s F_{WLC}(\alpha)d\alpha$ has to be introduced since free energy integrals are taken with respect to the zero-stretch state, whereas the liberated segment already has a prestrech. The entropic elasticity and H-bond energy terms contribute separately to free energy summation. If the chain is considered fully-extended, the WLC theory cannot be used as it diverges at very high extension levels. Quantifying the strength with assumptions on initial extension state of the bonded chain is therefore not straightforward.

Proceeding as done previously but now with the prestretch accounted for, one can substitute $\alpha = x/\lambda$ and solve for the critical value of α that satisfies $A_{WLC}(\alpha_{cr}) + F(s-\alpha_{cr}) + \gamma_s - A_{FIX}(s) = 0, F_{WLC}(\alpha_{cr}) = F_{cr}$. Provided that s is known and within the range of values admissible to the WLC model, rupture is thermodynamically favorable. Our calculations indicate that for $E_b = 4$ kcal/mol , $\xi_p = 0.4$ nm, variations in s in the range $0 \leq s \leq 0.6$ yield results within the error range observed from the original beta-sheet strength theory presented in Section 3.3.2.

In the case that a large prestretch is assumed in the chain (e.g. s $\approx$ 1), the WLC theory may not be the most suitable elasticity function as the free-energy diverges for extended chains. This scenario, specific to beta-sheet systems, would also involve the possibility that stick-slip motion is the governing mechanism of failure where H-bond reforming occurs after sliding. In this case one can evaluate the critical condition as

$2A_{WLC}(\alpha_{cr})+F(s-2\alpha_{cr})+\gamma_s-2A_{FIX}(s)=0$ and extrapolate results to the case with (e.g. $s \approx 1$). These results, which are shown in Figure 3-6 are in excellent agreement with our original prediction, illustrating the robustness of the main contributions of this work. Overall, the results shown in Figure 3-6 here and also presented in Section 3.3.2 agree well with experiments (Figure 3-4) on beta-proteins.

It should be noted that the number of H-bonds present and the loading geometry will generally affect the strength of the assembly. The predictions reported here correspond to the maximum strength of a large cluster of H-bonds loaded uniformly, illustrating an upper strength limit beyond the size-dependent regime.

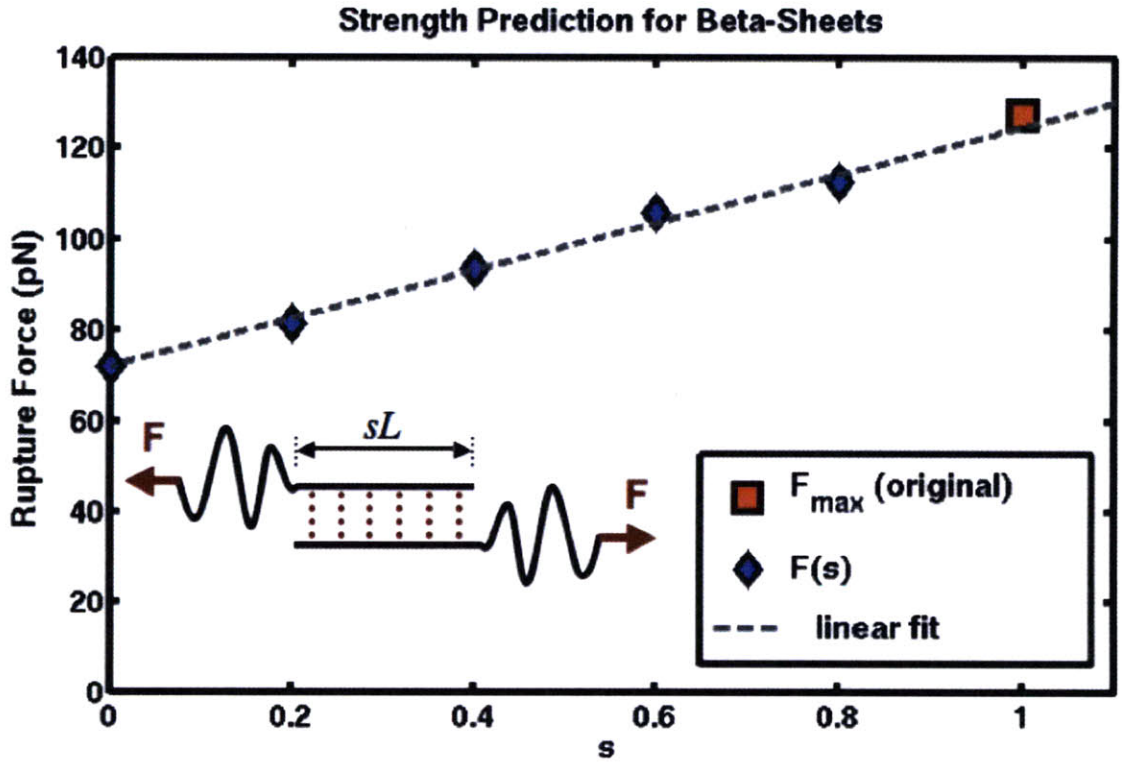

Figure 3-6: Strength prediction for varying values of s, including an extrapolation to s=1 (double stranded slip case relevant to beta-sheet topologies). The linear extrapolation is done for $0 \leq s \leq 0.8$, for values admissible to the WLC model. The value for $s \approx 1$ approaches the prediction made by our original model (see strength predictions in Section 3.3.2). The inset shows the double strand shear condition and parameter s, defined as the ratio of end-to-end length to contour length for the attached segment.

3.3.4 Influence of H-bond cluster size

MD simulations on the simplistic system suggest that H-bonds exhibit limited cooperation under external force. The equilibrium theory presented in the previous section complements this idea by showing that the rupture force of a beta-sheet with numerous H-bonds is finite. Theoretical derivation illustrates an extreme case of infinitely many H-bonds loaded extremely slowly and uniformly; this is essentially a scenario that allows maximum possible cooperation of the H-bonds.

The finite strength limit indicates that indeed a limited number of H-bonds actually participate in the rupture process and contribute to strength. Calculation of the critical number of H-bonds that break simultaneously from the strength prediction would be of great interest as this would yield a critical length scale for beta-strands beyond which further mechanical stability can't be achieved by additional bonds. The goal here is to link the strength limit to the number of bonds broken at the initiation of the rupture by using the phenomenological Bell model (see Chapter 2), in order to provide a link between external rupture force and energy barrier to unfolding. This idea is similar fracture process zone estimation in cracks propagating in a specimen.

Let's recall that the rupture force of the protein assembly is given as:

$$F_{\text{break}} = \frac{k_B T}{4\,\xi_p}\,\left[(1-\alpha_{cr})^{-2} + 4\,\alpha_{cr} - 1\right]. \tag{3.18}$$

In order to calculate the number of H-bonds that participate in a unit fracture event, the process at the rupture front where an *unknown* number of H-bonds, N_{cr}, break simultaneously should be considered. The strength of this cluster of H-bonds can be modeled by a statistical theory based on Bell's model [12]. The key parameter is the energy barrier: The energy barrier rises N_{cr}-fold when N_{cr} H-bonds break simultaneously. Thus the rupture force of this system is

$$F_{break}^{local}(N_{cr}) = \frac{1}{x_B}\,\left[k_B T \ln\left(\frac{1}{\omega\tau}\right) + E_{HB}^{0} N_{cr}\right], \tag{3.19}$$

where $\omega = 1 \times 10^{13}\ s^{-1}$ (natural frequency of bond vibration), τ is the characteristic time scale for H-bond dissociation, and x_B is the applied pulling distance at the

moment of bond rupture.

The parameter $x_B = 4$ Å corresponds to the distance required to pull the strand to initiate permanent rupture of bonds based on geometric considerations. The characteristic time scale τ= 20 ps corresponds to the time scale at which H-bond rupture occurs. This value is chosen according to experimental [213] and computational [1] results, and characterizes the dynamics of H-bond rupture, which is much faster than the loading rate in the asymptotic limit.

At the moment of rupture, the local rupture force given by eq. (3.19) must equal the critical rupture load F_{break}, given by eq. (3.19). Setting $F_{break}^{local}(N_{cr}) = F_{break}$, thus enables calculation of the number of H-bonds that break simultaneously in the unit rupture event,

$$N_{cr} = \frac{k_B T}{E_{HB}^0}\left[\frac{x_b}{4\,\xi_p}\,\left[(1-\alpha_{cr})^{-2} + 4\alpha_{cr} - 1\right] - \ln\left(\frac{1}{\omega\tau}\right)\right]. \tag{3.20}$$

A numerical prediction for N_{cr} is obtained by setting T= 300 K and $\gamma_s = 0.94$ kcal/mol/Å^{-1} (for $E_{HB}^0 = 2.83$ kcal/mole from the TEAR mode MD simulation), and ξ_p= 0.4 nm from previous experimental studies ([194, 172, 46, 173, 213]. The prediction of the critical number of H-bonds that break concurrently is $N_{cr} = 3.1$. The corresponding predicted maximum energy barrier (see Figure 3-5(d)) for this case is $E_{b,\max} = N_{cr}E_{HB}^0 = 8.77$ kcal/mol, with an asymptotic strength limit (at vanishing pulling rates) of 127 pN.

This is the most important result of this derivation: The value $E_{b,\max} = N_{cr}E_{HB}^0$ represents the highest energy barrier that can be achieved by a uniformly loaded H-bond assembly, providing an upper limit for its strength. Moreover, the prediction for the energy barrier is very close to what is observed in the SHEAR mode (E_b^{SHEAR}= 9.64 kcal/mol $\approx E_{b,\max}$ =8.77 kcal/mol) in MD simulations. There exists an intrinsic upper limit of 3-4 H-bonds that can break concurrently, illustrating that the conventional assumption of uniform shear loading of H-bonds in beta-strand fails once the number of H-bonds in an assembly exceeds this critical value.

The shear strength of beta-sheets can be calculated as a function of the strand

length, in the asymptotic limit (the analysis of the geometry of beta-sheets reveals that 1 H-bond $\approx$ 1 residue is a good approximation). The maximum rupture force of the sheet is then given by as

$$F_{ult}(N) = \max(F_{break}^{local}(N), F_{break}), \tag{3.21}$$

where N is the number of H-bonds in the assembly.

Equation (3.21), hereon abbreviated as the beta-strand strength model (BSSM), expresses the fact that any beta-sheet that employs less than the critical number of H-bonds N_{cr}, features a rupture force that is described by eq. (3.19), that is, the strength of increases with a increasing number of participating H-bonds. If N $=N_{cr}$, the rupture force equals F_{break}, the asymptotic strength limit of an assembly of H-bonds.

The shear strength equals the force required to break all of the bonds in the strand, $F_{ult}(N)$, divided by the contact area between the strands. Here it is assumed that the thickness T, of a beta-strand equals 5 Å, and its length is given by $N \times L_{x,0}$. The rupture force divided by the area $NL_{x,0}T$ gives the theoretical shear strength estimate for a beta-sheet,

$$\sigma_{\max}(N) = \frac{F_{ult}(N)}{NL_{x,0}T}. \tag{3.22}$$

For short strand lengths, since the rupture force does not increase with strands longer than the critical length scale, the shear strength of longer strands is much lower than short ones, scaling as

$$\sigma_{\max}(N) \sim \frac{1}{N}, \tag{3.23}$$

for large $N > N_{cr}$. The implications of this result is discussed in the following section.

3.4 Experimental validation

3.4.1 Direct comparison with AFM experiments

The fracture model model presented in the previous section can be used to explain the physics of bond rupture events observed in AFM or optical tweezers experiments. Here the focus is on interpreting and explaining the characteristic rupture behavior observed in the classical AFM study by Fernandez et al. on the titin I27 domain, with a single set of parameters. This study considers a protein domain with two differently sized clusters of H-bonds (cluster I with 2 H-bonds and cluster II with 6 H-bonds, lying below and above the critical number of H-bonds, N_{cr}) (see Figure 3-7).

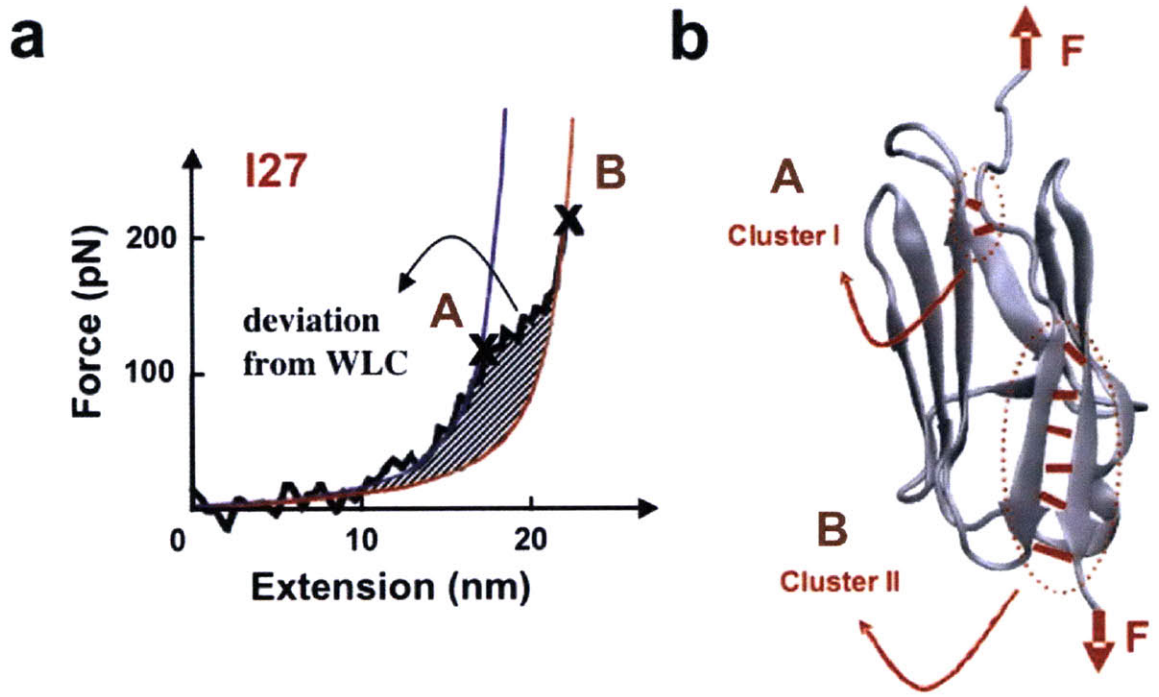

Figure 3-7: . Example force extension profiles of beta- sandwich protein structures, as obtained from experimental analyses (rupture marked with a cross). Force-extension profiles of I27 domain shown in panel (a) reveal a two-step process of unfolding, corresponding to rupture of two separate clusters of H bonds (plots redrawn based on data from Ref. [163]. The curves reveal that a single WLC model alone is not capable of describing the entire deformation range, and that there exist a transition from one WLC curve to a second one at a force level of approximately 110 pN.

Through a careful investigation of experimental data, Fernandez and his colleagues identified multiple force peaks and deviations from the WLC fit [163] to the force

extension profiles of the I27 domains. The unfolding of a single domain of the titin molecule occurs in a two-step process, involving first a hump barrier at which the first cluster of H-bonds break, and secondly a maximum rupture force at which the entire protein domain unfolds, followed by a rapid decay of the force. When the first cluster of H-bonds is removed by a proline mutation that breaks the cluster (that is, cluster I has zero H-bonds), the characteristic force peak and deviation of this cluster is not observed. The geometry of I27 domain and the behavior found in experiments on I27 and I32 domains are reviewed in Figure 3-7(a-b). Since the structure of the I27 domain is well known and its mechanical response has been characterized in detail by several AFM studies and molecular dynamics simulation (MD) [194, 156, 163]), it is a suitable benchmarking problem for validation of the BSSM.

We apply the BSSM to the structure shown in Figure 3-7(b). H-bond dissociation energy $E_{HB} = 5.05$ kcal/mol and persistence length $\xi = 0.4$ nm, according to the experimental values reported for I27 in Ref. [163]. Once these two parameters are fixed, the complete force-extension curve for this structure can be predicted without any additional fitting parameters. A direct comparison of the predicted energy barriers, strength and elasticity curve with the AFM experiment is shown in Figure 3-8(a-c). The initial loading of the entire protein domain follows a WLC-behavior until cluster I (N = 2) breaks. Rupture occurs at 120 pN according to BSSM. This strength prediction is based on the energy barriers predicted by the Bell model (concept illustrated in Figure 3-8(a), and discussed in Chapter 2), since for this protein domain $N_{cr} = 2.41 > N = 2$. This rupture force is in close agreement with the experimental value of 108 pN. After rupture of cluster I, the contour length of the protein increases by the amount of 6.6 Å (corresponding to the free chain length exposed due to the rupture of 2 H-bonds). More load can be sustained by the strongest cluster in the protein consisting of 6 H-bonds, hence the force continues to increase on the shifted WLC curve (due to the increased contour length) until the rupture of cluster II, which leads to complete unraveling.

The key question to address is, how much additional force can cluster II (6 H-bonds) resist at the point of rupture, and how this compares with cluster I (2 H-

bonds). The discussion of this issue illustrates the controversy associated with the current understanding. If one takes the same Bell formulation as done previously for 2 H-bonds (for which a good agreement was observed between theory and experiment), rupture should occur at 471 pN (see Figure 3-8(b)). However, the experimentally observed value is between 190-220 pN (see Figure 3-8(b-c)), significantly lower than this prediction, leading to a controversy in the interpretation of this phenomenon. According to the current understanding, this inconsistency can only be addressed by empirically selecting a second set of model parameters that describe the breaking of cluster II.

The issue is resolved with the realization that the rupture process for cluster II is governed by a different mechanism and must be described by the free energy release criterion and not the Bell model. The BSSM predicts that the maximum number of bonds that can break simultaneously for this protein domain is $N_{cr} = 2.41$, suggesting that the rupture force should saturate beyond 3 H-bonds to a value of approximately 156 pN. This finding explains the lower than expected rupture strength of cluster II, and why the Bell model is not capable of predicting the strength of cluster II.

The theoretical framework developed here applies for near-equilibrium pulling rates where the constant force assumption of the energy balance criterion is valid. Although AFM experiments are carried out at relatively slow deformation rates, non-equilibrium processes may still be significant. It is observed that the experimental force peak of 190-220 pN for the second cluster turns out to be higher than the BSSM prediction of 156 pN. It has been established that peak forces in experiments are highly rate dependent [75, 165, 219, 2]. The instantaneous loading rates in the experiments after the rupture of the first cluster may be significantly higher than equilibrium conditions, and may provide an explanation for this discrepancy. Furthermore, the flat transition predicted by BSSM, as a result of the assumption that the force remains constant during H-bond rupture, has also not been observed by this experiment. Rather, force seems to ramp up slightly at this point, indicating that indeed the equilibrium condition has not been reached. Future AFM experiments at slower pulling rates and higher resolution may provide better validation for the theoretical

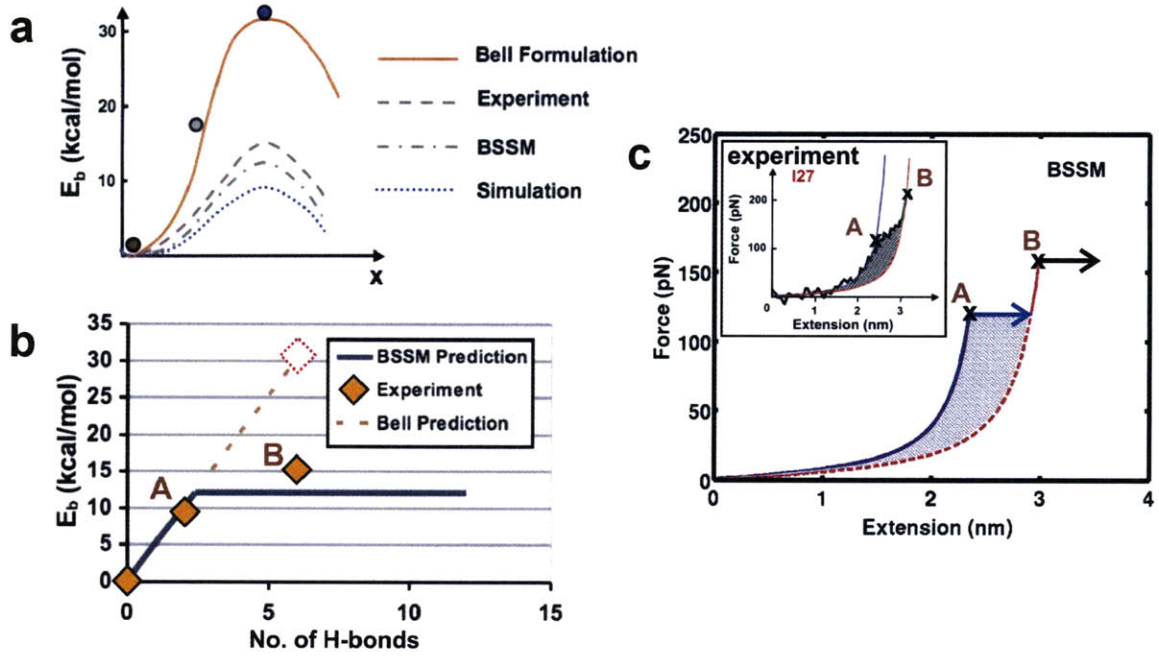

Figure 3-8: Comparison of the height of the overall energy barrier as a function of the number of H bonds in a beta strand, N. Calculation of the height of the effective energy barrier for six H bonds loaded in shear, based on experimental data, Bell theory, BSSM, and MD simulation results are shown in panel (a). Panel (b) shows the rupture strength of a beta-sheet (point of deviation from the WLC model), as a function of the number of interstrand H bonds, N. While the Bell model (homogeneous shear assumption) predicts a continuous increase in force with increasing cluster size, BSSM predicts saturation after three H-bonds in agreement with experimental data from I27 (geometry shown in Figure 3-7(b)). With a minimal number of constant parameters (persistence length of the polypeptide ξ_p and H-bond dissociation energy E_{hb}) and structural information (H-bond cluster size N), BSSM is able to predict the strength of different domains of a beta protein during an unfolding experiment. The dotted circle predicts the strength according to the Bell model; apparently it is much too large compared with the experimental observation. Panel (c) illustrates the force-extension behavior of I27 under external loading, experimental data [163] (inset) and direct comparison with elastic behavior predicted by BSSM theory (blue and red curves, rupture marked by a cross). Extension and relaxation processes are described by the WLC equation (eq. (3.4)), and the fracture force is predicted by Bell theory (eq. (3.19)) or for larger clusters by the Griffith energy balance condition (eq. (3.21)).We note that the length scales in the experimental results (shown in the inset of (c)) have been normalized by the number of tandem repeats stretched experimentally, to obtain the extension length scales corresponding to rupture of a single domain. The shaded area corresponds to the energy dissipated by the rupture of two H bonds. Note the indication of points A and B in this plot and in panel (b), for the geometry shown in Figure 3-7(b).

predictions presented here. Despite the disagreement, this simple model describes the overall force-extension behavior well.

Figure 3-8(a) summarizes the energy barrier predictions for the rupture of cluster II based on BSSM predictions, experimental values, as well as simulation results for a model three strand beta-sheet system [128, 131, 129]. It should be noted here that according to BSSM model, the energy dissipated by cluster rupture (that is, by overcoming of the energy barrier) depends on the force level and the change in contour length of the system rather than individually on initial and final states, in agreement with experiments [194, 172].

This comparison confirms that BSSM is capable of explaining key events in unfolding of proteins. The only input parameters are the H-bond dissociation energy and persistence length, which show limited variability for proteins [194, 195, 80, 172, 154, 173].

3.4.2 Comparison with proteomics data

Since the number of H-bonds in a beta-strand is proportional to the number of residues ($L_{cr} \sim N_{cr}$, assuming that the H-bonds are arranged in a linear geometry as shown in Figure 3-9 (a)), numerical prediction for N_{cr} leads to a critical geometric strand length L_{cr}: Strands beyond L_{cr} are prone to localization of deformation and rupture of H-bonds that reduces their efficacy, since not all of the H-bonds participate in the rupture process and thus contribute to the strength (Figure 3-9(a)). The physical basis for this limitation is the entropic elasticity of the protein backbone (an intrinsic property of protein structures) as well as the characteristic energy of H-bonds (also an intrinsic property of protein structures).

Notably, this finding explains recent proteomics data [180] obtained by analyzing the strand lengths of a wide range of beta-sheet structures. Figure 3-9(b) illustrates that shorter beta-strands are more prevalent [180]; strands that employ less than five residues are most common, and the prevalence decays sharply after this point.

Further, it is found that the shear strength correlates closely to the prevalence of the strand length (see Figure 3-9(b)). The key conclusion of this analysis is that the

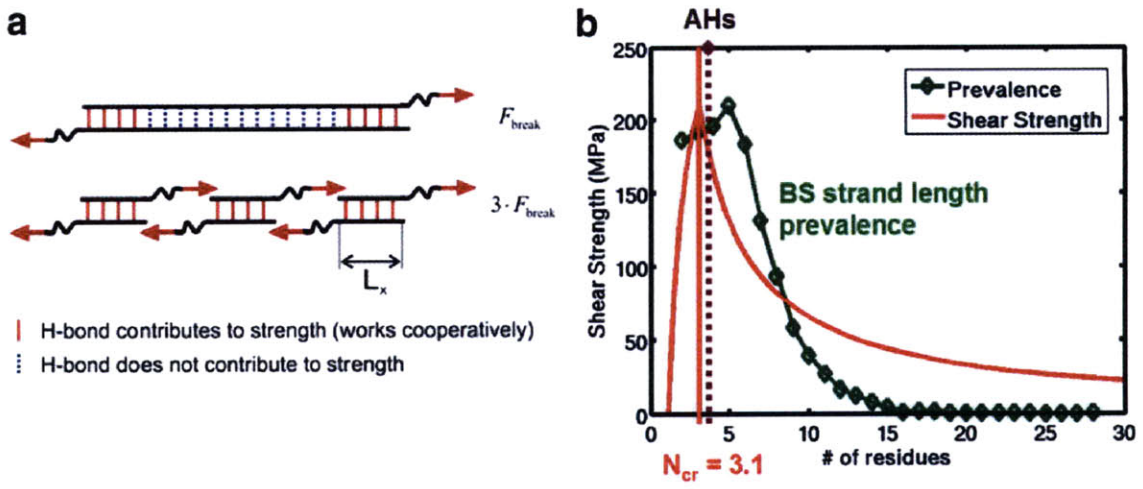

Figure 3-9: Size effects, shear strength and prevalence of strand length of beta-sheets. Panel (a): This plot illustrates the difference of the strength of a single, long beta strand vs. a combination of multiple small strands. In the upper plot, only H-bonds at the boundary are being stretched and contribute to the strength. In the lower plot, all H-bonds throughout the entire structure contribute to the strength, making the overall structure three times stronger. Panel (b): Shear strength of beta-sheets as a function of strand length, and prevalence of beta-sheet strand length as reported in [180]. The highest shear resistance is found at a characteristic length scale of 3.1 residues. Beyond this length scale, the shear strength drops rapidly. The plot of the prevalence over the strand length illustrates that shorter beta-strands are more prevalent [180]; in particular, strands that employ less than five residues are most common, and the prevalence decays sharply after this point. The dotted line marks the number of H-bonds in alpha-helical protein domains, where each convolution features 3.6 residues per turn.

evolutionary driving force for the selection of strand lengths in beta-sheet structures may be the maximization of the mechanical (and hence, thermodynamical) stability. This hypothesis is strongly supported by the fact that both the prevalence curve and the shear strength show a similar behavior (Figure 3-9(b)).

The intrinsic strength limit presented here applies to individual protein domains and can only be overcome by using structural hierarchies. Indeed, such hierarchical structures are commonly found in many mechanically strong biological materials such as spider silk and muscle tissue.

Moreover, the theoretical development reported here is not limited to beta-strands. As long as the system of interest contains an assembly of H-bonds that are loaded uniformly, the theoretical derivation reported here is valid. Therefore, this finding also explains the characteristic structure of alpha-helical proteins, which feature 3.6

H-bonds (assuming 1 residue $\approx$ 1 H-bond) per turn, which are loaded in parallel and that break concertedly (as recently shown in [1]). The characteristic number of H-bonds per turn in AHs closely resembles the strength limit predicted by the fracture theory. In other words, increasing the number of H-bonds per turn does not have an effect in increasing the mechanical stability of the protein. Similarly, beta-helix protein structures also feature a helical arrangement of ultra-short beta-strand segments with each less than 5 H-bonds. Figure 3-10 compares characteristic length scales observed in common protein secondary structures and the theoretical prediction presented here.

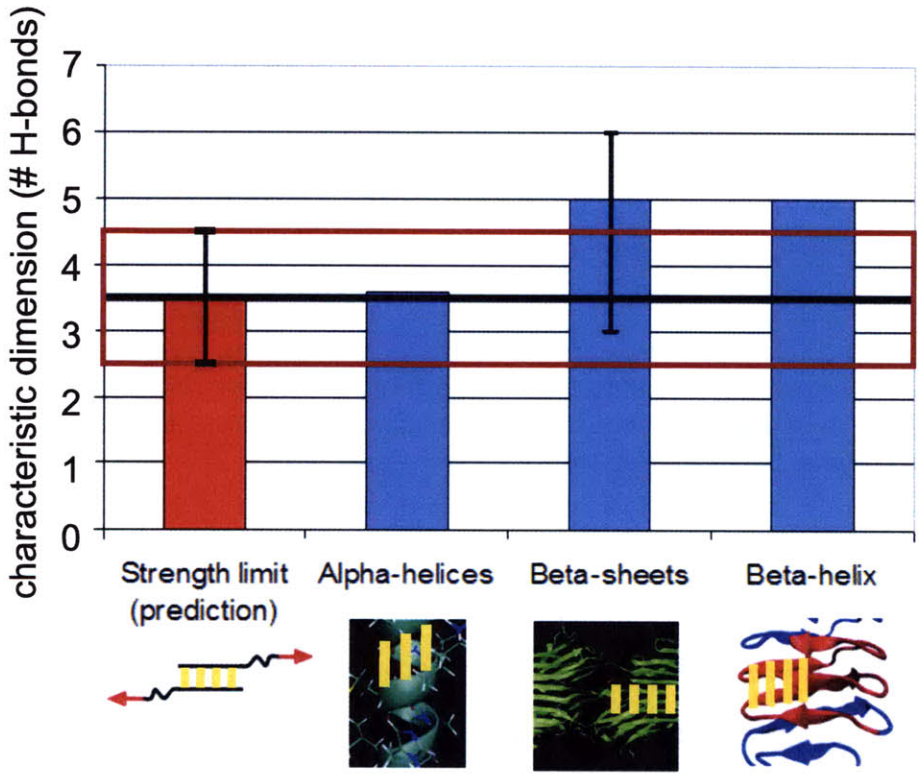

Figure 3-10: Characteristic dimensions of common protein secondary structures. This figure compares the characteristic dimensions of alpha-helices, beta-sheets and beta-helices to the strength limit. Since the theoretical derivation considers uniform deformation of H-bonds with no particular specificity to geometry, it may also apply to other protein structures where geometric confinement leads to higher mechanical stability. The fact that 3.6 H-Bonds per convolution exists on alpha-helices and beta-sheets on the sides of helices occur in clusters of 4 may be indicative of such a biological concept.

The fact that the results of this model only depend on fundamental properties of protein structures underlines the significance of the findings. The only input parame-

ters in this model are the persistence length ξ_p of the polypeptide chain, as well as the dissociation energy of a H-bond, E^0_{HB} (γ_s is calculated directly from E^0_{HB}). Both parameters can be determined reliably from either experiment or atomistic simulation. The parameter E^0_{HB} typically depends on the solvent, which influences the strength of H-bonds in the system [66]. Variation of the bond strength and predicted rupture strength can be estimated from Figure 3-11(a). Figure 3-11(b-c) suggests that the value of N_{cr} does not change significantly with variation in the parameters ξ_p, E^0_{HB} and τ. The main conclusion of this analysis is summarized in Figure 3-11(d), which is related to the biological material design concept outlined in this section. As a result of utilizing weak bonds and flexible chains that behave entropically (ξ_p in the order of a nanometer), cooperative rupture behavior at high force can be achieved for biopolymeric systems. Increasing bond strength naturally increases material strength, but reduces cooperation, leading to brittle behavior where bonds behave individually.

3.5 Discussion and conclusions

A variety of models for the fracture mechanics of ceramics and metals have been reported over the past decades, involving detailed descriptions of dislocation and crack extension mechanisms. However, similar advances for biological protein materials have thus far remained elusive. The results reviewed in this section for the first time describe a rigorous fracture mechanics approach to describe the fundamental bond rupture events in protein materials. In analogy to dislocation nucleation and propagation in ductile metals, the breaking of H-bonds represents a fundamental unit mechanism of materials failure.

It has been shown that the rupture strength of H-bond assemblies is governed by geometric confinement effects, suggesting that clusters of at most 3-4 H-bonds break concurrently, even under uniform shear loading of a much larger number of H-bonds. This universally valid result leads to an intrinsic strength limitation that suggests that shorter strands with less H-bonds achieve the highest shear strength.

As reported in Figure 3-4, the prediction close to 127 pN is in excellent agreement

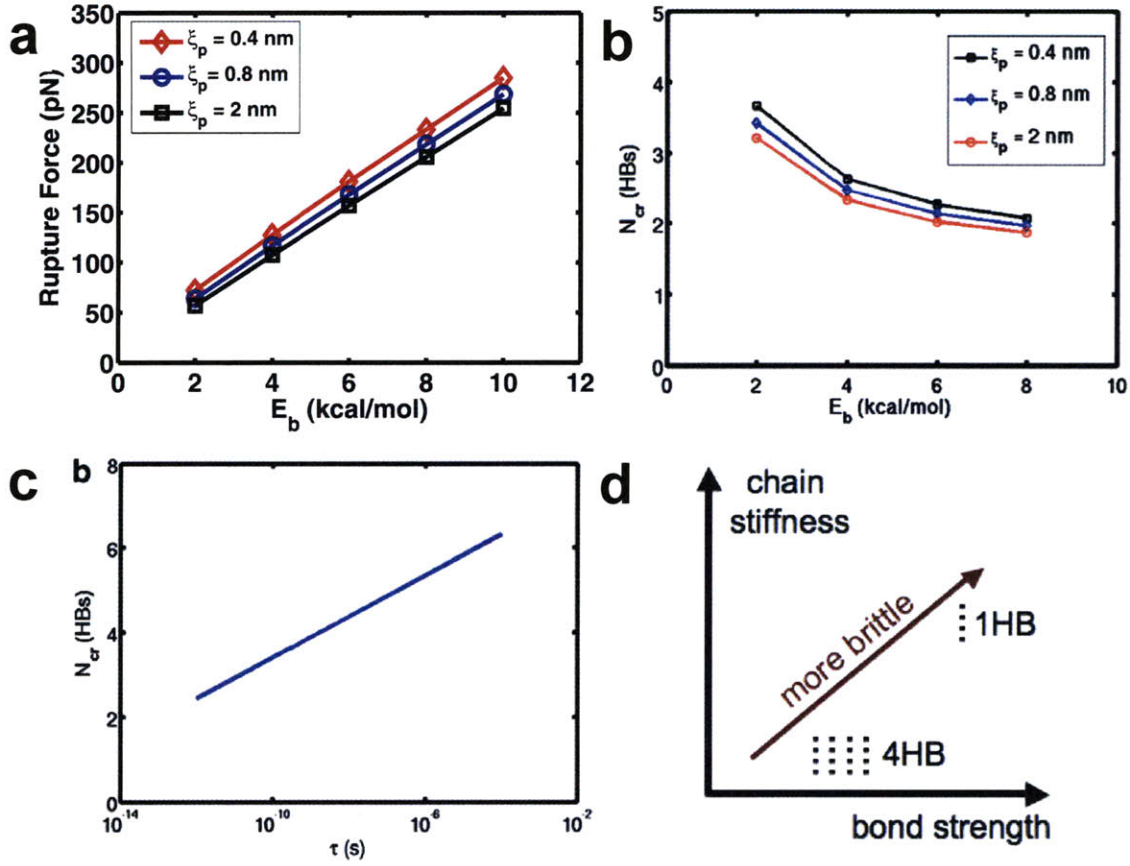

Figure 3-11: Sensitivity analysis for the critical number of H-bonds, N_{cr} (adapted from [129]). This plot illustrates the sensitivity of N_{cr} due to variations of the key physical parameters used in the length-scale derivation. This analysis illustrates the effect of uncertainties in these parameters on N_{cr}. Panel (a): N_{cr} is plotted as a function of E_b and ξ_p and it is evident that $N_{cr} \approx 3 \pm 1$ H-bonds over the domain of a range of measured values for variations in the H-bond energy and for variations in the persistence length. Panel (b): There is also a weak logarithmic dependence of N_{cr} on the characteristic time of H-bond rupture, τ. The critical number of H-bonds $N_{cr} \approx 4 \pm 2$ H-bonds over eight orders of magnitude of time-scale (here $E_b = 2.83$ kcal/mol and $\xi_p = 0.4$ nm), suggesting that thermal contributions are less important due to the high energy barrier. These results illustrate the significance and universality of this finding and support that the results obtained are robust. Panel (d) summarizes the key concept that emerges from this analysis, weak bonds and flexible chains lead to cooperative failure at high strength.

to a wide range of experimental data, and may explain the convergence to a finite limiting value at small deformation speeds. The size dependence analysis becomes of particular interest when compared with recent experimental proteomics data, suggesting a correlation between the shear strength and the prevalence of beta-strand lengths in biology (see Figure 3-9(b) and Figure 3-10). Shorter strands are more common in nature, and they exhibit higher mechanical (and hence thermodynamical) stability according to this prediction. This hypothesis is confirmed by direct large-scale full-atomistic MD simulation studies of beta-sheet structures in explicit solvent. It is also in agreement with experimental results that show a non-linear, diminishing increase of strength with increasing H-bond cluster size [163]. The finding suggests that the intrinsic strength limitation of H-bonds is overcome by the formation of a nanocomposite structure of H-bond clusters, thereby enabling the formation larger, much stronger beta-sheet structures. Such hierarchical features are commonly observed in strong materials such as spider silk and muscle fibers. The results could have interesting implications for the design mechanically robust, strong biomimetic or peptide based self-assembled materials. The model further confirms that fracture mechanics concepts, previously primarily applied to macroscale fracture phenomena, can also be directly applied at nanoscale, and can be used to describe failure mechanisms in protein materials. Similar approaches can be used to lay the foundation for the development of plasticity and strength models of biological protein materials.

Chapter 4

Nanomechanics of beta-solenoid structures

This chapter is focused on the nanomechanical properties of β-helices, where MD technique is employed to study tensile and buckling behavior of these protein nanotube type structures. A snapshot of the molecular structure of this protein material is shown in Figure 4-1. The β-helix is a protein structure that has been discovered relatively recently [254]. This protein structure represents a helical nanotube that consists of beta-sheet walls. The beta-helix is considered to belong to a novel, different type of secondary structure class called beta-solenoids. Beta-sheets and beta-solenoids make up all beta-structured protein motifs discovered so far, and share similar chemical, structural and mechanical properties.

Due to its recent discovery, this protein structure has not been studied as extensively as α-helices or β-sheets or many other protein structures, and therefore, there exists virtually no understanding of its mechanical behavior. Other than the work presented here, no experimental or computational studies of the mechanics of this protein structure have been reported up to date.

As reported in Ref. [130, 133], MD simulations that employ a fully atomistic model with an explicit solvent, and very slow pulling rates over extraordinarily long time-scales reaching the fraction of a micro-second (250 ns) are carried out to simulate the tensile deformation of this motif. This is followed by a series of compressive

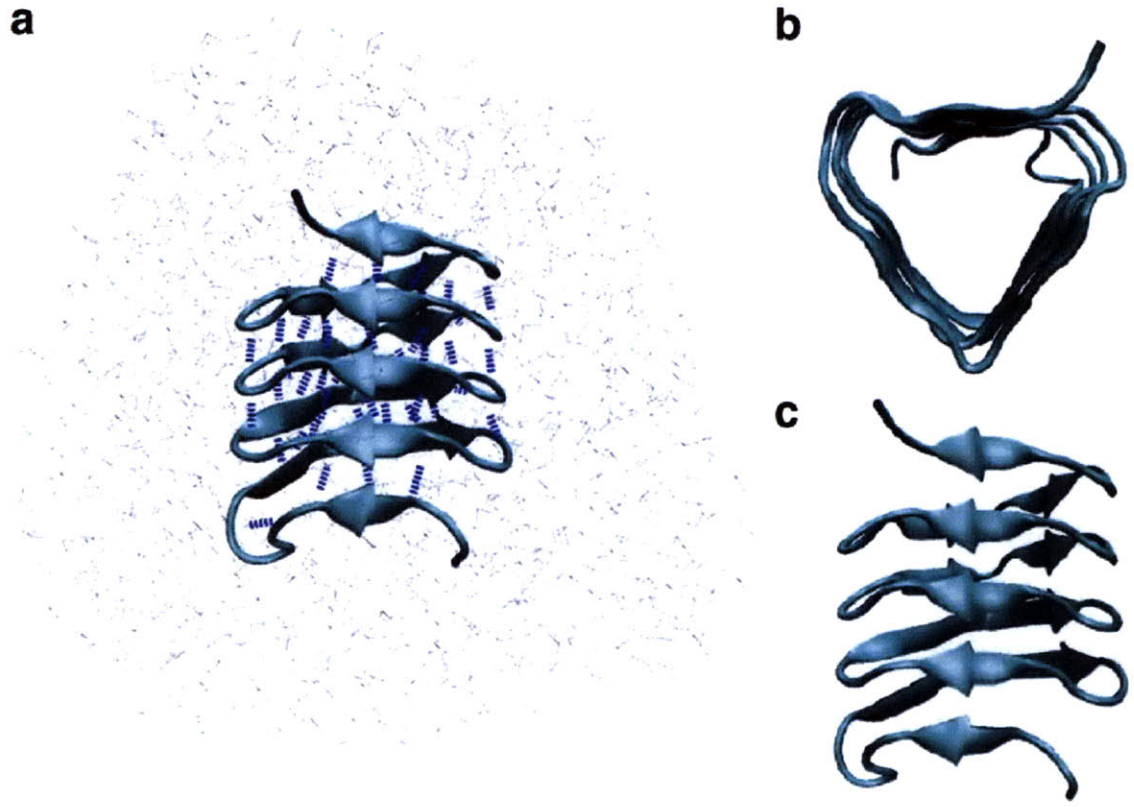

Figure 4-1: Full atomistic model (panel a): The structural coordinates of the single-beta-helix protein are obtained from the Protein Data Bank (PDB ID 1fwy) and solvated with a skin of water. Hydrogen bonds that are oriented parallel to the helix axis and perpendicular to the strands enable this system to be stable in this conformation. The beta-helix is a nanotube-like structure (panel b) with a triangular cross-section formed by three strands on the edges. This particular protein structure is classified as a single-stranded left handed beta-helix, consisting of a single polypeptide chain that creates a helix with approximately four to five beta-strands per side (panel c).

loading molecular dynamics simulations of the single and triple beta-helix structures, where it will be illustrated that this protein motif can withstand extremely large compressive loads, far exceeding the tensile strength. Size-dependent failure modes under compressive loading are summarized in a deformation map, and rate dependence of failure is discussed. These findings illustrate the potential of the beta-helix protein motif as an inspiration for nano-scale materials applications, ranging from stiff nanotubes to self-assembling peptide based fibers inspired by amyloids.

There is very little known about the structure-property-function link of beta-solenoids. One of the main goals of the analysis presented in this chapter is to illustrate the applicability of the simulation approaches to a recently discovered structure, for which little or no experimental data exists. Predictive modeling approaches such as the one presented here may provide valuable insight into structure-function-property relationships of other protein structures and may motivate experimental studies associated with disease aspects or materials science applications of proteins. The analyses presented here elucidate the intriguing mechanical properties of this important self-assembling protein structure, which may be eventually be used for developing exceptionally strong and elastic fibers that can surpass the strength of steel, as in the case of spider silk [247].

4.1 Background on beta-solenoids

Despite being discovered only recently as a protein structure, the β-helix has already gained tremendous interest in the scientific community. This is largely attributed to the belief that the β-helix is the fundamental structural unit of certain kinds of amyloid fibers, in particular those associated with prion proteins [250, 136]. Amyloids are caused by uncontrolled aggregation of proteins in tissue and organs during the course of many different fatal diseases such as Alzheimer's, Huntington's, prion diseases and type 2 diabetes [58]. It is not clearly known whether the amyloid fibrils form as the result of a mechanism to convert more toxic oligomers to larger, less harmful aggregates, or if they are the pathological agents themselves [53]. However, it

is now generally accepted that large amounts of amyloid deposition in tissue may be a cause of organ malfunction. There is also growing evidence suggesting that amyloids play a central role in pathogenesis of at least some of the aforementioned diseases [137]. Investigation of this new protein structure is therefore of great importance for the medical and biophysical community as it may reveal crucial insight into the formation and assembly of these fibers and may provide ideas for novel remedies to diseases affecting millions. Recent experimental studies have revealed the molecular structure of several proteins that contain this motif, as a result, atomistic models for the single-helix and triple-helix beta-solenoids have now been established, enabling all-atom simulations [254, 121].

Amyloids are most widely known for the pathological conditions that they are directly or indirectly associated with. From a materials scientist's perspective, however, amyloid formation is a fascinating materials growth phenomenon that can be put in use for beneficial applications [58]. Engineered self-assembling amyloid fibrils have been used as nano-wire templates, taking advantage of their patterned nanotubular structure [191, 205]. Amyloids are also found in some natural adhesives [167], and in spider silk in silk glands [125, 216], which has attracted some interest into possible structural applications of these materials and their interesting mechanical properties [124, 85]. Notably, amyloids are found to be extremely sturdy protein structures that can often not be disintegrated by the organism itself, suggesting a particularly stable structural arrangement and as a consequence, possibly large resistance against mechanical unfolding. These aspects motivate the studies of the mechanical properties of this protein structure.

Although a single structural motif such as the β-helix may not be an accurate molecular level description for all the different kinds of amyloids discovered so far, it is a model that captures all essential structural and biochemical features common to amyloids and is therefore an excellent representative system to begin a rigorous, nanomechanical characterization of amyloids and similar fibrous materials.

4.2 Single beta-helices in tension and compression

Single beta-helices form when a single polypeptide chain is coiled into a wide helix, formed by beta-strands separated by loop regions of variable length. Each side of the helix is an anti-parallel beta-sheet structure and helices generally contain two or three-sheets (sides). For the triple-sheet beta-helices investigated here, no specific sequence pattern has been identified so far [20].

All-atom MD simulations are performed using segments from the crystal structure of a protein obtained from Protein Data Bank [14] (PDB ID 1fwy). The segments used for the simulations form the characteristic beta-helix structure. For investigating mechanics of beta-helix structures, only the C-terminal domain of this protein is taken, which adopts a left-handed parallel beta-helix motif (residues 250-328).

The force displacement profiles obtained from simulations of the beta-helix model reveal that the mechanical behavior of this molecule is different from a Hookean spring. An important feature is that the mechanical response is quite asymmetric with respect to tension and compression, as shown in Figure 4-2(a-b). The maximum force observed in compression happens at approximately 17% engineering strain and is six times larger than the peak force observed in tension at approximately 800% strain (Figure 4-2(c)). Up until 17% strain, both systems behave more or less linearly, allowing comparison of stiffness values in small deformation regime. The elastic moduli computed using both hollow and solid cross-section assumptions show that the β-helix in compression is approximately 18 times stiffer than in tension, as shown in Figure 4-2(d).

The peaks, rises and valleys in the complex non-linear force response curve can only be understood by studying the details of the individual bond rupture events. Most of the helix loops uncoil with hydrogen bonds breaking more or less sequentially. Comparatively long time scales of the simulation and the coupling between tension and torsion in the helix allow the molecule to reorient its strands during the course of the simulation. In the later stages of unfolding, a sheet with five hydrogen bonds aligns itself parallel to the loading direction, as opposed to the general perpendicular

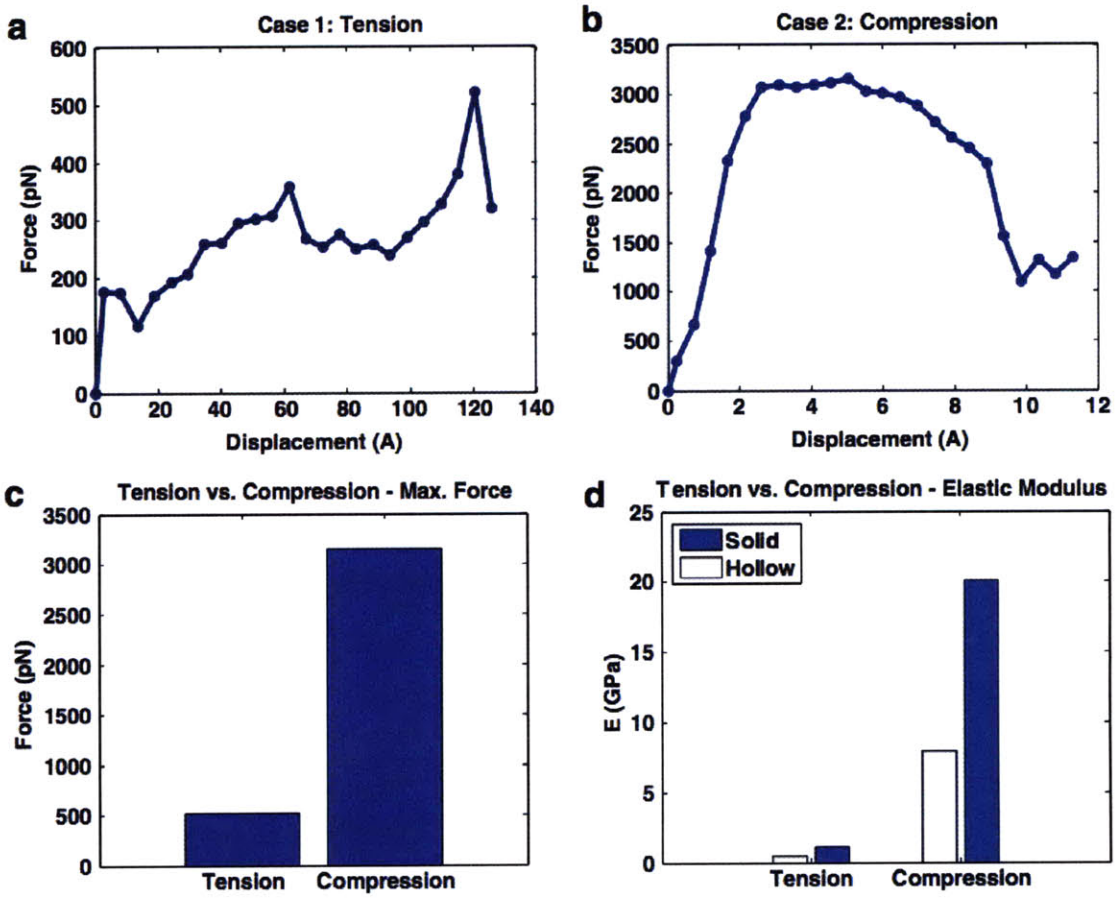

Figure 4-2: Tensile vs. compressive strength (panels a and b). The mechanical behavior of the beta-helix is rather different from a Hookean spring, since the response is strongly asymmetric in tension and compression. The maximum force observed in compression is at $\approx$ 17 % engineering strain, and is six times larger than the peak force observed in tension at $\approx$ 800% strain (panel c). Up to $\approx$ 17% strain, both systems behave more or less linearly, allowing comparison of stiffness values in small deformation regime. The elastic moduli computed using both hollow and solid cross-section assumptions show that the beta-helix is approximately 18 times stiffer in compression than in tension (panel d).

orientation of the strands. In this case, five hydrogen bonds in this sheet deform homogeneously under shear, cooperatively resisting further stretching like a hook-up point. Simultaneous rupture of these bonds corresponds to the force peak observed at approximately 120 Å pulling displacement. Simultaneous vs. sequential rupture events are illustrated in Figure 4-3.

The force vs. displacement profile of the β-helix reveals distinct deformation mechanisms under tensile loading, which are marked in Figure 4-4(a). Here, the same averaging scheme is used, with a different number of averaged points (bins) to illustrate different levels of detail in force response curve (different curves in Figure 4-4(a)) to enable a more careful analysis of the force-extension curve. The first peak (A) corresponds to the unfolding of the first loop at the fixed end. Unfolding of this loop relaxes the structure, resulting in a decrease in force (B). The regime that follows this is characterized by steady uncoiling of loops at the far end (C). The peak at approximately 60 Å pulling extension corresponds to the unfolding of the first loop at the near end, namely the N-terminal of the protein (D). Following this, the molecule relaxes again and steadily uncoils at a lower force value (E). The force begins to rise again due to reorientation of the helix to load hydrogen bonds in parallel and reaches a peak force value of approximately 600 pN (F), when approximately 5 hydrogen bonds on a single side break simultaneously.

The direct correlation between the number of bonds that break and the mechanical response can be quantified by plotting the number of ruptured bonds as a function of extension ($N_{HB}(x)$), as shown in Figure 4-4(b). To more rigorously analyze the bond rupture events as a function of deformation, a 9th order polynomial function is fitted to the bond break history data. The fit is depicted in Figure 4-4(c). The first derivative of this function, plotted in Figure 4-4(d) relates linearly to the rate of bond rupture events since the protein is pulled at a constant pulling velocity (thus $N_{HB}(t) \sim N_{HB}(x)$ since $x \sim t$ at constant $\dot{x}$).

The valleys correspond to periods where the backbone is being stretched out without any or with few hydrogen bonds breaking, and the peaks indicate simultaneous rupture of a larger number of hydrogen bonds. This can be confirmed by comparing

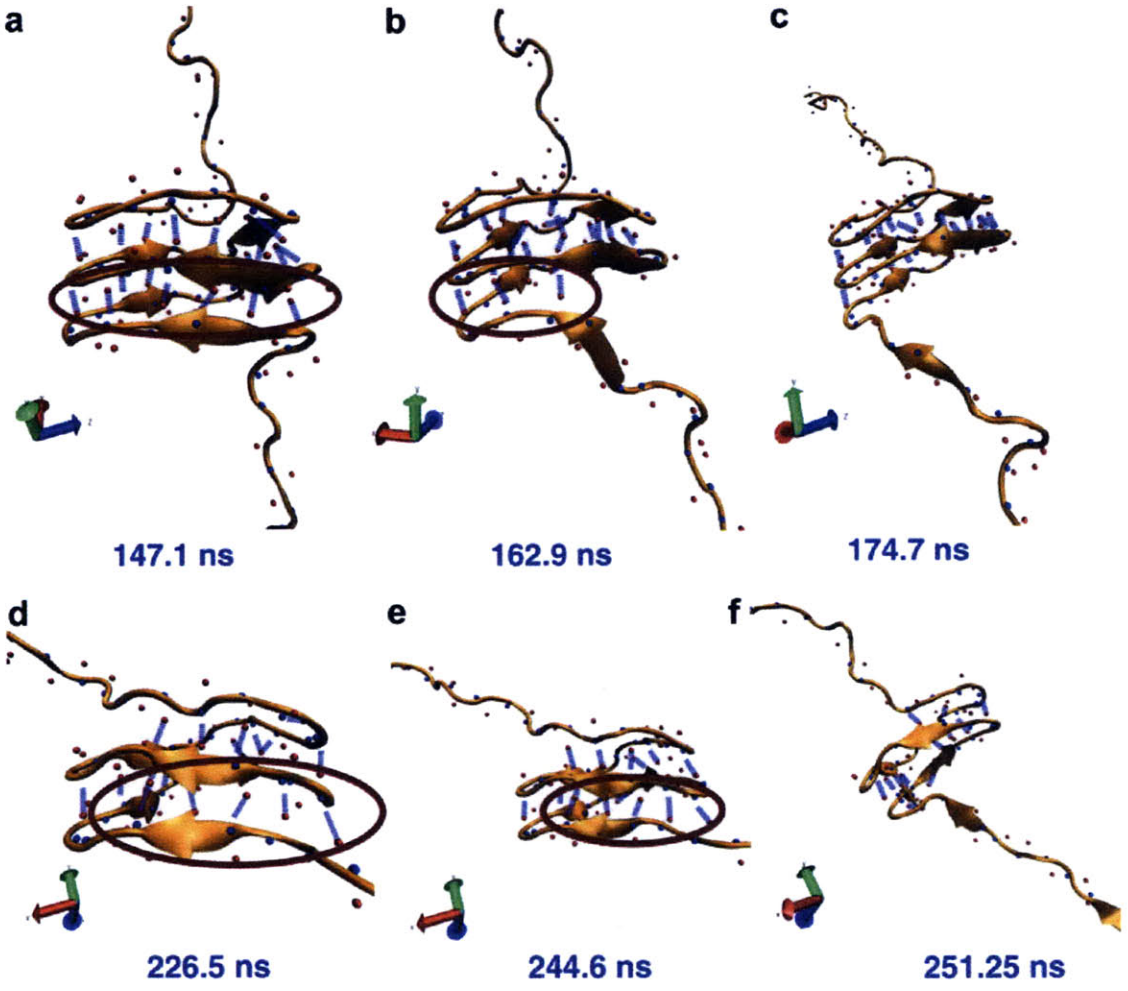

Figure 4-3: Tensile deformation - sequential vs. simultaneous rupture. The peaks, rises and valleys in the complex, non-linear force response curve can be better interpreted by studying individual bond rupture events. While most of the helix loops uncoil with hydrogen bonds breaking more or less sequentially (panels a-c), comparatively long time-scales of the simulation and the coupling between tension and torsion in the helix allow the molecule to reorient its strands during the course of the simulation. At this point, five hydrogen bonds in a sheet deform rather homogeneously under shear, cooperatively resisting further stretching like a strong traction point (panels d-f). Simultaneous rupture of these bonds corresponds to the largest force peak obtained over the course of the 250 ns time span of the simulation.

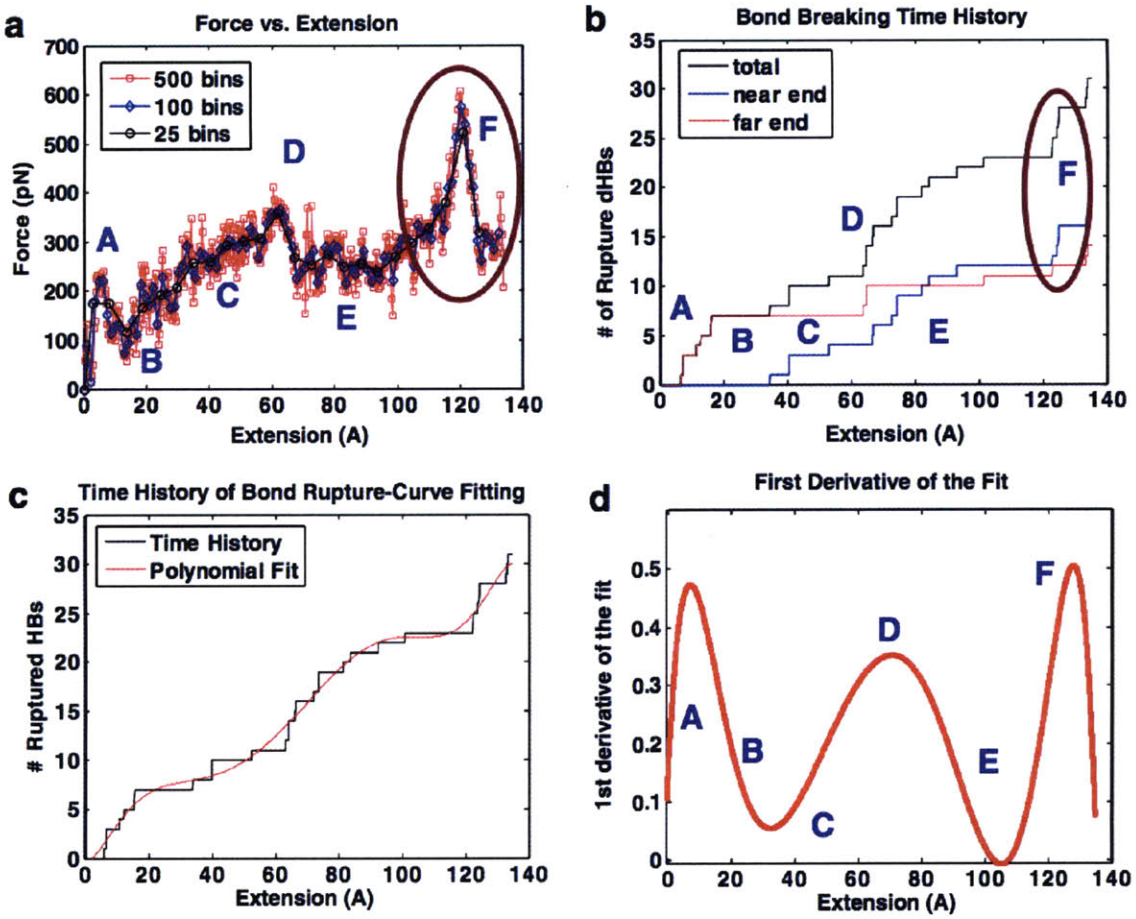

Figure 4-4: Relationship between bond rupture events and mechanical signature. The force vs. displacement profile of the beta-helix reveals distinct deformation mechanisms under tensile loading (panel a). Here, the same averaging scheme is used with different number of bins to illustrate the differences in level of detail associated with this process. The first peak in force-displacement curve (A) corresponds to the unfolding of the first loop at the fixed end. Unfolding of this loop relaxes the structure, resulting in the low force point (B). The regime that follows this is characterized by steady uncoiling of loops at the far end, with no more than 1-2 hydrogen bonds breaking at the same time (C). Peak force at $\approx$ 60 Å extension corresponds to the unfolding of the first near end loop (D). Following this, the molecule relaxes again and steadily uncoils with lower resistance (E). The force begins to increase again when the helix is reoriented and hydrogen bonds are loaded in parallel, and reaches a crest with a force value of $\approx$ 600 pN (F) when five hydrogen bonds on a single side break simultaneously. The direct correlation between the number of bonds that break and the mechanical response can be quantified by plotting the number of ruptured bonds as a function of extension (panel b). To further analyze the bond rupture events as a function of time, a ninth order polynomial function is fitted onto the bond data (panel c). The slope of this function relates to the speed of bond breaking events. The troughs correspond to periods where the backbone is being stretched out without any hydrogen bonds breaking, and the peaks indicate simultaneous rupture of a large number of bonds (panel d).

Figure 4-4(d) with Figure 4-4(a). It is apparent that each peak in the force-extension curve corresponds to a peak in the slope of the H-bond rupture history $N_{HB}(t)$, suggesting that the instantaneous strength of the beta-helix is proportional to the H-bond breaking rate,

$$F(t) \sim \frac{\partial N_{HB}(t)}{\partial t}. \tag{4.1}$$

This result may be useful for the development of constitutive relations or strength theories of protein structures.

In compression simulations, it is observed that failure occurs by sliding out of an entire single strand, corresponding to a local fracture scenario at a critical compressive forceP_{fr} = 3,154 pN.

In order to gain understanding of the mechanical behavior at larger length scales, a simplistic continuum model for the compressive strength of the beta-helix is proposed here. Schematic views, boundary conditions and the dimensions of the simplified hollow triangular section are shown in Figure 4-5. It is assumed that localized failure as observed in MD simulations competes with Euler buckling of the structure. Using the simplified model illustrated in Figure 4-5, the Euler buckling load [235] for this specific geometry is calculated as

$$P_{cr,0} = \frac{\pi^2 EI}{4L_0^2}, \tag{4.2}$$

where L_0 is the original length of the helix and the factor of four in the denominator derives from fixed-free boundary conditions. For the length L_0 used in the simulations, the Euler buckling load is $P_{cr,0}$ = 6,474 pN. This value is larger than the maximum compressive force value observed from simulation, denoted P_{fr}=3,154 pN, which is observed for a localized shear failure of the nanotube. Thus it is expected that the dominating failure mode for the geometry used in the MD simulations is not Euler buckling. This prediction is in good agreement with the MD simulation results that show that local failure is the governing mechanism. However, as the length of the molecule is increased, it is expected that Euler buckling, rather than localized failure,

will be the dominating deformation mechanism.

Using the peak force from SMD as an estimate for the maximum strength of the helix for a local fracture scenario, the critical molecular length beyond which Euler buckling [235] governs is calculated using this value as the critical buckling load: $L_{cr} = \frac{\pi}{2}\sqrt{\frac{EI}{P_{fr}}} \approx 2.15$ nm.

Based on these results, the compressive strength of the beta-helix as a function of its length can be plotted, as shown in Figure 4-6(a). The analysis reveals that buckling governs at length scales exceeding only a few nanometers. The resulting compressive strength of the helix is determined by the governing mechanism, given by

$$P_{\max}(L) = \min[\, P_{fr}, P_{cr}(L)\,] \tag{4.3}$$

where P_{fr} is the maximum force observed in local fracture deformation mode and $P_{cr}(L)$ is the buckling load as a function of length. As expected, strength diminishes with increasing length. It should be noted that note that this simple calculation is based on the assumption that the local failure strength P_{fr} is independent of the molecular length. The persistence length [17] of this structure, ξ_p, can also be estimated as

$$\xi_p = \frac{EI}{k_B T}, \tag{4.4}$$

where k_B is Boltzmann's constant, and T is temperature, given as 300 K, and EI is determined from material and structural properties determined from all-atom model and simulation results. At length scales exceeding ξ_p, even thermal vibrations could cause significant deformation in the molecule, so the behavior is like an ultra-thin wire. For the beta-helix structure, this length scale is found to be in the order of a few micrometers (ξ_p= 1.425 μm). The length scales estimated here can be utilized to construct a simple deformation map, indicating regions of different mechanical behavior as a function of the protein nanotube's length, as presented in Figure 4-6(b).

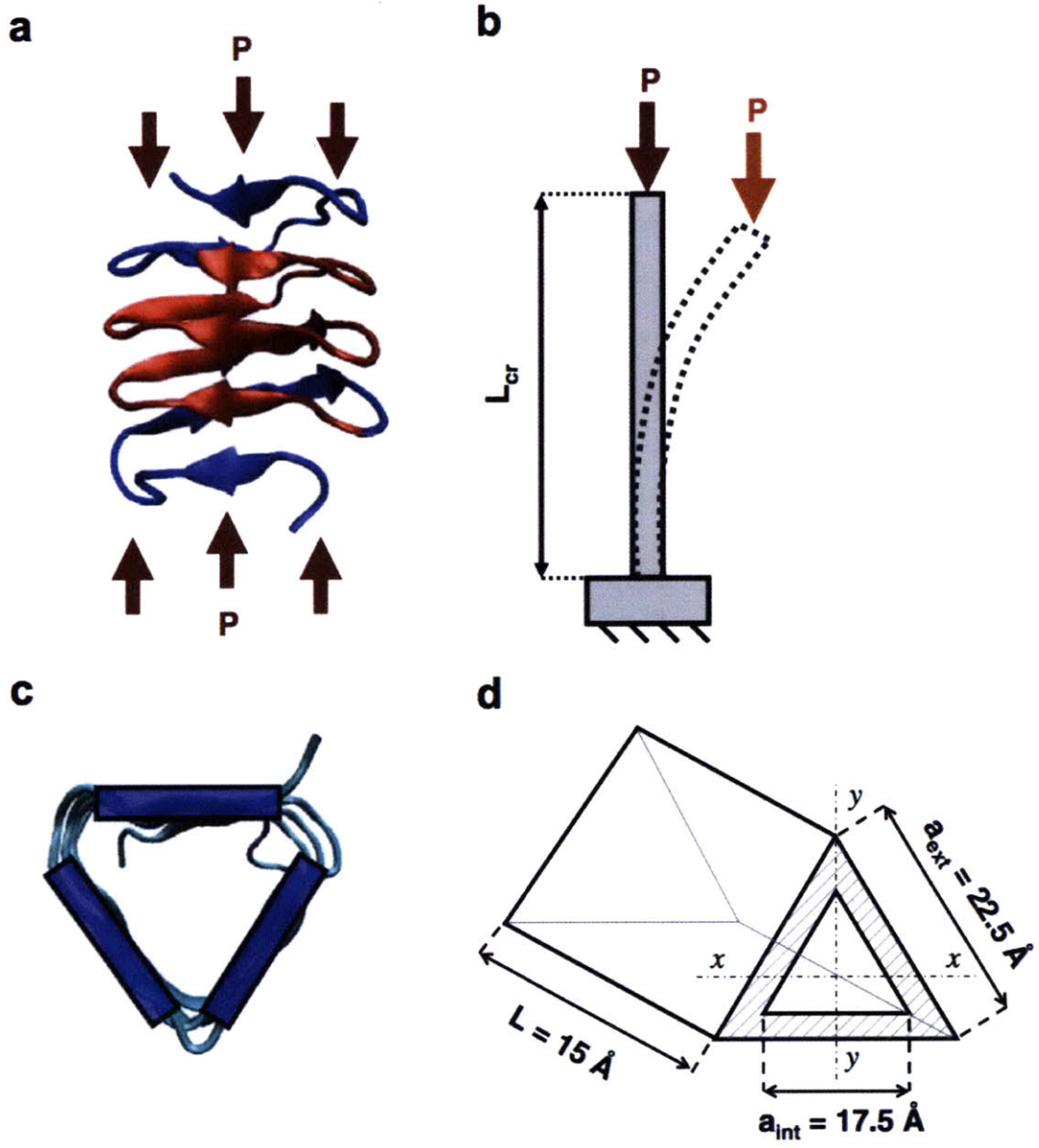

Figure 4-5: Schematics of the compression studies in an engineering analysis. The molecular structure is subjected to uniform compression load on each Cα atom on the first loop at the N-terminal (panel a). All of the Cα atoms are fixed in the last loop, namely C-terminal of the domain. The boundary conditions (panel b) are idealized based on the SMD setup and are used for buckling analysis. The cross-section of the beta-helix is triangular, with a strand on each side (panel c). The dimensions of the helix can be approximated from VMD to build a simplistic model. In this case, the helix is modeled as a hollow triangular section. All calculations are repeated for a solid section assumption as well (panel d).

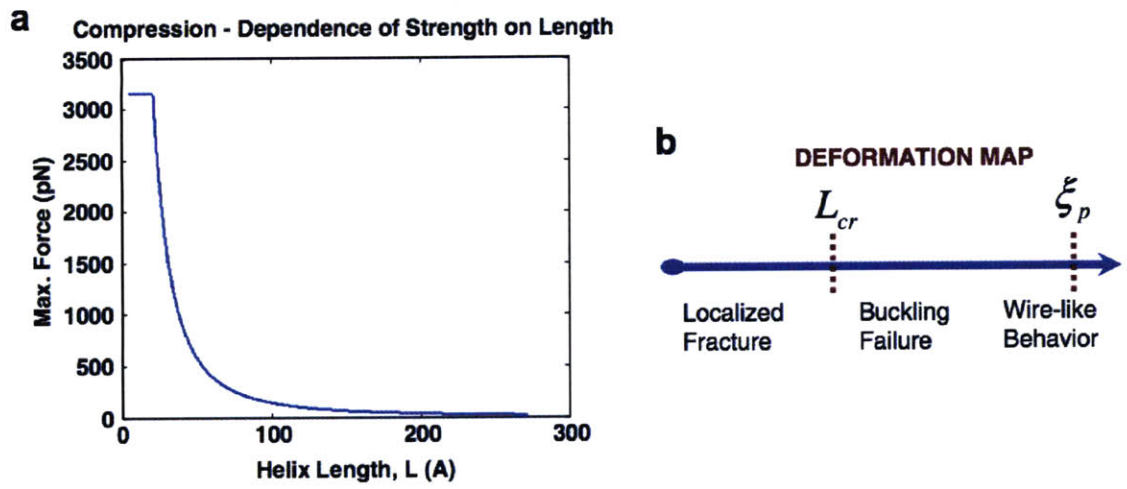

Figure 4-6: Compression analysis of results: in SMD simulations, failure occurs by sliding out of a single strand, corresponding to a local fracture scenario. The max force value observed for this case is lower than the Euler buckling load, further supporting that local failure is the governing mechanism. The maximum force observed from SMD is used as an approximation for the fracture strength for the 'short column', and the critical length beyond which buckling governs is calculated based on this value. Using these results, the strength of the helix as a function of its length is plotted (panel a). The persistence ξ_p length can also be determined, identifying the length scale at which thermal vibrations can cause significant deformation in the molecule. These results are summarized in a deformation map (panel b), indicating different regions of mechanical behavior as a function of length. Very short helices are sturdy and fail locally, longer helices are governed by buckling effects, and extremely long helices behave like wires with negligible capacity for sustaining compressive loads.

From SMD simulations and continuum studies, one can conclude that in compression; very short helices below a few nanometer lengths are sturdy and fail locally. Longer helices are governed by buckling effects, and extremely long helices beyond a molecular length on the order of a micrometer behave like wires with negligible capacity for sustaining compressive loads. This suggests that the nanotube length must be limited to a critical length scale on the order of a few nanometers in order to provide most optimal resistance to compressive deformation. It is noted that a similar transition of the behavior has been observed for carbon nanotubes, albeit at different values of the critical length scales [43].

MD simulations reveal that the beta-helix nanotube structure is extremely extensible and can sustain tensile deformation up to $\approx$ 800 % without rupture of the covalently bonded protein backbone. As the β-helix structures are similar to β-spirals

found in extensible domains found in elastin and spider capture silk [239, 11], simulations illustrate that extension over tens of nanometers and rupture of dozens of hydrogen bonds during deformation may be the molecular basis of the extraordinary toughness of these materials, in line with previous hypotheses [217], as the protein structure deforms while dissipating a significant amount of energy.

Recent experimental and MD studies have revealed that proteins with a shear topology have been found to provide exceptional resistance to forced unfolding due concurrent breaking of multiple hydrogen bonds upon loading parallel to strand orientation [198, 251, 23, 94, 72]. The analysis shown in Figure 4-4(d), further validates this claim. It is also observed that during the course of an ultra-long time scale MD simulation that extends to fractions of microseconds, significant unfolding and reorientation in the molecule can take place as a result of external force. This observation may provide insight into how beta-sheet rich domains in amyloid-like spider silk in the silk gland may be transformed to a shear topology due to forces experienced in spinning and elongational flow and may illustrate the impact of this reorientation on mechanical strength. The processing of protein fibers to generate such shear topologies could provide a means to enhance the mechanical resistance.

Usually, there is little motivation to study proteins under uniform compressive loads, as most proteins studied mechanically so far have found applications in fibers, materials that are predominantly loaded in tension. Thus there are very few studies in the literature that focus on this mode of deformation. The inspiration for the compressive study reported here derives from the triple beta-helix structure in the gp5 domain in the needle-like cell-puncture device of bacteriophage T4 [121]. Extreme responsiveness under compression shown here is a truly unique feature of the beta-helical topology among protein structures identified so far. The hollow triangular cross section of the beta-helix is thereby very efficient for serving under compression as it can sustain significant loads without buckling, while at the same time providing a highly dissipative, yet reversible mechanical behavior under tensile stretching. The methods presented here can perhaps be used to develop a critical length scale for the optimal resistance under compressive deformation for bacteriophage T4 cell puncture

device. A detailed investigation of this structure is presented in the next section.

4.3 Triple beta-helices and the cell-puncture needle application

This section focuses on a systematic nanomechanical characterization of the triple-beta-helix structure of the cell puncture device needle of the bacteriophage T4 virus, using molecular dynamics (MD) simulations. The behavior of this structure is studied under compressive loading, as in the case of physiological conditions, and investigate the rate dependent strength and stiffness of the molecule over four orders of magnitude in deformation rates.

The coordinates for the molecular model (monomer) are obtained from the Protein Data Bank [14], PDB file ID 1k28. The trimer is built by rotating the monomer structure around the helical axis and creating duplicates in a single coordinate file. The molecular assembly is truncated at the end of the gp5C domain and protein structure and coordinate files of this triple-beta-helix domain is created using psfgen tool in NAMD using CHARMM topology and parameter files [169, 160]. The assembly is then minimized for 10,000 steps and equilibrated in TIP3 water box of $6 \times 6 \times 12$ nm for 1.2 ns, with periodic boundary conditions. Simulations are carried out in an NPT ensemble using the Langevin piston Nose-Hoover method implemented in the NAMD software [169]. A Langevin piston with a target pressure of 1.01325 bar (1 atm) is used. The Langevin piston temperature is set at 310 K throughout the simulation. The stability of the protein structure without load applied is verified from hydrogen bond dynamics as well as RMSD data obtained from the molecular dynamics trajectory.

The Steered Molecular Dynamics (SMD) approach [154] with a constant compression velocity is used as the protocol for simulating the force-induced deformation of the protein. An SMD spring constant as k =1 kcal/mol/Å^2 and a varying SMD displacement rate of v = 0.1 to 25 m/sec). Compared with other SMD studies [219], the lowest effective loading rates used here (kv) are 10 to 100 times slower, reaching

simulation times exceeding 20 ns. For the system size (approximately 40,000 atoms) and time-scales of interest, these simulations require parallel computing resources. The simulations are carried out on a Linux cluster at the Laboratory for Atomistic and Molecular Mechanics, as well on the BlueGene supercomputer at the San Diego Supercomputing Center.

The boundary conditions for compression simulations consist of fixing the first two convolutions of the protein structure at one end, while applying the moving harmonic constraint to atoms in two convolutions on the other end of the molecule. This approach leads to uniform loading conditions throughout the cross section. The fixity and moving harmonic constraints are applied to Cα atoms of all three molecules of the triple-helix.

A different simulation protocol is used to estimate stiffness of the system at quasi-static loading. In this simulation method, referred to as small-perturbation simulations (SPS), the deformation velocity is kept constant initially until a desired displacement level is reached. At this point, the SMD velocity (initially 0.1 m/s, the slowest accessible rate) is reduced to zero, and the rate dependent effects on force disappear while the end-to-end length oscillates around a displaced equilibrium point. The system is then equilibrated and ratio of the long-time average of the change in end-to-end length to the net force acting on the system is taken as the representative stiffness of the system at quasi-equilibrium loading.

Molecular visualization of trajectories are performed using the VMD program [111], with H-bond cutoff distance and angle of 4 Å and 40° respectively. Post-processing is carried out using MATLAB, and averaging scripts are used for force-displacement plots and curve fitting tools for studying the rate dependence behavior. The best fit for rate-dependent plots is selected based on goodness criteria of the fit (e.g. R^2 values, confidence bounds) from a wide range of mathematical functions.

The topology of the cell-puncture device and the atomistic structure of the needle domain are shown in Figure 4-7(a). The (gp5C)3 domain of the molecular assembly represents the triple-helix beta-solenoid topology that is used to puncture through the outer cell membrane as part of the infection process. Following puncture, the

gp5 lysozyme domains shown in Figure 4-7(a) enzymatically breakdown the cell wall, after which the puncture device needle structure disassembles and viral DNA is injected into the cell through a larger cylindrical protein nanotube in the tail sheath. The protein assembly consists of three intertwined polypeptide chains with identical folds that are symmetric rotations of each other around the helical axis. The triangular helical cross section consists of beta-sheets on each of the three sides with a slight twist of $\approx$ 3 degrees per strand, as can be observed from the cross-sectional view shown in Figure 4-7(b). Atomistic simulations are performed to observe how the beta-helix topology of the needle behaves under a compressive loading scenario similar to physiological conditions. The main goal of the simulations is to understand the mechanical properties of this specific topology and investigate whether it could sustain large forces required to puncture through a cell membrane. Figure 4-7(c) illustrates the setup for compression simulations, where the standard SMD protocol is implemented to apply compressive forces to the molecule at a constant loading rate.

First, the compressive strength of the structure is characterized at different loading rates. This is an important aspect, as the simulations are carried out at much faster rates than experiments and extrapolation to slower rates is not straightforward. Figure 4-8(a) illustrates typical deformation profile of the triple-beta-helix under compression (results shown here from simulation with deformation speed $v = 25$ m/s). Significant "elastic" compression of the molecule can be observed before failure, in particular at the faster loading rates. As the structure is symmetric and an equilateral triangle at all cross-sections, there is no general weak axis based on the initial topology. This is because an equilateral triangle has the same moment of inertia along any axis passing through the center of mass, and hence the bending rigidity is uniform in all possible buckling directions (see Section 4.4). Moreover, further uniformity and interesting torsional properties are perhaps achieved through the true helical form of the molecular structure. It is observed that failure occurs when the structure buckles with respect to a random "weakened" axis due to perturbations stemming from non-heterogeneous evolution of the topology under external force, H-bond rupture mediated by water molecules and thermal vibrations.

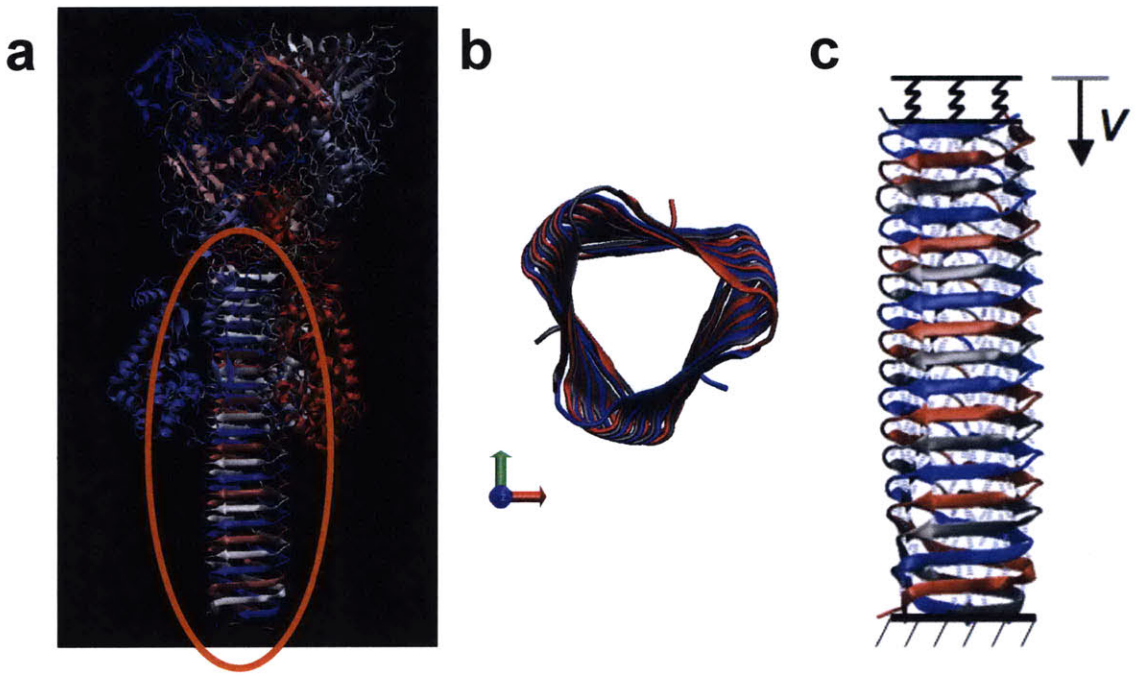

Figure 4-7: Schematics and simulation setup for the triple-beta-helix. Panel (a) illustrates the molecular structure of the cell puncture device of bacteriophage T4 virus and the needle generally known as the gp5C domain encircled in red. The protein assembly consists of three polypeptide chains with identical folds that are symmetric rotations of each other around the helical axis. The triangular helical cross section rotates 3 degrees with every convolution, as can be observed from the cross-section view shown in panel (b). Atomistic simulations are performed to observe how the beta-helix topology of the needle behaves under compressive force, to clarify whether it could sustain large forces required to puncture through a cell membrane. Panel (c) illustrates the setup for compression simulations, where the standard SMD protocol is implemented to apply compressive forces to the molecule with a constant loading rate.

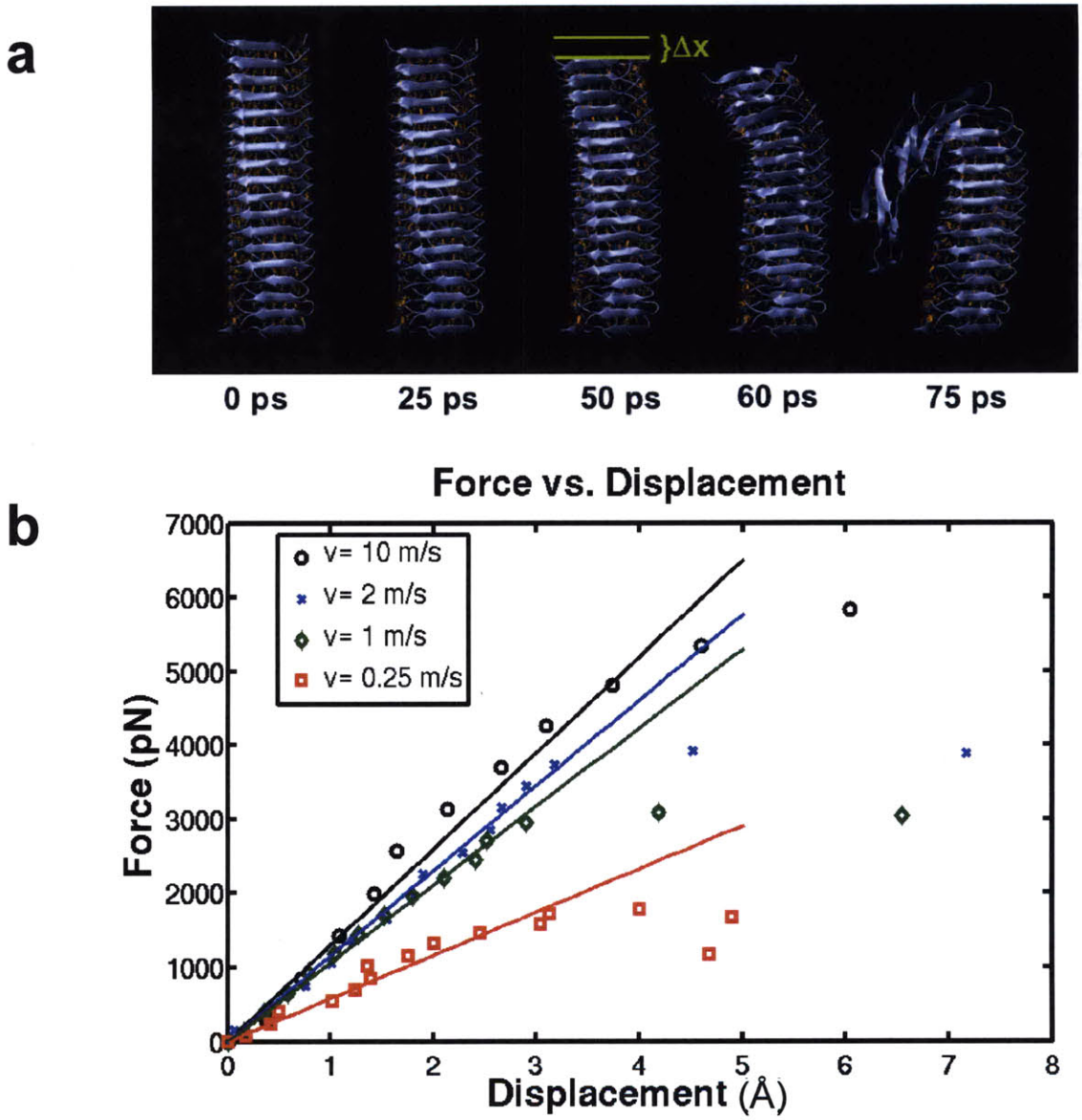

Figure 4-8: Snapshots of deformation, force extension plots. Panel (a) illustrates typical deformation profile of the triple-beta-helix under compression (results from simulation with deformation speed $v = 25$ m/s). Significant compression of the molecule can be observed before failure, in particular at the faster loading rates. As the structure is symmetric and an equilateral triangle at all cross sections, there is no general weak axis based on the initial topology. Failure occurs when the structure buckles with respect to a random weak axis due to non-heterogeneous evolution of the topology under external force, H-bond rupture mediated by water molecules and thermal vibrations. Force extension profiles for different loading rates are shown in panel (b). The initial force extension behavior can be estimated by a linear regime followed by nonlinear plastic-like behavior toward failure.

Force extension profiles for different loading rates are shown in Figure 4-8(b). The initial force extension behavior can be estimated by a linear elastic regime followed by nonlinear behavior towards failure, reminiscent of the buckling curves observed for most engineering materials. Rate-dependent elastic modulus of the system can be extracted from small deformation measurements where nonlinear instability effects are negligible, such that such that $E = kL/A$, where k is the initial stiffness, $L = 7.7$ nm is the length, and $A = 90$ Å^2 is the cross-sectional area of the molecule assuming a hollow section with 5 Å thickness representing the beta-sheet walls of the nanotube.

The peak force form each simulation can be plotted in force vs. log (v) space to illustrate rate dependent mechanical behavior over roughly four orders of magnitude in time-scales. Figure 4-9(a) illustrates the rate-dependent failure force (approximated as the peak force from the simulation) of the triple-beta-helix as a function of the deformation speed. A power-law fit describes the behavior best, and can also be verified as the points lie on a straight line in a log-log curve. While most of the data lies in the nanoNewton force range, failure forces on the order of a few hundred picoNewtons are expected for this specific topology at slower, experimental deformation rates based on the power law description. Figure 4-9(b) illustrates dependence of the initial stiffness on the loading rate, where a different behavior from the power law is observed since the stiffness seems to converge to a finite value at increasingly fast pulling rates. Figure 4-9(c) shows that the peak force observed from simulations corresponds to the point at which the number of H-bonds in the system begins to decrease, leading to structural instability.

The stiffness of the molecule is expected to converge to a finite value also at exceedingly slower rates, representing the quasi-static long time limit. Since time-scales of typical MD runs are much shorter than those in an ideal equilibrium-loading scenario, a different simulation protocol is adopted to estimate stiffness of the system at quasi-static rates. In these simulations (small-perturbation simulations, SPS), the thermal motion of the molecule around a displaced equilibrium point is investigated. Figure 4-10(a) and (c) show the loading scenario for the system as a function of time, while panels (b) and (d) of Figure 4-10 illustrate the change in end-to-end length of

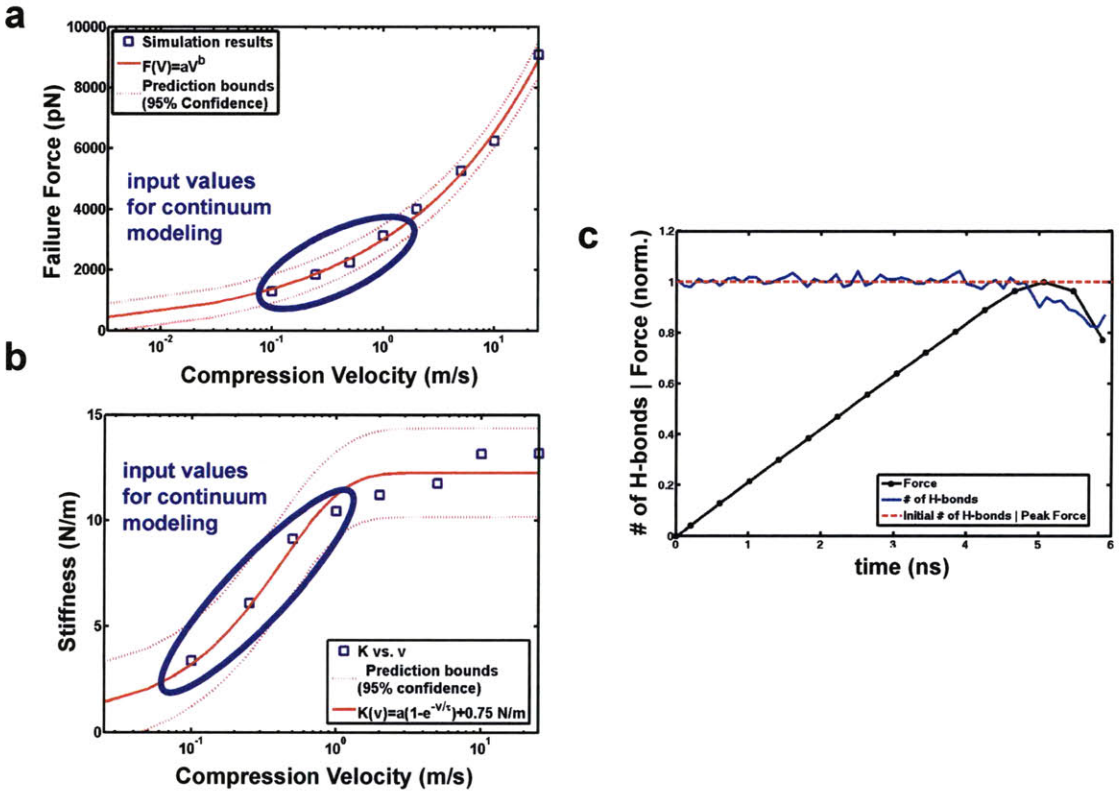

Figure 4-9: Rate dependence of force and stiffness. Panel (a) illustrates the rate-dependent failure force of the cell- puncture needle as a function of the deformation speed. We observe that a power-law fit describes the behavior best, and can also be verified as the points lie on a straight line in a log log curve. While most of the data lies in the nanoNewton force range, failure forces in the order of a few hundred picoNewtons are expected for this specific topology at slower, experimental deformation rates. Panel (b) illustrates that the stiffness is also rate-dependent, but its scaling is different from the power law behavior, as the stiffness is observed to converge to a finite value at increasingly fast pulling rates. Based on the analysis on small perturbation studies, it is predicted that a similar convergence is also likely in long time limit, slow deformation rates. Hence an exponential fit with a constant term makes most physical sense and also fits the data best. According to this, at faster rates, where the time-scale is too short for the system to respond to force (e.g., system is "frozen", or "shocked"), the stiffness values around 12 N/m. At slower rates thermal vibrations and water-mediated H-bond rupture leads to softer response. In the long time limit, system will converge to its static loading stiffness, measured here to be approximately 0.75 N/m. An exponential function with constant term captures this rate-dependence effect over all time-scales. Panel (c) illustrates that the peak force corresponds a moment shortly after the decrease in the total number of H-bonds in the system, suggesting that H-bond rupture is a key event defining the onset of instability.

the system due to the external force for two different levels of strain. The estimate for long-time limit yields 0.75 N/m for both cases, independent of the extent of deformation, supporting the result for the linear elastic fit for the initial deformation. This value is used as the constant term for the exponential fit to stiffness shown in Figure 4-9(b).

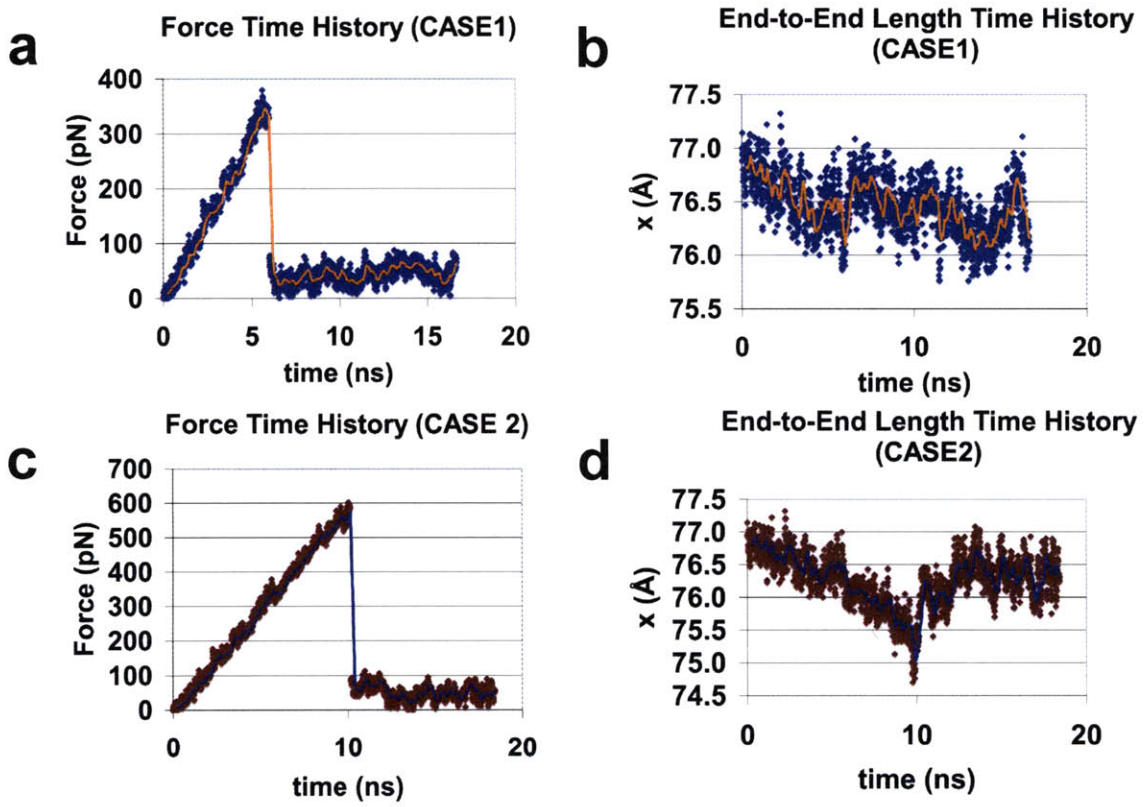

Figure 4-10: Small perturbation simulations (SPS). Panels (a) and (c) show the loading scenario for the system as a function of time, while (b) and (d) illustrate the change in end-to-end length of the system due to the external force. In the SPS, deformation velocity is kept constant initially, until a desired displacement level is reached. At this point, the SMD velocity (initially 0.1 m/s, the slowest accessible rate) is reduced to zero, and the rate-dependent effects disappear. The system is then equilibrated for 8-10 ns and the averaged ratio of displacement to force is taken as the representative stiffness of the system at static loading. The estimate for long-time limit turns out to be 0.75 N/m for both cases, independent of the extent of deformation, supporting the linear elastic fit for initial deformation.

In the range of loading rates studied, it is observed that the deformation profile of the system resembles those observed in typical buckling of columns, albeit the actual

dynamics of the system appears to be much more complex. Mapping of the initial stiffness to the peak force is challenging without further validation from experiments. However, a straightforward continuum type description of the rate dependent failure load based on the initial stiffness value can be formulated using Euler-Bernoulli beam buckling formulation. Since the exact boundary conditions of the system is unknown, the parameter α that denotes end conditions of an idealized beam is determined based on the values for force and stiffness for a small range of loading rates. The region of the fitting is denoted in Figure 4-9(a,b).

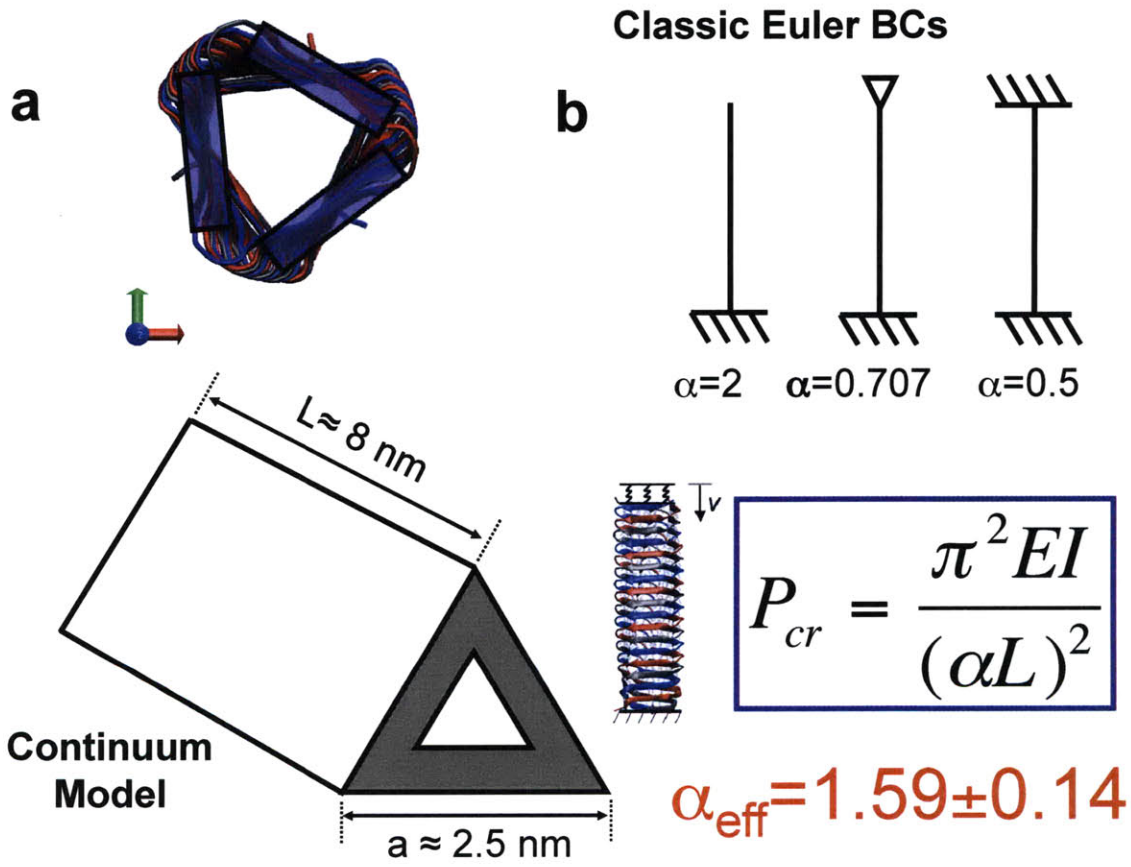

Figure 4-11: Schematics for continuum formulations. Panel (a) illustrates the simplified model of the structure for applying a continuum buckling theory to the system, taking into account rate-dependent initial stiffness of the system. Panel (b) illustrates different constants that need to be used based on boundary conditions of the compressive loading. Since the exact boundary conditions of the spring + beam system is unknown, the constant that gives the best mapping between initial stiffness and failure force is selected. The buckling formulation obtained can then be used to describe failure force at vanishing pulling rates based on the stiffness value obtained from small perturbation simulations.

Figure 4-11(a) illustrates the simplified model of the structure for applying a continuum beam-buckling theory to the system, taking into account the rate dependent initial stiffness of the system. The critical force that will lead to buckling can be estimated based on Euler theory [235] as:

$$P_{cr} = \frac{\pi^2 EI}{(\alpha L)^2} \tag{4.5}$$

where the parameter α denotes the effective length scaling factor based on the boundary conditions of the buckling mode. The moment of inertia I, and cross-sectional area A are estimated from the atomistic model with a hollow cross-section assumption. Figure 4-11(b) illustrates different constants that need to be used based on boundary conditions of the compressive loading. Since the exact boundary conditions of the system submerged in water is not precisely known in this case, a value for the constant α is selected to provide the best mapping between initial stiffness and failure force (taken here as the critical buckling load) prediction based on simulation results. The buckling formulation obtained can then be used to describe failure force at vanishing pulling rates based on the stiffness value obtained from small perturbation simulations. The selected region for stiffness and failure force values are shown in Figure 4-9(a) and (b) where the dimensionless buckling parameter α is found to be 1.59$\pm$.1425. The physical interpretation of this finding is that the boundary conditions adopted correspond to a case resembling fixed-free boundary conditions in typical Euler buckling with partial lateral restraints due to non-ideal loading conditions, fluid interaction or nonlinear material behavior.

Based on the constants obtained from the fitting to rate dependent regime, one can also provide a prediction for quasi-static loading rates. Taking the stiffness result from SPS simulations where $k_{SPS} = 0.75$ N/m, it is predicted that the failure load at near static deformation rates will be roughly 225 pN; a value that could be tested experimentally. The power law fit error bounds shown in Figure 4-9(a) agree well with this value. The fit can be modified to include this result as a constant term and matches well with the data from simulations, but more data points (in particular

at experimental rates) is necessary before an accurate asymptotical value can be proposed.

4.4 Solenoid cross-sectional shape its mechanical implications

The central matter addressed in this section is whether the triangulation of the core, as observed in certain amyloids and beta-solenoids (see Figure 4-12) [177, 121, 120, 136], may be a universal strategy to create structures with minimal amount of material use, yet maximum rigidity [130, 133]. This question is related to a century old problem in mechanics: the shape of the strongest column, or posing the question in the terminology of biology, the shape of the strongest filament. Lagrange, Weinberger, Clausen and others have studied this fundamental problem; and several solution approaches have been discussed controversially for the selection of the optimal cross-sectional shape [123]. Here a simple analysis is proposed, emphasizing the biological and nanomechanics relevance, and considering a select number of convex cross-sections observed in biological and engineering structures. The ideas put forth here can be investigated further with experimental techniques, where bending and compression tests or thermal fluctuation analysis of these domains can shed light on rigidity of fibrils and their link to materials science and biology applications.

The bending rigidity of a filament is linked to geometrical features of its cross-section such as the second moment of area (I) and the cross-sectional area (A). For a filament of given cross sectional area A, and length L, the ability to resist bending due to thermal or mechanical forces is directly proportional to EI, where E denotes elastic modulus of the material obtained by measuring the relationship between force and elongation in a tensile stretching test. For a given filament length and material properties (*i.e.* elastic modulus, *E)*, the radius of gyration defined as $r_G = \sqrt{I_{\min}/A}$, provides a measure of a cross-section's capacity to resist bending, where I_{min} denotes the smallest second moment of area along any axis passing through the centroid of

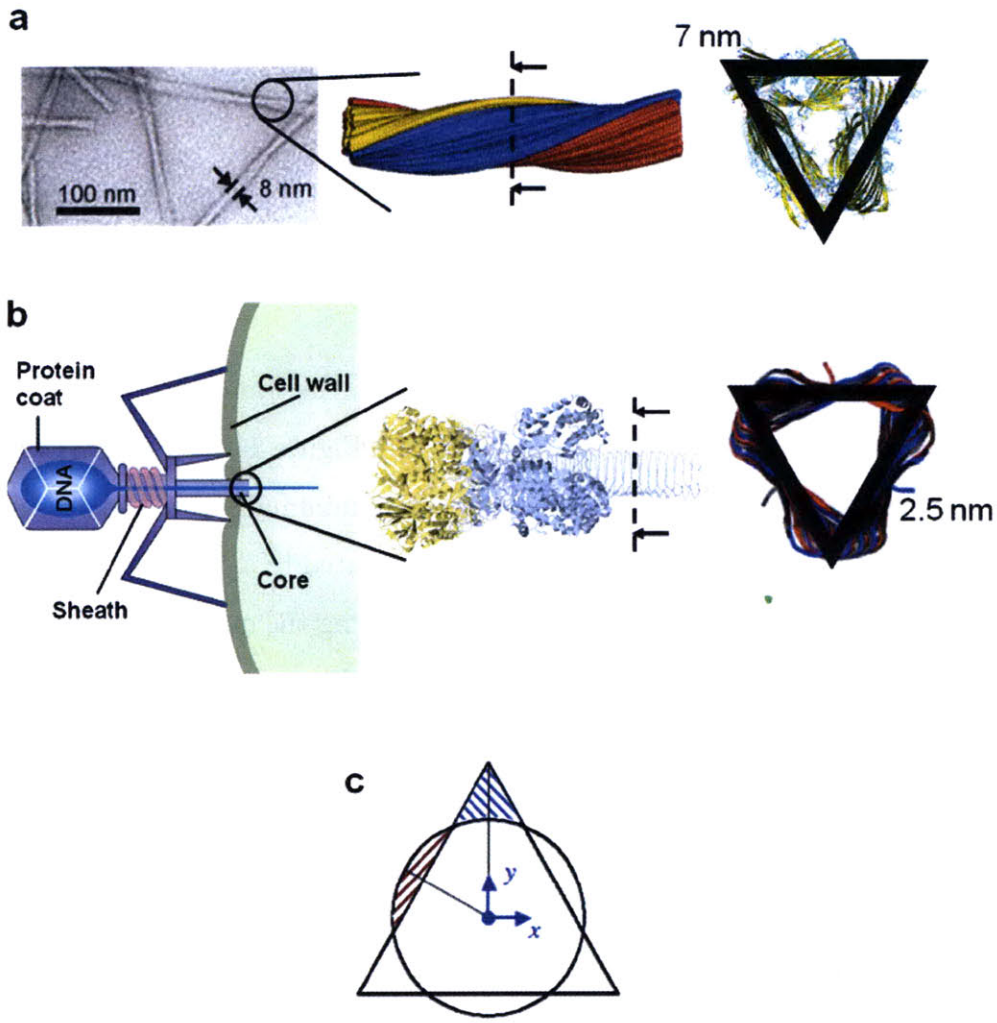

Figure 4-12: Triangular core of amyloids and beta-solenoids. The figure illustrates the cross-sectional properties of triangular peptide nanostructures discussed in this section. An equilateral triangle has been proposed as the cross-sectional shape of the 3-cross structure of certain amyloids (a). Cell-puncture device needle of bacteriophage T4 virus (b) is triple beta-helix with an equilateral triangle as the cross section. For the amyloids, each side is approximately 7 nm in length, whereas the cell-puncture device needle is much smaller, with a side length of approximately 2.5 nm. Panel (c) illustrates triangular and circular cross-sections of identical area. Axes indicated could be used to calculate second moment of area, I. Due to the symmetry of the structures, I is the same along any axis passing through the centroid, but larger for the triangle than for the circle. Comparing the shaded areas that lie outside the shared region, one can easily see that the triangular cross section places more area away from the centroid, thereby dramatically increasing I for the same amount of material. Continuum theory predictions indicate that $\approx 21\%$ increase in bending rigidity can be achieved by triangulation. Panel (a, left) modified from PNAS, "Molecular structural basis for polymorphism in Alzheimer's β-amyloid fibrils", Vol. 105 (47), pp. 18349-18354, Copyright 2008, with permission from the National Academy of Sciences [177].

the cross-section. This axis is the weakest axis along which bending or buckling of the structure would be easier under ambient forces. As such, the parameter r_G is commonly used in structural mechanics as a key parameter for quantifying the failure strength of columns in buildings. Similarly, r_G is also used in polymer science to quantify the flexibility of a macromolecule, and is linked to the persistence length, representing the length beyond which direction vectors of a filament are uncorrelated. Persistence length is directly correlated to bending rigidity as $\xi_P = \frac{EI}{k_B T}$.

The second moment of area, I around an axis is defined as $I_{xx} = \int x^2 dA$. For the equilateral triangle, second moment of area is the same for all axes passing through the centroid and is given as [90]:

$$I_{xx} = I_{yy} = \frac{bh^3}{36} = \frac{hb^3}{48}. \tag{4.6}$$

To compare a triangular cross-section with a circle, let's consider an area equality criterion that gives corresponding dimensions for circle and triangle:

$$\pi r^2 \equiv \frac{bh}{2} \tag{4.7}$$

where b,h denote the base and height of the equilateral triangle, and r is the radius of the circle with equivalent area. For an equilateral triangle,

$$h = \frac{\sqrt{3}b}{2} \tag{4.8}$$

and therefore the equivalence in eq. (4.7) becomes :

$$r^2 = \frac{\sqrt{3}b^2}{4\pi} \tag{4.9}$$

In the case of an equilateral triangle, the minimum second moment of area is given as:

$$I^{\Delta}_{\text{min}} = \frac{\sqrt{3}b^4}{96} \tag{4.10}$$

where b denotes the length of a side. For a circle, second moment of area is again the same for all axes passing through the centroid and is given as [90]:

$$I_{xx} = I_{yy} = \frac{\pi r^4}{4} \tag{4.11}$$

Considering an area (and for constant length, volume) equality, one can evaluate the minimum moment of area of a circle as :

$$I^{\mathrm{O}}_{\mathrm{min}} = \frac{3b^4}{64\pi} \tag{4.12}$$

One can then compute the ratio $I^{\Delta}_{\mathrm{min}}/I^{\mathrm{O}}_{\mathrm{min}}$ to compare which section for given cross-sectional area, A, yields higher minimum second moment of inertia. In this case, it is observed that the triangle beats the circle by 21%. In typical Euler buckling formula for columns, failure load is directly related to I_{min} such that:

$$P_{cr} \sim I_{\mathrm{min}}/A \sim r_G \tag{4.13}$$

where r_G is the radius of gyration. Hence by maximizing I_{min}, a triangular column makes use of material in a better fashion to yield stiffer structures. The argument is also valid for overall bending rigidity, leading to more persistent fibers. One can also show that the same reasoning applies for hollow sections provided that the wall thickness is large enough to prevent shell buckling.

The cross sections of circle or regular convex polytopes discussed above, such as equilateral triangle, square and hexagon, have the advantage that the rigidities for bending in different planes containing the axis are the same. For other shapes like rectangle with disparate length and width, the bending resistance is then limited by the mode with lower rigidity.

Continuum analysis suggests that filaments with an equilateral triangle as the cross-section have the greatest radius of gyration than any other convex cross-sectional shape for a constant cross-sectional area, including a circle. The advantage of the triangle that leads to 21% higher rigidity is its capability of distributing material further from the principal axes, away from the center, to provide a more efficient

resistance to bending.

This analysis has been extended by studying several different cross-sectional shapes using the finite element method (FEM) simulations [134]. Five types of cross-sections have been studied: triangle, square, hexagon, circle and triangle with rounded corners (considering that the corresponding biological structures with a sharp corner would have too high potential energy), all with equal area. Triangle is observed to perform best among all sections as it has the highest minimum moment of inertia, supporting that it is the ideal convex regular polytope that maximizes rigidity per material usage.

The results presented here are scale-independent and universal as they are purely geometrical, and can in principle be applied at all scales, including to building design, micromechanical systems, or nanostructures. The twisted triangular core of the cell-puncture needle, as well as amyloid filaments and many other fibrous protein topologies [16] may suggest that this concept also emerges as a design principle in biology, be it as a result of spontaneous aggregation as in amyloids, or as part of an evolutionary strategy as in the case of T4 virus. An investigation of the mechanical properties of these biological nanofibers merits further studies that focus on the universal geometric features of these structures and their link to their mechanical properties. Ideas that stem from such investigations could lead to more efficient designs of materials and structures, ranging from nanostructures to extreme designs in architecture such as skyscrapers and space elevators where novel materials with exceptional mechanical properties are crucial.

4.5 Discussion and conclusion

This section summarized findings from molecular dynamics simulations and atomistically informed continuum modeling of nanomechanics of the beta-helix protein motif, a key building block of amyloids and self-assembling peptide based fibers. This protein nanotube motif is highly extensible under tensile deformation; sustaining strains up to %800. Under compression, the structure is very robust and exhibits high re-

sistance before failure occurs. Through continuum approaches, the length-dependent deformation map of beta-helices under compression was illustrated.

The approach reviewed here illustrates how molecular dynamics simulations can reveal crucial information about biological structures when experimental data is not readily available. Such information can be used to build predictive models that explain materials behavior and structure property function relationships for proteins and protein based materials.

So far nanomechanical characterization of protein materials has focused on individual proteins, in an attempt to explain their specific biological functions or structural features from a biophysical perspective. Difficulties in interpretation and generalization of results from unfolding experiments or simulations of complex large-scale individual proteins are evident. For this reason, molecular modeling of particularly simplistic protein structures that play a universal and crucial role in nano-scale deformation mechanisms of many protein materials is currently the only reliable option for shedding light on deformation mechanisms, size-effects and other important aspects that can not easily be generically defined for more complex systems. Experimental studies at these length scales are extremely challenging with current technologies, therefore predictive models that can generalize nano-scale MD simulation results are indispensable for explaining experimental findings at larger scales.

Simulation studies and the analysis of the underlying mechanisms have shed light on the underlying atomistic mechanisms. The simulations have provided important insight on the rate dependent stability of this molecular assembly, illustrating that weak interactions such as H-bonds and electrostatic repulsion between charged side-chains govern the response to external force. This could be seen as similar to other biological materials, such as cartilage, where such weak interactions at the molecular scale also contribute significantly to the elastic modulus [44].

Further, it has been shown that an exponential fit with a constant term fits the rate-dependent stiffness data best, which makes physical sense as well, since constant molecular stiffness values are expected at both extremely slow (quasi-static) and fast (shock) loading rates. The analytical model enables us to predict the stiffness under

a wide range of conditions and time-scales, including shock loading and quasi-static loading. A plausible explanation for this rate-dependence behavior is that at faster rates, where the time-scale is too short for the system to respond to force (e.g system is "frozen", or "shocked"), the system has no time to readjust itself, water molecules have limited influence on H-bond dynamics, and thermal vibrations of the molecule are limited to very short timescales in which the dominant dynamic modes of the system can not be sampled adequately, hence the system reacts in a stiffer manner. At slower rates, thermal vibrations and water mediated H-bond rupture leads to a softer response. In the long time limit, system will converge to its static loading stiffness, estimated here to be approximately 0.75 N/m, roughly fifteen times lower than the stiffness values observed under shock loading conditions. An exponential function with a constant term captures this rate-dependence effect well over all time-scales in an analytical model.

Biological molecule	***d* (nm)**	**ξ_P (nm)**	**Reference**
Polypeptide Chain	0.5	0.4	[194, 173]
Tropocollagen	2	11-16	[228, 36]
Actin Filament	6-8	3000-17000	[91, 168]
Intermediate Filament	10	1000-3000	[168]
Single beta-helix	3	900-1400	[130]
Triple beta-helix	3	800-14,000	[133]
Amyloid	4-7	6,000-50,000	[218, 176]
Microtubule	25	1,000,000 - 8,000,000	[91]

Table 4.1: Diameter and persistence length of biological filaments.

The beta-helix topology exhibits very high mechanical stability under compressive loads, reaching strength values close to nanoNewtons at fast deformation rates. The extreme rigidity of the molecule is most evident when it is compared with other biomolecular and inorganic nanostructures. Table 4.1 compares the diameter and persistence length of single biomolecules as well as biomolecular assemblies. Further experimental studies, both *in vivo* and *in vitro* are required to investigate how the specific topology of the cell-puncture device is suitable for its function and whether the molecular assembly can be optimized further by chemical and structural modifi-

cations, for example for the purpose of creating multifunctional, bioactive or mechanically rigid protein nanotubes. Key contributions of this work are listed below:

- This work illustrates that the stiffness and compressive strength of the molecule are strongly rate dependent (as it is the case of most biological matter). In particular at fast deformation rates this may be relevant in puncture impact on the surface of a cell. This study indicates that the cell puncture device has a very high persistence length to diameter ratio, compared with other biomolecules, and can exhibit stiffness values much higher than that of a cell membrane (for instance, one study [61] shows $k_{CELL} \approx 0.27$ N/m for *E. coli*, much less than $k_{HELIX} = 0.75$-12 N/m as observed in this study).

- This is the first computational study at atomistic resolution that provides a detailed analysis of the compressive mechanical behavior of a single biological molecule, in this case a beta-solenoid type topology. The results and methods reported here can be readily applied to other protein structures, in particular to amyloids and prion proteins that show a similar molecular architecture. The mechanical compliance of protein deposits with cells and surrounding tissue at oligomer and fibril length scales may have important implications for many pathological conditions such as neurodegenerative diseases.

- This study shows that the critical buckling load based on the Euler-Bernoulli beam theory could possibly be used to predict buckling phenomena at the nanoscale when coupled with atomistic information, in this case rate dependent stiffness of the molecule. The agreement with failure force values obtained from molecular dynamics simulations supports this point. Investigation of the length-dependent stability of this particular topology either experimentally or by simulation is a necessity to determine whether or not Euler buckling is actually a valid model for this system.

- A basic structure-property-function relationship can now be proposed for this protein topology. The cell puncture needle is suitable for its function since: (i)

The hollow triangular cross section is ideal for compressive loading as it has no particular weak axis and maximizes moment of inertia per unit area (compared with other cross-sections such as square or circle), increasing the failure force dramatically. (ii) The structure utilizes weak H-bonds in compression to achieve significant compressive strength, an energetically efficient way of generating force to penetrate the cell membrane and a simpler method compared to more complex assembly mechanisms such as biomineralization.(iii) The stiffness of the molecule is significantly higher than that of cells, for all deformation rates studied, making this system ideal for serving as a puncture needle, (iv)The length of the molecule is long enough to span through the cell membrane but not significantly longer, allowing the molecule to have enough strength to serve its function without being susceptible to buckling at lower force levels. Longer molecules would fall short of this capability, as the buckling force scales inversely with the square of the length according to Euler theory.

From a materials scientist's perspective, generic properties of chemical interactions, bond rupture mechanisms and deformation modes are key input parameters that one needs to build multi-scale models that can predict properties of bulk materials. In light of this, the fundamental, simplistic approach presented here for beta-helices is extremely important for explaining universal deformation mechanisms, size effects and chemical / thermodynamical properties of biological materials and structures. Simulation methods coupled with theoretical frameworks such as those reviewed here illustrate the state-of-the-art in nano-scale / multi-scale modeling techniques for materials applications. The findings point out the potential of using the beta-helix protein motif for a variety of nano-scale materials science applications, ranging from stiff peptide nano-tubes to self-assembling peptide based fibers inspired by amyloids. These results illustrate that beta-helix structures appear as a promising candidate for many functional nanomaterials that may be used for structural applications, tissue scaffolds and non-invasive drug delivery systems inspired by the bacteriophage T4.

Chapter 5

Size dependence of the strength and toughness of beta-sheet nanocrystals in silk

Silk features exceptional mechanical properties such as high tensile strength and great extensibility making it one of the toughest materials known. The exceptional strength of silks, exceeding that of steel, arises from beta-sheet nanocrystals that universally consist of highly conserved poly-(Gly-Ala) and poly-Ala domains. This is counterintuitive, since the key molecular interactions in beta-sheet nanocrystals are H-bonds, one of the weakest chemical bonds known. This chapter summarizes findings from a series of large-scale molecular dynamics simulations, revealing that beta-sheet nanocrystals confined to a few nanometres achieve higher stiffness, strength and mechanical toughness than larger nanocrystals. The analysis presented illustrates that through nanoconfinement, a combination of uniform shear deformation that makes most efficient use of H-bonds, and the emergence of dissipative molecular stick-slip deformation lead to significantly enhanced mechanical properties. These findings explain how size-effects can be exploited to create bioinspired materials with superior mechanical properties in spite of relying on mechanically inferior, weak H-bonds.

5.1 Background on spider silk and significance of nanocrystals

Silk is a biological protein fiber with exceptional tunable mechanical properties such as high tensile strength and great extensibility, making it one of the toughest and most versatile materials known [247, 212, 11]. Since its development in China thousands of years ago, silk has been utilized throughout history not only as luxury fabrics but also in various technological applications including parachutes, medical sutures, and more recently, tissue regeneration [244]. Variations in synthesis conditions such as temperature, reeling speed and pH lead to silk-based peptide fibrils or protein aggregates with rather different structural and mechanical properties [233, 68, 188, 150]. However, despite decades of research in silk production and mechanical properties, the mass-production of silk and biomimetic materials remains a challenge, particularly due to silk's unique features that can only be achieved by the controlled self-assembly of the macromolecular constructs with molecular precision at the nanoscale.

Recent investigations revealed that anti-parallel beta-sheet crystals at the nanoscale, consisting of highly conserved poly-(Gly[G]-Ala[A]) and poly-Ala repeats found in both commercial and spider silk [103], play a key role in defining the mechanical properties of silk by providing stiff orderly cross-linking domains embedded in a semi-amorphous matrix that consists predominantly of less orderly beta-structures, 3_1 helices and beta-turns (Figure 5-1(a))[234, 240, 151]. Beta-sheet nanocrystals, bonded via assemblies of H-bonds [128, 131, 129], have dimensions of a few nm and constitute $\approx$ 10-15% of the silk volume, while with less orderly extended structures the beta-sheet content can exceed 50% for spider and silkworm silks [96, 200, 68]. When silk fibers are exposed to stretch, beta-sheet nanocrystals reinforce the partially extended and oriented macromolecular chains by forming interlocking regions that transfer the load between chains under lateral loading, similar to their function in other mechanical proteins [194, 163, 23, 72, 149, 151, 226, 37]. Thereby, beta-sheet nanocrystals provide cohesion between the long polypeptide strands, enabling the amorphous domains to stretch significantly. Eventually, fracture of beta-sheet

nanocrystals occurs at large deformation and large loads, where the typical loading at individual beta-sheet nanocrystals is lateral [103]. Recent experiments demonstrated that when the size of beta-sheet nanocrystals is reduced by moderating the reeling speed or by metal infiltration, silk displays enhanced toughness and greater ultimate strength [233, 68, 150], exceeding that of steel and other engineered materials. This is particularly intriguing, since H-bonds comprise the chemical bonds that underlie the beta-sheet nanocrystals structures, and are one of the weakest chemical bonds known. The issue of how H-bonded silk threads can reach such great strength and toughness to overcome the limitations of their inferior building blocks is a question of fundamental importance to understanding the behavior of a broader class of beta-sheet rich biological protein materials (e.g. amyloids, beta-solenoids, virulence factors).

To address this issue, the investigation of the properties of silk beta-sheet nanocrystals requires an accurate representation of the protein's amino acid sequence and chemical interactions from a bottom-up perspective. This renders explicit water molecular dynamics a suitable tool, as it provides an accurate description of the physical mechanisms governing self-assembly (*association*) and failure (*dissociation*) of polypeptide chains at submicron length-scales [26, 207, 219, 158, 37]. Here the focus is on beta-sheet nanocrystals with the sequence from *Bombyx mori* silk [81] as a model system, resembling the structure of nanocrystals found in silks from most spider and silkworm species due to the extremely high sequence conservation of the crystalline regions (consisting of $[GA]_N$ or $[A]_N$ repeats with 4-12 residues).

5.2 Simulation setup and theoretical background

The methods used in this Chapter are briefly overviewed in this section. A mix of MD simulation and continuum theory is used here to quantify the size-dependent strength and toughness of beta-sheet nanocrystals.

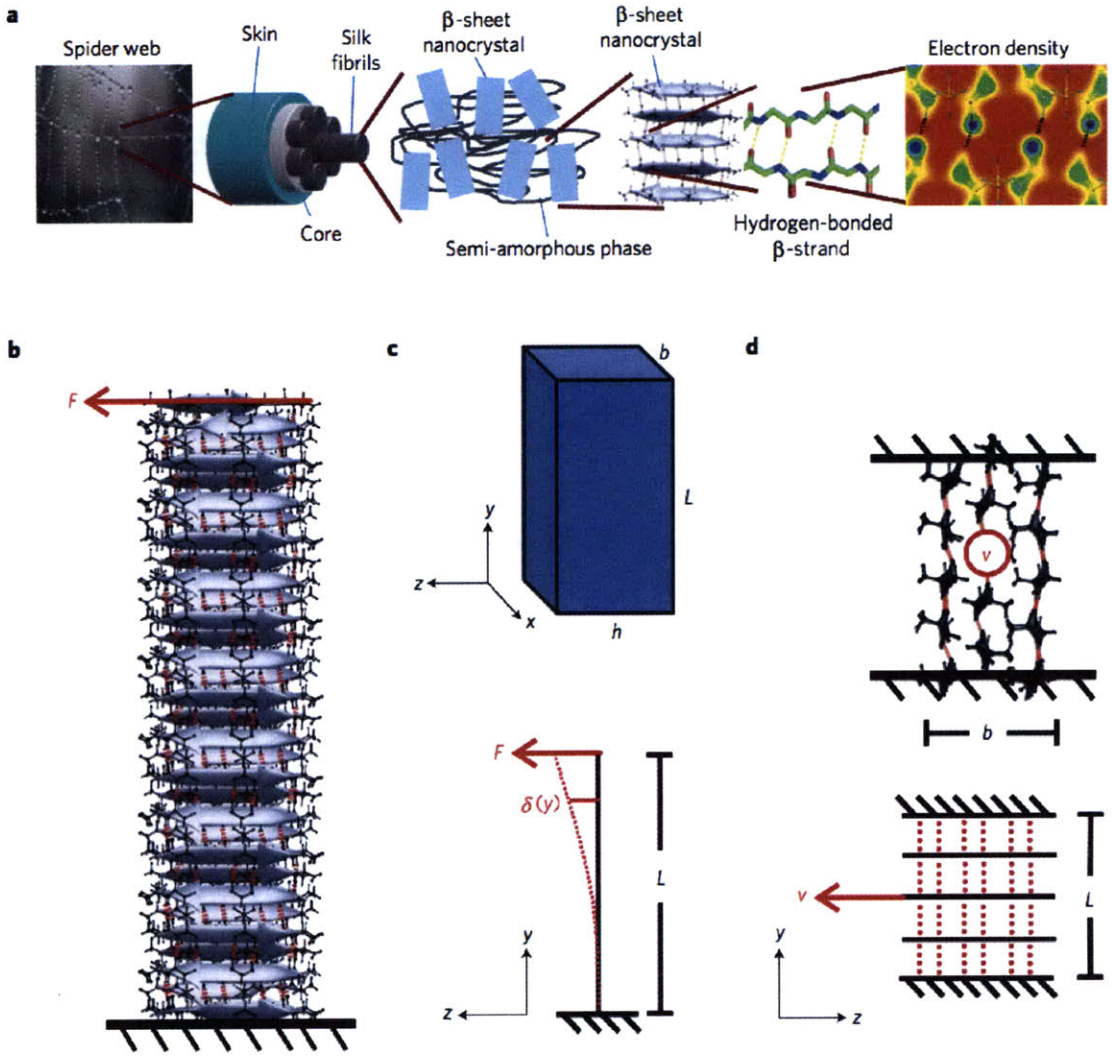

Figure 5-1: Simulation setup and theoretical considerations. Panel(a) illustrates a schematic of the hierarchical spider silk structure. Panel (b) shows the atomistic structure of the silk beta-sheet nanocrystal obtained from the Protein Data Bank (identification code 2slk) and replicated to build beta-sheet nanocrystals of different size. In the first set of simulations, the beta-sheet nanocrystal is subject to loading conditions similar to a cantilever beam with a constant tip loading, used to identify the bending rigidity and other structural properties. This loading mimics the characteristic lateral loading relevant to silk mechanics. Panel (c) shows the schematic representation of the beta-sheet nanocrystal and definition of coordinates used here (upper part, where parameters b and h describe geometric parameters related to the number of sheets and the length of strands in the nanocrystal, and L the size of the nanocrystal in the y-direction). The lower part shows the geometry of the bending study and defines the displacement variable. Setup for pull-out simulations, where in order to characterize fracture resistance of the nanocrystals, the central strand of the middle sheet is pulled out with constant velocity, while the top and bottom strands are restrained, as shown in panel (d).

5.2.1 MD simulation setup

In order to examine the key mechanical parameters of the silk beta-sheet nanocrystals as a function of size, two sets of computational experiments are carried out, namely bending and pull-out simulations. The choice of this setup is motivated by a desire to create a model system to quantify the size-dependent lateral stiffness of silk beta-sheet nanocrystals under lateral loading, which is the key loading condition of nanocrystals in silk [211, 173, 215]. In the bending scenario (Figure 5-1(b)), a constant lateral force is applied at one end of the nanocrystal while the other end remains fixed, resembling a cantilever beam with tip loading (see Figure 5-1(a)). This setup mimics biological assays employed in single molecule experiments to assess mechanical properties (e.g. elastic moduli) or the persistence length of biofilaments [174, 140]. In the pull-out scenario (Figure 5-1(c)), the centre strand of the assembly is pulled out while the outermost ends of the nanocrystal are fixed, in order to assess the deformation and fracture behavior of the system at large forces. The pull-out setup complements the first one by revealing the strength, fracture toughness and molecular failure mechanisms of nanocrystals as a function of their size.

The coordinates for the molecular model of silk are obtained from the Protein Data Bank with identification code 2slk [81], where different size nanocrystals used here are built by truncating the system to four strands and replicating the system by shifting the molecules at a proper distance [81]. The protein structure and coordinate files are created using the psfgen tool in NAMD [169] using CHARMM topology and parameter files [160]. The assembly is then minimized and equilibrated in a TIP3 explicit water box. Simulations are carried out in an NPT ensemble (Langevin piston Nose-Hoover method). The piston target pressure is set to 1.01325 bar (1 atm), and the temperature is 300 K. The stability of the beta-sheet nanocrystal without load applied is verified from H-bond dynamics, as well as RMSD data obtained from the molecular dynamics trajectory.

Simulations under lateral loading (Figure 5-1(b)) are carried out using the constant force module in NAMD. One end of the nanocrystal is fixed by constraining the

motion of the Cα-atoms at on the bottom strand of each sheet. The other end of the nanocrystal is deformed by tip loading, where a constant force is applied to the terminal Cα-atom of each strand at the top. For simulating the pull-out scenario (Figure 5-1(d)) Steered Molecular Dynamics (SMD) method is used [219], with a constant pulling velocity. The boundary conditions consist of fixing Cα-atoms on the top and bottom strands with an SMD spring constant k =10 kcal/mol/Å^2 and a displacement rate $\dot{x}$ =0.0005 Å/ps, in line with other recent SMD studies [219]. For the pullout simulations, the absolute value of the force levels will differ from experimental studies due to rate effects. Based on earlier studies comparing molecular dynamics simulations of beta-sheets with single molecule experiments [219, 128] it is estimated that the strength values are lowered by a factor of six at experimental rates, leading to pull-out strengths in the range of 250-400 pN. Indeed, the values incorporating rate effects are in agreement with AFM experimental results of the strength of spider silk nanofibers [173]. The comparison of the influence of the crystal size on strength and toughness, the focus of the study reported in this paper, is independent of the specific rate as confirmed by additional simulations carried out at varying SMD strain rates (v = 0.0005 – 0.02 Å/ps). Whereas the measured values of the rupture force vary with rate, the length-scale in which the stick-slip mechanism emergences is observed to be rate-independent.

The strain distribution analysis is carried out from the trajectories obtained from SMD simulations. The bond strain is calculated based on the average initial bond lengths and average length over a short time period just before first bond rupture. Measurements are taken from H-bonds below and above the strand and bond positions, and lengths are averaged at each point to provide a single trend across the cross-section. Bond strain is defined as the average initial length over the final, and cutoff is defined as 3 Å (H – O distance). A smoothing spline curve is fitted to the data to show the general trend for each case. The overall trend does not depend on this choice of fitting procedure.

5.2.2 Continuum theory formulations

Here, continuum beam theory to analyze the bending deformation of the system. By considering pure bending, the deformation of the tip is given as:

$$\delta_{tip} = \frac{PL^3}{3EI} \tag{5.1}$$

where P is the total applied force, E is the elastic modulus, and I is the second moment of area around the bending axis. By defining an effective stiffness $k_{eff} = F/\delta_{tip}$, the size dependence of the stiffness can be compared with the results obtained from atomistic simulations (for small forces k_{eff} is independent of the force value). Since the pure bending theory fails to describe the results of these simulations, an extended beam model that includes the effects of shear deformation is considered [56]. Incorporating shear effects, eq. (5.1) becomes:

$$\delta_{tip} = \frac{PL}{D_T} + \frac{PL^3}{3D_B}. \tag{5.2}$$

The effective stiffness is given as:

$$k_{eff}(L) = \frac{P}{\delta_{tip}} = \left(\frac{L}{D_T} + \frac{L^3}{3D_B}\right)^{-1}. \tag{5.3}$$

Equation (5.2) is compared with the pure bending formulation neglecting shear contribution in Figure 5-2(a). From the analysis of the best-fit curves, it is observed that while the pure bending formulation can explain data for large crystal sizes, it fails to provide a good overall description of the stiffness-size relationship, in particular for small crystals.

The ratio of the first and second terms in eq. (5.2) quantifies the relative importance of shear contributions in the deformation for a given beam with constant length and material properties. This parameter is defined as the shear contribution ratio s, which is given by

$$s(L) = \frac{3D_B}{L^2 D_T} = \frac{1}{4}\frac{Eh^2}{GL^2} \sim \frac{h^2}{L^2} \tag{5.4}$$

where the terms $D_B = EI$ and $D_T = GA_s$ denote the bending and shear rigidity. In this notation, E and G stand for elastic and shear moduli, respectively, and A_s is the cross-sectional area.

Best-fit curves for individual displacement profiles are obtained using results from the tip-displacement analysis, where values for D_B and D_T are taken as an initial estimate and further refined for each case within the 95% confidence range (roughly corresponding to $\pm 50\%$ maximum variation in D_B and D_T). The displacement profile is then computed as:

$$\delta(y) = \frac{Py}{D_T} + \frac{P}{D_B}\left(L\frac{y^2}{2} - \frac{y^3}{6}\right) \quad 0 \leq y \leq L, \tag{5.5}$$

where y denotes the distance from the fixed end along the length of the structure. The data points depicted in Figure 5-3(b) show the s values obtained for each case based on the refined D_B and D_T values from displacement profile curves. Deviations in these parameters are largely due to thermal vibration noise as well as nonlinearities arising from out-of-plane deformations, which influence small and large systems variably. To further confirm the size dependence of shear contributions, a pure shear (linear) deformation profile is fitted to the smallest system to quantify how well it predicts the tip displacement as a function of size. The tip deviations are calculated as the difference between the tip displacements calculated from the refined fit vs. the pure shear fit.

Cross sectional dimensions of $b = 1.37$ nm and $h = 2.23$ nm are used in the calculation of the moduli values, based on atomic coordinates. The elastic and shear moduli of the material can then be estimated from D_B and D_T . It should be noted here that since the system is highly anisotropic, Poisson's ratio cannot be directly derived from these values, which is highly specific to the plane of deformation studied.

5.3 Size-effect of beta-sheet nanocrystals: Predictions and experimental validation

Figure 5-2(a) shows snapshots from bending studies where the constant force applied laterally at the tip scales with the nanocrystal size, ranging from 100-375 pN. A softening of the system in larger nanocrystals is evident from snapshots of deformation, which increase as the size of the nanocrystal increases. The scaling of the stiffness is first compared with a classical Euler-Bernoulli beam theory commonly applied to biological filaments and nanostructures [56]. This model shows disagreement when the stiffness of the nanocrystal is plotted against size, and fails to properly describe all of the simulation results shown in Figure 5-3(a). This is because shear deformations are neglected in this classical beam theory, as it only considers tensile and compressive stresses in a beam's cross-section. Upon inclusion of shear correction [56], the agreement between simulation and beam theory is excellent, as shown in Figure 5-3(a). The comparison of the simulation results with the two models suggests that the disagreement arises from the shear contributions to deformation, which is neglected in the conventional theory but is apparently critical in describing the deformation mechanisms in beta-sheet nanocrystals.

Moreover, the series of nanomechanical experiments enables us to identify the distribution of strains in the nanocrystal. The relative importance of shear contributions in the deformation of a beta-sheet nanocrystal of size L is given by the shear contribution ratio

$$s(L) = \frac{3D_B}{L^2 D_T} = \frac{1}{4}\frac{Eh^2}{GL^2} \sim \frac{h^2}{L^2} \tag{5.6}$$

where the terms $D_B = EI$ and $D_T = GA_s$ denote the bending and shear rigidity. In this notation, E and G stand for elastic and shear moduli, respectively, and A_s is the cross-sectional area. For a rectangular cross section (Figure 5-1(c)), $I = bh^3/12$ and $A_s = bh$. The shear contribution ratio s plotted a function of size shown in Figure 5-3(b) illustrates that as the size of the nanocrystal gets smaller, the loading scenario

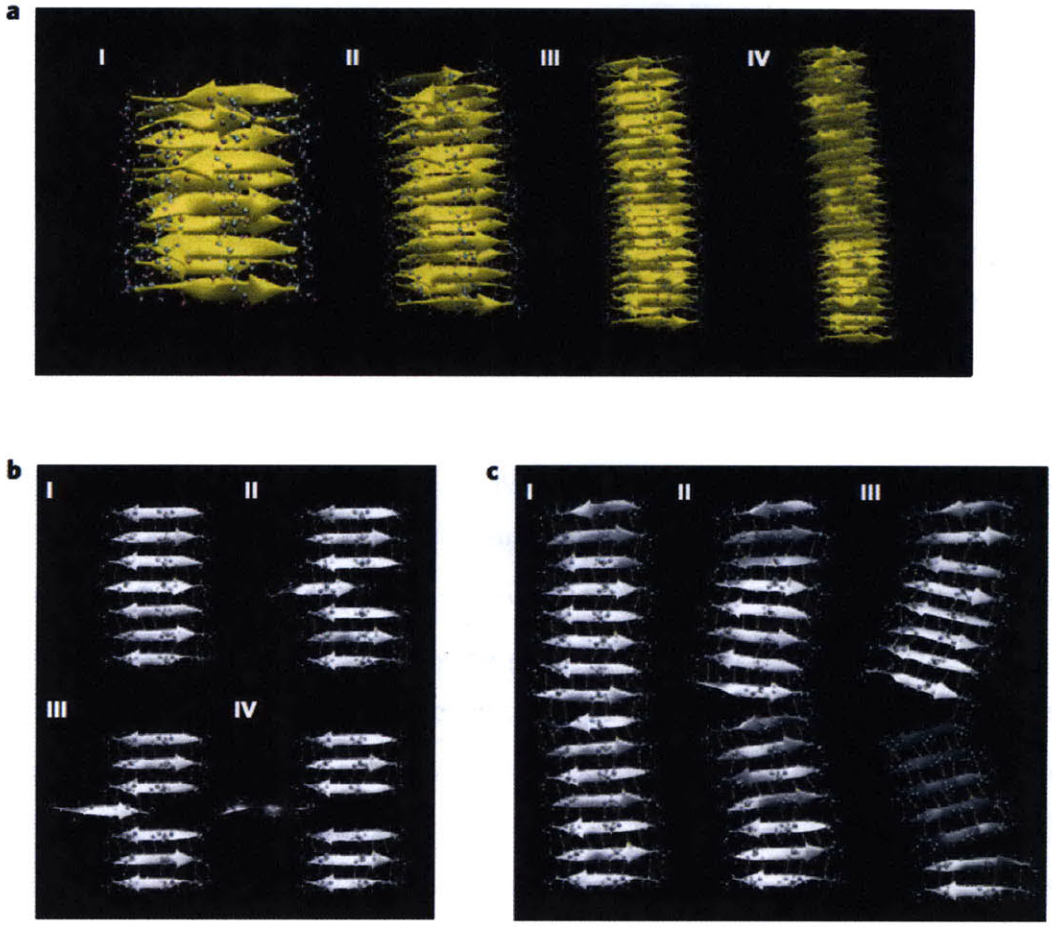

Figure 5-2: Snapshots of deformation profiles and failure mechanisms of silk beta-sheet nanocrystals. Panel (a) exhibits the size-dependent elastic deformation of beta-sheet nanocrystals, based on simulations in the bending setup under lateral loading (Figure 5-1(b,c)). Smaller structures show a linear displacement profile, in line with shear governed deformation, whereas a nonlinear profile with a significantly higher curvature is observed for longer structures, illustrating a transition from shear to bending governed deformation. Panels (b,c) illustrate fracture mechanism during pull-out simulations (lateral loading condition shown in Figure 5-1(d)). The beta-sheet nanocrystal sizes investigated here range from $\approx$ 2-7 nm. Panel (b) shows a visualization of the failure mechanism in a beta-sheet nanocrystal employing 7 strands ($L = 2.83$ nm), with snapshots taken just before and after failure by rupture of H-bonds. Panel (c) depicts the molecular failure mechanism of the largest beta-sheet nanocrystal studied ($L = 6.56$ nm), with snapshots taken just before and after failure by rupture of H-bonds. The main difference in the failure mechanism between the cases shown in panels (b,c) is that longer structures fail by significant bending, which leads to a crack like flaw formation due to non-uniform tensile deformation of H-bonds. In the case of shorter structures as shown in panel (b), the system responds more rigidly (stiffer), and H-bonds break via a stick-slip motion due to uniform shear loading. The stick-slip failure mode leads to significantly enhanced energy dissipation and multiple force peaks in the force-displacement curve (see Figure 5-4(a)).

becomes predominantly shear ($s > 1$), whereas for large beta-sheet nanocrystals pure bending controls deformation ($s < 1$). The existence of this important size-effect is independent of the particular boundary conditions employed here and is a result purely of the geometry and material properties of beta-sheet nanocrystals.

This change in stress distribution inside the nanocrystal has major implications on how individual H-bonds are loaded. When $s < 1$, H-bonds are stretched in tension (*i.e.*, pulled in the bonding direction). By contrast, when $s > 1$, H-bonds are being sheared (*i.e.*, pulled orthogonal to the bonding direction). For the smallest system considered here (L=1.87 nm), the shear contribution is twice of that of bending, whereas for the largest system (L=7.04 nm), shear contributes less than 10% to the total deformation. By defining a critical shear contribution ratio $s^* = 1$, a a critical shear transition length scale can be identified as $L^* \approx 2.5$ nm. Since both D_B and D_T scale linearly with the number of sheets (*i.e.*, dimension b along the x-axis as shown in Figure 5-1), the shear contribution ratio s is independent of variations of b.

In addition to the analysis of the significance of shear contributions as a function of the nanocrystal size, the elastic and shear moduli of the material can be estimated from D_B and D_T . Based on this analysis, the shear modulus G = 4.6 GPa and the elastic modulus E =22.6 GPa. Density Functional Theory (DFT) calculations confirm this result, leading to G = 10.32 GPa and E = 36.45 GPa. The DFT results are slightly higher, as expected for static calculations at zero temperature. Most importantly, these findings agree very well with reported experimental values of the elastic modulus of spider silk nanocrystals, where values in the range of E=16-28 GPa have been reported [181, 143] (see Table 5.1 for an overview). This direct comparison between simulation results and experiments provides validation for this simulation approach. The shear modulus of silk beta-sheet nanocrystals has not been tested directly, but results from torsion experiments on silk fibers that suggested shear moduli of 2.38 GPa for *N. Clavipes* dragline silk and 3.81 GPa for *B. mori* silk [141], in a similar range as the results obtained here.

The change in the distribution of strains in a nanocrystal as a function of its size is expected to have important repercussions on their ultimate fracture behavior under

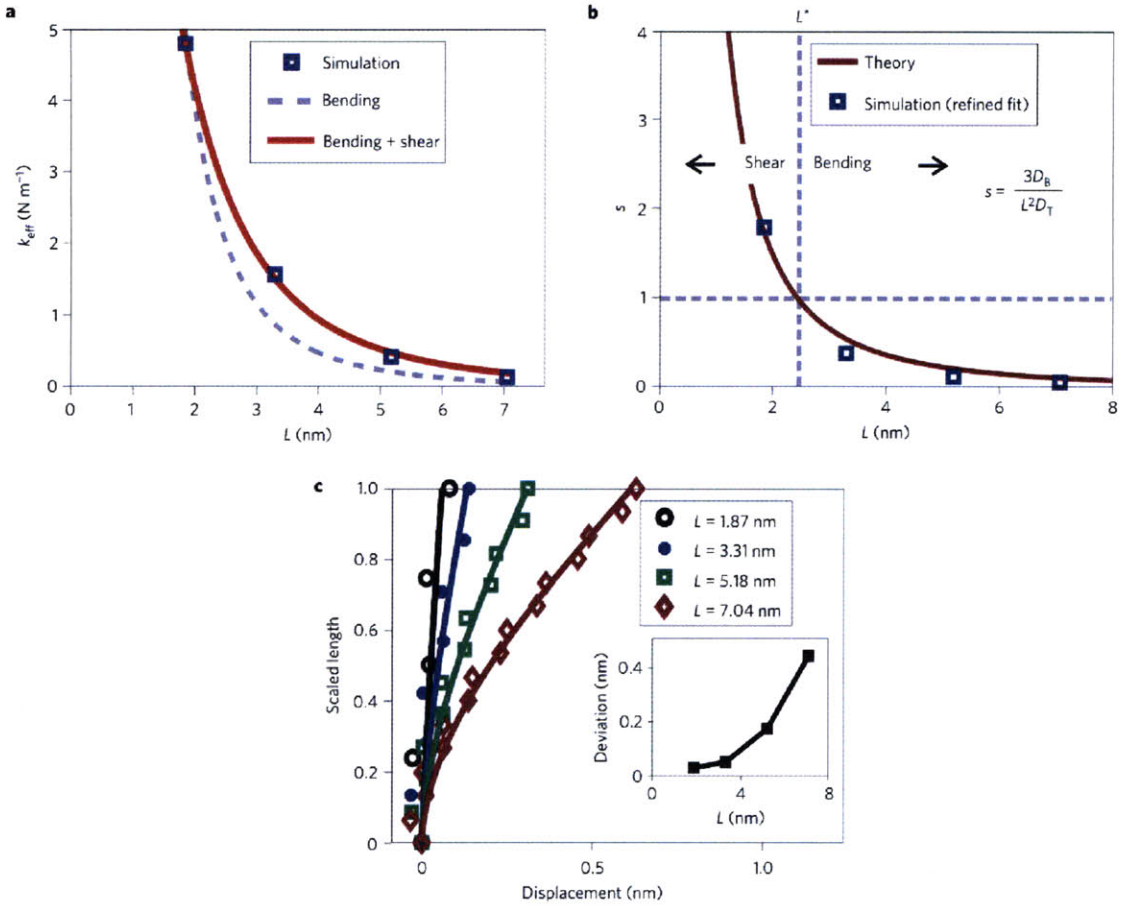

Figure 5-3: Size dependence of the stiffness, and bending versus shear contributions as a function of beta-sheet nanocrystal size. Effective stiffness of the beam, calculated from the ratio of force applied to the observed time average of displacement at the tip (panel (a)). When only bending deformation is taken into account, large deviations between the model and simulation results are observed. When shear contributions to deformation are incorporated into the model, the agreement with continuum theory is remarkable. Ratio of bending to shear contributions (defined as the shear contribution ratio s), as a function of beta-sheet nanocrystal size L (panel (b)). For the smallest system studied, the shear contribution is twice of that of bending, whereas for the longest system, shear contributes less than 10% to the total deformation. Displacement profiles for each case, where the continuous curves illustrate the continuum theory predictions that include both bending and shear (panel (c)). Data points in panel (b) show the s ratios obtained for each case, based on the refined D_B and D_T values from displacement profile curves. The inlay in panel(c) shows an illustration of the deviation of the tip displacement by considering the difference between the model that incorporates shear and bending and a model that incorporates only shear. The plot shows that the deviation is small for small beta-sheet nanocrystals, confirming the hypothesis that shear dominates at small nanocrystal sizes. In contrast, the deviation increases to rather large values at large nanocrystal sizes, showing that shear dominated deformation breaks down once the crystal size exceeds a critical dimension.

Approach	Young's Modulus (GPa)	Shear Modulus (GPa)
MD ([135])	22.6	4.6
Experiment ([143, 181])	16-28	2-4*
DFT ([135])	36.45	10.32
*Experimental measurements of the shear modulus were done at the fiber scale		

Table 5.1: Elastic properties of beta-sheet nanocrystals: MD, experiment and DFT comparison.

extreme loads. This is because in large nanocrystals H-bonds are loaded in tension, but are loaded in shear in small nanocrystals. Since H-bonds are known to be significantly weaker in non-uniform tension than in uniform shear [23], they thus feature a greatly varied capacity to withstand deformation. To test this hypothesis, the failure mechanisms of beta-sheet nanocrystals are investigated by carrying out pullout simulations (Figure 5-1(c)), utilized here to examine the ultimate strength, elastic energy storage (resilience) and energy dissipation (toughness) capacity of beta-sheet nanocrystals under variations of the beta-sheet nanocrystal size L. Force-displacement curves are shown in Figure 5-4(a) for varying nanocrystal sizes. As can be inferred from Figure 5-4(b-c), both the initial stiffness and the ultimate strength (maximum force peak) reach significantly larger values as the system size decreases below ≈ 3 nm. Notably, the toughness is also maximized for systems beyond ≈ 3 nm as shown in Figure 5-4(c). The length scale at which the changes in the material behavior occurs is strikingly similar to the critical crystal size estimated above, $L^* \approx 2.5$ nm, corroborating the hypothesis that the change in the strain distribution in the nanocrystals has major implications on their fracture behavior as the capacity of H-bonds to resist mechanical loads is altered.

The molecular mechanisms that lead to this behavior are evident from force-displacement graphs and an analysis of the molecular trajectories. In the case of small systems, the molecular assembly is stiff, and the initial rupture is followed by a stick-slip motion as the strand slides and reforms H-bonds. This leads to multiple force peaks in the mechanical response (see Figure 5-4(a)), which significantly increases the total dissipated energy. Similarly, the resilience of the system, defined

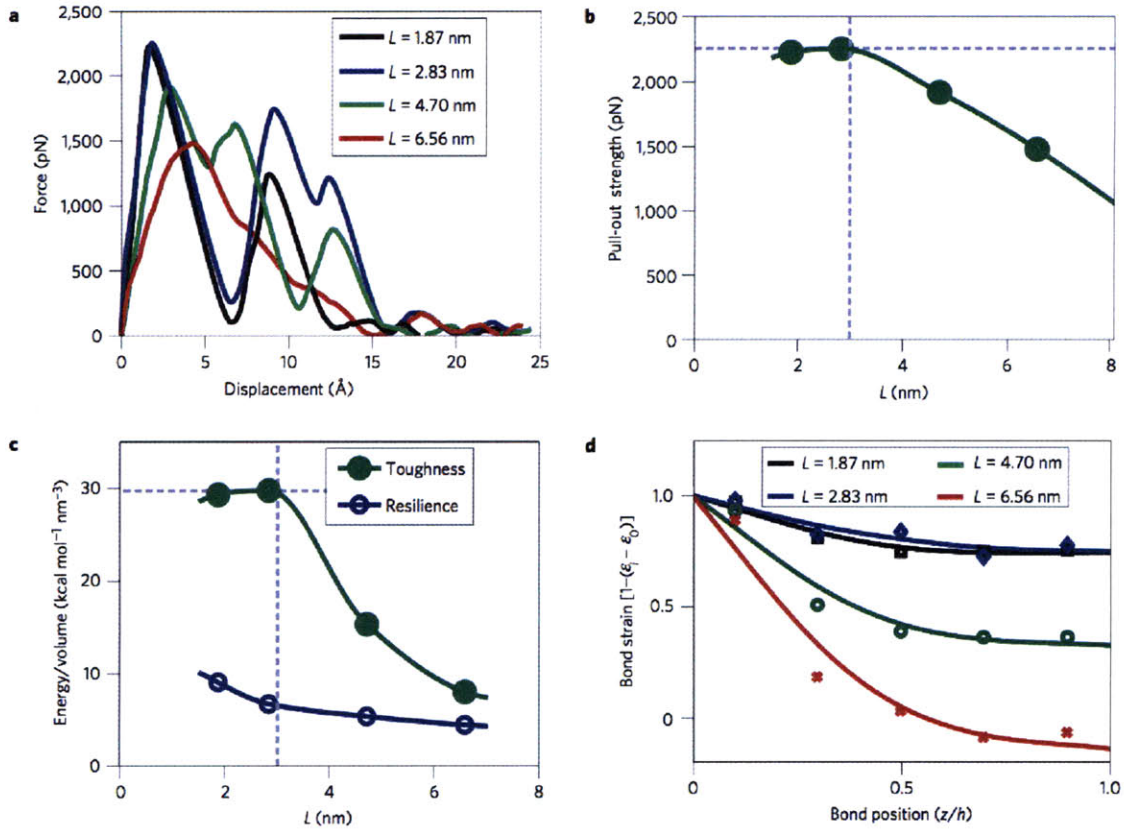

Figure 5-4: Strength, toughness, resilience and strain distribution in beta-sheet nanocrystals as a function of crystal size. Figure summarizes results from pullout simulations. Panel (a) shows the force-displacement profiles. Smaller systems exhibit multiple force-peaks due to a slip-stick motion that can only take place when the interacting surfaces in beta-sheet nanocrystals, adjacent beta-strands, are rigidly stabilized and act cooperatively. In this case the stiffer response of the system leads to higher pull-out forces and thus greater strengths (panel (b)) and significantly enhanced energy dissipation due to the slip-stick failure mechanism. The significance of this atomistic mechanism is evident from panel (c), where the representative toughness of the system is calculated as the area under the force-displacement curve. The results suggest that smaller beta-sheet nanocrystals are capable of dissipating significantly more energy during failure, which results in a high fracture toughness. Similarly, the resilience of the system - defined as the energy stored just before the initial rupture of H-bonds - is also greater for smaller size systems. In panel (d), an analysis of internal strain distribution in the beta-sheet nanocrystal is shown, as a function of crystal size. The plot shows the normalized H-bond strain over z/h, for different nanocrystal dimensions (load applied at z/h=0). This data clearly shows that the strain distribution becomes more homogeneous for small nanocrystals, but shows significant strain localization in large nanocrystals. The dashed lines illustrate the length-scales below which cooperation of the H-bonds is ensured through shear dominated loading, in agreement with bending simulations presented in Figure 5-3.

here as the elastic energy stored in the linear regime preceding the first rupture event, also becomes greater as the size is decreased, as shown in Figure 5-4(c). The molecular mechanism that leads to this change is directly evident from Figure 5-4(d), which shows the strain distribution of H-bonds along the section of the pulled strand preceding failure. As the size of the nanocrystal is reduced below $\approx$ 3 nm, the strain distribution becomes increasingly uniform and approaches an almost constant value throughout the entire section of the nanocrystal. The constant level of strain throughout small nanocrystals leads to homogeneous shear failure, in agreement with the observations reported in Figure 5-2(b). In contrast, the strain concentration that emerges at the loading point in large nanocrystals results in the formation of crack-like flaws that initiate catastrophic failure even at small loads, as shown in Figure 5-2(c).

An important implication of the model involves the length of individual beta-strands, represented by the parameter h (see Figure 5-1(c)). The strand length h is proportional to the number of amino acids in a beta-strand, since $h = NL_0$ (where L_0 $\approx$ 3 Å is the Cα-distance along a beta-strand, and N denotes the number of amino acids in a beta-strand). Since $s \sim h^2/L^2$ (eq. (5.4)), the shear contribution ratio s is expected to be higher for nanocrystals composed of longer beta-strands, leading to a greater shear contribution for a given nanocrystal size L. This scaling implies that modulating the beta-strand length could result in more rigid, tougher and more resilient structures. Specifically, the model predicts that for any given nanocrystal length L, a minimum strand length h could be found to reach a desired high level of s.

However, as shown in earlier studies the effectiveness of each beta-strand to carry homogeneous load under shear is limited, and by itself is a function of the strand length h [128, 131, 129]. While small groups of H-bonds in beta-strands can indeed work cooperatively under homogeneous shear, this mechanism breaks down beyond a critical number of H-bonds, at $N^* \approx 4$ (or equivalently, a critical strand length $h^* = N^*L_0$), due to a competition between entropic elasticity of the chain and the energetics of H-bonds that will lead to localized failure within each beta-strand

[128, 131, 129] (Figure 5-5(a) for a visualization of this effect). Most importantly, noting this limitation on h, the nanocrystal size cannot be increased arbitrarily, and a critical nanocrystal size L^* emerges at which it provides the best overall mechanical performance. An esimate for the critical dimensions can be given as $h^* \approx$ 1-2 nm and $L^* \approx$ 2-4 nm, defining length-scales at which all H-bonds in the nanocrystal gain a strong character through cooperativity (Figure 5-5). Notably, these critical dimensions are in agreement with experimental evidence that suggest that poly-Ala and poly-(GA) repeats typically span 4-12 amino acids (*i.e.*, in the range of $h = 1-3$ nm). It is very important to note here that the cooperativity of H-bonds in the formation of beta-sheet nanocrystals differs from cooperativity of H-bonds under mechanical load. This is due to the fact that the latter depends strictly on the stress-state of the H-bonds as shown here, which is also true for other beta-sheet structures [130, 23]

5.4 Discussion and implications for materials design

The main conclusion of this Chapter is that the nanoscale confinement of beta-sheet nanocrystals in silks plays a fundamental role in achieving great stiffness, resilience and fracture toughness at the molecular level of the structural hierarchy of silks, suggesting that smaller is stronger and tougher. In contrast to conventional belief, cooperative failure of H-bonds cannot be presumed *a priori*. Rather, the existence of cooperativity depends quite strongly on the size of the crystals and breaks down once beta-sheet nanocrystals exceed a critical size. Smaller beta-strand nanocrystals provide a greater stiffness and fracture resistance, as they are predominantly loaded in uniform shear, which leads to cooperative rupture of H-bonds and stick-slip energy dissipation mechanisms (Figure 5-4). Similar stick-slip mechanisms exist in other biological materials (e.g. wood or bone), and are also observed in metals in the form of dislocations that provide ductility [122, 84, 148, 128, 131, 129, 102]. Further contributions to the macroscale behavior of silk could arise from higher order effects due to

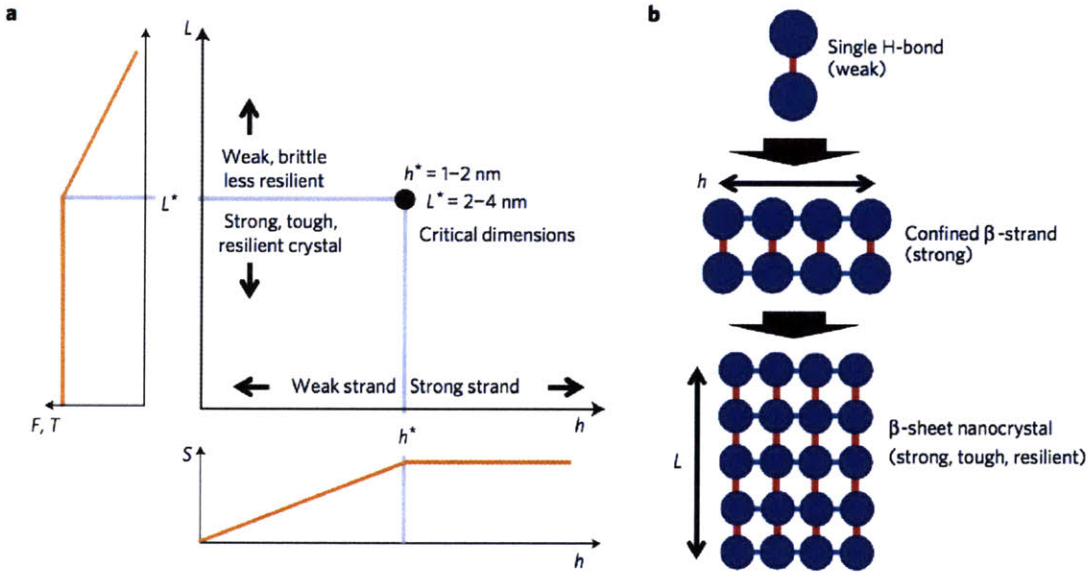

Figure 5-5: Hierarchical effects in the architecture of spider silk nanocrystals. The formation of confined beta-sheet nanocrystals with critical strand length h^* and critical nanocrystal size L^* provides maximum strength, toughness and stiffness. Panel (a) shows the schematic phase diagram to show the interplay of the parameters h and L in defining the properties of nanocrystals (S = schematic plot of the strength of a beta-strand as a function of strand length h [128, 131, 129], F = strength of nanocrystal as a function of crystal size L, T = toughness of nanocrystal as a function of crystal size L; both plotted schematically based on the results shown in Figure 5-4). While increasing the number of H-bonds in a beta-strand increases its mechanical stability for small numbers of H-bonds, the effect does not continue for beta-strands that contain more than ≈ 4 H-bonds, resulting in a plateau of the force S for lengths in excess of h^*. The physics behind this is a localization of deformation, similar to localization of shear slip in a crystal when dislocations are formed [122, 84, 148, 128, 131, 129, 102]. Panel (b) depicts a schematic illustration of how hierarchical structure formation in the strand length h and nanocrystal size L lead to the formation of high performance beta-sheet nanocrystals that combine strength, toughness and resilience despite being composed of structurally inferior, weak building blocks, H-bonds.

the hierarchical structure of the material [233, 84, 148, 102]. Since bending is a fundamental mode of deformation of beta-sheet nanocrystals that can be thermally excited [178], these findings could also be important for their thermodynamical stability.

From a slightly different point of view, the generation of a nanocomposite structure (Figure 5-5(b)) can be regarded as the deliberate placement of defects, which effectively results in flaw-tolerance [86]. Larger beta-sheet nanocrystals are softer and fail catastrophically at much lower forces due to crack-like flaw formation, as visualized directly in Figure 5-2. This catastrophic breakdown leads to rapid disintegration of silk fibers, which is further enhanced due to easier access of competing water molecules to H-bonds that facilitate rupture [219, 184]. This implies that silks with larger nanocrystals are weaker, in agreement with experimental studies [68]. Specifically, these results explain experimental findings that the reduction of the nanocrystal dimensions below 3 nm increases the ultimate strength and the modulus multiple-fold [68]. The mechanism underlying these experimental observations could be the brittle nature of larger beta-sheet nanocrystals. Initiating failure of small beta-sheet nanocrystals requires much higher forces and a significantly larger amount of mechanical energy. Furthermore, smaller crystals feature a self-healing ability until complete rupture occurs, which is attributed to the capacity of H-bonds to reform during stick-slip deformation. This mechanism protects backbone H-bonds from exposure to surrounding water. In the overall mechanical behavior of silk, strong, tough and mechanically resilient beta-sheet nanocrystals provide effective cross-links that contribute to the extraordinary macroscopic tensile strength and toughness of silks.

The approach illustrated here is generally applicable for predicting material properties of macromolecular nanocrystals and the shear transition length scale directly from atomistic simulation. The method uses generic parameters that can be calculated from experiments or atomistic simulations, and is in principle applicable to a broad range of nanostructures, including synthetic materials (e.g. polymer, ceramic or metal fibres). Moreover, the significance of shear contributions in nanocrystals is also relevant to protein fibres with a similar chemical makeup, such as amyloids, beta-solenoids, protein nanotubes or other fibrous beta-proteins that employ beta-strands

and may be linked to their significant stiffness and bending rigidity [140]. Indeed, there is evidence that shear effects are important in other biological filaments such as actin bundles and microtubules [54, 174]. The stiffness and stability of these structures may be improved slightly by controlling side-chain packing and steric zipper formation (e.g. poly-Ala vs. Gly-Ala repeats). but the scaling of the shear contributions is not be expected to vary considerably since Gly-to-Ala substitutions have been shown to have marginal influence on the mechanical properties of silk nanocrystals [253].

The size effect revealed here elucidates an efficient strategy to overcome the intrinsic brittleness and mechanical weakness of H-bonds by confining structures at controlled length-scales, which guarantees uniform deformation and concerted failure (Figure 5-5). Another consideration is that, given the seemingly simple nature of the amino acid composition of silk beta-sheet nanocrystals, their capacity to resist mechanical perturbation arises not from their specific chemical features but rather from universal features of the protein backbone combined with the high level of control over their structural dimensions. The utilization of weak H-bonds under nanoconfinement illustrates how a weakness is turned into a strength. Most engineered materials rely on strong (e.g. covalent) bonding, which requires considerable energy use during material synthesis that can also lead to catastrophic failure once bonds break. In contrast, the use of weak H-bonding facilitates self-assembly at moderate temperatures and provides a built-in capacity to self-heal since broken bonds can be reformed.

The application of these findings to the design of synthetic materials could provide us with new material concepts based on inexpensive, abundant constituents and facilitate the development of effective cross-linking domains. Since we are not limited technologically by natural building blocks, the potential to enhance biological materials beyond their natural capacity exists, perhaps through the combination with carbon nanostructures. Other opportunities include the incorporation of mutability, to develop materials whose mechanical properties can be controlled by external cues such as temperature, pH, magnetic or electric fields.

Chapter 6

Nanostructure and molecular mechanics of spider dragline silk protein assemblies

Spider silk is a self-assembling biopolymer that outperforms most known materials in terms of its mechanical performance. While experimental studies have shown that the molecular structure of silk proteins has a direct influence on the stiffness, toughness and failure strength of silk, no molecular level analysis of the nanostructure and associated mechanical properties of silk assemblies have been reported. This section focuses on a report of the atomic-level structures of MaSp1 and MaSp2 proteins from the *N. Clavipes* dragline spider silk sequence, obtained using replica exchange molecular dynamics, and subjected to mechanical loading for a detailed nanomechanical analysis. The structural analysis reveals that poly-alanine regions in silk predominantly form distinct and orderly beta-sheet crystal domains, while disorderly regions are formed by glycine rich repeats that consist of 3_1-helix type structures and beta-turns. The structural predictions are validated against experimental data based on dihedral angle pair calculations presented in Ramachandran plots, alpha-carbon atomic distances, as well as secondary structure content. A selection of structures is subject to mechanical loading, which reveals that distinctly different hydrogen bonded regions and the type of secondary structure control the mechanical response of silk

at the nanoscale. Both structural and mechanical characterization results show excellent agreement with available experimental evidence. These findings set the stage for extensive molecular-level investigations of silk, which may contribute towards an improved understanding of the source of the strength and toughness of this biological superfiber.

6.1 Background on spider silk ultrastructure and mechanics

Chemistry and nanoscale features of biological materials are crucial to understand the source of their mechanical properties such as strength, failure mechanisms and elasticity [84]. In protein materials, the primary structure of macromolecules, consisting of a linear sequence of individual amino acids describes the chemical specificity of the interactions at the molecular level, which give rise to formation and failure characteristics of the material [37, 148]. Spider silk is an extraordinary material that surpasses most synthetic fibers in terms of toughness through a balance of ultimate strength and extensibility [247, 212, 11, 214, 233]. The source of silk's unique properties has been attributed to the specific secondary structures of proteins found in the repeating units of spider silk proteins [103], which assemble into a hierarchical structure as shown in Figure 5-1(a).

Experimental studies have thus far primarily focused on developing a mapping between repeating sequence units of spider silk and the basic structural building blocks of fibrils. Two distinct proteins are typically found in dragline silks with similar sequence across species [88]. One of the most studied silk from spiders, *N. Clavipes* dragline silk contains MaSp1 and MaSp2 proteins, with different repeat units and possibly distinct mechanical functions [108, 24, 103]. MaSp1 contains glycine (Gly or G) rich Gly-Gly-X (GGX) repeats with poly-alanine (Ala or A) and GA domains, where X typically stands for alanine, tyrosine, leucine, or glutamine. Whereas MaSp2 also contains poly-Ala domains, it has a repeat unit with high proline content in

the form of (GPGQQ/GPGGY). The proline rich segments are intrinsically twisted which inhibit edge-to-edge aggregation of strands, thereby controlling the location and size of beta-sheet nanocrystals in silk. These segments, bearing sequence resemblance to elastin, are thought to form beta-spiral or type II beta-turn structures that provide extensibility through hidden length formation and control the unique thermomechanical properties of spider silk [93, 204, 103]. Earlier studies have suggested that MaSp1 is more dominant in the composite morphology of the spider dragline silk than MaSp2, with a ratio of approximately 3:2 or higher, depending on the species [24, 106, 97, 220]. Recent investigations revealed that anti-parallel beta-sheet crystals at the nanoscale, consisting of highly conserved poly-(Gly[G]-Ala[A]) and poly-Ala repeats found in both commercial and spider silk [103], play a key role in defining the mechanical properties of silk by providing stiff orderly cross-linking domains embedded in a semi-amorphous matrix with less orderly structures [151, 234, 240]. Earlier studies have shown that hydration level and solvent conditions such as ion content and pH play a role in the structure and mechanical properties of silk proteins[188, 64]. For instance, a unique aspect of silk fibers is their capacity to exhibit a dramatic reduction in length upon hydration; a phenomenon known as supercontraction [211, 241].

Beta-sheet nanocrystals that employ dense network of hydrogen bonds [129, 135] have dimensions of a few nm and constitute at least 10-15% of the silk volume, while with less orderly extended structures the beta-sheet content can be much higher for most silks [68, 96, 200]. The existence of 3_1-helices, and beta-turn or beta-spiral conformations has been suggested as models for the amorphous domains [151, 181, 234]; however, no definite atomistic level structural model has yet been reported. It is anticipated that novel statistical mechanics approaches [185], experimental methods, such as x-ray diffraction and scattering [197, 238], solid-state NMR [241, 108, 214] and Raman spectroscopy [151, 200, 201], combined with pioneering multiscale atomistic modeling methods such as those based on density functional theory [135, 184] or molecular dynamics [26, 42, 135, 158] will provide more insight into the atomic resolution structure for complex materials such as spider silk. However, due to the lack of current large-scale atomic resolution models, the links between genetic makeup,

chemical interactions and structure, as well as their associated macroscale mechanical properties remain obscure. Earlier atomistic studies have focused solely on the well-defined crystalline regions consisting of poly-Ala (spider silk) or poly-Gly-Ala repeats (commercial silk) [253, 135], and studies on the links between structural and nanomechanical features of the nanocomposite structure of silk have remained a challenge.

6.2 Simulation approach

In this work, atomistic simulations are carried out to identify nanostructural models of spider silk proteins, with the goal to develop a link between genetic sequence and resulting mechanical properties. The challenges of reaching native (equilibrium) structures within the time-scales accessible to conventional molecular dynamics simulation require enhanced sampling methods such as replica exchange molecular dynamics (REMD) [224] . REMD method (see Chapter 2) is used to investigate the structures formed by assemblies of short segments of MaSp1 and MaSp2 proteins. Along with other protein structure prediction approaches [19, 258], REMD is considered to be an effective tool for investigating folding and aggregation of proteins, as it reduces the likelihood of kinetic trapping at non-native states [203]. Through a fast search of the conformation space at high temperatures and more detailed investigation at low temperatures, it allows the system to overcome energy barriers and local minima corresponding to non-native structures [78, 166, 189, 192] of proteins and allows identifying native protein structures from the amino acid sequence, with atomistic resolution. The overall approach is summarized in Figure 6-1(a).

The computational approach used here consists of two steps, (i) structure identification with validation against experimental results, and (ii) mechanical stretching of the resulting structures, with subsequent analysis of mechanical properties and deformation mechanisms.

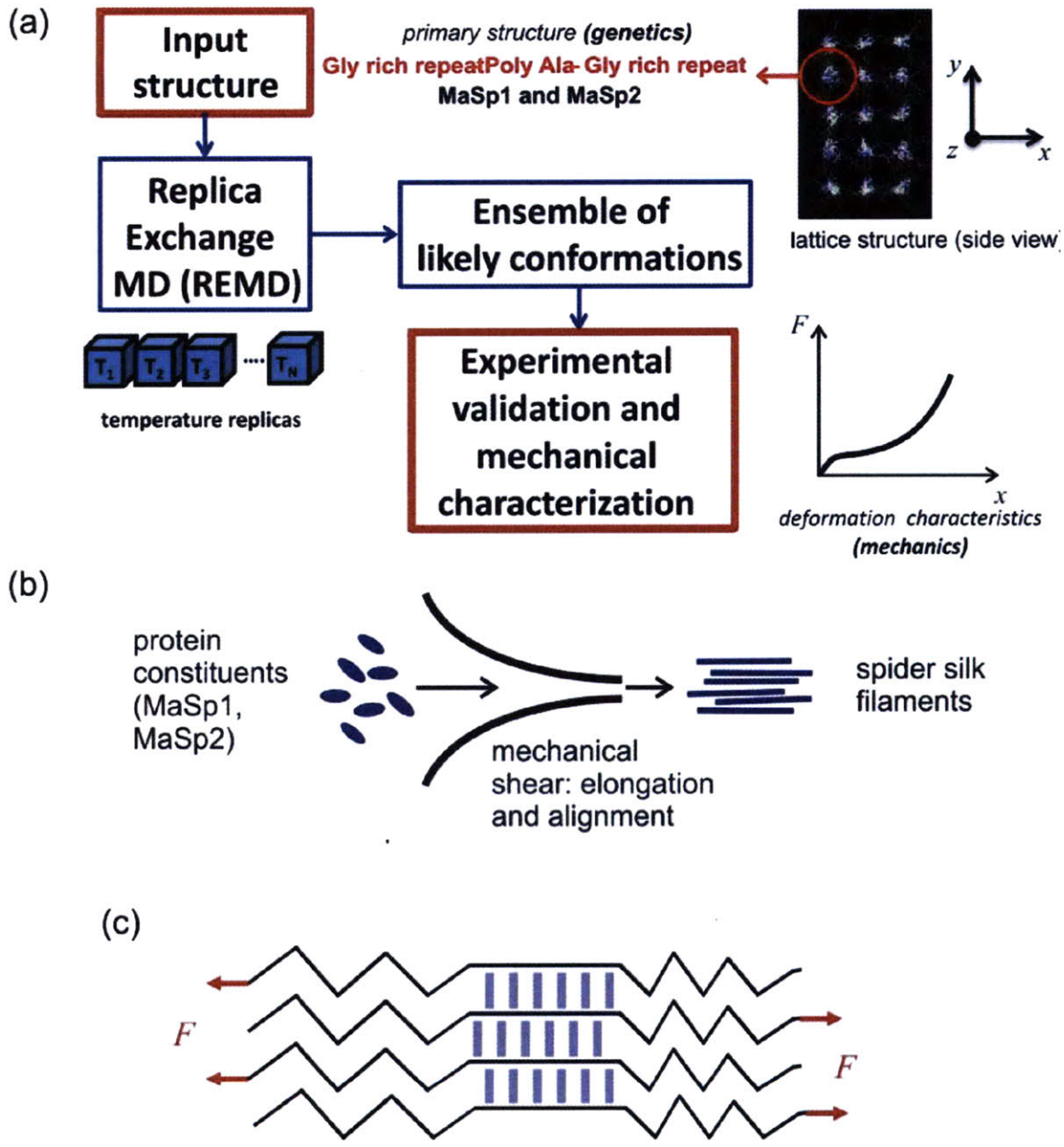

Figure 6-1: Simulation protocol and representative structure results. Panel (a) summarizes the approach taken here to identify the nanostructure of spider silk proteins, here focused on the MaSp1 and MaSp2 silk sequences from the *N. Clavipes* spider. Monomers representing sections of the MaSp1 and MaSp2 proteins (containing a poly-alanine repeat in the center) are used as the basic building block. Replica exchange simulations are carried out at multiple temperatures, and an ensemble of most likely, final structureas are compared with experimental evidence. Panel (b) illustrates the natural process of silk assembly (and fiber formation) during which silk proteins are subject to shear. The natural process of shearing and alignment of protein monomers motivates the choice of the initial geometry shown in panel (a). Panel (c) shows the mechanical loading condition employed in these simulations.

6.2.1 Structure identification

Primary protein structures (polypeptide sequences) are created with the amino acid sequence of the *N. Clavipes* MaSp1 and MaSp2 proteins, which constitute the majority of the silk's core [108]. The MaSp1 sequence is (in one-letter amino acid codes):

GGAGQGGYGGLGSQGAGRGGLGGQGAG**AAAAAA**

GGAGQGGYGGLGSQGAGRGGLGGQGAG.

The MaSp2 sequence is:

GPGQQGPGGYGPGQQGPGGYGPGQQGPSGPGS**AAAAAAAA**

GPGQQGPGGYGPGQQGPGGYGPGQQGPSGPGS

The monomers consist of two glycine-rich repeating units surrounding a poly-alanine segment to represent a single monocrystal. MaSp1 protein consists primarily of GGX type repeats followed by a GA and poly-Ala region. MaSp2 on the other hand has proline-rich GPGQQGPGGY repeats, and a poly-Ala region. The starting configuration is a lattice structure consisting of anti-parallel arrangement in one direction and parallel arrangement in the other. Previous findings on poly-alanine aggregation suggest that anti-parallel orientation in the hydrogen bonding direction and parallel stacking in the side-chain direction lead to stable beta-sheets [157], hence such an arrangement is considered here to be a good starting point for obtaining assembled structures by silk proteins. In the initial setup, each strand is separated by 10 Å in a square lattice (see Figure 6-1(a)). The simulations start from an extended conformation, which is very relevant in the processing of silk where elongational flow in the spinning duct leads to stretching and alignment of monomers in the concentrated dope, as shown in recent experimental work [188] (see Figure 6-1(b) for a schematic). It should be noted that higher temperature replicas can allow wide sampling around this basic orientation, where strands can rearrange in anti-parallel or parallel fashion as the strands can diffuse within the lattice. At high temperature replicas, weak interactions between the strands can be broken; i.e. the system melts and reforms a wide range of structures that can be investigated in detail at lower temperatures. A key hypothesis here for using an initial orderly lattice arrangement

is that edge-to-edge aggregation of strands - similar to amyloids [125] -, is the driving force for formation of mono-crystals, rather than local folding of the backbone onto itself through short turn structures. This is supported by recent experiments, which suggest that large extensional forces during spinning, and high concentration of the dope are requirements for crystal formation ([188]). This assumption is reasonable in the context of a monocrystal study as pursued here, and does not rule out the possibility of the formation of folded intramolecular contacts at longer length scales. Experimental evidences also suggest that strand orientation in the proximity of the crystals is aligned more or less parallel the fiber axis, also supporting this initial configuration [201, 240].

Structure prediction simulations are carried out with Langevin dynamics using CHARMM [25] and the EEF1.1 force field with a Gaussian effective solvent energy function [147]. The REMD protocol is setup using the MMTSB toolset [77]. A simulation time step of 2 fs is used by employing the SHAKE algorithm for hydrogen atoms. Solvent friction is added via a Langevin friction term that allows for high mobility and conformational sampling. While the EEF1.1 model has some particular modifications and simplifications on solvent, side-chain and hydrogen bond interactions, it has the benefit of being orders of magnitude faster than other implicit or explicit solvent models, and the solvent volume exclusion model is particularly attractive for large-scale assembly processes as in the case of spider silk. Since force fields are generally parameterized for room temperature calculations, ensemble structures belonging only to the lowest temperature replica are used, and higher temperature replicas serve only the function of overcoming kinetic trapping and fast conformational search in the REMD scheme.

Initial structures are oriented along the main chain axis using built-in functions in CHARMM (see geometry shown in Figure 6-1). Long initialization runs are performed achieve distinct starting configurations to enhance better sampling in the production run. This is followed by a production run starting from the final configurations of the replicas from the initialization run and using an exchange time step of 2 ps to allow the relaxation of the system. Starting from the final configurations of the

initial run, simulations are carried out on 64 replicas distributed over the 300-650 K temperature range (the high temperature replicas ensure that the protein can sample a wide range of conformations, specifically including those that resemble the initial structural ordering of strands). Each replica is simulated for a total of 10 ns, corresponding to a total simulation time of 640 nanoseconds.

An ensemble of structures is analyzed from the last 1,000 exchanges of the production run from the lowest temperature replica, and is used to extract the lowest energy structures (coordinates of all structures generated from the simulations are available upon request). Representative final structures are selected based on a clustering algorithm based on mutual similarity according to root mean square deviation for each structure. The K-means clustering algorithm is used here (from the MMTSB toolset [77]) to identify the largest clusters that have approximately 10% or more presence in the selected set, corresponding to the top 5 clusters. From the selected clusters, structures closest to the cluster center are selected as the representative models. Analysis on dihedral angles and alpha-carbon distances are carried out using the complete data set to achieve better statistical representation. The secondary structure content is calculated in Visual Molecular Dynamics (VMD) [111] using the STRIDE algorithm.

6.2.2 Mechanical analysis

Mechanical stretching tests are done using a constant loading rate of 2 pN applied every 20 ps, that is, a loading rate of 0.2 N/s. Half of the chains are randomly selected to be pulled in $+z$ direction while the other half is pulled in $-z$, to impose the characteristic lateral loading of the crystals as experienced in the native silk structure (see schematic shown in Figure 6-1(c)).

Force-extension plots are based on the forces applied, and the measured distance between the center of mass of pulled atoms is computed. The analysis on hydrogen bond dynamics is computed using .tcl scripts.

6.3 REMD and mechanical stretching simulation results

This section summarizes the findings from REMD simulations on example spider silk building blocks. Structural predictions are presented and compared with experimental evidence. The focus is primarily on dihedral angles, since this approach allows for a direct quantitative comparison that is independent of the size of the system studied. Since the majority of silk consists of glycine and alanine amino acids, a comparison of the dihedral angles of glycine and alanine residues with experimental data on these amino acids in spider silk provides the most reliable validation for the simulations.

6.3.1 Predicted representative structures and experimental validation

For MaSp1 alanine residues (see Figure 6-2(a)), it is observed that the most common phi-psi angle value is around (-150,135) in excellent agreement with experimental findings that suggested (-135,150), corresponding to a beta-sheet structure. For glycine (see Figure 6-2(b)), the results show a symmetry around the origin, exhibiting a wide distribution around approximately (+/-75,-/+75), in agreement with experimental findings around (+/-60,-/+135) [240] that also have symmetry. Similar results are observed for MaSp2 (Figure 6-2(c-d)), however, a wider distribution of glycine dihedral angles is evident from the Ramachandran plot.

While mapping of the poly-alanine regions to beta-sheet conformation is straightforward from the data on alanine residues for both structures, the glycine regions require more data to be able to distinguish them within common protein structures. A close look at proline residues in MaSp2 shows peaks around (-60,-30) and (-60,120), corresponding to type I and type II beta-turns that would incorporate adjacent Gly residues. From the ensemble of structures obtained from MaSp1 replica exchange simulations, alpha-carbon atom distances are computed between glycine residues that are three residues apart, to characterize the basic repeat unit length of the GGX

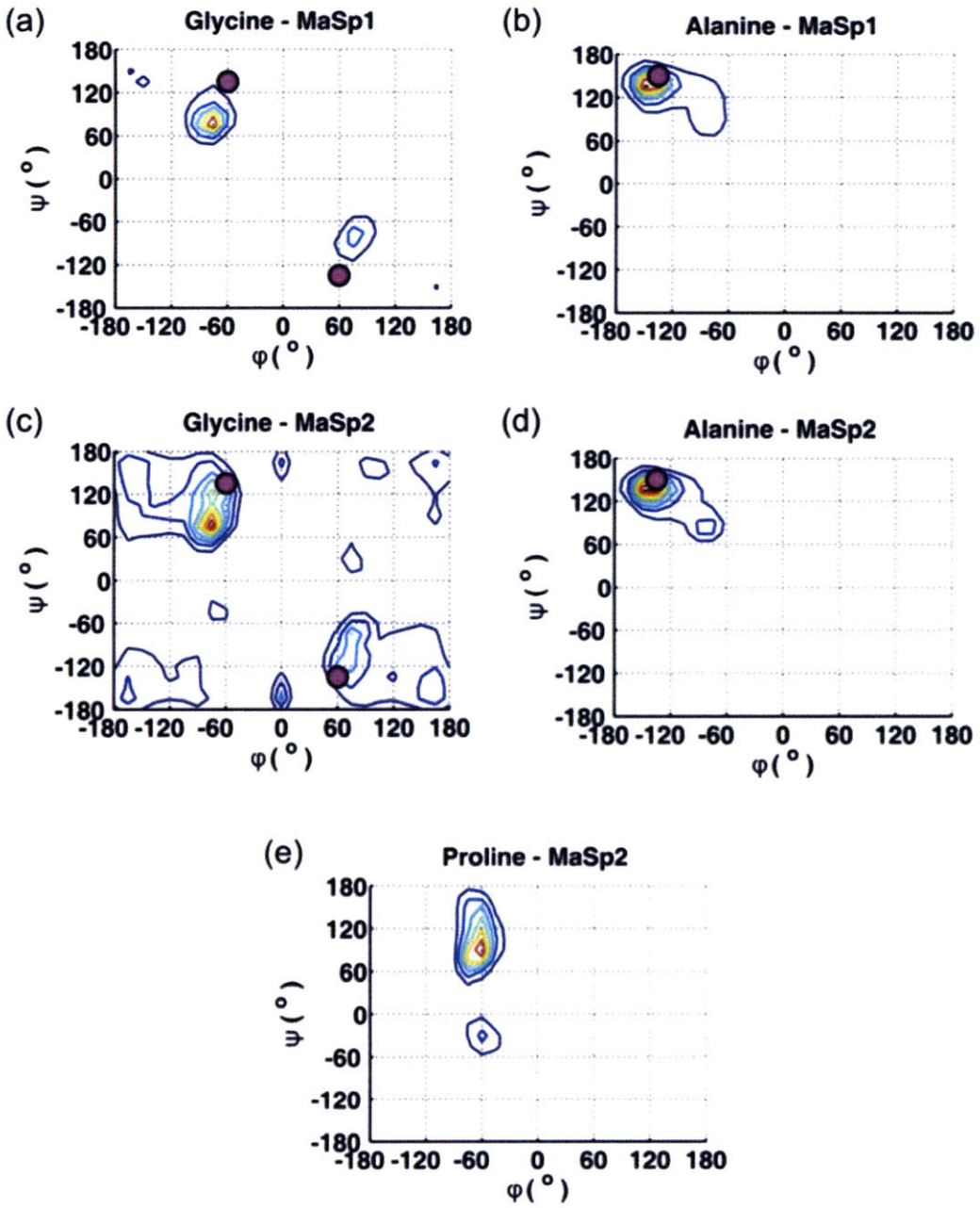

Figure 6-2: Comparison of resulting structures for MaSp1 and MaSp2 with experimental data for validation. Panels (a-d): For Ala residues, the most common phi-psi angle value is around (-150,135) in excellent agreement with experimental findings that suggested (-135,150) that correspond to a beta-sheet structure. For Gly residues, a wide distribution around approximately (+/-75,-/+75) is observed with symmetry around the origin, in line with experimental findings around (+/60,-/+135) that also show a symmetric distribution. The wider distributions in MaSp2 may be due to a more amorphous structure caused by high proline content. Panel (e): Dihedral angle distribution of Proline residues cluster around (-60,-30) and (-60,120), which correspond to type I and type II beta-turn conformations respectively.

domains. The probability distribution of C$\alpha(i)$-C$\alpha(i+3)$ distances measured from atomic coordinates has a distinct peak around 9 Å. This value is greater than most known beta-turn structures and falls short of the anti-parallel beta-sheet conformations that typically exceed 10 Å. These domains have interchain hydrogen bonding that is less oriented and sparser. Based on these observations, it is also evident that these disorderly structures in MaSp1 resemble the characteristics of 3_1-helices, which are the most likely conformations taken by the Gly-rich domains, supporting earlier hypotheses reported in the literature [144, 240]. The density of hydrogen bonding in these regions is significantly reduced compared with the orderly network in beta-sheet nanocrystals, suggesting that moderating the number and orientation of intermolecular interactions leads to the characteristic heterogeneous structural arrangement observed in silks.

For illustrative purposes, a representative collection of structures obtained from the simulations is shown in Figure 6-3 where percentages of different secondary structures are illustrated in each sub-panel (I-V). For both MaSp1 and MaSp2, a significant percentage of residues are found to be in semi-extended, disorderly conformations in agreement with the discussion put forth based on experimental studies [188]. These results suggest that poly-alanine regions have an extremely high propensity for aggregating into crystalline beta-sheet structures by inter-strand hydrogen bond formation and intersheet stacking in the side-chain direction. Along with Glycine regions that also form extended regions, the beta-sheet percentage ranges from 28% to 55% for the sequences studied in this work. These results compare well with recent NMR studies on dragline silk, which have indicated 34% beta-sheet content [117]. This highly orderly domain is dispersed within the glycine-rich repeat units, which are still fairly oriented but much less-orderly, forming more amorphous structures. A key finding from the secondary structure content analysis is the lack of any alpha-helix conformation of the MaSp1 and MaSp2 silk constituents, supporting a wide range of experimental evidence that ruled out this conformation [108, 116, 117, 240]. Instead, results presented here suggest that disorderly structures resembling 3_1-helices and beta-turns dominate non beta-sheet conformations in these proteins. The higher

content of proline in MaSp2 leads to more disorderly structures in amorphous regions and well-defined beta-sheet crystal regions. This is evident from both the lower ratio of beta-sheets in this sequence, and also the shorter lengths of beta-strands when comparing MaSp1 structures in Figure 6-3(a) with those of MaSp2 in Figure 6-3(b).

Altogether, the structure identification and comparison with available experimental results suggest that the models obtained from the simulations resemble those found in native spider silk. Overall, crystal structures consisting of poly-Ala repeats have the size of 2-3 nanometers in the chain direction, with partial beta-sheet domains in semi-crystalline regions, particularly for MaSp1. Stacks of 2 to 4 sheets (1-2 nm) and up to 8 strands per sheet (2-4 nm) are observed in simulations (but these values may increase if the number of chains considered is increased). The amorphous domains are semi-extended, and the average lateral length per residue for both structures emerges to be around 2 to 3 Å, which is shorter than fully extended structures that would require more than 3 Å length. Larger folds, where sheets are formed by polypeptide strands with self-interactions are not observed here, but such morphologies cannot be ruled out considering that actual spider silk sequences are much longer and therefore more flexible than the short segments studied in this work. This study sets the stage now to explore the mechanical properties of both the MaSp1 and MaSp2 structures using ultra-large-scale molecular dynamics simulations, which is discussed in the next section.

6.3.2 Molecular mechanics of dragline silk protein assemblies

This section summarizes findings from constant loading rate mechanical shear simulations on the selected structures, where randomly selected strands are pulled in opposite directions to mimic the relevant lateral (shear) loading conditions of small crystals in spider silk. The resulting force-displacement curves from these simulations are shown in Figure 6-4(a-b). Both MaSp1 and MaSp2 show a characteristic smooth curve with three regimes, where the relative moduli of these regimes depend on the secondary structure content. . For molecules with high turn ratio and low beta-sheet content (see MaSp1-IV and MaSp2-II), an initial stiff regime, followed by

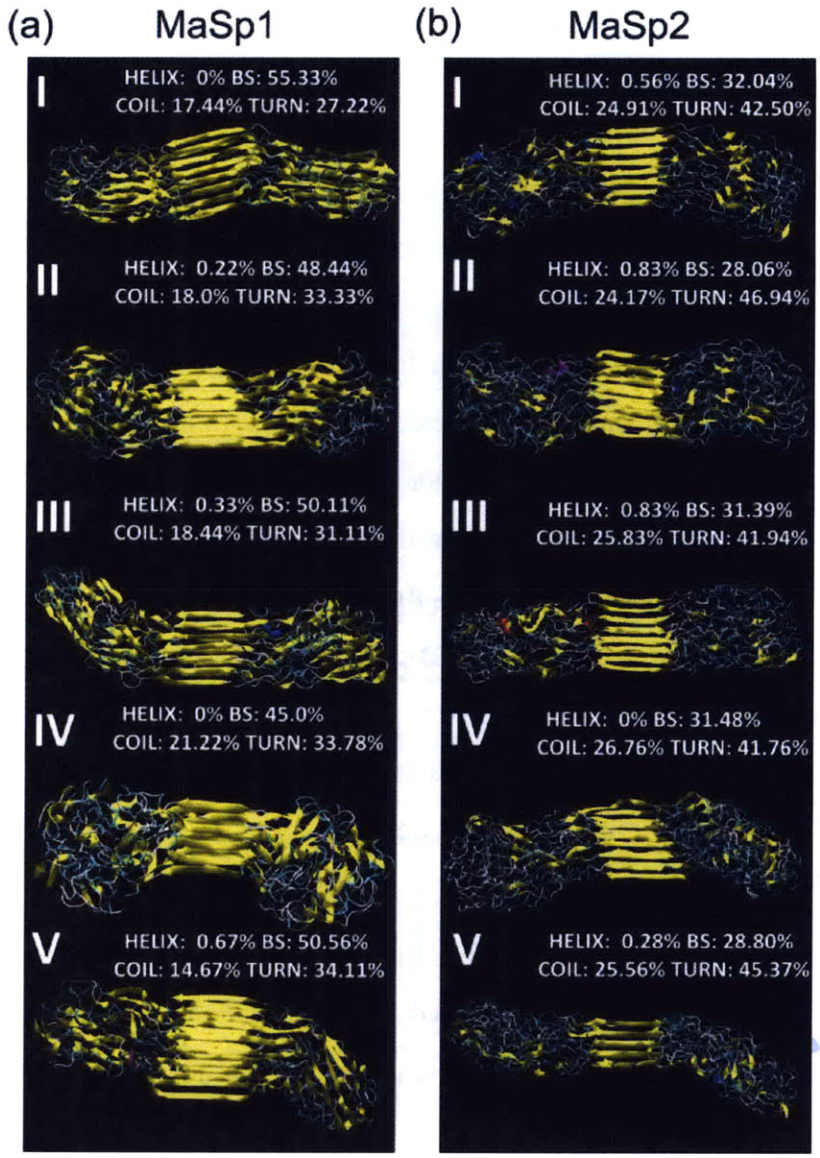

Figure 6-3: Secondary structure distribution of selected replicas. Representative replicas, selected according to clustering method, and the ratio of different common secondary structures are presented here. Panels I-V indicate the five most likely structures selected for each sequence. Panel (a): The majority of the structures are observed to be in beta-sheet or beta-turn conformation, where for MaSp1 beta-sheet content is higher due to the lack of proline residues that reduce chain aggregation in sheet form. The relative content of secondary structure controls the mechanical properties of protein materials; where greater crystallinity typically means greater strength and turn structures provide hidden length required for extensibility and toughness. Panel (b) illustrates representative structures from replica exchange simulations. Percentages of different secondary structures are illustrated in each sub-panel (I-V). The coloring is based on structural configuration, where yellow represents beta-sheet and extended structures. Insets illustrate the stacking formation of the beta-sheets in the poly-Ala regions. The results consistently illustrate that poly-Ala regions form highly orderly beta-sheet crystals whereas the Gly rich repeat units are less-orderly, forming more amorphous domains.

a softer regime, followed by a very stiff regime leading to failure are observed. On the other hand, systems with very high beta-sheet content (MaSp1-I, MaSp2-I) show a monotonically stiffening force-extension response. This suggests that the characteristic yielding behavior at the molecular level of the hierarchy of silk is controlled by the ratio of turn to beta-sheet structures, where a higher turn ratio leads to the emergence of the yielding phenomenon. This is an alternative means of achieving high initial stiffness in comparison with for instance MaSp1-I structure, which has more beta-sheet content and therefore exhibits higher initial stiffness but much less extensibility. MaSp1 also exhibits a larger deviation in the initial end-to-end length, owing to the wider secondary structure distribution. The effect of this is evident from the analysis displayed in Figure 6-4(c), which illustrates the yielding behavior of both structures and much larger variation in response in MaSp1, where increasing the turn content and inter-chain hydrogen bonding increases the initial stiffness and extensibility of the assembly, thereby improving the toughness as well. The failure strength of MaSp1 and MaSp2 seem to be similar in these simulations, as shown in Figure 6-4(d).

The force extension curves obtained here also show minor deviations from the inextensible chain models commonly used for polymer materials. At low forces, this is due to the rearrangement and rupture of bonds in the amorphous domains. At high force, the crystal morphology can change due to large shear stresses as well as transverse compression, where the latter causes buckling and the collapse of the sheet into a smaller, more compact formation.

Some of the structural transformations and failure mechanisms can be observed from the trajectories obtained from stretching simulations of the spider silk assemblies. Figure 6-5(a-b) show snapshots of deformation from MaSp1 and MaSp2 simulations. As evident from the snapshots, amorphous domains stretch significantly with applied force, and a transition from turn to beta-sheet structures are observed for both MaSp1 and MaSp2 (regime just before the point marked by a red arrow in Figure 7). The percent change seems to be larger for MaSp2, since in their unstretched configuration, proline residues act as beta-sheet breakers. A key observation is that failure of the

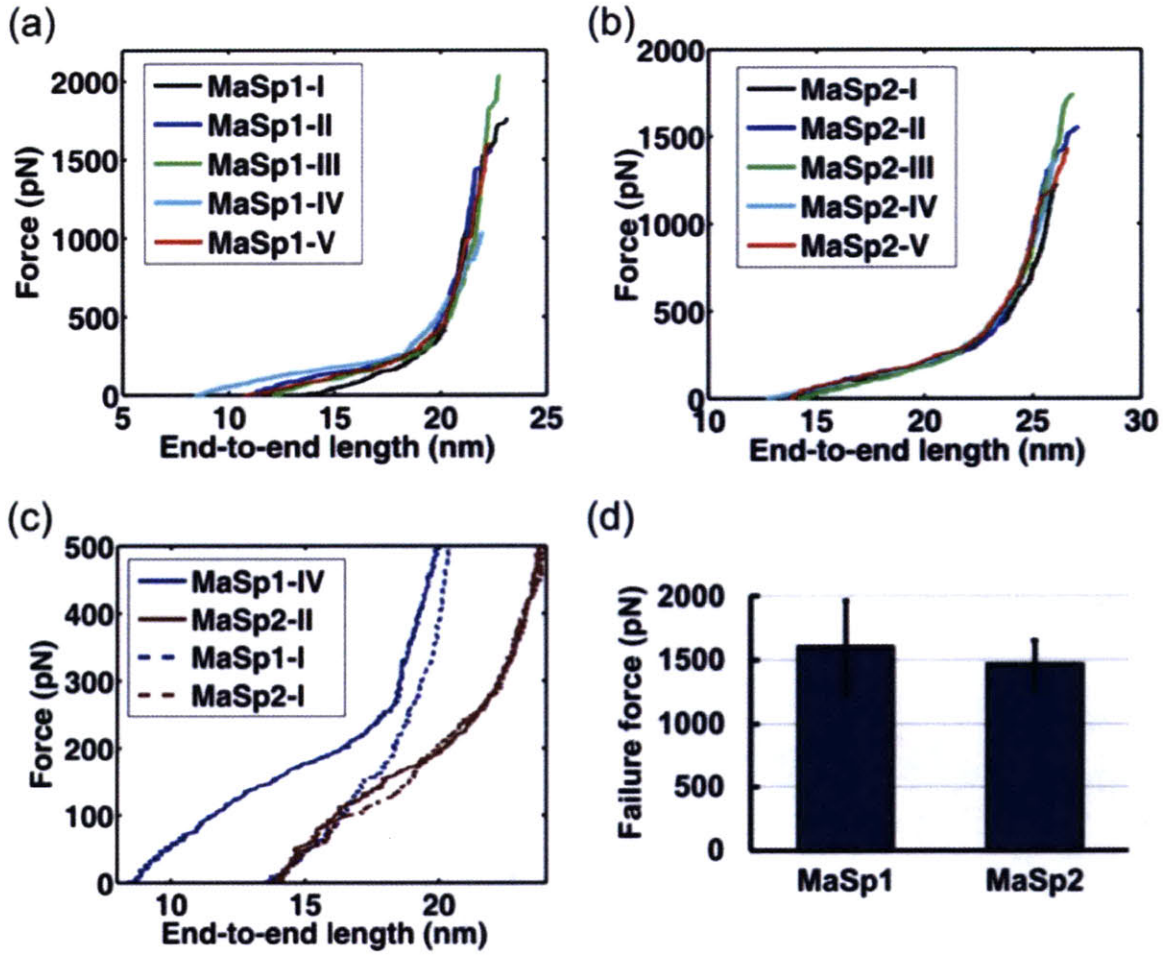

Figure 6-4: Force-displacement curves for selected structures. Figure shown illustrates the response of selected structures to shear forces applied to alternating strands. Force values shown are force applied per polypeptide strand. Panels (a-b) illustrate curves obtained from MaSp1 and MaSp2. The forces cause tensile stretching of the strands, where a strain stiffening behavior is evident once the chain reaches a certain length, independent of the chain's initial stretch state. The responses are similar for MaSp1 and MaSp2, however, it depends on the secondary structure content of the system. As shown in panel (c), MaSp1 structures have a large variation on beta-sheet vs. turn content, which leads to distinctly different mechanical responses. Solid lines shown cases having the largest turn content, whereas dashed lines illustrates structures with more beta-sheets. As the turn ratio increases, an initial stiff regime is observed followed by softening, followed by stiff bond stretching regime. For extended structures, the initial stiff regime disappears, and the typical strain stiffening behavior of polymer chains can be observed. The source of this difference is the existence of denser hydrogen bonding in amorphous regions due to turn formation, which leads to higher stiffness and energy dissipation for structures containing more turns. Lower variation of turn and beta-sheet content in MaSp2 leads to the reduced variation of the mechanical response for this structure. As can be inferred from panel (d), the failure strength of both structures is similar. Since the failure strength is controlled largely by the crystals, which have almost identical sequence in both cases, the morphology of the crystal plays most likely the key role in determining the failure strength.

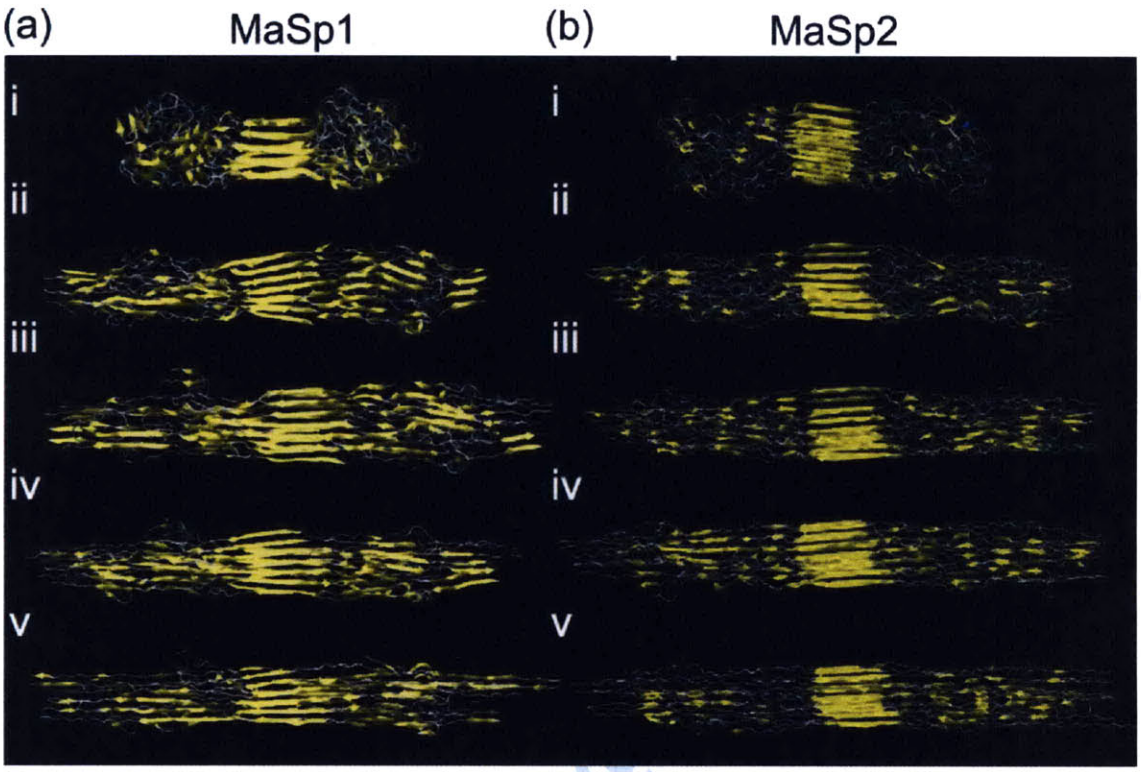

Figure 6-5: Stretching and structural transformation of the proteins. The figure illustrates the stretching behavior of the amorphous domains and crystals under shear forces for MaSp1 (panel (a)) and MaSp2 (panel (b)). Panels (i-v) illustrate the time sequence of events during the stretching simulations. As evident from the time sequence of snapshots (i-v, notation different here to differentiate time sequence and predicted structures shown in Figure 6-3), amorphous domains stretch significantly, and a transition from turn to beta-sheet structures are observed. The GGX repeats in MaSp1 are capable of forming beta-sheets during stretching, whereas this is observed to a lesser extent in MaSp2. A key observation is that failure of the system happens by sliding of strands with respect to each other upon breaking of the hydrogen bonds and side-chain contacts in the crystalline domain. This typically occurs at the interface region with solvent at the boundary of the crystal, leaving part of the crystal intact even after failure of the structure.

system happens by sliding of strands with respect to each other, which can occur only upon breaking of the hydrogen bonds in the crystalline domain. In the MD simulations, failure occurs at the interface region with solvent at the boundary of the crystal, leaving part of the crystal intact even after failure.

Representative plots are shown from MaSp1 and MaSp2 structures with highest turn content are shown in Figure 6-6. A comparison of Figure 6-6(a) and Figure 6-6(b) shows the strain softening behavior that occurs when the turn/beta-sheet ratio is high in both structures. Upon initial yielding, number of hydrogen bonds in the system and also the turn content decreases, as shown in Figure 6-6(a) and Figure 6-6(b) respectively for MaSp1 and MaSp2 proteins. Further stretching leads to a much stiffer regime due to stretching of the covalent backbone (regime after the point marked by a red arrow in Figure 6-6), which initiates rapid rupture of many hydrogen bonds in the amorphous domains. The system fails upon breaking of hydrogen bonds in the crystal, and the subsequent sliding of strands. The constitutive sigmoidal behavior of the force-displacement graphs observed here for both structures agrees qualitatively with the micro-scale response of spider silk.

6.4 Development of a constitutive law

This section focuses on the development of a constitutive law for spider silk's building blocks based on atomistic simulation results presented. The goal with this approach is to provide a fundamental description of spider silk mechanics using a bottom-up perspective, and thereby elucidate the design strategy behind the making of silks. For this reason, a simple coarse-grained model whose parameters are directly informed from atomistic results can be utilized to extend the findings from atomistic simulation. A combination of beta-sheet nanocrystals and semi-amorphous regions is modeled by two nonlinear springs in serial arrangement. Each constituent has a mechanical signature informed from full atomistic simulation (see Section 6.3) [135, 127, 132], as shown in Figure 6-7. This model, albeit simple, is capable of describing the key features of the nanomechanics of spider silk, without the introduction of any exper-

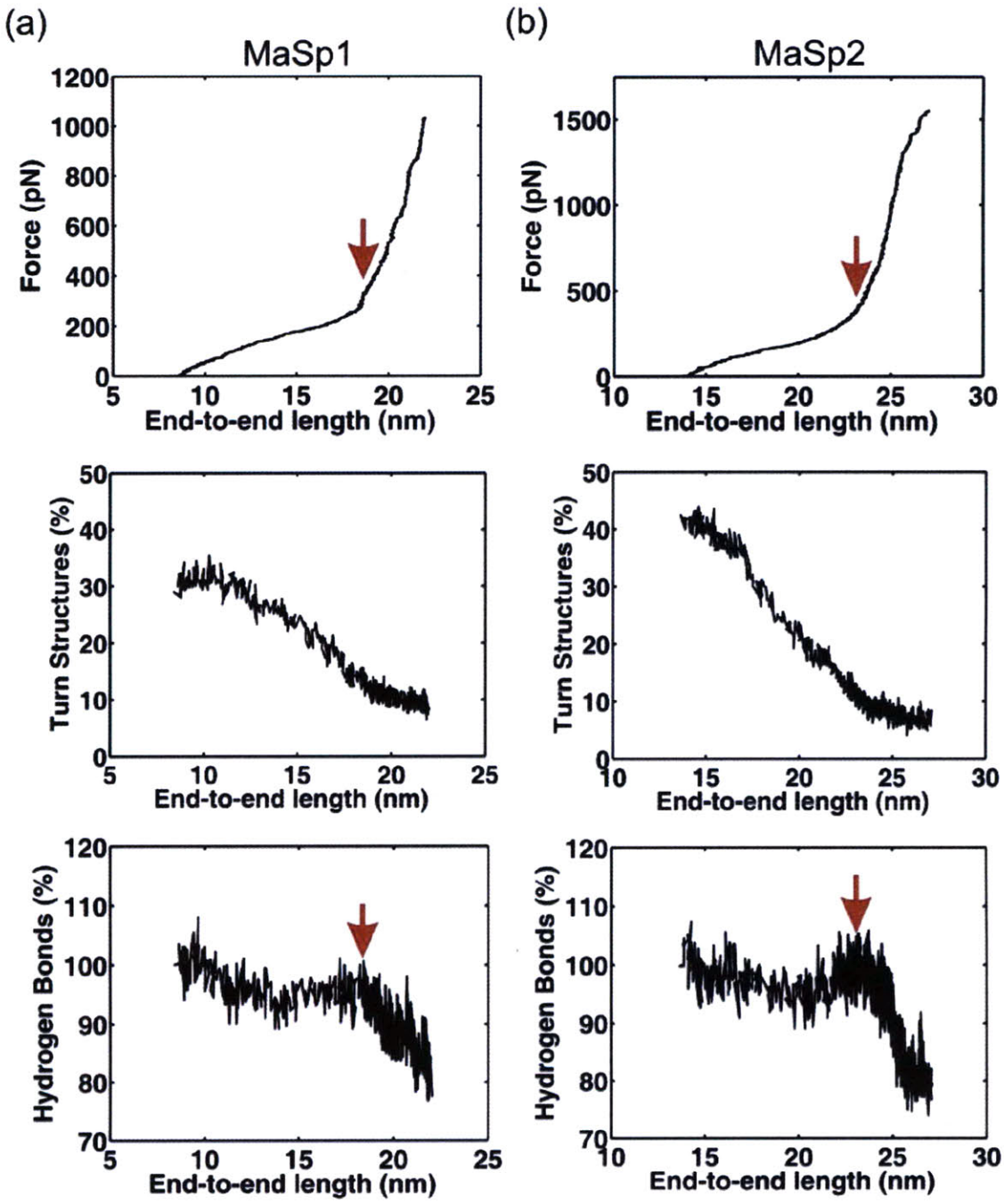

Figure 6-6: Representative plots are shown from MaSp1 and MaSp2 structures with the highest turn content. Comparison of panel (a) and panel (b) shows the strain softening behavior that occurs when the turn/beta-sheet ratio is high in both structures. Upon initial yielding, the number of hydrogen bonds in the system and also the turn content decreases. Further stretching leads to a much stiffer regime, which corresponds to rapid rupture of many hydrogen bonds (shown with a red arrow) in the non-crystalline domains. The system fails upon breaking of hydrogen bonds in the crystal, and sliding of beta-strands. The characteristic sigmoidal force-extension behavior observed here for both structures shows resemblance to the macroscale response of spider silk.

imental parameters. The serial spring assumption is reasonable at the nano-scale, since in the sequence of spider silk, crystalline domains are followed immediately by the glycine-rich repeats that form the semi-amorphous regions, forming a serial constitutive unit that is the fundamental building block of more complex hierarchies [214]. Since the basis of the model is atomistic simulation, the spring constants developed take into account the effects of solvent and molecular friction directly, and additional parameters for viscosity and other effects are therefore not considered in this work.

The constitutive relationship of the two domains in silk, derived from molecular dynamics simulations, is highly non-linear [135, 127, 132]. It is approximated here with a multi-linear function, where different spring constants are used to describe the strain dependent effective stiffness of the protein domains. Similar approaches have been taken in other atomistic fracture mechanics studies [40, 31].

The stiffness of the system under tensile stretch is given by a serial spring combination and given as

$$k_{\mathrm{T}} = k_{1i}k_{2i}/(k_{1i} + k_{2i}), \tag{6.1}$$

where k_{1i} and k_{2i} are the strain-dependent spring constants of semi-amorphous and crystalline domains, respectively. This results in the following expression for the force of the system as a function of deformation:

$$F_{\mathrm{T}} = k_{\mathrm{T}}\Delta L_{\mathrm{T}}, \tag{6.2}$$

where the total deformation is $\Delta L_{\mathrm{T}} = \Delta L_1 + \Delta L_2$, in which case ΔL_1 and ΔL_2 are the deformations of the semi-amorphous and beta-sheet nanocrystal domains, respectively. The parametersΔL_i are defined as

$$\Delta L_i = r_{x,i} - r_{0,i}, \tag{6.3}$$

where $r_{x,i}$, $r_{0,i}$ are the length of the domains at a given strain, and their initial length, respectively. The following sections explain how the spring constants of the semi-amorphous and beta-sheet nanocrystal domains are obtained and implemented

in the model.

Atomistic simulations of the mechanical properties of the semi-amorphous region of spider silk have been presented in Section 6.3 [135, 127, 132]. A representative volume element containing 15 polypeptide chains was considered in deriving the constitutive mechanical law, in order to obtain good statistics and an accurate account of the secondary structure distribution of the domain. The effective force-extension behavior derived from the large molecular assembly was then normalized for the force, area and length per single polypeptide chain, to develop an appropriate constitutive law. The length of the amorphous domain is taken here as 60 Å based on the atomistic simulation model. The results from these analyses are used directly to determine the parameters of the present coarse-grained model representing a single amorphous domain and a single beta-sheet nanocrystal. These simulations revealed a characteristic three-stage deformation pattern, where an initial stiff regime is followed by a yielding point and a long plateau, and eventual significant stiffening as the polypeptide's backbone is being stretched [135, 127, 132]. This behavior is associated to the presence and the breaking of secondary structures such as 3_1-helices and beta-turns, which is rich in intra-chain and inter-chain hydrogen bonding [233, 240, 185]. At larger strains, the structure enters a final high-stiffness regime, characterized by the stretching of covalent bonds along the protein backbone.

Values for the stiffness of the three different regimes of the semi-amorphous domain are extracted from atomistic simulation data [135, 127, 132]. Fitting the simulation curve with a tri-linear function, the following regimes of stiffness are obtained as a function of deformation:

$$k = \begin{cases} k_{11} & \text{if } \Delta L_1 < \Delta L_{11} \\ k_{12} & \text{if } \Delta L_{11} \leq \Delta L_1 \leq \Delta L_{12} \\ k_{13} & \text{if } \Delta L_1 > \Delta L_{12} \end{cases} \tag{6.4}$$

where the values for k_{1i}, as well as the transition deformations ΔL_{11} and ΔL_{12} are summarized in Table 6.1. The resulting force-deformation law is shown in Figure 6-7(a).

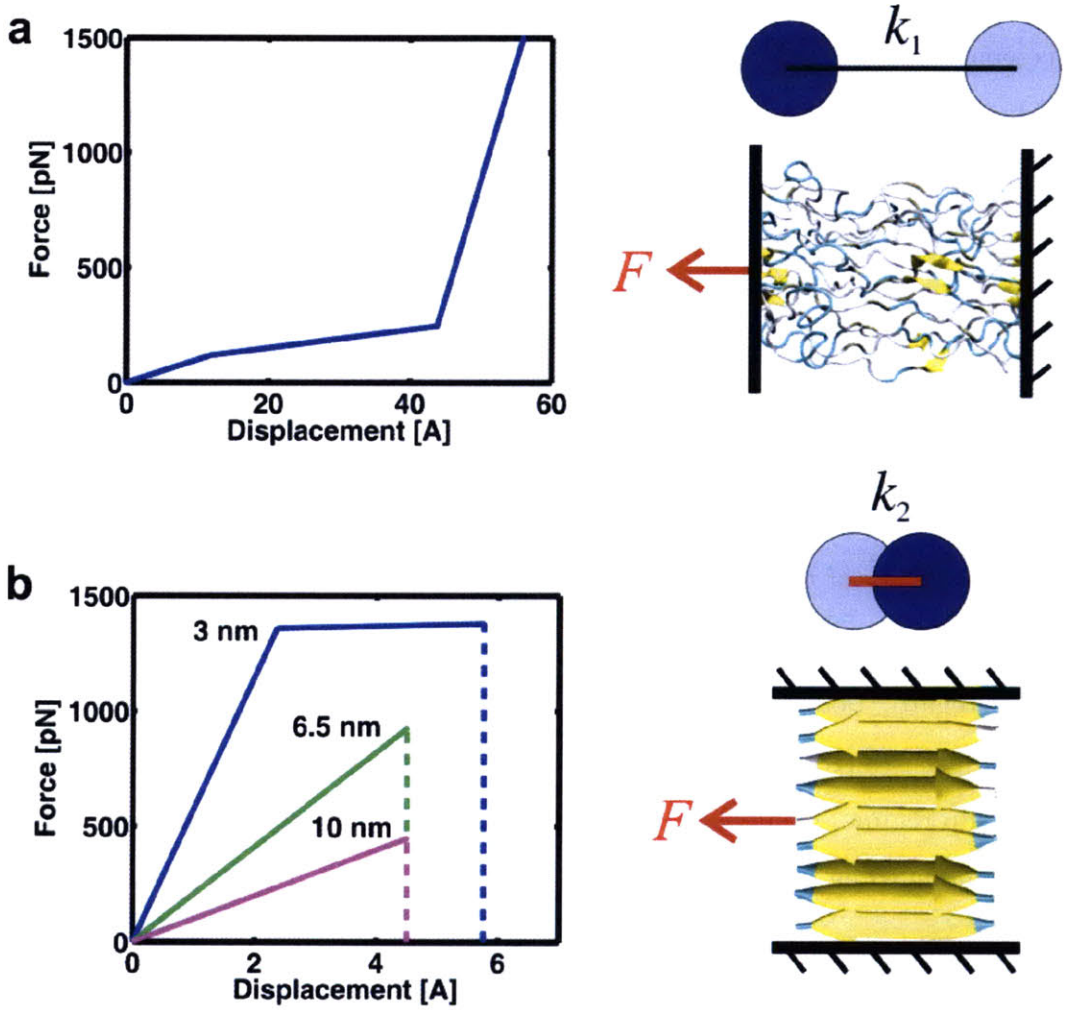

Figure 6-7: Constitutive behavior of the two elements represented in the coarse-grained model of silk (obtained from full atomistic molecular dynamics simulation). Panel (a) shows the behavior of the glycine rich semi-amorphous domain. Panel (b) depicts the behavior of beta-sheet nanocrystals under lateral loading. The smallest beta-sheet nanocrystal size shows the characteristic stick-slip phenomenon due to repeated breaking and reformation of H-bonds (modeled here as an elastic-plastic yielding behavior).

Model parameter	Value
k_{11} (pN/Å) – Initial regime	9.9
k_{12} (pN/Å) – Intermediate regime	3.96
k_{13} (pN/Å) – Final regime	103.84
ΔL_{11} (Å) – First transition point	12.0
ΔL_{12} (Å) – Second transition point	43.8

Table 6.1: Parameters for the semi-amorphous region in the mesoscale model.

The beta-sheet nanocrystal is also modeled as a nonlinear spring, where the force-displacement characteristic is informed from atomistic simulation results [135, 127, 132]. The beta-sheet nanocrystal dimension in the fibril direction is approximated to be 30 Å for all beta-sheet nanocrystal sizes (since it is defined by a fixed number of alanine residues in the beta-strand length direction that does not change). From atomistic simulations [135, 127, 132], it is observed that the properties of the beta-sheet nanocrystals vary as a function of crystal size, where small crystals are stiffer, dissipate more energy through a stick-slip mechanism and fail at higher force. The model presented here considers the effect of the variation of the size of beta-sheet nanocrystals on the mechanical behavior, by scaling the stiffness, strength and energy dissipation capacity of the beta-sheet nanocrystals according to size-effects observed in explicit water atomistic simulations.

In this regard, beta-sheet nanocrystals of different size are prescribed distinct mechanical features [181, 173, 143] in this model. In particular, the model includes a final stick-slip regime, which is observed only in the small-crystal case (3nm), whose stiffness is modeled as follows:

$$k = \begin{cases} k_{21} & \text{if } \Delta L_2 < \Delta L_{21} \\ k_{22} & \text{if } \Delta L_2 \geq \Delta L_{21} \end{cases} \tag{6.5}$$

where ΔL_{21} is the beta-sheet nanocrystal transition point (as a function of deformation) as defined in Table 6.2 and shown in Figure 6-7(b). In this sense, the crystal behavior is elastic-plastic, to take into account the additional energy dissipated by the stick-slip behavior. In the case of large crystals, the spring constant is the same over the whole regime of deformations and the mechanical behavior is elastic and brittle until rupture. In all cases, the force applied to the system drops to zero when $\Delta L_2 > \Delta L_{22}$, that is, when the maximum strength is reached and the crystal breaks. The spring constants are calculated by dividing the maximum tensile strength by the softening point for the 3 nm case, and by the breaking point for the large-crystal cases. The second softer regime for the small-crystal case is assumed to have a constant stiffness equal to 1% of the initial one, approximating the stick-slip behavior

observed in atomistic simulations [135]. The calculation of the breaking point is done by maintaining the dissipated-energy proportion between beta-sheet nanocrystals of different size. Explicit simulations suggest that a 3nm crystal is approximately three times tougher than a 6.5 nm-crystal. In the context of the bilinear spring model [135], this gives a breaking point value of 5.8 Å. It should be noted here that the atomistic calculations on strength and effective stiffness are based on the pullout force required to separate a single strand from the crystal, to be consistent with the normalization for a single polypeptide strand.

Model parameter	**Beta-sheet nanocrystal size**		
	3 nm	**6.5 nm**	**10 nm**
k_{21} (pN/Å) – Initial stiffness	576	205.5	99.5
k_{22} (pN/Å) – Second-regime stiffness	5.76	N/A	N/A
ΔL_{21} (Å) – Softening point	2.36	N/A	N/A
ΔL_{22} (Å) – Breaking point	5.8	4.5	4.5
$F_{\max}$ (pN) – Maximum tensile strength	1360	925	448

Table 6.2: Parameters for the beta-sheet nanocrystal in the mesoscale model.

The ratio of the maximum tensile strength between beta-sheet nanocrystals of different size [135] is derived directly from atomistic simulation for the 3 nm and 6.5 nm beta-sheet nanocrystals [135], while it is extrapolated for the 10 nm case. To be consistent with the force values of the semi-amorphous domain, the maximum strength is directly calculated from simulations in implicit solvent for a 3 nm-crystal system; and the large beta-sheet nanocrystal strengths are then determined using the strength ratio from explicit simulations. The softening point for the 3 nm beta-sheet nanocrystal is extracted from explicit simulation results, as well as the breaking point for the 6.5 nm beta-sheet nanocrystal. In the 10 nm case, the breaking point is instead assumed to be the same as the 6.5 nm case, given the fact that with increasing beta-sheet nanocrystal dimension the breaking point does not vary significantly as shown in [135].

Stress values are calculated by considering a constant radius of 5 Å along the whole system length and for the three cases of beta-sheet nanocrystal size, such that

$\sigma = F/A$, where σ is the computed stress, F is the force per chain and A is the area of a chain. This cross-sectional area is is an approximation for the effective area of a single silk polypeptide chain. Accordingly, all other calculations have been normalized for a single-chain system.

A MATLAB script provided in Appendix A is used for calculating the stress-strain behavior of the serial spring model discussed in this work. Strain is applied at constant steps to the overall structure and the deformation on single elements is computed. Mechanical toughness is calculated measuring the area under the force-extension curve (until structural failure) by means of a trapezoidal numerical integration. Data postprocessing is also performed using MATLAB scripts.

6.5 Implications of nanostructure on micromechanics

The model can now be used to simulate the mechanical deformation of silk according to the tensile loading condition shown in the inlay of Figure 6-8(a), focusing first on the system with a beta-sheet nanocrystal size of 3 nm that reflects natural silks. The resulting stress-strain curves display the characteristic shape observed in silks, that is, it displays an early yield point, leading to a significant softening that is followed by a severe stiffening effect. A detailed analysis reveals four distinct regimes of deformation, where each of them associated with specific atomistic mechanisms. The initial regime is characterized by a relatively high tangent modulus (1.07 GPa), owing to the stretching of semi-amorphous regions rich in H-bonds, in the form of 3_1-helices and beta-turns [233, 240, 185].

Rupture of the hydrogen bonds in the semi-amorphous domains leads to yielding at strain values of around 13%. The tangent modulus of this softer regime is lower, around 0.4 GPa. During this plateau regime, protein chains gradually align along the pulling direction [68, 200]. At a strain value of 48%, the stress-strain curve enters a high-stiffness regime (with a much higher tangent stiffness of around 10 GPa);

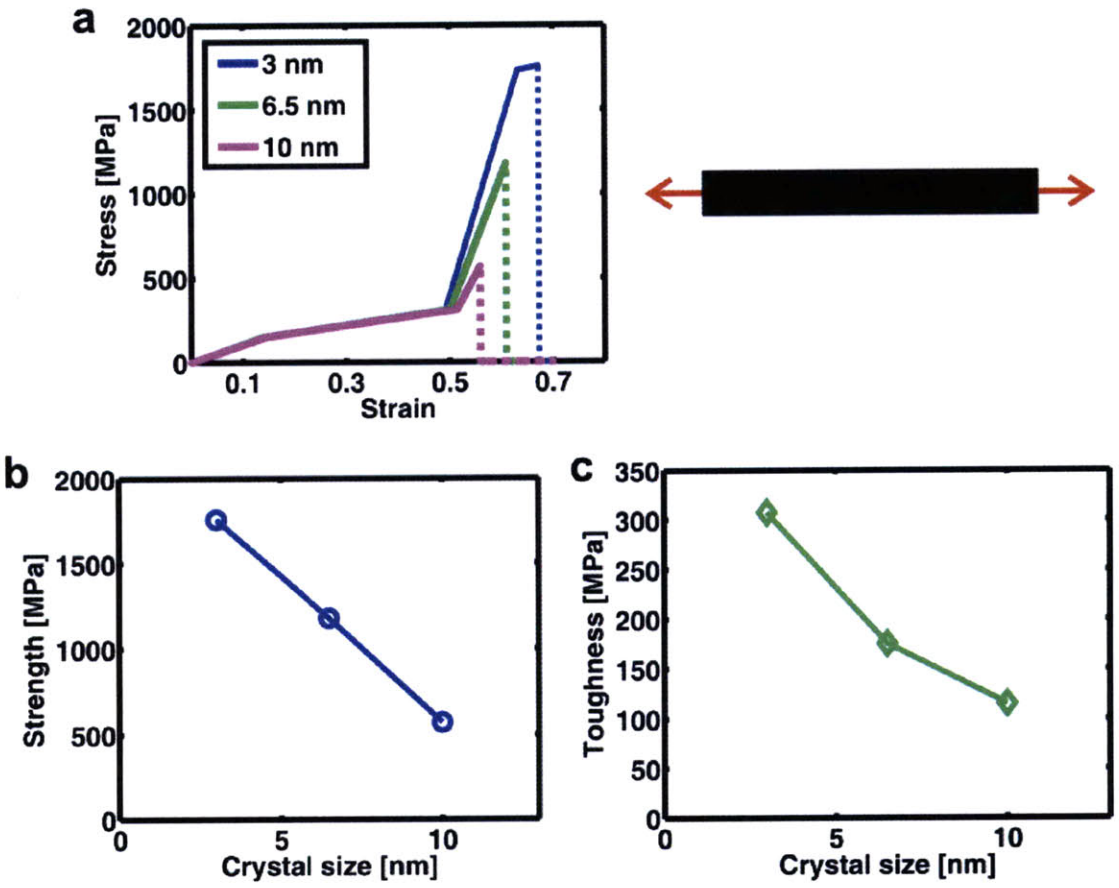

Figure 6-8: Stress-strain response of a silk fibril under tensile loading, for varying beta-sheet nanocrystal size, and analysis of strength and toughness dependence on crystal size. In panel (a), the plot depicts the stress-strain response of spider silk based on different beta-sheet nanocrystal sizes, ranging from 3 to 10 nm. The behavior after rupture is displayed with dotted lines. The results reveal dependence of the stress-strain response on the beta-sheet nanocrystal size. Panel (b) shows the variation of strength with beta-sheet nanocrystal size. The plot illustrates that silk fibers employing larger beta-sheet nanocrystals have a diminished strength of 1,178 and 570 MPa for the 6.5 and 10 nm crystal cases, compared with the small-crystal (3nm) system, which breaks at 1,757 MPa. Panel (c) depicts the variation in the toughness (toughness modulus) of the silk constitutive unit as a function of beta-sheet nanocrystal size. Enhanced mechanical properties of small beta-sheet nanocrystals play a governing role in the overall behavior. An increase in the toughness (modulus) from 115 MPa to 175 MPa and 307 MPa is observed when beta-sheet nanocrystal size is reduced from 10 nm to 6.5 nm and 3 nm.

at this point the semi-amorphous region has been completely stretched out and the beta-sheet nanocrystals sustain larger strains. This data shows that the beta-sheet nanocrystals play a significant role in defining the mechanical behavior of silk only at high-deformation levels, while the initial behavior is mainly governed by deformation of the semi-amorphous phase.

An interesting event observed in the stress-strain plot is the brief softening regime immediately prior to failure, which occurs at around 63% strain and is due to the stick-slip failure mechanism of the beta-sheet nanocrystals [135]. This final high-stress regime contributes to the overall toughness as it accounts for approximately 23% of the total energy dissipated. When the applied force reaches the maximum tensile strength, at strain values of 66.7%, beta-strands completely slide out at a failure stress of 1,757 MPa. This results in a drop of stress-levels, characterizing the ultimate failure of the material via molecular mechanisms activated at the nano-scale level.

The model can be used to carry out a systematic variation of the beta-sheet nanocrystal size, in order to quantify the dependence of the overall mechanical behavior on the crystal size. The motivation for this analysis is to validate whether earlier hypotheses that show that small changes in the crystal size translates to altogether different overall mechanical response in spider silk [135, 68]. The results of the mechanical analysis are shown in Figure 6-8(a), illustrating the stress-strain curves for varying beta-sheet nanocrystal sizes. The most important finding is the observation that the size of beta-sheet nanocrystals – at otherwise completely identical conditions – severely affects the mechanical response. The analysis shown in Figure 6-8(a) reveals that larger-crystal systems (6.5 nm and 10 nm beta-sheet nanocrystals) have a behavior that deviates significantly from the reference small-crystal case, especially at high levels of deformation. Silk fibrils with larger beta-sheet nanocrystals break at significantly lower stress values, and also show a shorter (60% and 55% strain, respectively) and softer third regime, with a tangent modulus of 8 GPa and 5.7 GPa, respectively. The initial and intermediate regimes, however, are comparable to the case with the smallest beta-sheet nanocrystal, where the transition points and

stiffness values do not vary substantially between the two systems.

As a result of these changes, even a slight increase in beta-sheet nanocrystal size leads to a significant loss of strength and toughness of the system (Figure 6-8(b-c)). With a value of 1,178 MPa, the 6.5 nm-crystal case shows a decrease of approximately 33% with respect to the reference system. The drop is even larger for the 10 nm-crystal case, which breaks at 570 MPa, at 67% less than the reference case. The drop in maximum stress and the absence of a high-stress final stick-slip regime further lead to toughness values that are considerably lower than the reference small-crystal case (see Figure 6-8(c)). A decrease of 43% and 63% are measured for the 6.5 nm and the 10 nm case, in comparison to the toughness value for the reference small beta-sheet nanocrystal case.

Measurement of the relative strains in both domains shed light on the deformation mechanisms linked to beta-sheet nanocrystal size. The analysis reveals that the contribution of the beta-sheet nanocrystals to deformation tends to decrease significantly as the size of beta-sheet nanocrystals is increased. In the small-crystal system, beta-sheet nanocrystals start to play a significant role once the semi-amorphous region begins to stiffen at around 50% strain, and dominates deformation when the stick-slip mechanisms of beta-sheet nanocrystal deformation is triggered. In the larger-crystal case, the beta-sheet nanocrystal contribution increases more gradually, and reaches a maximum just before the system breaks, that is, shortly after the semi-amorphous region enters the covalent hard-stretching regime. The breaking point is reached earlier in large-crystal systems, when semi-amorphous regions are less stretched (around 75%, compared to 92%). This observation suggests that the change of the beta-sheet nanocrystal size prevents the material to take full advantage of the entire potential of the semi-amorphous regions in terms of extensibility and energy dissipation capacity.

The results of this study are overall in good agreement with experimental data, where a similar variation of the beta-sheet nanocrystal size and its impact on larger-scale mechanical properties was reported recently [68], showing a drastic drop in toughness when the reeling speed decreases and beta-sheet nanocrystal size increases accordingly. The strain values in the three systems investigated here are higher than

those found in experimental studies (between 40% and 60%). This phenomenon can be explained based on two observations. First, the model considers an ideal and rather simplified structure that completely lacks the statistical variability and structural defects, and is perfectly aligned with the fiber axis. This generally results in enhanced strength and extensibility in comparison to experimental results. Second, it must be underlined that physiologically spun silks (*i.e.,* the small-crystal case) undergo a substantial pre-stretching at the orifice [68, 175]. However, the current model does not include this effect and consequently, an overestimation of the stretching capacity is expected. Future models could directly include variability of structure and defects in silk, and model the effect of pre-stretching to allow for a better comparison with experimental data (this could, for example, be achieved by the development of a two-dimensional model).

6.6 Discussion and conclusion

The results from atomistic simulations on MaSp1 and MaSp2 protein segments of the spider dragline silk from *N. Clavipes* have been reported in this Chapter. This study illustrates that the REMD simulation method establishes an atomistic basis for a wide variety of findings from experimental approaches of studies of the silk nanostructure. Key finding from secondary structure and dihedral angle analysis have shown that poly-Ala regions in MaSp1 and MaSp2 form beta-sheet crystals, whereas the glycine rich regions form semi-extended 3_1-helix type structures as well as beta-turns (Figure 6-2). No evidence for alpha-helix or beta-helix structure formation is observed in the simulations for both MaSp1 and MaSp2.

These results confirm that MaSp1 and MaSp2 form different nanostructures. MaSp1 tends to form more orderly extended structures with beta-sheet domains, whereas turn structures and a greater level of amorphousness are observed in MaSp2, due to the natural twist of the proline segments, which limits the crystalline fraction of the protein assemblies. The existence of semi-extended domains in the so-called amorphous matrix may be the molecular source of the large semi-crystalline fraction

observed in silks, and also form the basis of the so-called "pre-stretched" molecular configuration. The presence of less-dense hydrogen bonding in glycine-rich regions compared with alanine-rich regions indicates at least a two-phase system with complementary mechanical functions: extensibility and fracture strength. This is more evident in MaSp2, which has apparent features resulting from the proline rich sequence, such as higher disorder and well-defined crystal regions. Of particular importance for this effect is the lack of amide bonds and the torsional twisting at proline sites, which controls morphology development as the proline residues have varying capacity to form hydrogen bonding at different extension states. In the actual morphology of dragline silk, both proteins can be observed at varying amounts depending on the species. The current study focused on isolated systems of MaSp1 and MaSp2, but the same simulation protocol can be applied to understand the structure and behavior of mixed systems, and the possible role of proline in controlling intermolecular contacts between MaSp1 and MaSp2. Overall, the results reported here are strongly supportive of experimental findings on the spider dragline silk that suggest similar structural characteristics [234, 240].

The mechanical stretching simulations hint towards a sigmoidal constitutive behavior as observed in the macroscale response of spider silk, as shown in Figure 6-4(c). More analysis and upscaling of these results will have to be done to make a rigorous link between the nanoscale simulation results and macroscale experiments. Yet, these findings provide insight into the nanoscale deformation mechanisms, as shown in the simulation snapshots in Figure 6-5 and the detailed analysis in Figure 6-6. The initial softening of the system is evidently linked to the hydrogen bond breaking in the amorphous domain. Strain stiffening is observed once the covalent chains are highly extended after rupture of a large number of hydrogen bonds in the non-crystalline domains. The eventual failure of the system occurs when strands begin to slide across each other due to severed hydrogen bonds and side-chain interactions in the crystal. The failure strength of the molecular assembly shows less variation between sequences; rather, it is observed to depend more on the morphology of the crystal and the amorphous structures. Based on these simulations, it is evident that

mechanical stretching of the strands may play a role in controlling crystal size, where compression in the transverse direction causes collapse of the crystals into a more compact formation. This may be an important consideration for the influence of mechanical strain on the morphology of the silk crystals [96].

In summary, a framework for predicting the nanostructure of spider silk using atomistic principles is proposed with this study. The method can be widely applicable not only to different types of silks but could also to other biopolymers, and thereby provide a link between genetics and material properties. This materiomics framework provides a powerful tool to explore fundamental structure-property relationships of complex biological materials such as spider silk. Atomistic calculations resolve some of the controversies regarding the structure of the amorphous domains in silk, by shedding light on the semi-extended, well-oriented and more sparsely hydrogen bonded structures in amorphous domains. The results reported here, combining REMD structure identification with experimental validation and a subsequent nanomechanical analysis of fundamental deformation mechanisms, is the first of its kind for spider silk, and only the beginning. Future studies could be focused on improved structure prediction and a wide variation of conditions during mechanical analysis.

The most important question answered by the serial spring model derived from atomistics is that the severe change in mechanical properties of spider silk under relatively small variations of the size of beta-sheet nanocrystals can be explained solely based on structural effects (Figure 6-8). These results show that the confinement of beta-sheet nanocrystals to the nanoscale is essential for the superior mechanical properties of silks, as this is crucial to reach high extensibility and high levels of stress. These findings also relate the characteristic yielding point in the stress-strain curve, observed universally for many silks, to the onset of failure of semi-amorphous regions (see Figure 6-7(a)). As an overall result, these results show how small-crystal systems are more efficient in guaranteeing the required cross-linking strength, necessary for the semi-amorphous chains to fully extend and to enter a high-stiffness covalent regime. The capacity to sustain large tensile force as well as extension enhances the strength

and energy dissipation ability of the material.

These findings may set the stage for more extensive full-atomistic mechanics studies on silk domains that will contribute towards an improved understanding of the source of the strength and toughness of this biological superfiber, where the structures identified here could be subjected to mechanical loading under varying conditions, such as changes in solvent or pH, and even under variations of the amino acid sequence of the constituting protein domains. Future investigations could also focus on much larger, mixed MaSp1/MaSp2 systems in a variety of conditions or on the impact of mechanical constraints on the assembly and initial structure formation.

Chapter 7

Conclusion

This chapter summarizes key findings of this thesis, and discusses the significance of the results obtained from theoretical considerations and molecular dynamics simulations on beta-structures in protein materials. An overview of the capabilities of the beta-sheet strength model, presented in Chapter 3, is provided. The predicted size-scaling of the strength of H-bond clusters is reiterated, and implications for materials design are further discussed. Results from the beta-solenoid structures with a triangular core are summarized, and directions for future research are recommended in the second part of this Chapter. Emphasis is given for impact and the outlook of computational research on spider silk nanostructure, based on the results presented in Chapter 5 and 6 of this thesis.

7.1 Summary of key findings and significance

Beta-structures discussed in this thesis show great promise for the future of materials design. Self-assembly is the hallmark of how materials are synthesized in biology, in particular in the context of proteins, which form spontaneously at room or body temperature. Beta-structures have the capacity to form very orderly, fibril and nanotube structures through edge-to-edge aggregation, forming arrays of H-bonds with controlled orientation and size [257]. This structure is the basis of rigid amyloid fibrils, viral structures such as the cell-puncture device needle, and reinforcing nanocrystals

found in spider silk [120]. Mechanical context of the beta-structures was evident from some of the earlier studies at very small model systems [72, 23, 156, 194], but a rigorous study looking into effects of size, geometry and rate at hierarchical length-scales had been missing in literature. This thesis aimed to address this point, and thereby bridge the gap between the biophysical studies of beta-structures and a a generic, macro-level continuum description of the mechanics of biopolymers.

Protein materials such as spider silk nanocrystals, the cell-puncture device needle of T4,or amyloids point towards unique opportunities for creating strong materials that have structural and mechanical properties defined by weak bonds. Chapter 4 focused on the niche application of beta-solenoids used for generating compressive forces in the biological context. The unique triangular core, maximizing rigidity per cross-sectional area, and the tightly intertwined beta-helix topology of the cell-puncture needle suggests that exceptional mechanical properties such as stiffness and failure strength can be obtained from molecules that self-assemble at room temperature. Decades of research on climate change suggests that replacing our current energy intensive infrastructure and building materials with more sustainable, biodegradable substitutes inspired from natural systems will be extremely important to solve future problems in energy and sustainability. Lessons learned from fundamental research on physical biology will play a major role in understanding how to create structural materials from simple, reliable, and generally available polymeric building blocks.

Chapter 3 summarized findings on the size-scaling of the strength of H-bond cluster in polymers and proteins. A new theoretical framework (beta-sheet strength model), that takes into account only two parameters, namely H-bond strength and persistence length of chains, was proposed. The basic framework combined the fracture mechanics concept of energy balance to estimate strength, and worm-like chain model for elasticity to describe a constitutive law at the single molecule level. Several general predictions emerged as a result of this theory, which can be summarized as follows:

- Strength of key domains in proteins composed of H-bond assemblies at near equilibrium rates can reach a few hundred picoNewtons, but cannot exceed this

value without the presence of hierarchical structures. This is due to the fact that application of forces in the hundreds of picoNewtons leads to change in free-energy of the system that is adequate to cause rupture of a few hydrogen bonds, which leads to an instability that leads to the rupture of the entire domain, regardless of the number of bonds in the cluster.

- This local failure criterion immediately restrains the size of the H-bond cluster that will achieve this maximum strength, which appears to be in the order of 3 to 4 hydrogen bonds. This finding is in line with proteomics data suggesting that longer beta-sheets are strongly disfavored in biology due to their thermodynamical and mechanical instability. The prediction of strength and cluster size varies according to loading rate, loading geometry, as well as strength of bonding (according for instance to solvent conditions [66]).

- Beta-sheet strength model is capable of describing individual rupture events in a complex protein domain. This has been validated on the I27 domain in titin, where force peaks of rupture have been shown to depend on the size of H-bond clusters. Most importantly, the beta-sheet strength model (Chapter 3) suggests that there is a strength limit to entropic elasticity predicted by the WLC theory. In AFM experiments, this has typically been observed as deviations from the WLC curve described by a single set of parameters, however, the theoretical explanation for quantifying these key events had not been adequately addressed in the literature. The beta-sheet strength model is capable of predicting these deviations with a minimal number of input parameters, eliminating the need to map out the complete energy landscape via experiment or simulation solely for this purpose.

The importance of this size-effect was discussed in greater detail in Chapter 5, in the context of silk mechanics. Silk is one of the toughest and strongest materials known, which is directly related to the size of the beta-sheet nanocrystals that exist within the fibrils making up silk. These crystals form rigid cross-linking regions in the ultrastructure of the fibril through dense H-bonding. Yet, as shown through theory

and simulation, the cooperation of these bonds under mechanical force only occurs if the size of the crystal is, in both strand and H-bonding directions, limited to a few nm. This is due to the fact that longer strands suffer from the localized failure of H-bonds, and larger crystals are prone to bending under thermal and mechanical perturbation, which severely disrupts H-bond cooperation. The transition from bending to uniform shear governed deformation occurs to be very sharp in simulation; less than 1 nm change in size leads to large differences in strength and energy dissipation. Furthermore, smaller crystals can exhibit stick-slip failure, which increases energy dissipation through subsequent formation and breaking of H-bonds. Silk therefore benefits from toughening mechanisms similar to those observed in other biological materials (e.g. wood or bone), and in metals in the form of dislocations that provide ductility [122, 84, 148, 128, 131, 129, 102].The implication of this size-effect on the macro-scale is validated by tensile tests in silk, which also show a strong dependence of the fiber strength and toughness on nanocrystal size [150, 68, 135]. The work presented in this thesis provides a theoretical explanation for this phenomenon, which had remained elusive primarily due to the challenges of experimentally probing such small-scale features in silk. An interesting aspect of this study is the finding that silkworms and spiders have developed techniques for creating very tough nanocrystals from two of the most common, chemically simple amino acids (Glycine and Alanine) that can be procured or synthesized easily in most ecological and physiological settings.

Chapter 6 built on this size-effect concept by looking into the overall mechanical behavior of spider silk protein assemblies. The REMD method was utilized here to come up with coarse predictions of the nanostructure of silk proteins, in order to establish a constitutive law for silk at the nano and micro-scale. This is the first molecular level structure prediction attempt for silk, based entirely on simulation. The findings support a number of experimental findings on the link between the sequence and morphology of proteins that make up this outstanding fibrous material. Particularly, it is observed through simulation that poly-Alanine and poly-Glycine-Alanine domains form very orderly, densely H-bonded crystalline domains that, as

predicted in Chapter 5, control the fracture and large deformation mechanics of silk. Semi-amorphous domains consisting of GGX and GPGQQGPGGY repeats form wide range of sparsely H-bonded structures that provide extensibility to the silk. In particular, GGX repeats form semi-extended structures resembling the characteristics of 3_1-helices observed in poly-Glycines [59]. Proline-rich regions serve the role of beta-sheet breakers, limiting the size of nanocrystals in the strand direction. Both of these glycine-rich domains govern the yielding behavior of silk, which occurs due to rupture of H-bonds in the amorphous regions at less than 10% strain. The mechanical behavior is typically a trilinear behavior, where it is either monotonically stiffening (similar to WLC behavior) or stiff-soft-stiff, determined by rapid breaking of the H-bonds and presence of turn regions in the system. A very simplistic model consisting of multi-linear springs, with constants derived from atomistic results, was successful in qualitatively describing the macro-scale behavior observed in silk.

This thesis established a first step in understanding size-effects on mechanical strength of protein materials. It also presents a new direction in spider silk research, where large-scale full-atomistic and multi-scale modeling approaches complemented with experimental data will enable the discovery of Nature's secret recipe for ultra-strong and tough fibers.

7.2 Opportunities for future research

The previous section discussed the success of the theoretical model developed and size-effect predictions for beta-structures in protein materials, with relevance in particular to the mechanics of spider silk. The purpose of this section is to illustrate the shortcomings of the current work and propose directions for future research on beta-structures in protein materials.

The beta-sheet strength model, presented in Chapter 3 was very successful in describing a range of experimental observations using two simple parameters. The model is however developed for uniform, cooperative shear loading scenarios as explained in Chapter 3. This is a suitable assumption for explaining the maximum strength of

proteins, but naturally falls short of explaining all aspects of the protein unfolding problem. It should be noted,however, that the model could be modified for several other loading scenarios and geometries. This has already been done for alpha-helices and double-strand shear loading condition in beta-sheets (see Appendix B). Another consideration is that the worm-like chain model, while being widely accepted for polymer chain elasticity, is not the only model that could be used for this purpose. WLC model's main shortcoming is that it's an inextensible chain model, and predictions at very high force values diverge. The beta-sheet strength model is general and basic enough that any other elasticity model could be used for this purpose. Additional parameters can also be used to separately take into account solvation free energy of the strand, as well as the stochastic nature of bond breaking.

With regards to the modeling efforts in beta-solenoids, a systematic study on the size-dependent compressive strength of self-assembling triple-helices would be of great interest. A study, focusing on different length specimens, would validate the length-dependence of failure mechanisms predicted in Chapter 4. A careful analysis of the vibrational modes and failure mechanisms would be extremely beneficial for not only beta-solenoid proteins but also amyloids. While extremely challenging, it would be particularly interesting to identify and isolate the source of the size-dependent behavior in beta-sheets and beta-solenoids under tension and compression. Possible sources of these effects, some of which are discussed in detail in this thesis, could be variation of entropic contributions to elasticity, length dependent fundamental modes of vibration, or emergence of local vs. non-local failure of hydrogen bonds mediated by water and thermal motion. The development of a coarse-grained, Go-model type system that has high fidelity to atomistic results would be very beneficial for elucidating these mechanisms. Such a model could also shed light on the significance of the twist and triangular cross-section of the solenoids.

Most importantly, there is great potential for future work in understanding the nanostructure and mechanics of spider silk using atomistic simulation. The work presented in Chapters 3, 5, and 6 are a good starting point for this type of systematic characterization. Future work should particularly focus on understanding how silk

protein constituents self-assemble into particular folds that depend on sequence, and how this influences the initial modulus, yielding and ultimate fracture strength of silk. While the EEF1 model had the advantage of much faster sampling, enabling studying large-scale silk systems, more accurate all-atom models in explicit solvent could reveal more further insight into these systems, in particular on understanding the role of hydration, pH as well as other factors that control self-assembly and silk mechanics.

Another recommendation for future research is to find example generic polymers that could potentially utilize the size-effects predicted here for H-bonds. Simple coarse-grained models, representing crystalline and amorphous regions in a silk-like materials could be starting point for understanding what kind of intermolecular interactions are needed to control and form nanostructures that resemble those in silk. It should be noted, however, that such an approach is extremely challenging for several reasons. First, treatment of solvent is not straightforward for effective interaction energy based models, as opposed to hierarchical, structural models that already consider these effects implicitly in the constitutive relationships. Second, simple Go-type models greatly reduce the complexity of H-bond mechanics, as a result, distance based non-bonded interactions (e.g. Lennard-Jones) suffer from inadequate treatment of charge interactions and angular dependence of H-bond strength. A third issue is the classification of the strength of intermolecular interactions, simple coarse-grained models typically require some knowledge of the contact points in the native structure of the protein building blocks, which for the case of amorphous complex materials such as silk are obscure. Still, there is a great need for description of the mechanics of silks and other complex polymer materials at the meso-scale, and efforts to address this issue would nicely complement the work presented in this thesis.

Appendix A

Script library

Post-processing and setup scripts used for the research presented in this thesis are provided here for the reader's convenience.

A.1 .tcl scripts

A.1.1 Script to calculate H-bonds in VMD

```
set mol [molinfo top]
set prot [atomselect $mol "protein and (name N or name O)"]
set nf [molinfo $mol get numframes]
set log [open hbonds.txt w]

for {set i 0} {$i < $nf} {incr i} {
$prot frame $i
set nhb [llength [lindex [measure hbonds 4.0 40 $prot] 0]]
puts $log "$i \t $nhb"
puts "$i \t $nhb"
flush $log
}
```

A.1.2 Script to calculate phi-psi angles in VMD

```
# SHORT SCRIPT FOR GETTING PHI-PSI ANGLES

    set fp [ open "phi-psi.dat" w ]
    set sel [ atomselect top "resname GLY ALA and name CA" ]
    set n [ molinfo top get numframes ]

    for {set i 0 } { $i < $n } { incr i } {
        $sel frame $i
        $sel update
        puts $fp "\# frame: $i"

        set a [ $sel num ]
        for {set j 0 } { $j < $a } { incr j } {
            puts $fp "[expr $j + 1] [lindex [$sel get {resname phi psi}]
                $j]"
        }
    }

    $sel delete
    close $fp
```

A.1.3 Script to calculate secondary structure ratios in VMD

```
set TT [format "%5.2f" [llength [[atomselect top "name CA"] get resname
    ]]]
set HL [format "%5.2f" [llength [[atomselect top "helix and name CA"]
    get resname]]]
set BS [format "%5.2f" [llength [[atomselect top "betasheet and name CA
    "] get resname]]]
set RC [format "%5.2f" [llength [[atomselect top "coil and name CA" ]
    get resname]]]
set TR [format "%5.2f" [llength [[atomselect top "turn and name CA" ]
    get resname]]]

set HLP [format "%5.2f" [expr 100*$HL/$TT]]
```

```
set BSP [format "%5.2f" [expr 100*$BS/$TT]]
set RCP [format "%5.2f" [expr 100*$RC/$TT]]
set TRP [format "%5.2f" [expr 100*$TR/$TT]]

puts [format "HELIX: %5.2f " $HLP]
puts [format "BS    : %5.2f " $BSP]
puts [format "COIL  : %5.2f " $RCP]
puts [format "TURN  : %5.2f " $TRP]

#puts [format "PERCENT HL: %.2f BS: %.2f RC: " $HLP $BSP $RCP]
```

A.2 MATLAB scripts

A.2.1 Code for averaging force-extension curves from NAMD

```
function [t,F,R] = force_bin2(input,bin);
% input format: use the following awk commands to get input:
% more log.txt|grep SMD|./no_velocities.awk > data.q;
% awk 'NR > 7' data.q > input.q
% scp input.q username@clustername.mit.edu:/home/username/... "copy file
    "
% load input.q in matlab

% Input Format: Same as SMD TITLE
% 1. TS
% 2,3,4 CURRENT POSITION (CM of pulled atoms)
% 5,6,7 FORCE            (Force on SMD Spring)
% # of bins automatically implemented in code depending on max extension
    .
% binA = 1; % number of bins per Angstrom extension. default is 1.
I=mean(input(1:10,:),1)        % Values for timestep = 0;
N=size(input,1);     % Number of data points

% for j=1:N;
% RAW(j) =norm(input(j,2:4)-I(2:4)); % raw extension data
```

```
% end
% Emax = floor(max(RAW)) % maximum extension value, rounded down
% bin = binA*Emax;
 wind=floor(N/bin);  % Size of bins, remainder of data will be omitted

for i = 1:length(input);
    Xi(i) = norm(input(i,2:4)-I(2:4));
    Fi(i) = norm(input(i,5:7));
end

for i=1:bin;

    ini=(i-1)*wind+2;   %first bin pt.
    fin=i*wind+1;       %last bin point
    t(i)=mean(input(ini:fin,1),1);
    R(i)=mean(Xi(ini:fin));
    F(i)=mean(Fi(ini:fin));
end
%add the zero point that wasn't taken into the averaging scheme

t = [0 t];
R = [0 R];
F = [0 F];

%figure settings, changes defaults but will be reset when you quit

set(0,'defaultaxesfontsize',20);
set(0,'defaulttextfontsize',20);
set(0,'defaultaxesfontweight','demi');
set(0,'defaulttextfontweight','demi');
set(0,'defaultaxeslinewidth',4);
set(0,'defaultlinelinewidth',4);
set(0,'defaultlinemarkersize',6);

figure (1)
plot(R,F,'-ok')
```

```
title(['Force vs. Extension - # of bins = ',num2str(bin)])
xlabel('Extension (A)'),ylabel('Force (pN)')
figure (2)
plot(t*1E-6,F,'-ok')
title(['Time vs. Extension - # of bins = ',num2str(bin)])
xlabel('Time (ns)'),ylabel('Force (pN)')
figure (3)
plot(t,R,'-ok')
title(['Time vs. Extension - # of bins = ',num2str(bin)])
xlabel('Time (ns)'),ylabel('Extension (A)')
```

A.2.2 Serial spring model for spider silk

```
%Andrea Nova and Sinan Keten
%all lengths are in A

function monodimensional(type, color);      %type can be s (small) or l
   (large)                                          %l (large)

npoints=5000;   %how many points I have in the plot
strain=0.7;     %how much i strain the whole structure
crlength=30;    %initial length of the crystal
amlength=60;    %initial length of the amorphous region = 60 = (150-30)
   /2 [using MaSp2]
length=crlength+amlength;

if type=='s';
k11=576;        %all stiffnesses in pN/A
k12=0.01*k11;
r11=2.36;       %value in A
r12=5.8;

elseif type =='l';
k11=205.5;
k12=205.5;
```

```
r11=4.5;
r12=4.5;

elseif type =='xl';
k11=99.5;
k12=99.5;
r11=4.5;
r12=4.5;

end

length=crlength+amlength;

k21=9.9;
k22=3.96;
k23=103.84;
r21=0.2;    %fraction of initial length
r22=0.73;    %fraction of initial length

k2=k21;  %need to specify this for the first step of the cycle
k1=k11;

for i=1:npoints+1;
    dl(i,1)=(i-1)/npoints*strain*length;    % total deformation as a
        function of step

 if i==1;    %only for first step
    dl1(i,1)=k2/(k1+k2)*dl(i,1);    % deformation of each spring as a
        function of total length
    dl2(i,1)=k1/(k1+k2)*dl(i,1);
 else
    dl1(i,1)=dl1(i-1,1)+k2/(k1+k2)*(dl(i,1)-dl(i-1,1));
    dl2(i,1)=dl2(i-1,1)+k1/(k1+k2)*(dl(i,1)-dl(i-1,1));
 end
```

```
    if dl2(i,1)>=r21*amlength;   %sigmoidal spring reaches 1st
        transition point
        if dl2(i,1)>=r22*amlength;   %sigmoidal spring reaches 2nd
            transition point
        k2=k23; %assigment of the new k
        else
        k2=k22;
        end
    end

    if dl1(i,1)>=r11;   %crystal spring reaches transition point
       k1=k12;
    end

    k(i,1)=k1*k2/(k1+k2); %overall system stiffness in a series of
        springs

    if dl1(i,1)<=r12;
        if i==1;    %particular only for first step
        f(i,1) = k(i,1)*dl(i,1);
        else
        f(i,1) = f(i-1,1)+k(i,1)*(dl(i,1)-dl(i-1,1)); %force defined as
            incremental value
        end
    else f(i,1)=0;  %if crystal is broken the total force goes to zero

    end
end

 %conversion to stress; radius 5A; number in MPa:

sigma = f*1.2738853;
top=max(sigma)
maxstress=top;
location=find(sigma(:)==top);
initial=sigma(1:location);
```

```
dl5=dl(1:location);
sigma5=sigma(1:location);

plot(dl/length, sigma, [':',color], 'HandleVisibility', 'off');
hold on;
plot(dl5/length, initial, color);
set(gca, 'XTick', [0.1 0.3 0.5 0.7])

xlabel('Strain');
ylabel('Stress [MPa]');

toughness=trapz(dl5/length, sigma5)
```

A.3 Shell scripts

A.3.1 Basic shell script for setting up REMD with EEF1 on Teragrid Abe cluster

```
#!/bin/tcsh
#PBS -l walltime=00:10:00
#PBS -l nodes=6:ppn=4
#PBS -V
#PBS -N testjob

cd /cfs/scratch/users/keten/silk-REX3res

mvapich2-start-mpd

###setenv NP `wc -l ${PBS_NODEFILE} | cut -d'/' -f1`
#cat $PBS_NODEFILE | sort -u | awk ' { print $1, "8 ",dirh } ' dirh=/cfs
    /scratch/users/keten/silk-REX3res > hosts.abe

awk '{printf("%s 4 /cfs/scratch/users/keten/silk-REX3res \n",$1);}' <
    $PBS_NODEFILE|sort -u > hosts.abe
```

```
setenv MV2_SRQ_SIZE 4000

aarex.pl -n 20000 -hosts hosts.abe -charmmlog charmm.log -log server.log
    \
        -par archive,psf=out.psf,em_out.crd  \
        -mdpar prnlev=3,nogb,shake=1,shakemode='hyd',param=19x,eef1file
            =/cfs/scratch/users/keten/silk-REX3res/solvpar.inp \
        -mdpar lang=1,langfbeta=1.0,xpar=/cfs/scratch/users/keten/silk-
            REX3res/param19_eef1.1.inp \
        -mdpar echeck=99999999.0,xtop=/cfs/scratch/users/keten/silk-
            REX3res/toph19_eef1.1.inp \
        -mdpar explicit=0,ewald=0,cuton=7,cutoff=9,cutnb=10,dielec=rdie
            ,trunc=switch,dynupdimg=10,dynoutfrq=100,dynsteps=250 \
        -temp 24:300:700 final.*.pdb


mpdallexit
```

Appendix B

Extension of the beta-sheet strength model

This section focuses on extensions of the beta-sheet strength model presented in Chapter 3. Here, modifications of the theory for a different application is presented. In particular, application of the theory to alpha-helical structures will be briefly discussed.

B.1 Application to alpha-helices

The beta-sheet strength model was originally developed for beta-sheets, however, since the theory is based on energy balance concepts, it can be applied to a variety of loading conditions and geometries. This sections summarizes how the theory may apply to alpha-helical structures. The aim here is to find the critical force that will initiate rupture of H-bonds in an alpha-helix at quasi-equilibrium deformation rates. Starting again from the Griffith condition used to predict the onset of fracture[95], the free energy released by freeing polypeptide chains from their geometric confinement in helical convolutions must equal the energy required to break these H-bonds. The free energy balance condition at the onset of fracture requires that $G = -(A_2 - A_1 - F\delta)/d\lambda = \gamma_s$ where γ_s denotes energy released by rupture of H-bonds per unit crack advance, $F\delta$ is the work done by the external force on the system, and A_1 and A_2

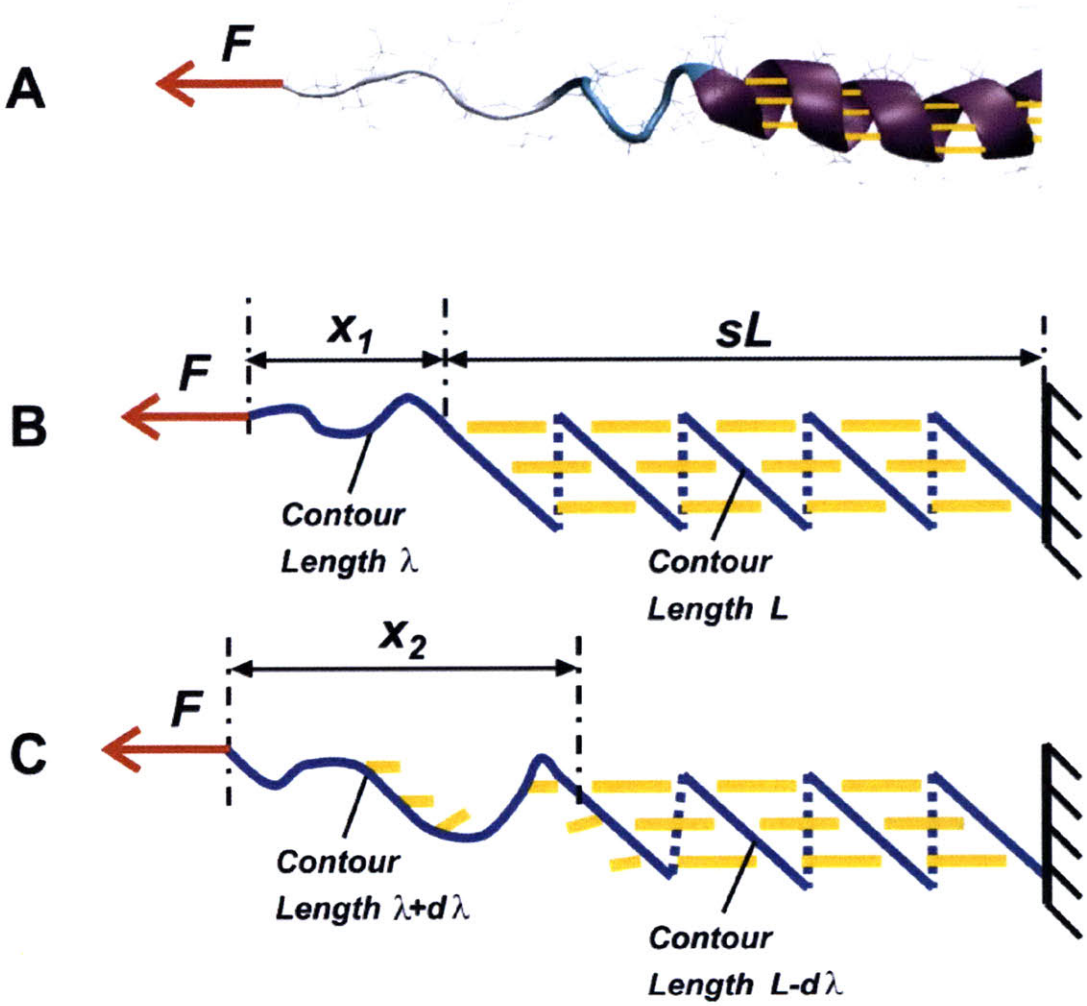

Figure B-1: Panel (a) depicts the atomistic-scale protein structure of a single alpha helix from a vimentin coiled-coil dimer. The helical backbone is stabilized by parallel arrangements of hydrogen bonds (yellow dashed lines). Panels (b) and (c) show a schematic model system of an alpha-helix strained by an external force before and after onset of rupture, showing the process of releasing a segment of backbone polypeptide due to the rupture of H-bonds, thereby increasing the contour length of the free end entropic chain by $d\lambda$

are the initial and final free energies of the protein backbone as determined from the worm-like chain elasticity theory. The free energy of the system before and after rupture is given as:

$$A_1 = \lambda A_{WLC} - \gamma_s L + L A_{FOLD} \tag{B.1}$$

and

$$A_2 = (\lambda + d\lambda) A_{WLC} - \gamma_s (L - d\lambda) + (L - d\lambda) A_{FOLD}, \tag{B.2}$$

where$A_{WLC} = \int_0^{\alpha} F_{WLC}(\alpha) d\alpha$ is the free energy state (energy per length) of the already unfolded free segments of the protein and $A_{FOLD} = \int_0^{s} F_{WLC}(\alpha) d\alpha$ is the free energy state of the folded segment of the chain. The additional term arises from the fact that in the original beta-sheet strength model formulation, the bonded region of the chain was assumed to be relaxed, similar to the free chain. However, since the chain is folded into a helix, it is in a semi-extended state, and entropic contribution due to this effect needs to be taken into account. Hereby α equals to the ratio of the end-to-end length of the free chain to its contour length $\alpha = x/\lambda$, equivalent to mechanical stretch, and the parameter $s = \lambda/L$ that denotes the ratio of the end-to-end length of the alpha-helix to its contour length, L (the physical meaning of this parameter is that it describes how much contour length is stored per unit length alpha-helix). Figure B-1 illustrates the geometry and variables used in the formulation. The energy contribution from the external force is given as:

$$\delta W_F = -F(\alpha - s) d\lambda. \tag{B.3}$$

Hence the critical condition for HB rupture can be given as:

$$A_{WLC}(\alpha_{cr}) + F(s - \alpha_{cr}) + \gamma_s - A_{FOLD}(s) = 0, \tag{B.4}$$

where α_{cr} is the critical stretch level that initiates rupture. The strength at the asymptotic regime (AR), then found by the WLC model, through $f_{AR} = F_{WLC}(\alpha_{cr})$, leading to:

$$f_{AR} = \frac{k_B T}{4\xi_P} \left[(1 - \alpha_{cr})^{-2} + 4\alpha_{cr} - 1 \right]. \tag{B.5}$$

The parameter γ_s describes the HB energy stored per unit length of AH and can be obtained from

$$\gamma_s = \frac{E_B^0}{L_0}, \tag{B.6}$$

where E_B^0 is the dissociation energy of a single bond and $L_0 = 0.33$ nm is the distance between adjacent H-bonds along the length of the helix. The parameter $s = 0.45$ can be estimated from atomistic simulations of the deformation mechanics of alpha-helices, where the unfolded length of the molecule can easily be calculated to find the ratio with initial end-to-end distance. These values are also in excellent agreement with the well-established alpha-helix pitch of 5.4 $\mathring{A}$ per convolution. It should be noted that the fracture model is independent of the size of the macromolecule and the helical domain. This is because the initial unfolded contour length does not influence the strength prediction. E_B^0 can be estimated from the MD simulation results, where the 3.6 H-bonds in one convolution break simultaneously, thus $E_b^0 = E_b^{SDM}/3.6 = 3.1$ kcal/mol, and therefore $\gamma_s = 0.91$ kcal/mol/$\mathring{A}$. Based on these two parameters, E_b^0 and ξ_P, the asympotic strength of alpha-helix domains can be estimated as approximately 189 pN. This finding is in agreement with experimental results on alpha-helices [145, 119].

Bibliography

[1] T. Ackbarow, X. Chen, S. Keten, and M. J. Buehler. Hierarchies, multiple energy barriers, and robustness govern the fracture mechanics of alpha-helical and beta-sheet protein domains. *Proc. Natl. Acad. Sci. U. S. A.*, 104:16410–16415, 2007.

[2] T. Ackbarow, S. Keten, and M. J. Buehler. A multi-timescale strength model of alpha-helical protein domains. *J. Phys.: Condens. Matter*, 21(3):035111, 2009.

[3] S. A. Adcock and J. A. McCammon. Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chem. Rev.*, 106(5):1589–1615, 2006.

[4] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell, Fourth Edition.* Garland, 2002.

[5] A. Arkhipov, P. L. Freddolino, K. Imada, K. Namba, and K. Schulten. Coarse-grained molecular dynamics simulations of a rotating bacterial flagellum. *Biophys. J.*, 91(12):4589–4597, 2006.

[6] I. Bahar and R. Jernigan. Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. *J. Mol. Biol.*, 266(1):195–214, 1997.

[7] P. Ball. *Made to Measure: New Materials for the 21st Century.* Princeton University Press, Princeton, N.J., USA, 1997.

[8] G. Bao and S. Suresh. Cell and molecular mechanics of biological materials. *Nat. Mater.*, 2(11):715–725, 2003.

[9] M. Bathe. A finite element framework for computation of protein normal modes and mechanical response. *Proteins: Struct. Funct. Bioinf.*, 70(4):1595–1609, 2008.

[10] Z. P. Bazant. Size effect on structural strength: a review. *Arch. Appl. Mech.*, 69(9-10):703–725, 1999.

[11] N. Becker, E. Oroudjev, S. Mutz, J. P. Cleveland, P. K. Hansma, C. Y. Hayashi, D. E. Makarov, and H. G. Hansma. Molecular nanosprings in spider capture-silk threads. *Nat. Mater.*, 2(4):278–283, 2003.

[12] G. I. Bell. Models for specific adhesion of cells to cells. *Science*, 200(4342):618–627, 1978.

[13] B. J. Berne and J. E. Straub. Novel methods of sampling phase space in the simulation of biological systems. *Curr. Opin. Struct. Biol.*, 7(2):181–189, 1997.

[14] F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. Protein data bank - computer-based archival file for macromolecular structures. *J. Mol. Biol.*, 112(3):535–542, 1977.

[15] P. Bjelkmar, P. Larsson, M. A. Cuendet, B. Hess, and E. Lindahl. Implementation of the CHARMM force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *J. Chem. Theory Comput.*, 6(2):459–466, 2010.

[16] S. M. Block, C. L. Asbury, J. W. Shaevitz, and M. J. Lang. Probing the kinesin reaction cycle with a 2d optical force clamp. *Proc. Natl. Acad. Sci. U. S. A.*, 100(5):2351–2356, 2003.

[17] D. Boal. *Mechanics of the Cell.* Cambridge University Press, Cambridge, UK, 2001.

[18] L. Bozec, J. de Groot, M. Odlyha, B. Nicholls, S. Nesbitt, A. Flanagan, and M. Horton. Atomic force microscopy of collagen structure in bone and dentine revealed by osteoclastic resorption. *Ultramicroscopy*, 105(1-4):79–89, 2005.

[19] P. Bradley, K. M. S. Misura, and D. Baker. Toward high-resolution de novo structure prediction for small proteins. *Science*, 309(5742):1868–1871, 2005.

[20] C.-I. Branden and J. Tooze. *Introduction to protein structure.* Garland Pub., New York, 2nd edition, 1999.

[21] D. W. Brenner, O. A. Shenderova, J. A. Harrison, S. J. Stuart, B. Ni, and S. B. Sinnott. A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons. *J. Phys.: Condens. Matter*, 14(4):783–802, 2002.

[22] K. Broberg. *Cracks and Fracture.* Academic Press, 1990.

[23] D. J. Brockwell, E. Paci, R. C. Zinober, G. S. Beddard, P. D. Olmsted, D. A. Smith, R. N. Perham, and S. E. Radford. Pulling geometry defines the mechanical resistance of a beta-sheet protein. *Nat. Struct. Biol.*, 10(9):731–737, 2003.

[24] A. E. Brooks, H. B. Steinkraus, S. R. Nelson, and R. V. Lewis. An investigation of the divergence of major ampullate silk fibers from Nephila clavipes and Argiope aurantia. *Biomacromolecules*, 6(6):3095–3099, 2005.

[25] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM - a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4(2):187–217, 1983.

[26] C. L. Brooks. Methodological advances in molecular-dynamics simulations of biological-systems. *Curr. Opin. Struct. Biol.*, 5(2):211–215, 1995.

[27] M. Buehler. Molecular nanomechanics of nascent bone: fibrillar toughening by mineralization. *Nanotechnology*, 18:295102, 2007.

[28] M. Buehler. *Atomistic modeling of materials failure.* Springer (New York), 2008.

[29] M. Buehler. Molecular architecture of collagen fibrils: A critical length scale for tough fibrils. *Curr. Appl. Phys.*, pages 440–442, 2008.

[30] M. Buehler. Nanomechanics of collagen fibrils under varying cross-link densities: Atomistic and continuum studies. *J. Mech. Behav. Biomed. Mater.*, 1(1):59–67, 2008.

[31] M. Buehler, F. Abraham, and H. Gao. Hyperelasticity governs dynamic fracture at a critical length scale. *Nature*, 426:141–146, 2003.

[32] M. Buehler, J. Dodson, P. Meulbroek, A. Duin, and W. Goddard. The computational materials design facility (CMDF): A powerful framework for multi-paradigm multi-scale simulations. *Mat. Res. Soc. Proceedings*, 894:LL3.8, 2006.

[33] M. Buehler, A. C. v. Duin, and W. Goddard. Multi-paradigm modeling of dynamical crack propagation in silicon using the reaxFF reactive force field. *Phys. Rev. Lett.*, 96(9):095505, 2006.

[34] M. Buehler and H. Gao. *Ultra large scale atomistic simulations of dynamic fracture, Handbook of Theoretical and Computational Nanotechnology.* American Scientific Publishers (ASP), 2006.

[35] M. Buehler, H. Tang, A. C. v. Duin, and W. Goddard. Threshold crack speed controls dynamical fracture of silicon single crystals. *Phys. Rev. Lett.*, 99:165502, 2007.

[36] M. Buehler and S. Wong. Entropic elasticity controls nanomechanics of single tropocollagen molecules. *Biophys. J.*, 93(1):37–43, 2007.

[37] M. Buehler and Y. Yung. Deformation and failure of protein materials in physiologically extreme conditions and disease. *Nat. Mater.*, 8(3):175–188, 2009.

[38] M. J. Buehler. Large-scale hierarchical molecular modeling of nanostructured biological materials. *J. Comput. Theor. Nanosci.*, 3(5):603–623, 2006.

[39] M. J. Buehler. Nature designs tough collagen: Explaining the nanostructure of collagen fibrils. *Proc. Natl. Acad. Sci. U. S. A.*, 103(33):12285–12290, 2006.

[40] M. J. Buehler and H. J. Gao. Dynamical fracture instabilities due to local hyperelasticity at crack tips. *Nature*, 439(7074):307–310, 2006.

[41] M. J. Buehler and S. Keten. Colloquium: Failure of molecules, bones, and the earth itself. *Rev. Mod. Phys.*, 82(2), 2010.

[42] M. J. Buehler, S. Keten, and T. Ackbarow. Theoretical and computational hierarchical nanomechanics of protein materials: Deformation and fracture. *Prog. Mat. Sci.*, 53(8):1101–1241, 2008.

[43] M. J. Buehler, Y. Kong, and H. J. Gao. Deformation mechanisms of very long single-wall carbon nanotubes subject to compressive loading. *J. Eng. Mater. Technol.*, 126(3):245–249, 2004.

[44] M. D. Buschmann and A. J. Grodzinsky. A molecular-model of proteoglycan-associated electrostatic forces in cartilage mechanics. *J. Biomech. Eng-T. ASME*, 117(2):179–192, 1995.

[45] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith. Entropic elasticity of lambda-phage DNA. *Science*, 265(5178):1599–1600, 1994.

[46] C. Bustamante, S. B. Smith, J. Liphardt, and D. Smith. Single-molecule studies of DNA mechanics. *Curr. Opin. Struct. Biol.*, 10(3):279–285, 2000.

[47] M. Carrion-Vazquez, H. B. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser, and J. M. Fernandez. The mechanical stability of ubiquitin is linkage dependent. *Nat. Struct. Biol.*, 10(9):738–743, 2003.

[48] J. Chen, C. Brooks, and J. Khandogin. Recent advances in implicit solvent-based methods for biomolecular simulations. *Curr. Opin. Struct. Biol.*, 18(2):140–148, 2008.

[49] J. Chen, W. Im, and C. Brooks. Balancing solvation and intramolecular interactions: Toward a consistent generalized born force field. *J. Am. Chem. Soc.*, 128(11):3728–3736, 2006.

[50] K. Chenoweth, S. Cheung, A. C. T. van Duin, W. A. Goddard, and E. M. Kober. Simulations on the thermal decomposition of a poly(dimethylsiloxane) polymer using the ReaxFF reactive force field. *J. Am. Chem. Soc.*, 127(19):7192–7202, 2005.

[51] K. Chenoweth, A. C. T. van Duin, and W. A. Goddard. ReaxFF reactive force field for molecular dynamics simulations of hydrocarbon oxidation. *J. Phys. Chem. A*, 112(5):1040–1053, 2008.

[52] S. Cheung, W. Q. Deng, A. C. T. van Duin, and W. A. Goddard. ReaxFF(MgH) reactive force field for magnesium hydride systems. *J. Phys. Chem. A*, 109(5):851–859, 2005.

[53] F. Chiti and C. M. Dobson. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.*, 75:333–366, 2006.

[54] M. M. A. E. Claessens, M. Bathe, E. Frey, and A. R. Bausch. Actin-binding proteins sensitively mediate f-actin bundle stiffness. *Nat. Mater.*, 5(9):748–753, 2006.

[55] B. Colombini, M. A. Bagni, G. Romano, and G. Cecchi. Characterization of actomyosin bond properties in intact skeletal muscle by force spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.*, 104(22):9284–9289, 2007.

[56] J. J. Connor. *Introduction to structural motion control.* Prentice Hall Pearson Education, Inc., Upper Saddle River, N.J., 2003.

[57] T. Courtney. *Mechanical behavior of materials.* McGraw-Hill, New York, NY, USA, 1990.

[58] D. L. Cox, H. Lashuel, K. Y. C. Lee, and R. R. P. Singh. The materials science of protein aggregation. *MRS Bull.*, 30(6):452–457, 2005.

[59] F. H. C. Crick and A. Rich. Structure of polyglycine ii. *Nature*, 176(4486):780–781, 1955.

[60] S. Cross, Y.-S. Jin, J. Rao, and J. Gimzewski. Nanomechanical analysis of cells from cancer patients. *Nature Nanotech.*, 2:780783, 2007.

[61] A. da Silva and O. Teschke. Dynamics of the antimicrobial peptide PGLa action on Escherichia coli monitored by atomic force microscopy. *World J. Microbiol. Biotechnol.*, 21(6-7):1103–1110, 2005.

[62] M. Dao, C. T. Lim, and S. Suresh. Mechanics of the human red blood cell deformed by optical tweezers. *J. Mech. Phys. Solids*, 51(11-12):2259–2280, 2003.

[63] A. A. Deniz, S. Mukhopadhyay, and E. A. Lemke. Single-molecule biophysics: at the interface of biology, physics and chemistry. *J. R. Soc. Interface*, 5(18):15–45, 2008.

[64] C. Dicko, F. Vollrath, and J. M. Kenney. Spider silk protein refolding is controlled by changing pH. *Biomacromolecules*, 5(3):704–710, 2004.

[65] H. Dietz and M. Rief. Elastic bond network model for protein unfolding mechanics. *Phys. Rev. Lett.*, 1(9):098101, 2008.

[66] L. Dougan, A. S. R. Koti, G. Genchev, H. Lu, and J. M. Fernandez. A single-molecule perspective on the role of solvent hydrogen bonds in protein folding and chemical reactions. *ChemPhysChem*, 9(18):2836–2847, 2008.

[67] J. Doyle. Rules of engagement. *Nature*, 446:860, 2007.

[68] N. Du, X. Y. Liu, J. Narayanan, L. A. Li, M. L. M. Lim, and D. Q. Li. Design of superior spider silk: From nanostructure to mechanical properties. *Biophys. J.*, 91(12):4528–4535, 2006.

[69] A. v. Duin, S. Dasgupta, F. Lorant, and W. Goddard. ReaxFF: A reactive force field for hydrocarbons. *J. Phys. Chem. A*, 105:9396–9409, 2001.

[70] D. M. Ebenstein and L. A. Pruitt. Nanoindentation of biological materials. *Nano Today*, 1(3):26–33, 2006.

[71] A. J. Engler, S. Sen, H. L. Sweeney, and D. E. Discher. Matrix elasticity directs stem cell lineage specification. *Cell*, 126(4):677–689, 2006.

[72] K. Eom, P. C. Li, D. E. Makarov, and G. J. Rodin. Relationship between the mechanical properties and topology of cross-linked polymer molecules: Parallel strands maximize the strength of model polymers and protein domains. *J. Phys. Chem. B*, 107(34):8730–8733, 2003.

[73] T. Erdmann and U. S. Schwarz. Stability of adhesion clusters under constant force. *Phys. Rev. Lett.*, 92(10):108102, 2004.

[74] T. Erdmann and U. S. Schwarz. Bistability of cell-matrix adhesions resulting from nonlinear receptor-ligand dynamics. *Biophys. J.*, 91(6):L60, 2006.

[75] E. Evans and K. Ritchie. Dynamic strength of molecular adhesion bonds. *Biophys. J.*, 72(4):1541–1555, 1997.

[76] E. A. Evans and D. A. Calderwood. Forces and bond dynamics in cell adhesion. *Science*, 316(5828):1148–1153, 2007.

[77] M. Feig, J. Karanicolas, and C. L. Brooks. Mmtsb tool set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph. Model.*, 22(5):377–395, 2004.

[78] M. Feig, A. D. MacKerell, and C. L. Brooks. Force field influence on the observation of pi-helical protein structures in molecular dynamics simulations. *J. Phys. Chem. B*, 107(12):2831–2836, 2003.

[79] R. P. Feynman. There's plenty of room at the bottom. *IEEE J. MEMS*, 1:60–66, 1992.

[80] T. E. Fisher, A. F. Oberhauser, M. Carrion-Vazquez, P. E. Marszalek, and J. M. Fernandez. The study of protein mechanics with the atomic force microscope. *Trends Biochem. Sci.*, 24(10):379–384, 1999.

[81] S. A. Fossey, G. Nemethy, K. D. Gibson, and H. A. Scheraga. Conformational energy studies of beta-sheets of model silk fibroin peptides .1. sheets of poly(ala-gly) chains. *Biopolymers*, 31(13):1529–1541, 1991.

[82] B. W. Fraser, P. Nuclear organization of the genome and the potential for gene regulation. *Nature*, 447(7143):413–417, 2007.

[83] P. Fratzl, H. S. Gupta, E. P. Paschalis, and P. Roschger. Structure and mechanical quality of the collagen-mineral nano-composite in bone. *J. Mater. Chem.*, 14(14):2115–2123, 2004.

[84] P. Fratzl and R. Weinkamer. Nature's hierarchical materials. *Prog. Mat. Sci.*, 52(8):1263–1334, 2007.

[85] T. Fukuma, A. S. Mostaert, and S. P. Jarvis. Explanation for the mechanical strength of amyloid fibrils. *Tribol. Lett.*, 22(3):233–237, 2006.

[86] H. Gao, B. Ji, I. Jager, E. Arzt, and P. Fratzl. Materials become insensitive to flaws at nanoscale: Lessons from nature. *Proc. Natl. Acad. Sci. U. S. A.*, 100(10):5597–5600, 2003.

[87] M. Gao, H. Lu, and K. Schulten. Unfolding of titin domains studied by molecular dynamics simulations. *J. Muscle Res. Cell Motil.*, 23(5-6):513–521, 2002.

[88] J. Gatesy, C. Hayashi, D. Motriuk, J. Woods, and R. Lewis. Extreme diversity, conservation, and convergence of spider silk fibroin sequences. *Science*, 291(5513):2603–2605, 2001.

[89] K. Gelse, E. Poschl, and T. Aigner. Collagens - structure, function, and biosynthesis. *Adv. Drug Delivery Rev.*, 55(12):1531–1546, 2003.

[90] J. M. Gere. *Mechanics of materials.* Brooks/Cole, Belmont, Calif., 6th edition, 2004.

[91] F. Gittes, B. Mickey, J. Nettleton, and J. Howard. Flexural rigidity of microtubules and actin-filaments measured from thermal fluctuations in shape. *J. Cell Biol.*, 120(4):923–934, 1993.

[92] W. Goddard. A perspective of materials modeling. In S. Yip, editor, *Handbook of Materials Modeling.* Springer, 2006.

[93] J. Gosline, M. Lillie, E. Carrington, P. Guerette, C. Ortlepp, and K. Savage. Elastic proteins: biological roles and mechanical properties. *Proc. R. Soc. B*, 357(1418):121–132, 2002.

[94] F. Grater, J. Shen, H. Jiang, M. Gautel, and H. Grubmuller. Mechanically induced titin kinase activation studied by force-probe molecular dynamics simulations. *Biophys. J.*, 88(2):790–804, 2005.

[95] A. Griffith. The phenomenon of rupture and flows in solids. *Phil. Trans. Roy. Soc. A*, 221:163–198, 1920.

[96] D. T. Grubb and L. W. Jelinski. Fiber morphology of spider silk: The effects of tensile deformation. *Macromolecules*, 30(10):2860–2867, 1997.

[97] P. A. Guerette, D. G. Ginzinger, B. H. F. Weber, and J. M. Gosline. Silk properties determined by gland-specific expression of a spider fibroin gene family. *Science*, 272(5258):112–115, 1996.

[98] C. Guzman, S. Jeney, L. Kreplak, S. Kasas, A. J. Kulik, U. Aebi, and L. Forro. Exploring the mechanical properties of single vimentin intermediate filaments by atomic force microscopy. *J. Mol. Biol.*, 360(3):623–630, 2006.

[99] T. Haliloglu, I. Bahar, and B. Erman. Gaussian dynamics of folded proteins. *Phys. Rev. Lett.*, 79(16):3090–3093, 1997.

[100] P. Hanggi, P. Talkner, and M. Borkovec. Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys*, 62(2):251341, 1990.

[101] F. Hanke and H. J. Kreuzer. Breaking bonds in the atomic force microscope: Theory and analysis. *Phys. Rev. E*, 74(3), 2006.

[102] M. A. Hartmann and P. Fratzl. Sacrificial ionic bonds need to be randomly distributed to provide shear deformability. *Nano Lett.*, 9(10):3603–3607, 2009.

[103] C. Y. Hayashi, N. H. Shipley, and R. V. Lewis. Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. *Int. J. Biol. Macromol.*, 24(2-3):271–275, 1999.

[104] S. Hayward and N. Go. Collective variable description of native protein dynamics. *Annu. Rev. Phys. Chem.*, 46:223–250, 1995.

[105] B. Heymann and H. Grubmuller. Dynamic force spectroscopy of molecular adhesion bonds. *Phys. Rev. Lett.*, 84(26):6126–6129, 2000.

[106] M. B. Hinman and R. V. Lewis. Isolation of a clone encoding a second dragline silk fibroin. Nephila clavipes dragline silk is a two-protein fiber. *J. Biol. Chem.*, 267(27):19320–19324, 1992.

[107] J. Hirth and J. Lothe. *Theory of Dislocations.* Wiley-Interscience, 1982.

[108] G. P. Holland, M. S. Creager, J. E. Jenkins, R. V. Lewis, and J. L. Yarger. Determining secondary structure in spider dragline silk by carbon-carbon correlation solid-state NMR spectroscopy. *J. Am. Chem. Soc.*, 130(30):9871–9877, 2008.

[109] J. Howard. *Mechanics of Motor Proteins and the Cytoskeleton.* Sinauer Associates, 2001.

[110] G. Hummer and A. Szabo. Kinetics from nonequilibrium single-molecule pulling experiments. *Biophys. J.*, 85(1):5–15, 2003.

[111] W. Humphrey, A. Dalke, and K. Schulten. VMD: Visual molecular dynamics. *J. Mol. Graph.*, 14(1):33, 1996.

[112] W. Hwang, S. G. Zhang, R. D. Kamm, and M. Karplus. Kinetic control of dimer structure formation in amyloid fibrillogenesis. *Proc. Natl. Acad. Sci. U. S. A.*, 101(35):12916–12921, 2004.

[113] C. Hyeon and D. Thirumalai. Measuring the energy landscape roughness and the transition state location of biomolecules using single molecule mechanical unfolding experiments. *J. Phys.: Condens. Matter*, 19(11):–, 2007.

[114] G. R. Irwin. Analysis of stresses and strains near the end of a crack traversing a plate. *J. Appl. Mech.*, 35:379–386, 1957.

[115] B. Isralewitz, M. Gao, and K. Schulten. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.*, 11(2):224–230, 2001.

[116] L. W. Jelinski. Establishing the relationship between structure and mechanical function in silks. *Curr. Opin. Solid State Mater. Sci.*, 3(3):237–245, 1998.

[117] J. E. Jenkins, M. S. Creager, R. V. Lewis, G. P. Holland, and J. L. Yarger. Quantitative correlation between the protein primary sequences and secondary structures in spider dragline silks. *Biomacromolecules*, 11(1):192–200, 2010.

[118] W. Jorgensen, D. Maxwell, and J. TiradoRives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118(45):11225–11236, 1996.

[119] M. Kageshima, M. A. Lantz, S. P. Jarvis, H. Tokumoto, S. Takeda, A. Ptak, C. Nakamura, and J. Miyake. Insight into conformational changes of a single alpha-helix peptide molecule through stiffness measurements. *Chem. Phys. Lett.*, 343(1-2):77–82, 2001.

[120] A. Kajava, J. Squire, and D. Parry. Beta-structures in fibrous proteins. *Fibrous Proteins: Amyloids, Prions And Beta Proteins*, 73, 2006.

[121] S. Kanamaru, P. G. Leiman, V. A. Kostyuchenko, P. R. Chipman, V. V. Mesyanzhinov, F. Arisaka, and M. G. Rossmann. Structure of the cell-puncturing device of bacteriophage T4. *Nature*, 415(6871):553–557, 2002.

[122] J. Keckes, I. Burgert, K. Fruhmann, M. Muller, K. Kolln, M. Hamilton, M. Burghammer, S. V. Roth, S. Stanzl-Tschegg, and P. Fratzl. Cell-wall recovery after irreversible deformation of wood. *Nat. Mater.*, 2(12):810–814, 2003.

[123] J. Keller. The shape of the strongest column. *Arch. Rational Mech.*, 5(1):275–285, 1960.

[124] M. S. Z. Kellermayer, L. Grama, A. Karsai, A. Nagy, A. Kahn, Z. L. Datki, and B. Penke. Reversible mechanical unzipping of amyloid beta-fibrils. *J. Biol. Chem.*, 280(9):8464–8470, 2005.

[125] J. M. Kenney, D. Knight, M. J. Wise, and F. Vollrath. Amyloidogenic nature of spider silk. *Eur. J. Biochem.*, 269(16):4159–4163, 2002.

[126] S. Keten, J. Bertaud, D. Sen, Z. Xu, T. Ackbarow, and M. J. Buehler. *Multiscale Modeling of Biological Protein Materials: Deformation and Failure.* Springer, 2010.

[127] S. Keten and M. Buehler. Nanostructure and molecular mechanics of dragline spider silk protein assemblies. *J. R. Soc. Interface*, under submission, 2010.

[128] S. Keten and M. J. Buehler. Asymptotic strength limit of hydrogen bond assemblies in proteins at vanishing pulling rates. *Phys. Rev. Lett.*, 100(19):198301, 2008.

[129] S. Keten and M. J. Buehler. Geometric confinement governs the rupture strength of h-bond assemblies at a critical length scale. *Nano Lett.*, 8(2):743–748, 2008.

[130] S. Keten and M. J. Buehler. Large deformation and fracture mechanics of a beta-helical protein nanotube: Atomistic and continuum modeling. *Comput. Method. Appl. Mech. Eng.*, 197(41-42):3203–3214, 2008.

[131] S. Keten and M. J. Buehler. Strength limit of entropic elasticity in beta-sheet protein domains. *Phys. Rev. E*, 78(6):061913, 2008.

[132] S. Keten and M. J. Buehler. Atomistic model of the spider silk nanostructure. *Appl. Phys. Lett.*, 96, 2010.

[133] S. Keten, J. Rodriguez Alvarado, S. Mft, and M. Buehler. Nanomechanical characterization of the triple beta-helix domain in the cell puncture needle of bacteriophage T4 virus. *Cell. Mol. Bioeng.*, 2(1):66–74, 2009.

[134] S. Keten, Z. Xu, and M. J. Buehler. Triangular core as a universal strategy for stiff nanofilaments. *in submission*, 2010.

[135] S. Keten, Z. Xu, B. Ihle, and M. J. Buehler. Nanoconfinement controls stiffness, strength and mechanical toughness of [beta]-sheet crystals in silk. *Nat. Mater.*, 9(4):359–367, 2010.

[136] A. Kishimoto, K. Hasegawa, H. Suzuki, H. Taguchi, K. Namba, and M. Yoshida. beta-helix is a likely core structure of yeast prion sup35 amyloid fibers. *Biochem. Biophys. Res. Commun.*, 315(3):739–745, 2004.

[137] R. Kisilevsky. Review: Amyloidogenesis-unquestioned answers and unanswered questions. *J Struct Biol*, 130(2-3):99–108, 2000.

[138] H. Kitano. Computational systems biology. *Nature*, 420(6912):206–210, 2002.

[139] H. Kitano. Systems biology: A brief overview. *Science*, 295(5560):1662–1664, 2002.

[140] T. P. Knowles, A. W. Fitzpatrick, S. Meehan, H. R. Mott, M. Vendruscolo, C. M. Dobson, and M. E. Welland. Role of intermolecular forces in defining material properties of protein nanofibrils. *Science*, 318(5858):1900–1903, 2007.

[141] F. K. Ko and J. Jovicic. Modeling of mechanical properties and structural design of spider web. *Biomacromolecules*, 5(3):780–785, 2004.

[142] A. Krammer, H. Lu, B. Isralewitz, K. Schulten, and V. Vogel. Forced unfolding of the fibronectin type iii module reveals a tensile molecular recognition switch. *Proc. Natl. Acad. Sci. U. S. A.*, 96(4):1351–1356, 1999.

[143] I. Krasnov, I. Diddens, N. Hauptmann, G. Helms, M. Ogurreck, T. Seydel, S. S. Funari, and M. Muller. Mechanical properties of silk: Interplay of deformation on macroscopic and molecular length scales. *Phys. Rev. Lett.*, 100(4):048104, 2008.

[144] J. Kummerlen, J. D. vanBeek, F. Vollrath, and B. H. Meier. Local structure in spider dragline silk investigated by two-dimensional spin-diffusion nuclear magnetic resonance. *Macromolecules*, 29(8):2920–2928, 1996.

[145] M. A. Lantz, S. P. Jarvis, H. Tokumoto, T. Martynski, T. Kusumi, C. Nakamura, and J. Miyake. Stretching the alpha-helix: a direct measure of the hydrogen-bond energy of a single-peptide molecule. *Chem. Phys. Lett.*, 315(1-2):61–68, 1999.

[146] T. Lau, A. Kushima, and S. Yip. An atomistic method for slow structural deformations. *IOP Conf. Series: Materials Science and Engineering*, 3:012002, 2009.

[147] T. Lazaridis and M. Karplus. Effective energy function for proteins in solution. *Proteins: Struct. Funct. Genet.*, 35(2):133–152, 1999.

[148] P. LeDuc and D. Robinson. Using lessons from cellular and molecular structures for future materials. *Adv. Mater.*, 19(22):3761–3770, 2007.

[149] E. H. Lee, M. Gao, N. Pinotsis, M. Wilmanns, and K. Schulten. Mechanical strength of the titin Z1Z2-telethonin complex. *Structure*, 14(3):497–509, 2006.

[150] S. M. Lee, E. Pippel, U. Gosele, C. Dresbach, Y. Qin, C. V. Chandran, T. Brauniger, G. Hause, and M. Knez. Greatly increased toughness of infiltrated spider silk. *Science*, 324(5926):488–492, 2009.

[151] T. Lefevre, M. E. Rousseau, and M. Pezolet. Protein secondary structure and orientation in silk as revealed by Raman spectromicroscopy. *Biophys. J.*, 92(8):2885–2895, 2007.

[152] A. Lehninger, D. L. Nelson, and M. M. Cox. *Lehninger Principles of Biochemistry*. W. H. Freeman, 5th edition, 2008.

[153] W. K. Liu, H. S. Park, D. Qian, E. G. Karpov, H. Kadowaki, and G. J. Wagner. Bridging scale methods for nanomechanics and materials. *Comput. Method. Appl. Mech. Eng.*, 195(13-16):1407–1421, 2006.

[154] H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.*, 75(2):662–671, 1998.

[155] H. Lu and K. Schulten. Steered molecular dynamics simulations of force-induced protein domain unfolding. *Proteins: Struct., Funct., Genet.*, 35(4):453–463, 1999.

[156] H. Lu and K. Schulten. The key event in force-induced unfolding of titin's immunoglobulin domains. *Biophys. J.*, 79(1):51–65, 2000.

[157] B. Y. Ma and R. Nussinov. Molecular dynamics simulations of alanine rich beta-sheet oligomers: Insight into amyloid formation. *Protein Sci.*, 11(10):2335–2350, 2002.

[158] B. Y. Ma and R. Nussinov. Simulations as analytical tools to understand protein aggregation and predict amyloid conformation. *Curr. Opin. Chem. Biol.*, 10(5):445–452, 2006.

[159] A. D. Mackerell. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.*, 25(13):1584–1604, 2004.

[160] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102(18):3586–3616, 1998.

[161] M. W. Mahoney and W. L. Jorgensen. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.*, 112(20):8910–8922, 2000.

[162] J. F. Marko and E. D. Siggia. Stretching DNA. *Macromolecules*, 28(26):8759–8770, 1995.

[163] P. E. Marszalek, H. Lu, H. B. Li, M. Carrion-Vazquez, A. F. Oberhauser, K. Schulten, and J. M. Fernandez. Mechanical unfolding intermediates in titin modules. *Nature*, 402(6757):100–103, 1999.

[164] S. L. Mayo, B. D. Olafson, and W. A. Goddard. Dreiding - a generic force-field for molecular simulations. *J. Phys. Chem.*, 94(26):8897–8909, 1990.

[165] R. Merkel, P. Nassoy, A. Leung, K. Ritchie, and E. Evans. Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature (London)*, 379(6714):50–53, 1999.

[166] N. Miyashita, J. E. Straub, D. Thirumalai, and Y. Sugita. Transmembrane structures of amyloid precursor protein dimer predicted by replica-exchange molecular dynamics simulations. *J. Am. Chem. Soc.*, 131(10):3438–+, 2009.

[167] A. S. Mostaert and S. P. Jarvis. Beneficial characteristics of mechanically functional amyloid fibrils evolutionarily preserved in natural adhesives. *Nanotechnology*, 18(4):044010, 2007.

[168] N. Mucke, L. Kreplak, R. Kirmse, T. Wedig, H. Herrmann, U. Aebi, and J. Langowski. Assessing the flexibility of intermediate filaments by atomic force microscopy. *J. Mol. Biol.*, 335(5):1241–1250, 2004.

[169] M. T. Nelson, W. Humphrey, A. Gursoy, A. Dalke, L. V. Kale, R. D. Skeel, and K. Schulten. NAMD: A parallel, object oriented molecular dynamics program. *Int. J. Supercomput. Ap.*, 10(4):251–268, 1996.

[170] H. Nguyen and C. Hall. Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *Proc. Natl. Acad. Sci. U. S. A.*, 101(46):16180–16185, 2004.

[171] H. Nguyen and C. Hall. Spontaneous fibril formation by polyalanines; discontinuous molecular dynamics simulations. *J. Am. Chem. Soc.*, 128(6):1890–1901, 2006.

[172] A. F. Oberhauser, P. E. Marszalek, H. P. Erickson, and J. M. Fernandez. The molecular elasticity of the extracellular matrix protein tenascin. *Nature*, 393(6681):181–185, 1998.

[173] E. Oroudjev, J. Soares, S. Arcidiacono, J. B. Thompson, S. A. Fossey, and H. G. Hansma. Segmented nanofibers of spider dragline silk: Atomic force microscopy and single-molecule force spectroscopy. *Proceedings of the National Academy of Sciences*, 99(90002):6460–6465, 2002.

[174] F. Pampaloni, G. Lattanzi, A. Jonas, T. Surrey, E. Frey, and E. L. Florin. Thermal fluctuations of grafted microtubules provide evidence of a length-dependent persistence length. *Proc. Natl. Acad. Sci. U. S. A.*, 103(27):10248–10253, 2006.

[175] P. Papadopoulos, J. Soelter, and F. Kremer. Hierarchies in the structural organization of spider silk a quantitative model. *Colloid Polym. Sci.*, 287(2):231–236, 2009.

[176] R. Paparcone, S. Keten, and M. J. Buehler. Atomistic simulation of nanomechanical properties of Alzheimer's A[beta](1-40) amyloid fibrils under compressive and tensile loading. *J. Biomech.*, 43:1196–1201, 2010.

[177] A. K. Paravastu, R. D. Leapman, W. M. Yau, and R. Tycko. Molecular structural basis for polymorphism in alzheimer's beta-amyloid fibrils. *Proc. Natl. Acad. Sci. U. S. A.*, 105(47):18349–18354, 2008.

[178] J. Park, B. Kahng, R. D. Kamm, and W. Hwang. Atomistic simulation approach to a continuum description of self-assembled [beta]-sheet filaments. *Biophys. J.*, 90(7):2510 – 2524, 2006.

[179] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, and P. Kollman. Amber, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.*, 91(1-3):1–41, 1995.

[180] S. Penel, R. G. Morrison, P. D. Dobson, R. J. Mortishire-Smith, and A. J. Doig. Length preferences and periodicity in beta-strands. antiparallel edge beta-sheets are more likely to finish in non-hydrogen bonded rings. *Protein Eng.*, 16(12):957–961, 2003.

[181] M. C. Philip, A. F. Stephen, A. A. Margaret, W. S. John, L. K. David, W. W. Adams, K. E. Ronald, M. David, and L. V. Deborah. Mechanical and thermal properties of dragline silk from the spider Nephila clavipes. *Polym. Adv. Technol.*, 5(8):401–410, 1994.

[182] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, 26(16):1781–1802, 2005.

[183] J. Ponder and D. Case. Force fields for protein simulations. *Adv. Protein Chem.*, 66:27–60, 2003.

[184] D. Porter and F. Vollrath. The role of kinetics of water and amide bonding in protein stability. *Soft Matter*, 4(2):328–336, 2008.

[185] D. Porter, F. Vollrath, and Z. Shao. Predicting the mechanical properties of spider silk as a model nanostructured polymer. *Eur. Phys. J. E*, 16(2):199–206, 2005.

[186] C. B. Prater, H. J. Butt, and P. K. Hansma. Atomic force microscopy. *Nature*, 345(6278):839–840, 1990.

[187] K. G. Ramachandran, G.N. Structure of collagen. *Nature*, 176:593595, 1955.

[188] S. Rammensee, U. Slotta, T. Scheibel, and A. R. Bausch. Assembly mechanism of recombinant spider silk proteins. *Proc. Natl. Acad. Sci. U. S. A.*, 105(18):6590–6595, 2008.

[189] F. Rao and A. Caflisch. Replica exchange molecular dynamics simulations of reversible folding. *J. Chem. Phys.*, 119(7):4035–4042, 2003.

[190] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff. Uff, a full periodic-table force-field for molecular mechanics and molecular-dynamics simulations. *J. Am. Chem. Soc.*, 114(25):10024–10035, 1992.

[191] M. Reches and E. Gazit. Casting metal nanowires within discrete self-assembled peptide nanotubes. *Science*, 300(5619):625–627, 2003.

[192] Y. M. Rhee and V. S. Pande. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys. J.*, 84(2):775–786, 2003.

[193] M. Rief, J. M. Fernandez, and H. E. Gaub. Elastically coupled two-level systems as a model for biopolymer extensibility. *Phys. Rev. Lett.*, 81(21):4764–4767, 1998.

[194] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub. Reversible unfolding of individual titin immunoglobulin domains by afm. *Science*, 276(5315):1109–1112, 1997.

[195] M. Rief, M. Gautel, A. Schemmel, and H. E. Gaub. The mechanical stability of immunoglobulin and fibronectin iii domains in the muscle protein titin measured by atomic force microscopy. *Biophys. J.*, 75(6):3008–3014, 1998.

[196] M. Rief, J. Pascual, M. Saraste, and H. E. Gaub. Single molecule force spectroscopy of spectrin repeats: Low unfolding forces in helix bundles. *J. Mol. Biol.*, 286(2):553–561, 1999.

[197] C. Riekel and F. Vollrath. Spider silk fibre extrusion: combined wide- and small-angle x-ray microdiffraction experiments. *Int. J. Biol. Macromol.*, 29(3):203–210, 2001.

[198] R. Rohs, C. Etchebest, and R. Lavery. Unraveling proteins: A molecular mechanics study. *Biophys. J.*, 76(5):2760–2768, 1999.

[199] C. Rountree, R. Kalia, E. Lidorikis, A. Nakano, L. v. Brutzel, and P. Vashishta. Atomistic aspects of crack propagation in brittle materials: Multimillion atom molecular dynamics simulations. *Annu. Rev. Mat. Res.*, 32:377–400, 2002.

[200] M. E. Rousseau, T. Lefevre, L. Beaulieu, T. Asakura, and M. Pezolet. Study of protein conformation and orientation in silkworm and spider silk fibers using Raman microspectroscopy. *Biomacromolecules*, 5(6):2247–2257, 2004.

[201] M. E. Rousseau, T. Lefevre, and M. Pezolet. Conformation and orientation of proteins in various types of silk fibers produced by Nephila clavipes spiders. *Biomacromolecules*, 10(10):2945–2953, 2009.

[202] B. Roux and T. Simonson. Implicit solvent models. *Biophys. Chem.*, 78(1-2):1–20, 1999.

[203] K. Sanbonmatsu and A. Garca. Structure of met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Proteins: Struct. Funct. Genet.*, 46(2):225–234, 2002.

[204] K. N. Savage and J. M. Gosline. The role of proline in the elastic mechanism of hydrated spider silks. *J. Exp. Biol.*, 211(12):1948–1957, 2008.

[205] T. Scheibel, R. Parthasarathy, G. Sawicki, X. M. Lin, H. Jaeger, and S. L. Lindquist. Conducting nanowires built by controlled self-assembly of amyloid fibers and selective metal deposition. *Proc. Natl. Acad. Sci. U. S. A.*, 100(8):4527–4532, 2003.

[206] H. A. Scheraga, M. Khalili, and A. Liwo. Protein-folding dynamics: Overview of molecular simulation techniques. *Annu. Rev. Phys. Chem.*, 58:57–83, 2007.

[207] I. Schwaiger, C. Sattler, D. R. Hostetter, and M. Rief. The myosin coiled-coil is a truly elastic protein structure. *Nat. Mater.*, 1(4):232–235, 2002.

[208] U. Seifert. Rupture of multiple parallel molecular bonds under dynamic loading. *Phys. Rev. Lett.*, 84(12):2750–2753, 2000.

[209] U. Seifert. Dynamic strength of adhesion molecules: Role of rebinding and self-consistent rates. *Europhys. Lett.*, 58(5):792–798, 2002.

[210] L. C. Serpell. Alzheimer's amyloid fibrils: structure and assembly. *Biochimica Et Biophysica Acta-Molecular Basis of Disease*, 1502(1):16–30, 2000.

[211] Z. Z. Shao and F. Vollrath. The effect of solvents on the contraction and mechanical properties of spider silk. *Polymer*, 40(7):1799–1806, 1999.

[212] Z. Z. Shao and F. Vollrath. Materials: Surprising strength of silkworm silk. *Nature*, 418(6899):741–741, 2002.

[213] S. Sheu, D. Yang, H. Selzle, and E. Schlag. Energetics of hydrogen bonds in peptides. *Proc. Natl. Acad. Sci. U. S. A.*, 100(22):12683–12687, 2003.

[214] A. H. Simmons, C. A. Michal, and L. W. Jelinski. Molecular orientation and two-component nature of the crystalline fraction of spider dragline silk. *Science*, 271(5245):84–87, 1996.

[215] J. Sirichaisit, V. L. Brookes, R. J. Young, and F. Vollrath. Analysis of structure/property relationships in silkworm (Bombyx mori) and spider dragline (Nephila edulis) silks using Raman spectroscopy. *Biomacromolecules*, 4(2):387–394, 2003.

[216] U. Slotta, S. Hess, K. Spiess, T. Stromer, L. Serpell, and T. Scheibel. Spider silk and amyloid fibrils: A structural comparison. *Macromol. Biosci.*, 7(2):183–188, 2007.

[217] B. L. Smith, T. E. Schaffer, M. Viani, J. B. Thompson, N. A. Frederick, J. Kindt, A. Belcher, G. D. Stucky, D. E. Morse, and P. K. Hansma. Molecular mechanistic origin of the toughness of natural adhesives, fibres and composites. *Nature*, 399(6738):761–763, 1999.

[218] J. F. Smith, T. P. J. Knowles, C. M. Dobson, C. E. MacPhee, and M. E. Welland. Characterization of the nanoscale properties of individual amyloid fibrils. *Proc. Natl. Acad. Sci. U. S. A.*, 103(43):15806–15811, 2006.

[219] M. Sotomayor and K. Schulten. Single-molecule experiments in vitro and in silico. *Science*, 316(5828):1144–1148, 2007.

[220] A. Sponner, E. Unger, F. Grosse, and K. Weisshart. Differential polymerization of the two main protein components of dragline silk during fibre spinning. *Nat. Mater.*, 4(10):772–775, 2005.

[221] A. Strachan, E. M. Kober, A. C. T. van Duin, J. Oxgaard, and W. A. Goddard. Thermal decomposition of RDX from reactive molecular dynamics. *J. Chem. Phys.*, 122(5):054502, 2005.

[222] A. Strachan, A. C. T. van Duin, D. Chakraborty, S. Dasgupta, and W. A. Goddard. Shock waves in high-energy materials: The initial chemical events in nitramine RDX. *Phys. Rev. Lett.*, 91(9):098301, 2003.

[223] S. J. Stuart, A. B. Tutein, and J. A. Harrison. A reactive potential for hydrocarbons with intermolecular interactions. *J. Chem. Phys.*, 112(14):6472–6486, 2000.

[224] Y. Sugita and Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314(1-2):141–151, 1999.

[225] T. Sulchek, R. W. Friddle, and A. Noy. Strength of multiple parallel biological bonds. *Biophys. J.*, 90(12):4686–4691, 2006.

[226] J. I. Sulkowska and M. Cieplak. Mechanical stretching of proteins - a theoretical survey of the protein data bank. *J. Phys.: Condens. Matter*, 19(28):–, 2007.

[227] Y. L. Sun, Z. P. Luo, and K. N. An. Stretching short biopolymers using optical tweezers. *Biochem. Biophys. Res. Commun.*, 286(4):826–830, 2001.

[228] Y. L. Sun, Z. P. Luo, A. Fertala, and K. N. An. Direct quantification of the flexibility of type I collagen monomer. *Biochem. Biophys. Res. Commun.*, 295(2):382–386, 2002.

[229] Y. L. Sun, Z. P. Luo, A. Fertala, and K. N. An. Stretching type II collagen with optical tweezers. *J. Biomech.*, 37(11):1665–1669, 2004.

[230] S. Suresh, J. Spatz, J. P. Mills, A. Micoulet, M. Dao, C. T. Lim, M. Beil, and T. Seufferlein. Connections between single-cell biomechanics and human disease states: gastrointestinal cancer and malaria. *Acta Biomater*, 1(1):15–30, 2005.

[231] K. Tai, F. Ulm, and C. Ortiz. Nanogranular origins of the strength of bone. *Nano Lett.*, 11:2520–2525, 2006.

[232] A. S. Tatham and P. R. Shewry. Elastomeric proteins: biological roles, structures and mechanisms. *Trends Biochem. Sci.*, 25(11):567–571, 2000.

[233] Y. Termonia. Molecular modeling of spider silk elasticity. *Macromolecules*, 27(25):7378–7381, 1994.

[234] B. L. Thiel, K. B. Guess, and C. Viney. Non-periodic lattice crystals in the hierarchical microstructure of spider (major ampullate) silk. *Biopolymers*, 41(7):703–719, 1997.

[235] S. Timoshenko and J. Gere. *Theory of Elastic Stability.* McGraw-Hill, New York, NY, USA, 1988.

[236] M. Tirion. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.*, 77(9):1905–1908, 1996.

[237] V. Tozzini. Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.*, 15(2):144–150, 2005.

[238] J. E. Trancik, J. T. Czernuszka, F. I. Bell, and C. Viney. Nanostructural features of a spider dragline silk as revealed by electron and x-ray diffraction studies. *Polymer*, 47(15):5633–5642, 2006.

[239] D. W. Urry and T. M. Parker. Mechanics of elastin: molecular mechanism of biological elasticity and its relationship to contraction. *J. Muscle Res. Cell Motil.*, 23(5-6):543–559, 2002.

[240] J. D. van Beek, S. Hess, F. Vollrath, and B. H. Meier. The molecular structure of spider dragline silk: Folding and orientation of the protein backbone. *Proc. Natl. Acad. Sci. U. S. A.*, 99(16):10266–10271, 2002.

[241] J. D. van Beek, J. Kummerlen, F. Vollrath, and B. H. Meier. Supercontracted spider dragline silk: a solid-state NMR study of the local structure. *Int. J. Biol. Macromol.*, 24(2-3):173–178, 1999.

[242] D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. Mark, and H. Berendsen. GROMACS: Fast, flexible, and free. *J. Comput. Chem.*, 26(16):1701–1718, 2005.

[243] P. Vashishta, R. Kalia, and A. Nakano. Large-scale atomistic simulations of dynamic fracture. *Comp. in Science and Engrg.*, pages 56–65, 1999.

[244] C. Vepari and D. L. Kaplan. Silk as a biomaterial. *Prog. Polym. Sci.*, 32(8-9):991–1007, 2007.

[245] J. F. V. Vincent. *Structural biomaterials.* Princeton University Press, Princeton, N.J., 1990.

[246] D. Voet and J. G. Voet. *Biochemistry.* J. Wiley & Sons, New York, 2004.

[247] F. Vollrath and D. P. Knight. Liquid crystalline spinning of spider silk. *Nature*, 410(6828):541–548, 2001.

[248] E. B. Walton, S. Lee, and K. J. Van Vliet. Extending Bell's model: How force transducer stiffness alters measured unbinding forces and kinetics of molecular complexes. *Biophys. J.*, 94(7):2621, 2008.

[249] W. Wang, O. Donini, C. M. Reyes, and P. A. Kollman. Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.*, 30:211–243, 2001.

[250] C. Wasmer, A. Lange, H. Van Melckebeke, A. Siemer, R. Riek, and B. Meier. Amyloid fibrils of the het-s(218-289) prion form a beta solenoid with a triangular hydrophobic core. *Science*, 319(5869):1523–1526, 2008.

[251] D. K. West, D. J. Brockwell, P. D. Olmsted, S. E. Radford, and E. Paci. Mechanical resistance of proteins explained using simple molecular models. *Biophys. J.*, 90(1):287–297, 2006.

[252] E. Williams. Some observations of Leonardo, Galileo, Mariotte and others relative to size effect. *Ann. Sci.*, 13(1):23–29, 1957.

[253] S. B. Xiao, W. Stacklies, M. Cetinkaya, B. Markert, and F. Grater. Mechanical response of silk crystalline units from force-distribution analysis. *Biophys. J.*, 96(10):3997–4005, 2009.

[254] M. D. Yoder, S. E. Lietzke, and F. Jurnak. Unusual structural features in the parallel beta-helix in pectate lyases. *Structure*, 1(4):241–251, 1993.

[255] C. B. Yuan, A. Chen, P. Kolb, and V. T. Moy. Energy landscape of streptavidin-biotin complexes measured by atomic force microscopy. *Biochemistry*, 39(33):10219–10223, 2000.

[256] Y. Yung, J. Chae, M. Buehler, C. Hunter, and D. Mooney. Cyclic tensile strain triggers a sequence of autocrine and paracrine signaling to regulate angiogenic sprouting in human vascular cells. *Proc. Natl. Acad. Sci. U. S. A.*, 106(36):5279–15284, 2009.

[257] S. G. Zhang. Fabrication of novel biomaterials through molecular self-assembly. *Nat. Biotechnol.*, 21(10):1171–1178, 2003.

[258] Y. Zhang. Progress and challenges in protein structure prediction. *Curr. Opin. Struct. Biol.*, 18(3):342–348, 2008.

[259] S. N. Zhurkov. Kinetic concept of the strength of solids. *Int. J. Fract. Mech.*, 1:311–323, 1965.