Mechanical Properties of Elastomeric Proteins

Ravi Kappiyoor

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Engineering Mechanics

Ishwar K. Puri, Chair

Daniel M. Dudek, Co-chair

Jianhua Xing

Mark Stremler

Shane Ross

December 2, 2013

Blacksburg, Virginia

Keywords: Protein elasticity, molecular dynamics, nanoscale material properties

Mechanical Properties of Elastomechanical Proteins

Ravi Kappiyoor

ABSTRACT

When we stretch and contract a rubber band a hundred times, we expect the rubber band to fail. Yet our heart stretches and contracts the same amount every two minutes, and does not fail. Why is that? What causes the significantly higher elasticity of certain molecules and the rigidity of others? Equally importantly, can we use this information to design materials for precise mechanical tasks? It is the aim of this dissertation to illuminate key aspects of the answer to these questions, while detailing the work that remains to be done.

In this dissertation, particular emphasis is placed on the nanoscale properties of elastomeric proteins. By better understanding the fundamental characteristics of these proteins at the nanoscale, we can better design synthetic rubbers to provide the same desired mechanical properties.

This work was supported in part by the MultiSTEPS NSF IGERT Grant.

Dedication

I dedicate this dissertation to my family. My parents have always provided me with financial and emotional support and my little brother has always been a good friend to laugh with. Without their support, I would not have been able to complete even a small portion of this work.

Acknowledgments

It would have been impossible to complete even a fraction of this work without the input of my advisor, Dr. Ishwar K. Puri. He has been a source of constant inspiration with his unparalleled scientific vision. Aside from his constant encouragement for independent thinking and novel approaches to solve existing scientific problems, he has impressed upon me the art of communicating science to society. While the current academic literature is rife with esoteric jargon, confusing and perplexing the average reader, he required all of our work to be clear and concise before even considering submission to a peer-reviewed journal. Further, he has been an excellent research mentor. Over the last three and a half years, there have been many times when I have overlooked key details that he quickly spotted and brought to my attention.

I would also like to thank my collaborator and former labmate Monrudee, who took me under her wing when I was still an undergraduate student in ESM just trying to figure out what research was. She was the one who first taught me the thought process involved with research at the bleeding edge, and mentored me in my first-ever research project. Her input

and guidance was instrumental in making me the researcher I am today.

Ganesh provided my first introduction to molecular dynamics (MD) studies. Without his guidance and mentorship, I would not have been unable to undertake several of the studies that I present in this dissertation.

Special thanks to my labmates: Aschvin, Austin, Souvik, and Suvojit for making my stay in the Multiphysics Research Group such an enjoyable one. Aside from all of the friendly and helpful conversations we have had in the lab, they have been very helpful in my personal life as well.

I thank my committee members for taking time out of their busy schedules to give valuable feedback on my research. I also thank the entire ESM staff, especially Anne-Marie, Linda, Lisa, and Sally for always going out of their way to provide me with all of the support I could have asked for.

Last, but certainly not least, I thank my parents for their support. Without their emotional and financial support, I would never have been able to even make it to graduate school, let alone finish it. They provided motivation whenever I was feeling down, and provided guidance whenever I was lost and confused. It is impossible to overstate how much they have influenced in making me who I am today.

Attributions

I would like to thank my collaborators for their input in my dissertation. Specifically, I would like to thank Drs. Balasubramanian, Dudek and Puri for their help and advice in my first work as a graduate student, determining the effect of charge on the mechanical properties of elastomechanical properties (Chapter 3 of this document). I would like to further thank Drs. Dudek and Puri for their insight and advice for my work in determining the importance of secondary structure on the properties of titin (Chapter 4 of this document). Finally, I wish to thank Dr. Puri for his advice and mentorship for the final two studies of this dissertation, the effect of cross-linking on mechanical properties (Chapter 5) and the effect of solvent on mechanical properties (Chapter 6).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In fiscal year 2012 alone, the company Firestone made a net profit of over $2.15 billion on tire sales, accounting for 84% of their total business and an increase of almost 28% over the same number last year. Over 1/3 of Americans today die due to heart failure. Without the ability to effectively patch the heart for a long period of time, this number is likely to grow in the future. What do these seemingly disparate examples have in common? They both involve materials that exhibit long range elasticity. My doctoral research focuses on discovering, at a molecular level, the underlying cause of this long range elasticity.

1.1 Long Range Elasticity

Long range elasticity is defined as the ability of a material to return to its original conformation even after a large deformation has been applied. We are quite familiar with a

variety of different materials that exhibit this type of elasticity, e.g. airplane tires, the human heart, etc. While we have been successful in taking advantage of these material properties to build a wide array of products, ranging from rubber bands to body armor, we have not yet been able to reproduce the full range of material properties seen in nature. For example, polybutadiene, the most resilient artificial rubber known to man, has a resilience of only 80%. This is much lower than the 99% resilience found in resilin, a protein found in insect cuticles [1]. It is the focus of my doctoral research to highlight the underlying causes, at a molecular level, of how to control the ability of certain molecules to remain elastic over long deformations.

To better understand the underlying causes for the variety of material properties in biological materials, we start from the bottom: atoms, the elementary form of matter. By understanding how individual atoms form into complex structures that yield vastly different material properties, we will be better prepared to understand the fundamental requirements for a highly elastic material, thereby enabling us to fabricate novel materials to exact specifications.

1.2 Nature as Inspiration

When we stretch and contract a rubber band a hundred times, we expect the rubber band to fail. Yet our heart stretches and contracts the same amount every two minutes, and does not fail. Why is that? What causes the significantly higher fatigue-lifetime of certain

molecules and the quick failure of others? Equally importantly, can we use this information to design materials for precise mechanical tasks? It is the aim of this dissertation to illuminate key aspects of the answer to these questions, while detailing the work that remains to be done.

Nature has a wide variety of materials to accomplish mechanical tasks. For example, elastin is responsible for maintaining the shape of arteries and lung alveoli [2], resilin maintains the structural stability in the wing joints of dragonflies [3], and spider silks are used as a protective layer for spider eggs and to create spider webs [4]. In order to perform this diverse set of tasks, organisms have evolved to produce elastomechanical materials having a broad spectrum of elastic properties: resilin is the most resilient rubber (i.e. it has the lowest hysteresis loss) known to man [5, 1], elastin is highly durable [6], and spider silk is one of the toughest materials ever discovered [7]. A number of molecular studies have been performed [8, 9] to obtain a better understanding of the origin of these remarkable properties. In this project, we look only at a small subset of this vast group, termed elastomeric proteins.

1.3 Elastomeric Proteins

Elastomeric proteins are a class of proteins that have high resilience and low stiffness [10]. They have evolved in a variety of biological systems and perform precise biological roles [5], such as the use of elastin in the human heart to provide high fatigue-lifetime [11]. Even within this small subset of materials, however, there is a vast array of structures and

functionality that remain to be explored. In this section, we highlight the questions that will be answered in this document as well as questions that would require future work.

One protein of particular interest is resilin, a protein found primarily in insect cuticles. Resilin has a long fatigue-lifetime and has a higher resilience than polybutadiene [1, 12, 13], making it of particular interest. Resilins have several different types and functions, depending on the insect in which it is found. For example, in fruit flies resilin provides structural support during repeated movements of their wings [1], while in cicadas [14] and moths [15], resilin is an energy storage device that efficiently stores and releases the vibrational energies in sound-producing organs. However, despite these vast differences in function, there are two common threads that connect all forms of resilin: (1) resilin tends to be highly polar [16], and (2) there is a four amino acid motif (sequence) that repeats in every resilin [17] - tyrosine (Y), glutamine (G), alanine (A), and proline (P), or YGAP in the FASTA format as detailed in Table 1.1 - where crosslinking between different strands of resilin is believed to occur. What is special about the four amino acids presented above? Similarly, what is special about polarity? Why is it that, though significant portions of resilin are different among various insects, those two characteristics remain conserved throughout the vast majority of insects?

Another elastomeric protein of interest, titin, is found in human muscle cells [18]. Titin is a large protein that spans the length of a half-sarcomere. However, only a very small portion of titin is functionally extensible [19]. This portion of titin consists of a region that has no obvious secondary structure - composed of primarily proline (P), glutamic acid (E), valine (V) and lysine (K) - called the PEVK domain, and a region that is immunoglobin-like and

folds into a specific secondary structure [18, 20, 21, 22, 23]. Of these two regions, is one or the other more responsible for the elasticity of titin? If so, how much of this is due to the difference in structure, and how much is due to the difference in amino acid content?

1.4 Computational Tools

Computational experiments can be undertaken for any potential research problem with a low cost of resources and time, something that physical experiments cannot replicate. Further, it is possible to perform parametric investigations of atomic level properties through the use of computational tools, illuminating novel features that can be used to tune material properties at the molecular level. In this dissertation, we are concerned with two such computational tools: classical molecular dynamics (MD) simulations and Car-Parrinello molecular dynamics (CPMD) simulations. MD simulations treat individual atoms as spheres, and employs an empirical potential energy function to determine the intermolecular interactions among the atoms. CPMD simulations treats the atomic nucleus and inner electrons as a single entity (termed the ionic center), and the valence electrons as a separate entity, and employs density functional theory to determine intermolecular interactions. A detailed explanation of these two computational tools is provided in Chapter 2 of this dissertation.

Table 1.1: Definitions of FASTA text format used

Single Letter Representation	Amino Acid
A	Alanine
B	Aspartic Acid
C	Cysteine
D	Aspartic Acid
E	Glutamic Acid
F	Phenylalanine
G	Glycine
H	Histidine
I	Isoleucine
K	Lysine
L	Leucine
M	Methionine
N	Asparagine
O	Pyrrolysine
P	Proline
Q	Glutamine
R	Arginine
S	Serine
T	Threonine
U	Selenocystteine
V	Valine
W	Tryptophan
Y	Tyrosine
Z	Glutamic Acid

1.5 Document Overview

This dissertation document spans seven chapters. Chapter 1 (the current chapter) is the introduction. Chapter 2 covers the methodology used for this work. Chapter 3 shows that higher side chain polarity on the amino acids of resilin decrease the stiffness and increase the extensibility of resilin. The results have been published in [24] - reproduced with the permission of The Royal Society of Chemistry. Chapter 4 shows that, with equivalent structure, proteins with differing amino acid content will have qualitatively similar mechanical properties. Chapter 5 shows the importance of the chemical composition of different cross-links. Chapter 6 shows that increasing the protein-solvent interaction energies greatly decreases the stiffness and increases the extensiblity of resilin. Finally, chapter 7 details the overarching conclusions and work that remains to be done in this field.

Chapter 2

Methodology

Computational experiments allow researchers to perform several experiments simultaneously, an advantage that physical experimental setups cannot provide. Further, computational experiments provide more complete control over the variables of interest. For example, if we were to study the effect of charge on mechanical properties (as is done in this work), computational experiments allow us to manually alter the charge on individual atoms while keeping everything else constant so that we isolate the effect of charge. In this work, we make use of both classical molecular dynamics (MD) as well as Car-Parrinello molecular dynamics (CPMD). The two methodologies are explained below.

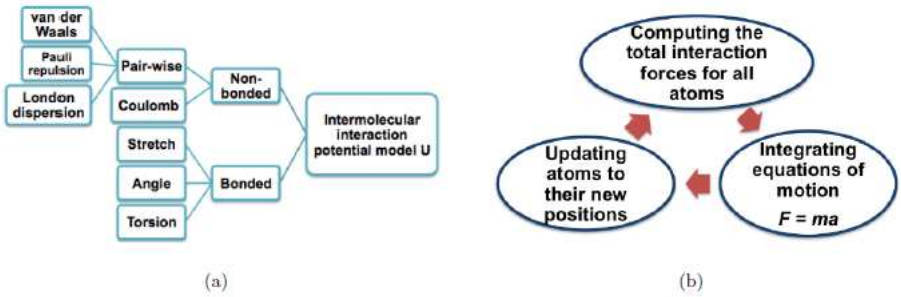

Figure 2.1: (a) The potential model is composed of several different types of interactions. (b) The standard MD simulation strategy.

2.1 Classical Molecular Dynamics

Classical molecular dynamics (MD) simulates the motion of individual atoms and molecules. The initial position and velocities of individual atoms is given as input, and trajectories are determined through the numerical solution of Newton's equations of motion for a system of interacting particles. The forces that act on each atom are determined by using Newton's Second Law:

$$m\frac{\partial^2 r}{\partial t^2} = F = -\frac{\partial U}{\partial r} \tag{2.1}$$

Here, U is an empirical potential energy function that is determined through experimental or quantum mechanical simulation data. This potential energy is composed of several terms: non-bonded interaction energy (e.g. van der Waals and coulombic interaction energies), bonded interaction energy, etc.

One of the chief tasks in setting up any MD simulation is the proper selection of which terms are important to model, and the functional form of each of those terms (called the force field). For example, we will consider two force fields: MM3 and OPLS-AA. The force field MM3 [25] (typically used for hydrocarbons) is defined as:

$$\begin{aligned} U_{MM3} = &\sum_{nonbonded} U_{nonbonded} + \sum_{bonds} U_{bond} + \sum_{angles} U_{angle} + \sum_{dihedrals} U_{dihedrals} + \sum_{impropers} U_{impropers} \\ &+ \sum_{bonds} \sum_{bonds'} U_{bond-bond} + \sum_{bonds} \sum_{angles} U_{bond-angle} \\ &+ \sum_{bonds} \sum_{dihedrals} U_{bond-dihedrals} + \sum_{angles} \sum_{dihedrals} U_{angles-dihedrals} \\ &+ \sum_{angles} \sum_{angles'} \sum_{dihedrals} U_{angles-angles-dihedrals} \end{aligned} \tag{2.2}$$

Here, $U_{nonbonded}$ refers to the van der Waals and coulombic interactions between two non-bonded atoms, U_{bond} refers to the potential energy due to a bond between two atoms, U_{angle} refers to the potential energy due to an angle formed between two bonds, $U_{dihedral}$ refers to the potential energy due to the dihedral angle formed between two connected angles, and $U_{improper}$ is a corrective term. The other terms refer to interactions along the molecule - e.g. $U_{bond-bond}$ calculates how the change in the length of one bond affects the length of other bonds. In the OPLS-AA force field [26], however, several of these terms are simply set to zero and we are left with:

$$U_{OPLS-AA} = \sum_{nonbonded} U_{nonbonded} + \sum_{bonds} U_{bond} + \sum_{angles} U_{angle} + \sum_{dihedrals} U_{dihedrals} \tag{2.3}$$

Clearly, then, if we expect interactions along the molecule to be an important part of the dynamics of the molecule we wish to simulate, then OPLS-AA is not a valid force field to

use. We would need to use MM3 (or other similar force fields) in order to get an accurate result from the simulation.

Aside from which terms are added, we see that even the functional form of some of the terms that are included is different. For example, we look at the $U_{dihedral}$ term of MM3 and OPLS-AA, respectively:

$$U_{dihedral-MM3} = K_1\,[1 - cos(\phi)] + K_2\,[1 - cos(2\phi)] + K_3\,[1 - cos(3\phi)] \tag{2.4}$$

$$U_{dihedral-OPLS-AA} = K_1\,[1 - cos(\phi - \phi_0)] + K_2\,[1 - cos(2(\phi - \phi_0))] + K_3\,[1 - cos(3(\phi - \phi_0))] + K_4\,[1 - cos(4(\phi - \phi_0))] \tag{2.5}$$

Here, K_i and ϕ_0 are parameters that need to be input. Aside from the extra term that we see in $U_{dihedral-OPLS-AA}$, we notice one key difference between the two force fields: in MM3, the potential energy is assumed to be at a minimum when the dihedral angle is $\phi = 0$, whereas in OPLS-AA, no such assumption is made; ϕ_0 is an input parameter that determines the location of the energy minimum. As such, MM3 is not a valid force field to use when the molecular geometry is not planar (i.e. the dihedral angle is non-zero at the energy minimum).

For the MD simulations presented in this work, we are interested in the simulation of proteins in an aqueous environment, which is precisely what the OPLS-AA force field was optimized for [26]. This potential energy can be written as:

$$U = \sum_{i,j>i} 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] + \sum_{i,j>i} \frac{q_i q_j}{C r_{ij}} + \sum_{bonds} K_b \left(b - b_0 \right)^2 + \sum_{angles} K_\theta \left(\theta - \theta_0 \right)^2 +$$

$$\sum_{dihedrals} \left(K_1 \left[1 - cos(\phi - \phi_0) \right] + K_2 \left[1 - cos(2(\phi - \phi_0)) \right] + K_3 \left[1 - cos(3(\phi - \phi_0)) \right] + \right.$$

$$\left. K_4 \left[1 - cos(4(\phi - \phi_0)) \right] \right) \tag{2.6}$$

Here, ϵ, σ, q_i, K_b, K_θ, K_i, b_0, θ_0, and ϕ_0 are all input parameters. ϵ is a measure of the strength of van der Waals interaction between two different atoms and σ is a measure of the length-scale on which this interaction is important. q_i and q_j are the charges on atoms i and j, respectively. K_b, K_θ, and K_i are stiffness values that determine the extent of the force holding a bond, angle, or dihedral, respectively, in place. b_0, θ_0, and ϕ_0 are the equilibrium bond length, angle, and dihedral angle, respectively.

Once the empirical potential energy function has been set, we must then decide on the ensemble to be used. Though there are several such ensembles that can be used, there are only three ensembles that are used commonly in MD simulations: the microcanonical ensemble, the canonical ensemble, and the isothermal-isobaric ensemble. These ensembles represent different physical assumptions about the environment of the system. These ensembles are taken into account as modifications of either the potential or kinetic energy of the system.

The microcanonical ensemble represents a system that is isolated in that it cannot exchange energy or particles with its environment, and cannot expand or contract in volume. It therefore corresponds to an adiabatic process with no heat exchange. For this ensemble,

then, there is no modification to the potential energy function required. As the macroscopic variables that are kept constant in the microcanonical ensemble are: (i) the number of particles in the system (N), (ii) the volume of the system (V), and (iii) the total energy of the system (E), the microcanonical ensemble is commonly called the NVE ensemble.

The canonical ensemble represents a system that is in thermal equilibrium with a heat bath. The system is allowed to exchange energy (but not particles or volume) with the heat bath, so that total energy in this system is not necessarily conserved. This is done through the use of a thermostat. Commonly used thermostats include the velocity-rescale thermostat, in which the kinetic energies of the atoms are directly altered to correspond with the correct temperature, and the Nose-Hoover thermostat, in which the heat bath is made an integral part of the system through the addition of an artificial variable associated with an artificial mass. As the macroscopic variables that are kept constant in the canonical ensemble are: (i) the number of particles in the system (N), (ii) the volume of the system (V), and (iii) the temperature of the system (T), the canonical ensemble is commonly called the NVT ensemble.

The isothermal-isobaric ensemble represents a system that is in thermal equilibrium with a heat bath that applies a fixed pressure on the system. As in the canonical ensemble, the system is allowed to exchange energy (but not particles) with the heat bath so that the total energy is not necessarily conserved. Further, a system in the isothermal-isobaric ensemble is allowed to grow or shrink in order to adapt to the pressure applied by the heat bath. The energy transfer is, as above, done through the use of a thermostat. The change in volume

is determined through the use of a barostat. Examples of barostats include the Berendsen barostat, which simply scales the size of the system every few timesteps to adapt to the pressure, and the Parrinello-Rahman barostat, in which a new variable is introduced to the potential energy to model the motion of the boundary of the system. As the macroscopic variables that are kept constant in the isothermal-isobaric ensemble are: (i) the number of particles in the system (N), (ii) the pressure of the system (P), and (iii) the temperature of the system (T), the isothermal-isobaric ensemble is commonly called the NPT ensemble.

2.2 Car-Parrinello Molecular Dynamics

As mentioned above, MD simulates the motions of individual *atoms*. This means that the information regarding the electronic structure of the atoms are not simulated, and are in fact taken into account by the potential parameters that are input for the simulation. In order for MD to give correct results, then, we must assume that the electronic structures of the atoms do not change over the course of the simulation. This means that simulations that involve a drastic change in temperature or pressure, or involve the formation of a covalent and/or ionic bond cannot be simulated through the use of MD simulations. Instead, we will make use of Car-Parrinello molecular dynamics (CPMD) [27]. In CPMD, as in MD, the initial atomic positions and velocities are given as input. Here, however, instead of considering each atom as a single unit, we split it up into two different components: (1) the valence electrons and (2) the nucleus of each atom and the non-valence electrons, which together will be called the

ionic mass center of each atom.

In principle, it appears to be very simple to amend the MD methodology and allow the simulation of the electronic configuration as well. By explicitly solving the electronic structure problem, and using this electronic structure to calculate forces on the ionic mass centers instead of an empirical potential energy function, we can easily accommodate the incorporation of valence electrons into our simulations. Unfortunately, however, attempts to follow this methodology have shown that atomic dynamics do not conserve energy unless a high degree of convergence is required in the electronic-structure calculations. This makes this solution scheme prohibitive, as the computational resources required to solve even a single electronic structure problem using density functional theory are very high [27]. Doing this for every timestep, then, for tens of thousands to millions of timesteps is impractical.

CPMD makes use of simulated annealing to minimize the electronic energy. While the usual approach to simulated annealing is the use of Monte Carlo techniques [27], it is not easy to use in the electronic structure problem. As electronic structure is normally described through the use of a set of orbitals which must be orthonormal, the move of each electronic variable is constrained. However, in Monte Carlo, the move must be random and achieving randomness in the presence of constraints is not straightforward [27]. In order to get around this, Car and Parrinello proposed that MD equations of motion be used instead [28, 29], with the constraints satisfied using the SHAKE algorithm [30].

Initially, the electrons are assumed to be in an arbitrary configuration. As the simulation progresses, this configuration 'rolls down' the energy surface, thereby gaining fictitious kinetic

energy (the 'motion' corresponds to motion in the electronic configuration space and is not a physical motion; therefore, the kinetic energy associated with this motion is not a physical kinetic energy). Fictitious kinetic energy is removed periodically from the system until the electrons finally come to rest at a potential energy minimum. While this process does not guarantee a global minimum has been reached, it has thus far achieved satisfactory results [27]. This process is run in parallel to conventional MD on the ionic mass centers.

2.3 Summary

The first part of this work - effect of charge on elastomechanical properties of resilin (chapter 3) and the effect of secondary structure on the mechanical properties of titin (chapter 4) - was done using MD simulations. The second part of this work - effect of cross-linking on the mechanical properties of resilin (chapter 5) and the effect of varying solvent on mechanical properties of resilin (chapter 6) - was done using CPMD simulations.

Chapter 3

Effect of Charge on Elastomechanical Properties of Resilin

3.1 Motivation

While there are several types of elastomeric proteins, attention has recently focused on resilin, a protein primarily found in insect cuticles, due to its long fatigue-lifetime and a resilience higher than polybutadiene [1, 12, 13]. Resilins have different amino acid sequences and perform different functions depending on the insect in which it is found. As such, the individual mechanical properties of different resilins are also dissimilar. However, there are two key similarities among all types of resilin: (1) resilin is hydrophilic [16] and (2) an four-amino acid motif consisting of tyrosine (Y), glutamine (G), alanine (A), and proline (P)

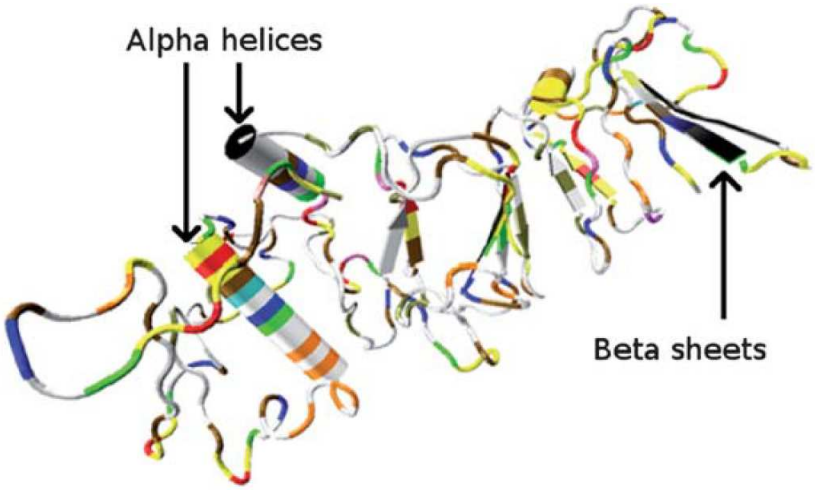

Figure 3.1: Homology model of resilin. Cylinders represent alpha helices and arrows represent beta sheets (some representative cases labeled). An accurate model of resilin would be expected to have little to no secondary structure (i.e. no alpha helices or beta sheets). However, there is a significant amount in this structure (obtained using the templates protease 1KAP and hydrolase 1GO8), making it inapplicable for correct quantitative measurements.

repeats several times throughout the protein [17]. It is unclear whether these similarities among resilins arise as a result of evolutionary processes, conserving the crucial aspects of each type of resilin for its intended task, or if these are simply the result of genetic drift. In this work, we examine the hydrophilicity of different resilin molecules in order to determine whether it has any effect on the mechanical properties.

3.2 Methodology

We employ a bottoms-up approach to investigate the mechanical properties of resilin through detailed atomistic simulations. We simulate four different motifs and characterize their mechanical properties in order to better understand the factors that provide resilin with an extraordinary elasticity. Additionally, a comprehensive understanding of the inherent mechanical properties of these motifs, which are the building blocks of resilin, can enable us to engineer specific types of synthetic bio-inspired elastomeric rubber targeted towards specific applications (e.g., efficient energy storage, or handling repeated movements and vibrations in mechanically moving devices).

Molecular dynamics simulations require an initial structure for the protein. Since a full strand of resilin is believed to be highly unstructured with little or no secondary structure [31, 32] we need to adopt a template protein that has high sequence identity with a particular resilin. We use the amino acid sequence of the resilin found in *Drosophila melanogaster* [1] and obtain a homology model [33]. We use two templates (protease 1KAP and hydrolase 1GO8) that have similar amino acid sequences and known structures to create a model structure for resilin. However, the presence of a significant amount of secondary structure (alpha helices and beta sheets) in this model, as shown in Figure 3.1, indicates that it has some error. While the model could be useful to make qualitative predictions, we instead consider only certain motifs that are of a short length and do not form secondary structures, since these should be able to provide more reliable insights into the mechanical properties of

resilin.

We use four different motifs, two commonly found in the fruit fly resilin (YGAPGGGNGGRPSDT and YGAPGGGDGNGGRPSSS) and the other two in mosquito resilin (YGAPAQTPSSQ and YGAPAPSRPSQQ). Additional simulations involving these motifs are conducted by (i) setting all the charges of every atom in the entire motif to zero (which, although nonphysical and unrealistic, can aid our understanding of the role that polarity has on the mechanical properties of resilin), (ii) setting all the charges of every atom in the entire motif to half of their natural value, and (iii) replacing each of the polar amino acids (except for the initial tyrosine (Y)) in the same motifs with a random nonpolar amino acid. A structure for these motifs is created using the freely available AmberTools program tLeap [34].

Our molecular dynamics (MD) simulations, using the GROMACS package [35] for all the structures thus prepared, undergo the same set of procedures through a leapfrog time-integration scheme with a time step of 0.002 picoseconds (ps), i.e.: (i) energy minimization (no time marching during this simulation step), (ii) a canonical NVT (constant volume and constant temperature) ensemble based simulation performed over 2×10^7 time steps with temperature controlled at 310 K by rescaling the molecular velocities every 1ps, (iii) two NPT (constant pressure and constant temperature) ensemble based simulations, each performed over 2×10^7 time steps, the first by velocity rescaling for thermostat and a Berendsen barostat to control the pressure at 1 bar, and the second using a Nose-Hoover thermostat and a Parrinello-Rahman barostat, both with coupling times of 1 ps, (iv) an NVE (constant volume and constant energy) equilibration, during which the motif is allowed to move freely,

for 5×10^7 time steps, (v) followed finally by a pulling simulation that holds the N-terminus of the motif (the amide group attached to the alpha carbon of the first amino acid) fixed while pulling on the C-terminus (the carboxyl group attached to the alpha carbon of the final amino acid) at a constant strain rate of 0.1 nm/ns. (Since the literature provides no evidence of the strain-dependency of resilin, we chose 0.1 nm/ns to provide good resolution within a reasonable time.) We used the OPLS AA force field [36] (which was used for similar studies of protein mechanical strengths [37]) for our simulations with Lorentz-Berthelot mixing rules employed for cross-interactions. The simulations were run in bulk water (using the SPC potential for water), with periodic boundary conditions. Coulombic interactions had a cutoff distance of 1.4 nm, while van der Waals interactions used a switch cutoff scheme with an inner and outer cutoff of 1.4 and 1.5 nm, respectively. Long range charge interactions were obtained using a Particle Mesh Ewald (PME) summation. Each result is an average of three different simulations.

3.3 Results and Discussion

The force versus displacement curve for the YGAPAQTPSSQ motif found in mosquito [17] resilin is shown in Figure 3.2(A). The motif has low stiffness and is capable of withstanding high strains since the force required to deform the motif is minimal (a maximum force of 13 pN is observed) and the motif does not undergo failure even after being subject to a very high strain (of 26.7%, or a deformation of 0.8 nm) that is similar to one that is

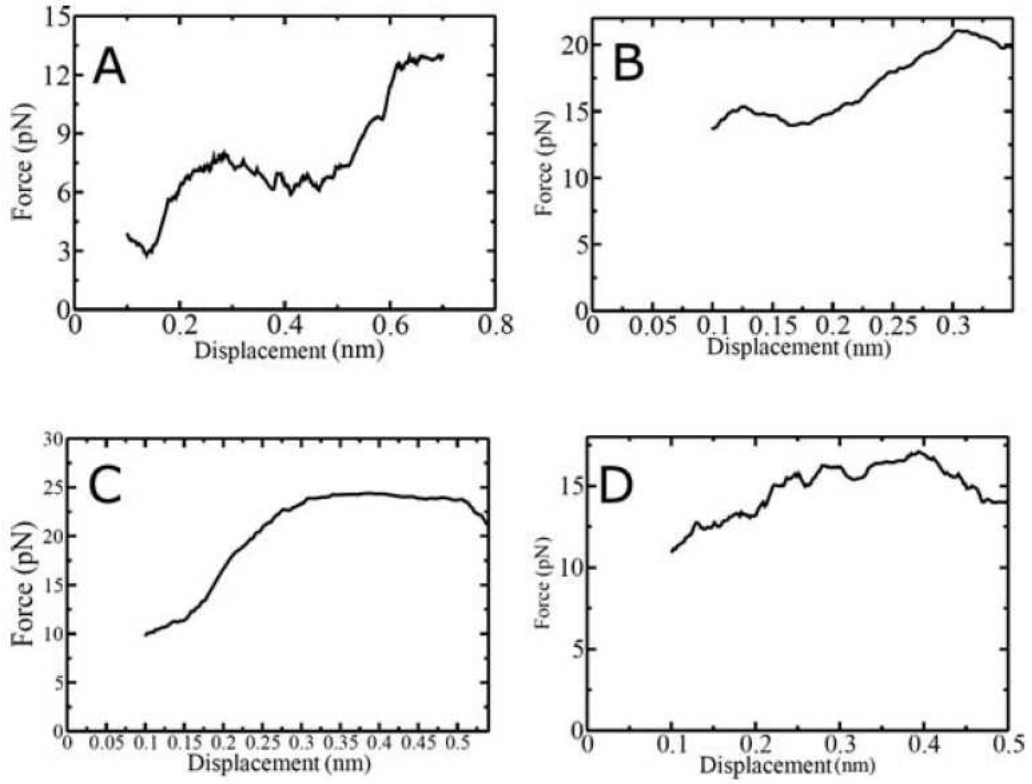

Figure 3.2: Force versus displacement curve for (A) naturally occurring YGAPAQTPSSQ, (B) the zero-charge counterpart (i.e., the same motif with all of the charges set to zero), (C) the half-charge counterpart (i.e., the same motif with all of the charges set to half their natural value) and (D) the nonpolar counterpart (i.e., a motif in which all of the polar amino acids, except the first Y, are replaced with a random nonpolar motif). The force data shown represents a moving average over the previous 100 force values obtained from the simulation. The naturally occurring motif experiences a maximum force of 13 pN, at which point it has deformed 26.7%. The zero-charge counterpart experiences a maximum force of 21 pN, and at which point it has deformed 10%. The nonpolar counterpart experiences a maximum force of 17 pN, at which point it has deformed 13.3%.

imposed on an entire resilin chain [1].

In order to understand the influence of charges on the mechanical properties of resilin, we simulate (i) YGAPAQTPSSQ with all charges on each side chain set to zero (we refer to this as the zero-charge counterpart), (ii) the same motif with all charges set to half their natural value (referred to as the half-charge counterpart) and (iii) YGAPAGWPVVG, where all of the polar amino acids (excepting the first tyrosine (Y)) are replaced with nonpolar amino acids (which we refer to as the nonpolar counterpart). The force versus displacement curves obtained for the above three cases are presented in Figures 3.2(B), 3.2(C) and 3.2(D), respectively. Figure 3.2(B) shows that the force required to deform the same motif (for which results are shown in 3.2(A)) by even 0.1 nm when the charges are removed is 14 pN, which is 1 pN more than the maximum force observed when charges are present. When simulated for longer times, the motif breaks at a deformation of only 0.3 nm, or a strain of 10%, with a force of slightly over 20 pN. Figure 3.2(C) shows a similar high initial stiffness for the half-charge counterpart, followed by a similar early failure at a deformation of 0.51 nm, or a strain of 17% with a force of 24 pN. Figure 3.2(D) shows that the nonpolar motif also exhibits a noticeably higher stiffness than the original motif, requiring more than 15 pN to stretch 0.4 nm for a strain of 13.33%, when it undergoes failure. Results obtained from simulations for other naturally occurring motifs and their zero-charge and nonpolar counterparts are presented in Table 3.1.

The results presented in Table 3.1 and graphically represented in Figure 3.3 show that all naturally occurring motifs have a lower stiffness (defined as the maximum force observed

divided by the maximum displacement observed) and higher extensibility. Since the only difference between the naturally occurring motifs, the zero-charge, half-charge and nonpolar counterparts is the polarity of the amino acid constituents (since the length differences between naturally occurring motifs and its counterparts are at most 1%), we hypothesize that the charges on the amino acids influence the elastic properties of resilins, with increased polarity indicating a higher extensibility and lower stiffness. Such a behavior is also apparent through comparisons between the different naturally occurring motifs. The two strands occurring in fruit fly resilin, though polar, contain a higher percentage of nonpolar amino acids than the strands found in mosquitoes, which contributes to their relatively higher stiffness and lower extensibility.

3.4 Conclusion

While several different resilins occur in different insects, all of them share two essential similarities. First, most resilins have the same repeating set of four amino acids (YGAP), and, second, they are all polar 9. Our objective was to determine if the polarity of the side chains plays a role in the mechanical properties of resilin. The results unambiguously reveal the strong correlation between the polarity of resilin and its superior mechanical properties. A larger percentage of polar amino acids within a motif provides both a higher resistance to fracture and lower stiffness. This dependence of the elastic behavior on the polarity is explained through the interactions between the amino acid side chains and water.

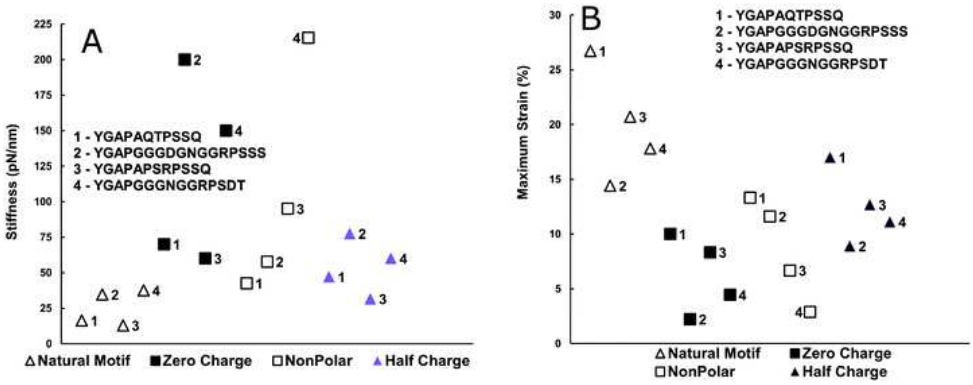

Figure 3.3: (A) The stiffness, defined as the ratio of the maximum force observed before failure to the maximum displacement observed before failure, of each naturally occurring motif (represented by triangles), the zero-charge counterparts (represented by filled squares), and the non-polar motifs (represented by empty squares). The natural motifs have a lower stiffness than their zero-charge, half-charge and nonpolar counterparts. (B) The maximum strains observed, defined as the ratio of the deformation observed before failure (or, if the motif did not undergo failure, the maximum deformation observed) to the initial length of the motif, for the natural motifs (represented by triangles), their zero-charge (represented by filled squares), and non-polar counterparts (represented by empty squares). The natural motifs have a higher extensibility than any of the counterpart sets of motifs.

Table 3.1: The maximum force and displacement recorded for each motif that was tested. The first column denotes the motif, the second describes the type of motif, and the third lists the maximum force for each motif. The maximum displacement and the corresponding strain show that the natural motifs are far more extensible than either of their two counterparts. The higher average number of hydrogen bonds formed for the natural motifs indicate that these have a stronger interaction with water. All of the natural motifs have a lower stiffness (defined as the ratio of maximum force to maximum displacement) than their other two counterpart sets. * indicates that the motif was unbroken throughout the simulation.

Motif	Type	Max. Force (pN)	Max. Displacement (nm)	Strain (%)	Hydrogen Bonds Formed	Stiffness (pN/nm)
YGAPAQTPSSQ	Natural	13	0.8*	26.7	47.195	16.25
YGAPAQTPSSQ	Zero-charge	21	0.3	10	2.951	70
YGAPAQTPSSQ	Half-charge	24	0.51	17	16.561	47.05
YGAPAGWPVVG	Nonpolar	17	0.4	13.3	30.745	42.5
YGAPGGGDGNGGRPSSS	Natural	22.5	0.65	14.4	51.176	34.62
YGAPGGGDGNGGRPSSS	Zero-charge	20	0.1	2.22	6.122	200
YGAPGGGDGNGGRPSSS	Half-charge	31	0.4	8.89	20.976	77.5
YGAPGGGCGAGGAPWWW	Nonpolar	30	0.52	11.6	37.902	57.69
YGAPAPSRPSSQ	Natural	8	0.62	20.7	43.000	12.9
YGAPAPSRPSSQ	Zero-charge	15	0.25	8.33	4.195	60
YGAPAPSRPSSQ	Half-charge	12	0.38	12.67	13.073	31.58
YGAPAPWCPWWI	Nonpolar	19	0.2	6.67	34.756	95
YGAPGGGNGGRPSDT	Natural	30	0.8*	17.8	51.244	37.5
YGAPGGGNGGRPSDT	Zero-charge	30	0.2	4.44	5.707	150
YGAPGGGNGGRPSDT	Half-charge	30	0.5	11.11	20.585	60
YGAPGGGIGGAPWCV	Nonpolar	28	0.13	2.89	36.317	215.4

The enhanced potential energy and hydrogen bonding due to charge interactions between the polar amino acid side chains and surrounding water molecules result in the improved mechanical properties of the hydrophilic naturally occurring motifs. These results have been published in [24] - reproduced with the permission of The Royal Society of Chemistry.

These results only apply to a single chain of resilin. There exist, as yet, no published studies on how the properties of resilin change as two or more chains cross-link. In order to be useful in a bulk environment, it will be important to understand how these results are affected in a cross-linked environment.

Chapter 4

Effect of Structure on Mechanical Properties of Titin

4.1 Motivation

Historically, it has been asserted that the specific functionality of a protein is determined in large part by its structure [38]. Since 1894, when Emil Fischer first proposed the "lock and key" model [39] stating enzymatic "locks" could only be opened by very specific protein "keys", this understanding has served us well. However, we now know that over 50% of all eukaryotic proteins contain large disordered regions within the overall structure [40]. How can we reconcile this information with the original model?

Starting from the 1980s, researchers began to hypothesize that these long disordered

regions are important in order to provide proteins with "flexibility" [38, 41]. While it is intuitively obvious that a disordered region will be more flexible than a region with a set structure, it remains unclear as to how we can characterize it. This is due to the lack of knowledge about the mechanics of these disordered regions [38]. How do these proteins react under stress? How do they interact with their environment? If we take a protein and do nothing else but change the structure, how does that affect its behavior? It is the goal of this paper to address all of these questions through examining an example of a protein with two structurally distinct domains, titin.

Titin is found in muscle cells, where its high extensibility prevents stretch-induced muscle damage [18]. While a single titin molecule spans the length of a half-sarcomere and connects the Z-line to the thick filaments within the sarcomere, the only section of titin that is functionally extensible is found in the I-band of muscles [19]. The extensible region of titin consists of two structurally distinct domains [18, 20, 21]. One region with no obvious secondary structure is composed primarily of the amino acids proline (P), glutamine (E), valine (V), and lysine (K), and is thus referred to as the PEVK domain [22]. The other is an immunoglobulin-like domain that folds into a specific secondary structure [23]. We investigate how these two topologically different segments of titin interact to create the mechanical properties of titin.

When resilin, an elastomeric protein that has little secondary structure [24, 31, 32] is photochemically crosslinked with GB1, an immunoglobulin-like domain [23], such that the composite material emulates the secondary structure of titin, experiments show that the

mechanical properties of titin are also reproduced [12]. Hence, we hypothesize that proteins that have a similar structure also have similar mechanical properties even if their amino acid composition is very different. Consequently, as a corollary, proteins with different structure have different mechanical properties even if their amino acid compositions are similar. We further conjecture that the source of the elasticity of a protein molecule such as titin, which contains structured and unstructured domains, arises from its unstructured domain once the molecule has been fully extended.

4.2 Methodology

We use molecular dynamics (MD) simulations to simulate four different configurations and characterize their mechanical properties. These configurations are: (1) the first 91 amino acids found in the unstructured PEVK domain, (2) the 91 amino acids from the proximal immunoglobulin-like domain of titin in its natural configuration (i.e. with its secondary structure intact), (3) the PEVK domain and immunoglobulin-like domain in sequence, referred to as the combined configuration, and, finally, (4) the immunoglobulin-like domain of titin with the secondary structure artificially removed, referred to as the structure-removed immunoglobulin domain. We choose only the first 91 amino acids in the PEVK domain instead of the entire molecule so as to have a fair point of comparison with the immunoglobulin-like domain in the structure-removed configuration as well as in the structured configuration once it has popped open. The initial configurations for the unstructured

PEVK and structure-removed immunoglobulin domain are generated using the freely available AmberTools program tLeap, while the natural immunoglobulin domain is obtained from the literature [23].

We use the GROMACS package to perform the MD simulations. Once the configurations have been prepared, the same five-step procedure is followed to simulate tensile tests for each setup. The procedure involves: (1) energy minimization, (2) NVT equilibration using a velocity-rescale thermostat to control the temperature at 310 K, (3) NPT equilibration using a velocity-rescale thermostat and a Berendsen barostat [42] to control temperate and pressure at 310 K and 1 bar, respectively, (4) NPT equilibration using a Nose-Hoover thermostat [43] and Parrinello-Rahman barostat [44] to control the temperature and pressure at 310 K and 1 bar, respectively, and (5) a pulling simulation that holds the N-terminus of the structure (which is the amide group attached to the alpha carbon of the first amino acid) fixed while pulling on the C-terminus (i.e., the carboxyl group attached to the alpha carbon of the final amino acid) at a constant displacement rate of 0.1 nm/ns.We use the OPLS-AA force field [36] and apply Lorentz-Berthelot mixing rules. The simulations are run in bulk water.

4.3 Results and Discussion

Figure 4.1 presents the results, as running averages over a displacement of 0.1 nm, for three different simulations. Figures 4.1(a) and 4.1(c) correspond to the unstructured PEVK domain and the combined configuration, respectively. These two simulations returned nearly

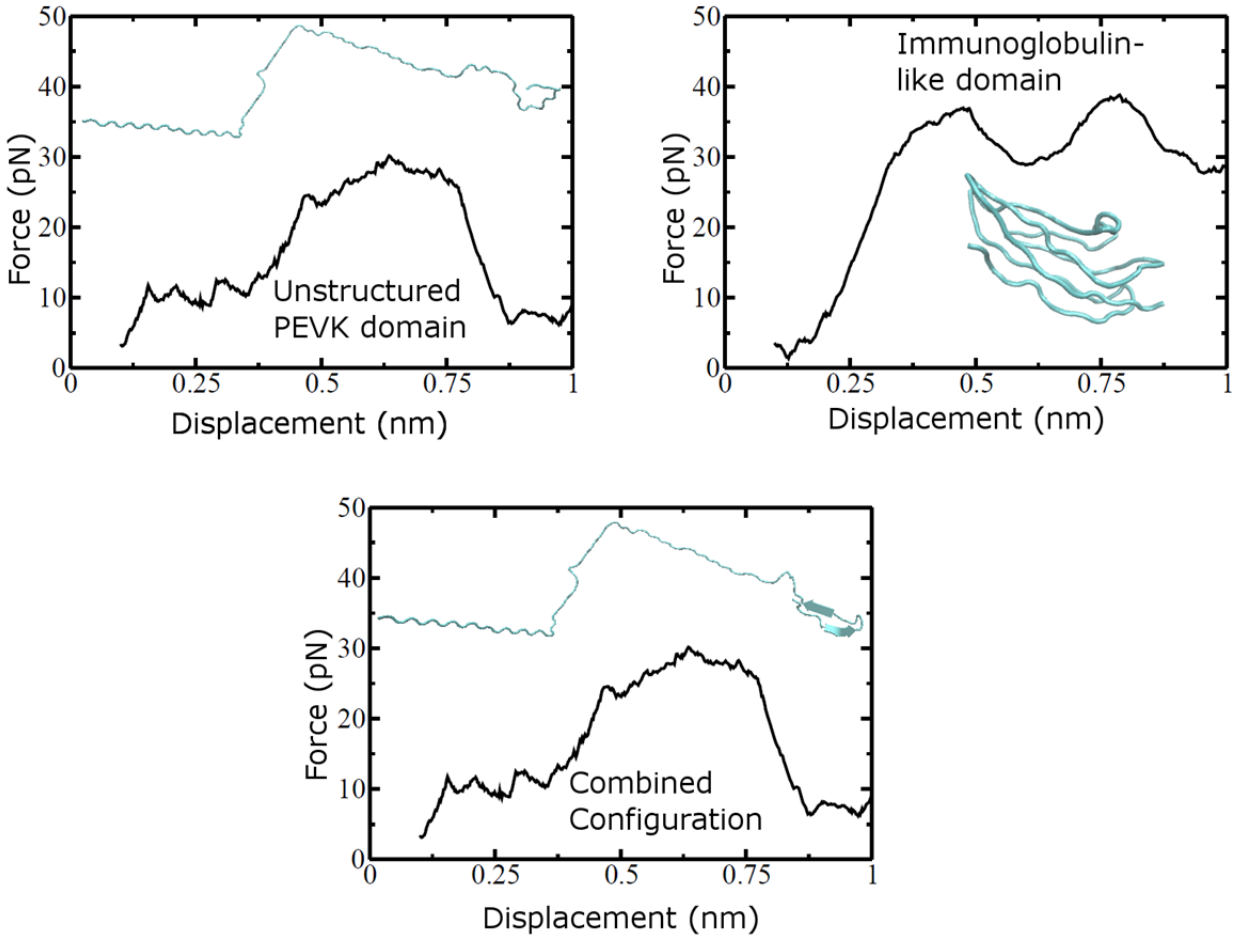

Figure 4.1: Force *vs.* displacement curves for the different configurations of titin. (a) The unstructured configuration of titin begins to yield at a ∼30 pN and a ∼0.625 nm extension, resulting in a stiffness of ∼48 pN/nm. (b) The structured configuration of titin has two yield points. The first of these is associated with the failure of the secondary structure, while the second corresponds to the failure of the material. The first yield point occurs at a force of ∼40 pN and an extension of ∼0.45 nm, i.e., it has a stiffness of ∼90 pN/nm. The second such point corresponds to a force of ∼40 pN and a extension of ∼0.8 nm, which results in a

stiffness of ∼50 pN/nm, which is virtually identical to that of the structured configuration. (c) The combined configuration behaves a similar manner to the unstructured configuration, confirming our hypothesis that the unstructured portion of titin bears primarily responsibility for the extensibility of titin.

identical results, confirming our hypothesis that the extensibility of titin originates in its unstructured PEVK domain. Figure 4.1(b) presents results for the immunoglobulin-like domain. The maximum displacement before failure for each of these tests is ∼0.8 nm, which is significantly lower than the measurements made by Li et al [45]. We attribute this difference to the high strain rates that we must impose on the system to conclude a simulation in a timely manner. For example, since the initial length of the combined configuration is ∼5 nm, a displacement rate of 0.1 nm/ns is equivalent to a strain rate of 2% per ns, i.e., 2×10^{9}% per s.

The results for the unstructured PEVK domain, shown in Figure 4.1(a), are to be expected. There is an initial flat response, i.e., the applied force produces a displacement simply as the molecule straightens itself. This is followed by a quasi-linear increase in force for increasing displacement, and finally a yield point at which the molecule begins to fail. The yield force for this configuration is ∼30 pN after the molecule has been extended to ∼0.625 nm. These values provide a stiffness, defined as yield force divided by displacement at yield, of ∼48 pN/nm.

The response of the immunoglobulin-like domain, shown in Figure 4.1(b), follows a same

force *vs.* displacement progression. Relatively little force is initially required to unravel the molecule. Then, there is a linear increase in force for increasing displacements followed by yielding. However, for this configuration, instead of continuing to decrease after the material yields, the force increases once again with increasing displacement and reaches a second yield point. The first yield point occurs due to the failure of the secondary structure while the second such point arises due to the failure of the molecule as a whole. Using the first yield point, i.e., when the force and displacement are ∼40 pN and ∼0.45 nm, provides a stiffness of ∼90 pN/nm, which is larger than for the unstructured PEVK domain. The yield force at the second yield point is a similar ∼40 pN, although the displacement has now grown to ∼0.8 nm. Here, the stiffness is ∼50 pN/nm, which is similar to that for the unstructured PEVK domain. This lends credence to our argument that, if the structure is similar, the material properties of proteins, even those with significantly different amino acid content, are also similar.

A tensile test on the combined configuration shows that, as the combined configuration stretches in the same manner as does the unstructured PEVK domain, the structured portion of the molecule remains essentially inert. Hence, the results of tensile tests on the combined configuration are virtually identical to those for the unstructured PEVK domain. This provides evidence for our hypothesis that the elasticity of the molecule, once fully extended, is derived from the unstructured portion of the molecule.

The simulation for the structure-removed immunoglobulin domain provided inconclusive results. While the initial displacement indicated that the molecule had a stiffness on par with

that for the unstructured configuration, the molecule underwent failure at a displacement of 0.1 nm. Using the running average scheme used for the other simulations, this provided only a single data point, which is insufficient for a conclusive analysis.

The tensile tests on the unstructured PEVK domain, immunoglobulin-like domain, and combined configurations demonstrate that the stiffness of titin has a strong dependence on its secondary structure. Specifically, Figures 4.1(a) and 4.1(b) show that, after removing secondary structure, the immunoglobulin-like domain and the unstructured PEVK domain have a virtually identical stiffness. This is a strong indication that it is secondary structure and not amino acid sequence that is instrumental in the mechanical properties of titin. After an initial straightening of linker regions that connect consecutive immunoglobulin-like domains, elongation is initiated in the PEVK domain, which provides elasticity to the molecule well before the immunoglobulin-like domains experiences a significant force.

The structure-removed immunoglobulin domain, however, failed far earlier than any of the other configurations. In order to reconcile this result, we also examined the coulombic interactions between the protein and the water. Previous work has deemed such interactions to be important for the mechanical properties of proteins [24] when the secondary structure is ignored, with greater coulombic interactions leading to a higher extensibility. The coulombic interactions between the structured/structure-removed immunoglobulin domains and water are $\sim$54000 kJ/mol, while the same interactions for the combined/unstructured domains and water are $\sim$30000 kJ/mol. The difference in coulombic interaction is the reason for the significantly lower extensibility of the structure-removed configuration. In other words,

changing the hydrophilicity of a protein greatly changes its extensibility but does not have a similar effect on stiffness. Stiffness is controlled through the structure of the protein. The secondary failure of the immunoglobulin-like domain provides additional evidence of the origin of stiffness. Once the structure of the immunoglobulin domain is straightened and broken, it is only capable of stretching ∼0.1 nm before failure, a result that is similar to that found for the structure-removed immunoglobulin domain.

Finally, we investigate the cause of the higher stiffness in the immunoglobulin-like domain before the failure of the secondary structure. We find that in the immunoglobulin-like domain, the amino acid side chains from an average of ∼52.439 hydrogen bonds with other amino acid side chains within the structure. In the unstructured PEVK domain, however, this number comes down to ∼1.392. The higher stiffness before failure of the immunoglobulin-like domain, then, is simply due to the increased number of hydrogen bonds formed within the structure.

4.4 Conclusion

The extensible region of titin consists of two structurally distinct domains, one containing a minimal secondary structure (termed the PEVK domain) and another that has high secondary structure (called the immunoglobulin-like domain). We simulated tensile tests on four different configurations of titin: (1) the first 91 amino acids found in the PEVK domain, (2) the immunoglobulin-like domain of titin in its natural configuration (i.e. with the

secondary structure intact), (3) the unstructured and structured configurations in sequence, and (4) the immunoglobulin-like domain of titin with the secondary structure artificially removed. The results show that the primary tensile component of titin, once fully extended, is the PEVK domain. Further, we find that, even with different amino acid sequences, the presence of like structure leads to similar mechanical properties. This is due to the high number of hydrogen bonds that are formed within compact structures leading to a higher stiffness, while the low number of hydrogen bonds in an unstructured domain leads to a much lower stiffness. This result reinforces the hypothesis that disordered regions provide greater flexibility and provides a molecular mechanism as to how this occurs.

Chapter 5

Effect of Cross-linking on Mechanical Properties of Resilin

5.1 Motivation

We showed in chapter 4 that one of the primary determinants of protein mechanical properties is the overall structure. We further showed in chapter 3 that, in the absence of structure, the electrical charge of the amino acid side chains played a significant role in the overall material properties. Both of these studies were done with only one chain, and therefore did not take into consideration the effect of the formation of cross-links. While we anticipate that cross-linking the molecule would increase the stiffness of the protein, the molecular basis for this increase in stiffness is unclear.

We have known for some time that the elasticity of polymer networks is dependent on the molecular mass between cross-links through the formula, $G = \frac{\rho RT}{M_c}$ [46]. This formula clearly indicates that the presence of extra cross-links will decrease the elasticity of a molecule. In deriving this theory, the chemical composition of the chains are neglected, and focus is instead placed on geometric constraints. While this theory has clearly stood the test of time in a bulk material setting, how applicable is it to the nanoscale? It is clear that the exact chemical composition of a chain will become increasingly important as we shrink down to a material on the scale of only a handful of chains. How much more important will it become? In order to answer this question, we focus on the cross-linking of one particular protein, resilin.

Since resilin was first discovered in 1960 [47], investigations have examined its structure and behavior at both the macroscale level and the scale of an individual protein. The resilience of several different types of resilin has been tested [1, 48, 49], as well as the effect of polarity of the amino acid side chains on molecular behavior [24]. The literature is still silent on the intermediate scale, i.e. at the level of two or three proteins that have been cross-linked. The cross-links that can be formed in resilin and their possible influence on its mechanical properties are unknown.

We investigate the cross-linking of two strands of resilin through atomistic simulations. We simulate two short repeat units found in mosquito resilin [17], specifically YGAPAQTP as specified in the FASTA format, that are 0.3 nm apart. The simulations are conducted for different environments to observe the cross-links that form. A tensile test is subsequently

simulated to determine how these cross-links affect mechanical properties.

5.2 Methodology

Initially, the two YGAPAQTP strands are solvated in water at three temperatures, i.e., 310 K, 360 K, and 410 K, which we refer to as the low, medium, and high temperature cases. While the medium and high temperature cases are aphysical, they thermally simulate the presence of enzymes that lower the activation energy required to create a cross-link. It is computationally expensive to simulate these enzymes *in situ*. Hence, the commonly used alternative (or reaction coordinates) requires a predefined awareness of the bonds that will occur. Instead, we simulate the influence of enzymes by increasing the temperature, since instead of lowering the activation energy, we now increase the energy available to the proteins. This simulation is followed by the tensile test. A structure of the motifs is created through the use of the AmberTools program, tLeap [34].

Once the structure is prepared, we use the CPMD package [50, 51] to perform our simulations. All simulations undergo the same procedure through a leapfrog time integration scheme with a time step of 5000 atomic time units ($\sim$0.1 fs). The procedure includes (i) wavefunction optimization to optimize the initial electronic configuration (but there is no time marching during this step), (ii) equilibration over 10000 time steps with the temperature controlled through a velocity rescaling thermostat over which we study bond formation, and (iii) a pulling simulation in which the alpha carbon of the N-terminus of one strand

is held fixed while the alpha carbon of the C-terminus of the other strand is pulled at a constant displacement rate of 0.1 nm per ps. All of the pulling simulations were done at 310 K in order to eliminate the effect of temperature on mechanical properties.

To observe bond formation, we follow the motion of the Wannier-function centers of each valence electron [52]. Bond formation is defined as the existence of a Wannier-function center within 0.175 nm of two distinct atoms [52]. Once formed, a bond is designated as broken if the associated Wannier-function center moves farther than 0.175 nm from either atom.

5.3 Results and Discussion

The initial configuration of all the temperature regimes tested is shown in Figure 5.2(A). After equilibration, during which the two strands come to a minimum energy state, we see the two strands of the protein move away from each other after the proline molecules, as shown in Figure 5.2(B). This is due to the presence of the hydrogen atoms connected to the glycine molecule on the outside of the chains forcing the proline side chains to rotate in between the two chains, thereby pushing them apart.

In all three temperature regimes, all bonding between the two resilin strands occurs in the initial four amino acids of each strand, i.e., the YGAP sequence, as it would in naturally occurring resilin [53]. For the low temperature case, we observe bond formation between the two tyrosine molecules in the sequence, producing a dityrosine bond, as shown in Figure 5.2(C). Once this structure is formed, we obtain the force *vs.* distance curve shown in

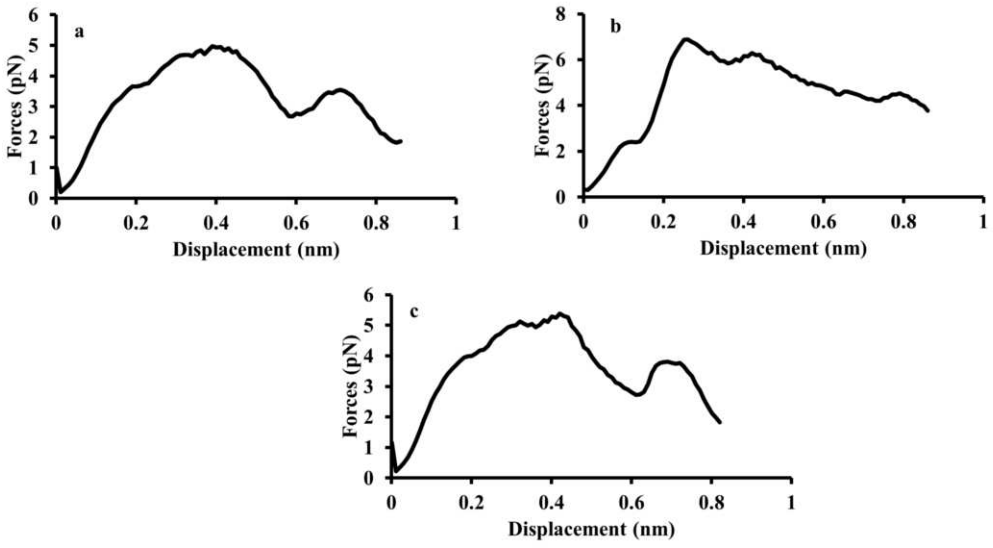

Figure 5.1: Force vs displacement curves for the different temperature regimes tested. (a) The low temperature case begins to yield at a ∼5 pN force and ∼0.4 nm extension, resulting in a stiffness of ∼12.5 pN/nm. (b) The medium temperature case has a yield point at a force of ∼7 pN and an extension of ∼0.25 nm, i.e., it has a stiffness of ∼28 pN/nm. The increased stiffness in this case is attributed to the presence of a bond between two glycine molecules. (c) The high temperature case yields at a force of ∼5.6 pN and an extension of ∼0.4 nm, resulting in a stiffness of ∼14 pN/nm. The increased stiffness in this case is attributed to the presence of a strong interaction between two proline molecules.

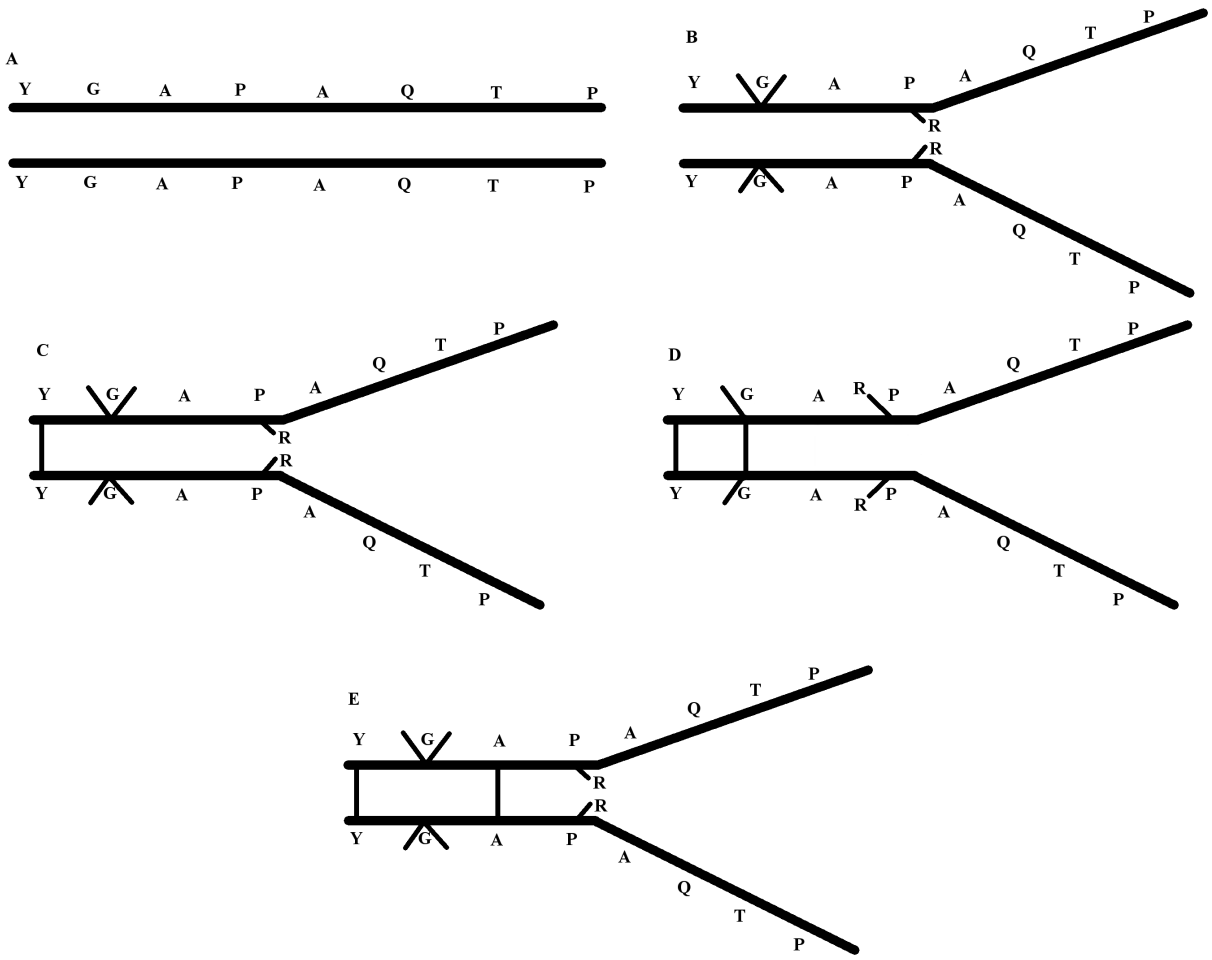

Figure 5.2: Different geometries of the cross-linked system. (A) For all of the temperature regimes tested, the initial configuration is simply two straight chains placed next to each other. (B) After the initial equilibration phase, all of the configurations assume the same shape. The two hydrogen atoms attached to each glycine molecule (represented by the two lines from each glycine molecule) provide a steric hindrance to the side chain of the proline molecule (represented by the R attached to each proline molecule). This causes the proline side chains to rotate such that they are now in between the two strands. This forces the

two chains to move apart from one another to accommodate the bulky proline side chains, giving us the configuration shown. (C) In the low temperature case, only a dityrosine bond is formed. (D) In the medium temperature case, both a dityrosine bond and a glycine-glycine bond is formed. This bond formation frees up the proline side chains to rotate out from in between the two chains. When this formation undergoes a tensile test, the chains straighten out and interact more closely with one another, greatly increasing their stiffness. (E) In the high temperature case, both a dityrosine bond and an alanine-alanine bond is formed. In this temperature regime, the hydrogen atoms attached to the glycine do not interact with one another, and the proline side chains are not free to rotate. This keeps the two strands of the protein away from each other, thereby decreasing intermolecular interactions and, as such, stiffness.

Figure 5.1(a). This figure shows that the low temperature case has a stiffness, defined as the maximum force experienced by the protein divided by the maximum displacement at that force, of 12.5 pN/nm. The low temperature case fails at a displacement of 0.4 nm.

For the medium temperature case, two bonds are formed as shown in Figure 5.2(D). The dityrosine bond observed in the low temperature case, as well as a cross-link between the two glycine molecules in the sequence. This latter bond occurs when the bond between the alpha carbon and the side chain, i.e., a hydrogen-atom bond, is broken in both glycines so that the two alpha carbons become directly linked. This bond greatly influences the mobility of the overall structure, which is presented through the force vs. distance curve for this case

shown in Figure 5.1(b). This figure provides a stiffness of 28 pN/nm and shows failure at 0.25 nm.

For the high temperature case, two bonds are once again formed as shown in Figure 5.2(E). As in the low and medium temperature cases, the dityrosine bond is produced relatively quickly. In this case, however, we do not observe a link between the two glycine molecules. Instead, there is an ionic interaction between the amide group of the alanine in one strand and the carboxyl group of the alanine in the other strand. This bond also influences the mobility of the overall structure, as shown in the force vs. distance curve in Figure 5.1(c). This case has a stiffness of 14 pN/nm and failure at 0.4 nm. In all three cases, failure occurs at the site of the cross-link, but not within either strand.

The highest stiffness and the lowest extensibility occurs for the medium temperature case. This is primarily due to the bond formed between the two glycine molecules. For the low and high temperature cases, the two hydrogen atoms attached to each glycine molecule prevents the side chains of the proline molecules from rotating out from in between the two chains. In the medium temperature case, however, a hydrogen atom from each of the two glycine molecules are removed and form a bond of their own, and the two alpha carbons are directly linked. This drastically reduces the steric hindrance felt by the proline side chains, allowing them to rotate out, thereby freeing the rest of the two chains to come back together. This greatly increases the intermolecular interactions between the two proteins, going from $\sim$2500 kJ/mol in the low and high temperature cases to $\sim$3500 kJ/mol in the medium temperature case. These results reinforce the hypothesis that high glycine and proline content is one of the

requirements to be an elastomeric protein [54]. Without these two amino acids interacting in exactly the manner they do naturally, we show that chain-chain interactions are higher, and therefore stiffness is higher.

The high temperature case also exhibits a higher stiffness and lower extensibility than the low temperature case. Here, instead of a glycine-glycine bond, we see the formation of an alanine-alanine bond. While this case also exhibits higher stiffness and lower extensibility than the low temperature case, it is for a different reason. Here, we do not see the rotation of amino acid side chains that is seen in the medium temperature case. Instead, it appears that the increased stiffness is simply due to the decrease of M_c, the molecular weight between cross-links, as would be predicted by the bulk theory. This result reinforces the importance of understanding the molecular basis of different cross-links. Some cross-links, such as those found in the medium temperature case, drastically reduce the molecular mobility of the protein, while others, such as those found in the high temperature case, do not and simply operate on the same principles as found in the bulk theory.

The cross-links formed in the resilins are found primarily in the YGAP sequence. The data show that the intermolecular interactions between the two strands of resilin are highest in this sequence. In fact, for the low and high temperature cases, intermolecular interactions between the two YGAP sequences are $\sim$2000 kJ/mol, while all other interaction energies between the two strands are $\sim$500 kJ/mol. For the medium temperature case, the gap between the ends of the chains begins to close, thereby increasing chain-chain interaction. However, even in this case, we note that the other interaction energies increase to $\sim$1500

kJ/mol , but intermolecular interactions between the two YGAP sequences are still higher at ∼2000 kJ/mol . This result reinforces the importance of the YGAP sequence for the structural stability of resilin.

5.4 Conclusions

In this chapter we find that, at a molecular level, it is not only the distance between cross-links that determines the mechanical properties. In fact, the exact type of cross-link that is formed can play a large role in the molecular mobility of the system, thereby drastically changing the mechanical properties. Further, we find that the YGAP repeat unit is important for cross-linking in resilin and confirm the importance of glycine and proline as required elements bestowing resilin, and possibly all elastomeric proteins, with elasticity.

Chapter 6

Effect of Solvent on Mechanical Properties of Titin

6.1 Motivation

It has recently been discovered that, in an aqueous environment, titin derives its low stiffness from the unstructured PEVK domain and not the structured immunoglobular domain. It has been hypothesized that this is due to the nature of the interactions between the highly polar PEVK domain and the polar environment, water. It is unclear how this would change if the protein were to be placed in a different solvent that is less polar. We hypothesize that, placed in a nonpolar solvent, the previously stiff immunoglobular domain will begin to have significantly higher interactions with the solvent, greatly decreasing the

stiffness. Simultaneously, the previously flexible PEVK domain will have significantly lower interactions with the solvent, dramatically increasing the stiffness.

To test our hypothesis, we study three different configurations of titin in two different solvents. Specifically, the three configurations are: (i) the first 20 amino acids of the PEVK domain, (ii) one of the beta sheets in the immunoglobular domain (which is composed of 20 amino acids), and (iii) these two structures combined in a linear fashion, hereafter referred to as the combined configuration. All three of these configurations are tested in both water and ethanol.

6.2 Methodology

The initial configurations for the PEVK domain and the combined configuration were generated using the freely available AmberTools program tLeap, while the natural immunoglobulin domain is obtained from the literature [23]. Once the configuration has been generated, it is then solvated using the freely available GROMACS program genbox.

Once the structure is prepared, we use the CPMD package [50, 51] to perform our simulations. All simulations undergo the same procedure through a leapfrog time integration scheme with a time step of 5000 atomic time units ($\sim$0.1 fs). The procedure includes (i) wavefunction optimization to optimize the initial electronic configuration (but there is no time marching during this step), (ii) equilibration over 10000 time steps with the temperature controlled through a velocity rescaling thermostat, and (iii) a pulling simulation in which

the alpha carbon of the N-terminus is held fixed while the alpha carbon of the C-terminus is pulled at a constant displacement rate of 0.1 nm per ps.

6.3 Results and Discussion

Figure 6.1 shows the results, as running averages over a displacement of 0.05 nm, for the three simulations run in water. Figures 6.1(a) and 6.1(c) correspond to the unstructured PEVK domain and the combined configuration, respectively, which returned nearly identical results. This confirms a previous study that has shown that the extensibility of titin originates in its unstructured PEVK domain. Figure 6.1(b) presents results for the immunoglobulin-like domain. The maximum displacement of all three of these segments is much smaller than those in the earlier study. This is primarily due to the much smaller segments used in this study as compared to those presented in Chapter 4 of this work. Figure 6.2 shows the corresponding results for the same simulations run in ethanol. Figure 6.2(a) and 6.2(b) correspond to the unstructured PEVK domain and the structured immunoglobular domain, respectively. The simulation of the combined configuration in ethanol has not yet been completed, and, as such, is not included in this work.

The results shown in Figure 6.1(a) for the unstructured PEVK domain are to be expected. The initial flat response is due to the molecule simply straightening itself out. This is followed by a quasi-linear increase in force with increasing displacement, followed by a yield point wherein the molecule fails. The yield force for this configuration is $\sim$7 pN and occurs after

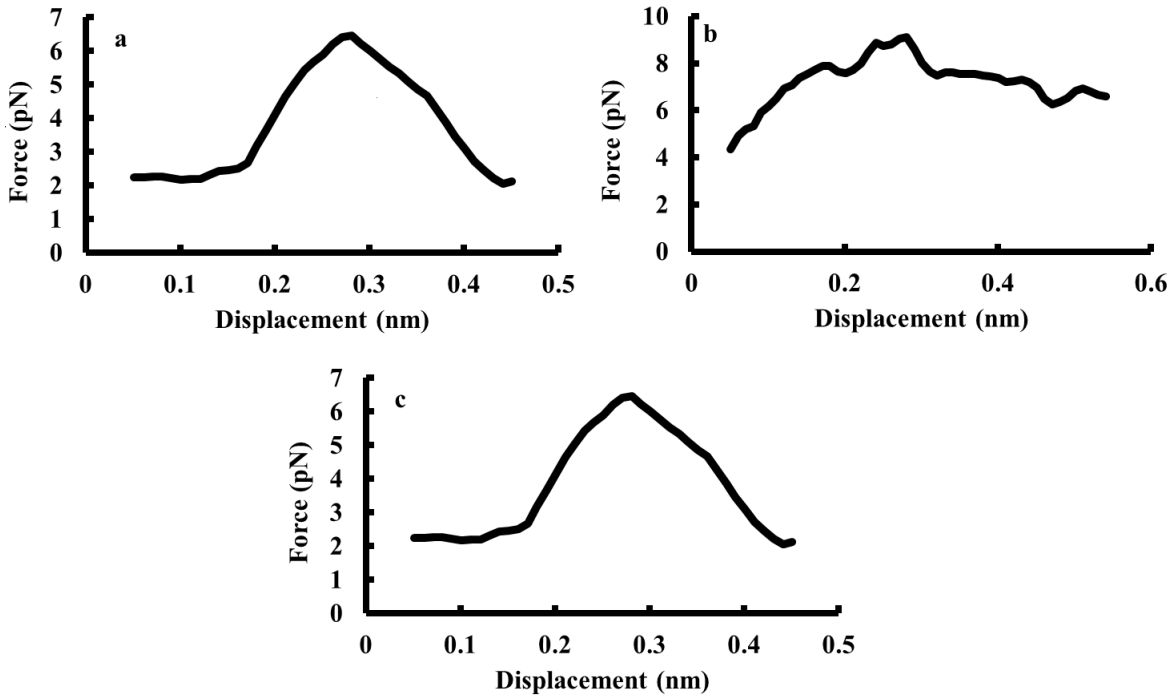

Figure 6.1: Force *vs.* displacement curves for the different configurations of titin in water. All of the data points represent a running average over 0.05 nm. (a) The unstructured configuration of titin begins to yield at a ∼7 pN and a ∼0.3 nm extension, resulting in a stiffness of ∼23.3 pN/nm. (b) The structured configuration of titin has two yield points. The first of these is associated with the failure of the secondary structure, while the second corresponds to the failure of the material. The first yield point occurs at a force of ∼10 pN and an extension of ∼0.2 nm, i.e., it has a stiffness of ∼50 pN/nm. The second such point corresponds to a force of ∼9 pN and a extension of ∼0.3 nm, which results in a stiffness of ∼30 pN/nm. (c) The combined configuration behaves a similar manner to the unstructured configuration, confirming our hypothesis that the unstructured portion of titin bears primarily responsibility for the extensibility of titin.

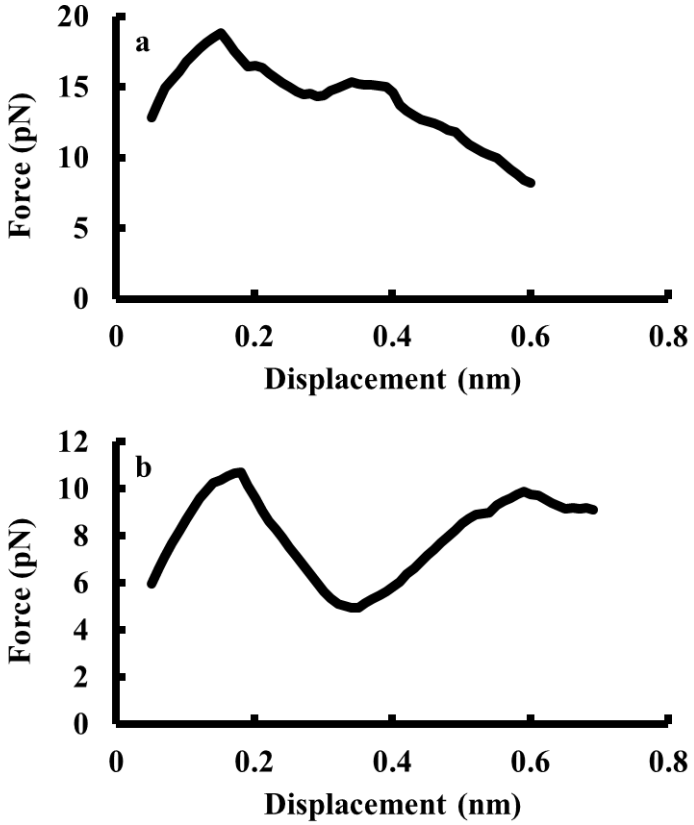

Figure 6.2: Force *vs.* displacement curves for the different configurations of titin in ethanol. All of the data points represent a running average over 0.05 nm. (a) The unstructured configuration of titin begins to yield at a ∼20 pN and a ∼0.15 nm extension, resulting in a stiffness of ∼133.3 pN/nm. (b) The structured configuration of titin has two yield points. The first of these is associated with the failure of the secondary structure, while the second corresponds to the failure of the material. The first yield point occurs at a force of ∼11 pN and an extension of ∼0.2 nm, i.e., it has a stiffness of ∼55 pN/nm. The second such point corresponds to a force of ∼10 pN and a extension of ∼0.6 nm, which results in a stiffness of ∼16.7 pN/nm.

the molecule has been extended by a distance of ∼0.3 nm. Using these values, we calculate a stiffness, defined as yield force divided by displacement at yield, of 23.3 pN/nm.

The results shown in Figure 6.1(b) follows the same general trend. I.e., there is an initial region of low force followed quickly by a linear increase in force with increasing displacements, and ending with yielding of the molecule. With this molecule, though, after a small dip, the force required to elongate the molecule begins to rise again and eventually yields a second time. The first yield point occurs simultaneously with the failure of the secondary structure present in the immunoglobular domain, and the dip in force is related to that portion of the molecule straightening out. The second yield point corresponds to the failure of the molecule as a whole. The first yield point occurs at a force of ∼10 pN and a displacement of ∼0.2 nm, giving a stiffness of 50 pN/nm, which is significantly higher than that of the PEVK domain. The second yield point, however, ∼9 pN and a displacement of ∼0.3 nm, for a stiffness of 30 pN/nm. This stiffness is similar to that of the PEVK domain in water.

Figure 6.1(c) shows that a tensile test of the combined configuration deforms in a manner almost identical to that of the unstructured PEVK domain while the structured portion of the molecule remains inert. As such, the results of the two tensile tests are virtually identical.

The results from Figure 6.1 reinforce the results from Chapter 4 of this work. As we found in that study, the unstructured portion of titin is primarily responsible for the elasticity of titin in water. Further, the second yield point of the structured immunoglobular domain has a similar stiffness to that of the unstructured PEVK domain, which indicates that secondary structure, and not exact amino acid sequence is responsible for stiffness.

The response of the unstructured PEVK domain when placed in ethanol is shown in Figure 6.2(a). We see immediately that there appears to be no initial straightening out region for this tensile test. Upon closer inspection, we see that this is due to the initial equilibrated configuration. In ethanol, unlike in water, the unstructured PEVK domain is virtually straight even before tensile testing has begun. We hypothesize that this is due to the inverse transition temperature of titin [8] (a temperature above which a protein will straighten out in order to allow crystallized water trapped within any folds to become free) being lower in ethanol than in water, causing the domain to straighten out.

Aside from the lack of an initial straightening out period, the tensile test behaves in a similar manner to that of the corresponding test performed in water, i.e., there is a quasi-linear increase in the force with the displacement followed by the yield point, which occurs at a force of $\sim$20 pN and a displacement of $\sim$0.15 nm. This gives us a stiffness of $\sim$133.3 pN/nm, significantly higher than the same value found in water.

The results shown in Figure 6.2(b) follow the same pattern as Figure 6.2(a). Once again, there is no initial straightening out of the domain as the initial equilibrated configuration is virtually straight before tensile testing has begun. As in the simulation in water, this domain shows two failure points: one for the secondary structure and another for the material itself. The first failure point occurs at a force of $\sim$11 pN and an extension of $\sim$0.2 nm for a stiffness of 55 pN/nm. The second such point occurs at a force of $\sim$10 pN and an extension of $\sim$0.6 nm, yielding a stiffness of $\sim$16.7 pN/nm.

In order to determine the fundamental cause of the observed differences in stiffness, we

look at the interaction energies between the solvent and the molecule. We find that, in water, the unstructured configuration has an interaction energy of ∼2000 kJ/mol with the water, whereas the same interaction energy for the structured configuration is only ∼1000 kJ/mol. In ethanol, however, this situation is reversed. The structured configuration has an interaction energy of ∼2500 kJ/mol with the ethanol while the same interaction energy for the unstructured configuration is ∼500 kJ/mol. This provides further evidence as to the importance of protein-solvent interactions on the mechanical properties.

6.4 Conclusions

Recent studies have shown that, in water, a polar solvent, the unstructured domain of titin is the primary source of elasticity while the structured domain of titin is inert. In this study, we present preliminary evidence that, in a nonpolar solvent, this is reversed. Further data is needed before it is possible to make a strong claim.

Chapter 7

Conclusion

7.1 Summary

In this dissertation, we focused on the atomic origin of the elastic properties of some of natures most elastic proteins. Specifically, we focused on four major topics: (i) the effect of polarity on the mechanical properties of resilin, (ii) the effect of secondary structure on the mechanical properties of titin, (iii) the effect of cross-linking on the mechanical properties of resilin, and (iv) the effect of solvent on the mechanical properties of titin.

The first study focused on the effect of artificially changing the charge on the amino acid constituents of resilin on its mechanical properties. We found that a significant portion of the elasticity of resilin is due to the interaction that it has with the surrounding solvent. By reducing the charge, we were reducing the ability of resilin to interact with the surrounding

water, thereby significantly decreasing the extensibility of the molecule.

The second study focused on the effect of alpha helices and beta sheets on the mechanical properties of titin. We found that the stiffness of titin depends primarily on the secondary structure, and does not depend as highly on the polarity of the protein, which determines the extensibility. By removing the secondary structure, it is possible to greatly decrease the stiffness of elastomeric proteins.

The third study focused on the effect of cross-link formation of two resilin repeat units on the mechanical properties of the protein. We found that, in resilin, cross-links occur primarily in the YGAP repeat unit. Further, we found that, if cross-links were to form in either glycine or proline, the stiffness and extensibility of the protein is affected greatly. This provides further evidence for the importance of glycine and proline on the mechanical properties of elastomeric proteins.

The fourth and final study is not yet fully completed. It focuses on the effect of changing the environment of the protein to determine the effect of solvent polarity on mechanical properties of titin. So far, results indicate that, in nonpolar solvents, the structured, immunoglobular domain of titin (also the nonpolar domain of titin) has a higher extensibility than the unstructured PEVK domain. The effect on the molecule as a whole has yet to be determined.

From the first and fourth study, we find that it is not merely the structural components of the protein that is relevant to the mechanical properties as a whole. Rather, the interactions

between the solvent and the protein play a significant role in the overall molecular mobility of the protein. Further, the second and fourth studies highlighted the important role that structure plays in the mechanical properties. An increased structure corresponds directly to a decreased molecular mobility that significantly increases the stiffness of any material. Finally, the third study showed us that the geometry of individual molecular interactions can play a large role in the mechanical properties. Overall, all of these studies point to one conclusion: Molecular mobility is dependent on more than simply the individual building blocks of the protein. The interaction of these building blocks with each other, as well as within the solvent is just as important, if not more so, on the overall mechanical properties of the protein.

7.2 Future Work

While the results from these studies provide novel insights into the molecular mechanisms used by nature to produce elastic proteins, there are still several unanswered questions. For instance, how important are these properties as we scale up the system that is studied? If we were to study a 100 amino acid sequence instead of a 17 amino acid sequence (as in the first study), would we find that the effect of polarity is now minimal? If so, how is this effect mitigated? If not, how can we use this information to design materials with a specific extensibility? Further, the bulk of this research was done on single chains, with only one study focusing on two chains. How would multiple cross-links between proteins, or even

cross-links between different types of proteins alter the results of these studies? *In vivo*, these proteins are part of a complex hierarchy that is designed to perform very specific tasks; how do the properties at the level of the individual chain affect those of the entire hierarchy?

There are also questions concerning the application of the results of these studies. Nature has created these elastomeric proteins through manipulation at an atomic level. As mankind does not yet possess the ability to manipulate at a fine scale, is it possible to use these results on a coarser scale? E.g., if instead of using a sequence of specific amino acids, we were to use larger blocks of material that had similar properties, would we get similar results? If so, is this principle expandable to other fields, so that we may be able to determine a method to create a material given only its material properties?

Chapter 8

Bibliography

[1] Christopher M Elvin, Andrew G Carr, Mickey G Huson, Jane M Maxwell, Roger D Pearson, Tony Vuocolo, Nancy E Liyou, Darren CC Wong, David J Merritt, and Nicholas E Dixon. Synthesis and properties of crosslinked recombinant pro-resilin. *Nature*, 437(7061):999–1002, 2005.

[2] Lisa D. Muiznieks, Anthony S. Weiss, and Fred W. Keeley. Structural disorder and dynamics of elastin. *Biochemistry and Cell Biology*, 88(2):239–250, 2010. PMID: 20453927.

[3] Torkel Weis-Fogh. A rubber-like protein in insect cuticle. *Journal of Experimental Biology*, 37(4):889–907, 1960.

[4] John M. Gosline, M. Edwin DeMont, and Mark W. Denny. The structure and properties of spider silk. *Endeavour*, 10(1):37–43, 1986.

[5] Peter R Shewry, Arthur S Tatham, and Allen J Bailey. *Elastomeric proteins: structures, biomechanical properties, and biological roles*. Cambridge University Press, 2003.

[6] Catherine M Bellingham, Margo A Lillie, John M Gosline, Glenda M Wright, Barry C Starcher, Allen J Bailey, Kimberly A Woodhouse, and Fred W Keeley. Recombinant human elastin polypeptides self-assemble into biomaterials with elastin-like properties. *Biopolymers*, 70(4):445–455, 2003.

[7] Sarah Rauscher and Régis Pomès. Structural disorder and protein elasticity. *Fuzziness*, pages 159–183, 2012.

[8] Bin Li, Darwin OV Alonso, and Valerie Daggett. The molecular basis for the inverse temperature transition of elastin. *Journal of molecular biology*, 305(3):581–592, 2001.

[9] Sinan Keten and Markus J Buehler. Atomistic model of the spider silk nanostructure. *Applied physics letters*, 96(15):153701–153701, 2010.

[10] John Gosline, Margo Lillie, Emily Carrington, Paul Guerette, Christine Ortlepp, and Ken Savage. Elastic proteins: biological roles and mechanical properties. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357(1418):121–132, 2002.

[11] Ivan Vesely. The role of elastin in aortic valve mechanics. *Journal of biomechanics*, 31(2):115–123, 1997.

[12] Shanshan Lv, Daniel M Dudek, Yi Cao, MM Balamurali, John Gosline, and Hongbin Li. Designed biomaterials to mimic the mechanical properties of muscles. *Nature*, 465(7294):69–73, 2010.

[13] Svend Olav Andersen. The cross-links in resilin identified as dityrosine and trityrosine. *Biochimica et biophysica acta*, 93:213, 1964.

[14] HC Bennet-Clark. Resonators in insect sound production: how insects produce loud pure-tone songs. *Journal of experimental Biology*, 202(23):3347–3357, 1999.

[15] Niels Skals and Anne Marrie Surlykke. Sound production by abdominal tymbal organs in two moth species: the green silver-line and the scarce silver-line (noctuoidea: Nolidae: Chloephorinae). *Journal of experimental biology*, 202(21):2937–2949, 1999.

[16] Manoj B Charati, Jamie L Ifkovits, Jason A Burdick, Jeffery G Linhardt, and Kristi L Kiick. Hydrophilic elastomeric biomaterials based on resilin-like polypeptides. *Soft matter*, 5(18):3412–3416, 2009.

[17] Svend Olav Andersen. Studies on resilin-like gene products in insects. *Insect biochemistry and molecular biology*, 40(7):541–551, 2010.

[18] Matthias Rief, Mathias Gautel, Filipp Oesterhelt, Julio M Fernandez, and Hermann E Gaub. Reversible unfolding of individual titin immunoglobulin domains by afm. *Science*, 276(5315):1109–1112, 1997.

[19] Ave Minajeva, Michael Kulke, Julio M Fernandez, and Wolfgang A Linke. Unfolding of titin domains explains the viscoelastic behavior of skeletal myofibrils. *Biophysical journal*, 80(3):1442–1451, 2001.

[20] L Tskhovrebova, J Trinick, JA Sleep, and RM Simmons. Elasticity and unfolding of single molecules of the giant muscle protein titin. *Nature*, 387(6630):308–312, 1997.

[21] Jen Hsin, Johan Strümpfer, Eric H Lee, and Klaus Schulten. Molecular origin of the hierarchical elasticity of titin: simulation, experiment, and theory. *Annual Review of Biophysics*, 40:187–203, 2011.

[22] Wolfgang A Linke, Michael Kulke, Hongbin Li, Setsuko Fujita-Becker, Ciprian Neagoe, Dietmar J Manstein, Mathias Gautel, and Julio M Fernandez. Pevk domain of titin: an entropic spring with actin-binding properties. *Journal of structural biology*, 137(1):194–205, 2002.

[23] Mark Pfuhl and Annalisa Pastore. Tertiary structure of an immunoglobulin-like domain from the giant muscle protein titin: a new member of the i set. *Structure*, 3(4):391–401, 1995.

[24] Ravi Kappiyoor, Ganesh Balasubramanian, Daniel M Dudek, and Ishwar K Puri. Elastomechanical properties of resilin. *Soft Matter*, 7(22):11006–11009, 2011.

[25] Norman L Allinger, Young H Yuh, and Jenn Huei Lii. Molecular mechanics. the mm3 force field for hydrocarbons. 1. *Journal of the American Chemical Society*, 111(23):8551–8566, 1989.

[26] William L Jorgensen and Julian Tirado-Rives. The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6):1657–1666, 1988.

[27] Dahlia K Remler and Paul A Madden. Molecular dynamics without effective potentials via the car-parrinello approach. *Molecular Physics*, 70(6):921–966, 1990.

[28] Richard Car and Mark Parrinello. Unified approach for molecular dynamics and density-functional theory. *Physical review letters*, 55(22):2471, 1985.

[29] Alain Polian, Paul Loubeyre, and Nino Boccara. Simple molecular systems at very high density. In *NATO ASIB Proc. 186: Simple Molecular Systems at Very High Density*, volume 1, 1989.

[30] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman JC Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327–341, 1977.

[31] GF Elliott, AF Huxley, and Torkel Weis-Fogh. On the structure of resilin. *Journal of Molecular Biology*, 13(3):791–IN13, 1965.

[32] Kate M Nairn, Russell E Lyons, Roger J Mulder, Stephen T Mudie, David J Cookson, Emmanuelle Lesieur, Misook Kim, Deborah Lau, Fiona H Scholes, and Christopher M Elvin. A synthetic resilin is largely unstructured. *Biophysical journal*, 95(7):3358–3365, 2008.

[33] Sebastian Kruggel and Thomas Lemcke. Generation and evaluation of a homology model of pfgsk-3. *Archiv der Pharmazie*, 342(6):327–332, 2009.

[34] David A Case, Thomas E Cheatham, Tom Darden, Holger Gohlke, Ray Luo, Kenneth M Merz, Alexey Onufriev, Carlos Simmerling, Bing Wang, and Robert J Woods. The amber biomolecular simulation programs. *Journal of computational chemistry*, 26(16):1668–1688, 2005.

[35] Berk Hess, Carsten Kutzner, David van der Spoel, and Erik Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation*, 4(3):435–447, 2008.

[36] George A Kaminski, Richard A Friesner, Julian Tirado-Rives, and William L Jorgensen. Evaluation and reparametrization of the opls-aa force field for proteins via comparison with accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B*, 105(28):6474–6487, 2001.

[37] Senbo Xiao, Wolfram Stacklies, Murat Cetinkaya, Bernd Markert, and Frauke Gräter. Mechanical response of silk crystalline units from force-distribution analysis. *Biophysical journal*, 96(10):3997–4005, 2009.

[38] Vladimir N Uversky and A Keith Dunker. Understanding protein non-folding. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1804(6):1231–1264, 2010.

[39] Raymond U Lemieux and Ulrike Spohr. How emil fischer was led to the lock and key concept for enzyme specificity. *Advances in carbohydrate chemistry and biochemistry*, 50:1–20, 1994.

[40] Ken A Dill and Justin L MacCallum. The protein-folding problem, 50 years on. *science*, 338(6110):1042–1046, 2012.

[41] Robert Huber and William S Bennett. Functional significance of flexibility in proteins. *Biopolymers*, 22(1):261–279, 1983.

[42] Herman JC Berendsen, J Pl M Postma, Wilfred F van Gunsteren, ARHJ DiNola, and JR Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81:3684, 1984.

[43] William G Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31(3):1695, 1985.

[44] M Parrinello and A Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*, 52(12):7182–7190, 1981.

[45] Hongbin Li, Wolfgang A Linke, Andres F Oberhauser, Mariano Carrion-Vazquez, Jason G Kerkvliet, Hui Lu, Piotr E Marszalek, and Julio M Fernandez. Reverse engineering of the giant muscle protein titin. *Nature*, 418(6901):998–1002, 2002.

[46] Leslie Ronald George Treloar. *The physics of rubber elasticity*. Oxford University Press, 1975.

[47] Torkel Weis-Fogh. A rubber-like protein in insect cuticle. *Journal of Experimental Biology*, 37(4):889–907, 1960.

[48] Russell E Lyons, Kate M Nairn, Mickey G Huson, Misook Kim, Geoff Dumsday, and Christopher M Elvin. Comparisons of recombinant resilin-like proteins: repetitive domains are sufficient to confer resilin-like properties. *Biomacromolecules*, 10(11):3009–3014, 2009.

[49] Guokui Qin, Amit Rivkin, Shaul Lapidot, Xiao Hu, Itan Preis, Shira B Arinus, Or Dgany, Oded Shoseyov, and David L Kaplan. Recombinant exon-encoded resilins for elastomeric biomaterials. *Biomaterials*, 32(35):9231–9243, 2011.

[50] Wanda Andreoni and Alessandro Curioni. New advances in chemistry and materials science with cpmd and parallel computing. *Parallel Computing*, 26(7):819–842, 2000.

[51] Jürg Hutter and Alessandro Curioni. Car–parrinello molecular dynamics on massively parallel computers. *ChemPhysChem*, 6(9):1788–1793, 2005.

[52] Nicola Marzari, Ivo Souza, and David Vanderbilt. An introduction to maximally-localized wannier functions. *Psi-K Newsletter*, 57:129, 2003.

[53] Svend Olav Andersen. Regional differences in degree of resilin cross-linking in the desert locust, *Schistocerca gregaria*. *Insect biochemistry and molecular biology*, 34(5):459–466, 2004.

[54] Sarah Rauscher, Stéphanie Baud, Ming Miao, Fred W Keeley, and Régis Pomès. Proline and glycine control protein self-organization into elastomeric or amyloid fibrils. *Structure*, 14(11):1667–1676, 2006.