Molecular Dynamics Analysis of Supercontraction in Spider Dragline Silk

by

Laura Batty

Bachelor of Engineering in Civil and Environmental Engineering
University of Edinburgh, 2012

Submitted to the Department of Civil and Environmental Engineering in Partial Fulfillment of the Requirements for the Degree of

MASTER OF ENGINEERING IN CIVIL AND ENVIRONMENTAL ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013

Signature of Author: .....
Department of Civil and Environmental Engineering
May 10, 2013

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

Accepted By: .....
Heidi M. Nepf
Chair, Departmental Committee for Graduate Students

Molecular Dynamics Analysis of Supercontraction in Spider Dragline Silk

by

Laura Batty

Submitted to the Department of Civil and Environmental Engineering
on May 10, 2013
in Partial Fulfillment of the Requirements for the Degree of
Master of Engineering in Civil and Environmental Engineering

Abstract

Spider dragline silk is a material that has evolved over millions of years to develop finely tuned mechanical properties. It is a protein-based fiber, used as the main structural component in spider webs and as a lifeline for the spider, and it combines strength and extensibility to give it toughness currently unmatched by synthetic materials. Dragline silk has the unusual tendency of shrinking by up to 50% when exposed to high humidity, a phenomenon called supercontraction. Supercontraction is thought to occur due to the association of water molecules with the amorphous region of silk proteins. The water molecules are believed to break the hydrogen bonds that connect the protein strands, causing a fundamental reorganization of molecular structure, which is manifested at the macro scale by a large retraction in length. However, the details of these mechanisms remain unknown and have not been directly demonstrated in prior research.

Here we use full-scale atomic modeling of spider silk using molecular dynamics to investigate the structure and properties of this material at a length scale that is not yet accessible by experimental methods. A model of spider silk protein is used to explore the phenomenon of supercontraction. Two classes of simulations with different models are performed, and in both cases the models show a reorganization of the molecular structure consistent with the theory of supercontraction, yet fail to show the dramatic change in size that is observed on the macro scale.

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

Acknowledgements

First and foremost, I should like to thank Professor Markus J. Buehler, for his direction, his help, his ideas and for taking a chance on a MEng student. Thank you to Tristan Giesa for his invaluable help, to Professor Carole C. Perry for her direction, and to Zhao Qin for his computing ability. Finally, a huge thanks to the remaining the LAMM residents, who were always willing to lend a hand: Dieter, Leon, Graham, Nina, Shu-Wei, Chun-Teh, David, Max, Anna, Tuukka, Baptiste and Arun.

Table of Contents

Abstract **3**

Acknowledgements **5**

1 Introduction **9**

1.1 Aim 9

1.2 Objectives 10

1.3 Natural Materials 10

1.3.1 Hierarchy 10

1.4 Natural Materials and Engineering 11

1.4.1 Why Research Natural Materials? 11

1.4.2 Nature as a Driver for Form 11

2 Background and Literature Review **13**

2.1 Water 13

2.1.1 The Water Molecule 13

2.1.2 Hydrogen Bonds 13

2.2 Proteins 14

2.2.1 Amino Acids 14

2.2.2 Hydrophobic and Hydrophilic Groups 16

2.2.3 Secondary Structure 17

2.3 Spider Dragline Silk 19

2.3.1 Molecular Structure of Dragline Silk 21

2.3.2 Mechanical Properties of Silk 23

2.3.3 Supercontraction 24

3 Computational Methods and Tools **27**

3.1 Why use Computational Methods 27

3.2 Tools 27

3.2.1 Molecular Dynamics 27

3.2.2 Visualization 28

3.2.3 CHARMM Force Field 28

3.2.4 Implicit and Explicit Water Models 28

3.3 Silk Models 29

4 Experimental Section **31**

4.1 Experiment 1: MaSp1 in a Vacuum 31

4.1.1 Hypothesis 33

4.1.2 Method 34

4.1.3 Results 35

4.1.4 Discussion 47

4.1.5 Outlook 48
4.2 Experiment 2: Glycine-Rich Linker Region 49
4.2.1 Hypothesis 50
4.2.2 Method 50
4.2.3 Results 52
4.2.4 Discussion 62
4.2.5 Outlook 63
5 Conclusion 64
6 References 67
7 Appendix 75
7.1 Experiment 1: MaSp1 in a Vacuum 75
7.1.1 NAMD Configuration File 75
7.1.2 Secondary Structure of a Trajectory 78
7.1.3 Average Position of Equilibrated Structure 79
7.1.4 Radius of Gyration and Gyration Tensor 80
7.2 Experiment 2: Glycine-Rich Linker Region (Snippet) 81
7.2.1 Configuration File for all Snippet Simulations 81
7.2.2 Distance of Water Molecules to Center of Protein 83
7.2.3 Distance of all Protein Atoms to Center of Protein 84

1 Introduction

Natural materials have evolved over millions of years to develop finely tuned mechanical properties to serve specific functions. Spider dragline silk is one such material. Used as the main structural component of the web, and as a lifeline for the spider, it must be strong, stiff and able to absorb energy without breaking. Its semi-crystalline molecular structure imparts it with both strength and extensibility that together create toughness currently unmatched by synthetic materials.

As a protein-based fiber, dragline silk is strongly affected by humidity levels. If wetted, an unrestrained dragline silk fiber will suddenly contract in length by up to 50%, a phenomenon called supercontraction. If the fiber is restrained, supercontraction will general significant stresses. Some argue that supercontraction is an evolutionary mechanism that keeps the web taut and responsive under added load. The manifestation of supercontraction on the macro scale must be due to fundamental changes in the silk's molecular structure, however the details of these mechanisms remain unknown and have not been directly demonstrated in prior research. The phenomenon has been observed and measured experimentally, but supercontraction has not been shown in atomic resolution.

Here we used full-scale atomic modeling of spider silk using molecular dynamics to investigate the structure and properties of this material at a length scale that is not yet accessible by experimental methods. Adding and removing water molecules to the system, and equilibrating the protein using molecular dynamics, the effects of water content on size, shape and structure of the spider silk protein can be measured. By creating conditions that would cause supercontraction on a macro scale, the response of the silk's molecular structure can be measured and compared to theoretical models of supercontraction.

1.1 Aim

The aim of the study is to investigate supercontraction on a full-scale atomistic model of spider dragline silk using molecular dynamics. Supercontraction is manifested by a dramatic shrinkage on the macro scale, and by a fundamental reorganization of the structure on a molecular scale. Exposing the model to large changes in water content is hypothesized to lead both to changes in the secondary structure and to bulk changes in the sample size.

1.2 Objectives

The study is broken down into two computational experiments. The first considers a full-resolution model of spider dragline silk protein MaSp1, previously built and equilibrated in a water box and hence thought to have already supercontracted. This experiment removes all the water and equilibrates the sample in a vacuum using molecular dynamics. The objective of this experiment is to investigate the effect of water on MaSp1 by removing it and observing the effects of its absence. The sample could expand as the secondary structure can assume a more ordered shape with the ability to reform hydrogen bonds – in effect, it could show a reversal of supercontraction.

The second experiment uses a much smaller sub-sample of MaSp1, from the hydrophilic amorphous glycine-rich region of the silk protein, and exposes it to varying water contents. This system is much smaller and has a vastly reduced computation time, which allows for more simulations. The objective of this experiment is to measure the effect of different hydration levels on a small sample of the amorphous region of MaSp1. It is expected that by increasing the water content, the sample assumes a less ordered secondary structure and decreases in volume. In addition, this small system may give insights on the role of water during protein folding, either in deciding the secondary structure or in leading the protein to maximize exposure of its hydrophilic residues to water.

1.3 Natural Materials

Natural materials, such as bone, silk, feathers, hooves, sponge spicules, shells, and innumerable others, are composed of relatively weak building blocks and bonds, yet often display excellent properties having evolved to serve specific biological functions central to a species' survival [1]. These properties, such as strength, robustness and adaptability are often achieved through a diversity of structural arrangements that span multiple scales; from molecular to macro, creating a complex assembly that allows inherently weak components to exhibit excellent properties [2-6].

1.3.1 Hierarchy

Biological materials look very different depending at which length scale they are being observed [6]. A primary structure, for instance a repeating pattern of amino acids in a protein chain, is folded into a secondary structure, which aggregates with other strands into a tertiary structure, and so on from the nano to the meso and the macro scale: this is structural hierarchy. Different length scales have different structures and functions [2, 7, 8]. Although at high structural levels biomaterials are complex, at the most fundamental levels they are generally simple and comprised of few building blocks [2, 3, 5, 6, 9, 10].

The hierarchical structural arrangements inherent to natural materials serve to stretch the design space from the relatively small number of basic constituents that are available such as C, Ca, N, O, H, P, Si). With different combinations across length scales, these universal building blocks create diverse biological functionality [1, 2, 11]. Hierarchy allows the seamless combination of both form and material [8]. The weak bonds between the blocks allows materials to explore a variety of structural states, and by being easily broken and reformed, impart natural materials with adaptability, toughness and flaw tolerance [2, 5, 9, 12]. Biological materials are able to sense new requirements and feed these into structural changes at distinct scales [1].

Understanding how the different levels of hierarchy work together is an essential step in characterizing the material and being able to predict how it will respond to external stimuli [13]. The material's properties on the macro scale such as deformation and fracture properties, directly depend on the material's microscopic structure [6]. Thus, multi-scale experimentation and modeling techniques are necessary to understand how structure and properties are linked [5, 6].

1.4 Natural Materials and Engineering

1.4.1 Why Research Natural Materials?

A main driver for the research into natural materials, from an engineering perspective, is developing heightened functionality with increasingly limited resources. Understanding natural materials could mean developing a new class of materials, with significantly less environmental impact and substantially improved properties to support and enhance technological advancement in areas such as medicine, energy and the environment [2, 7]. Not only are engineered materials made from finite resources, they usually require high energy input in the manufacturing and refining stage. The building blocks of natural materials, in contrast, are abundant and cheap. Natural materials make more efficient use of resources [14].

1.4.2 Nature as a Driver for Form

In structural engineering and architecture, form finding is a design tool that optimizes the shape of a structure to carry a given load. If the load is well defined, form finding can create extremely efficient, light structures and lead to considerable material savings. Nature has been doing this for millions of years. Since it costs an organism energy to create material, the process has been largely optimized, material use is minimized [12], shapes and forms have been created to efficiently serve and adapt to their functions [8]. These principles of efficient form for a given function can be applied to architecture and structural engineering [15].

The study of natural materials may also lead to an increased understanding of how different levels of hierarchy in an engineered structure interact with each other. Hierarchy is inherent to engineered structures, as shown in Figure 1.1. The understanding of how natural materials use almost arbitrary constituents in multiple levels of hierarchy may even lead to the use of weaker, less resource- and energy-intensive structural materials [2, 16], to the development of improved forms [16], and the creation of more resilient structures that are less likely to fail catastrophically.

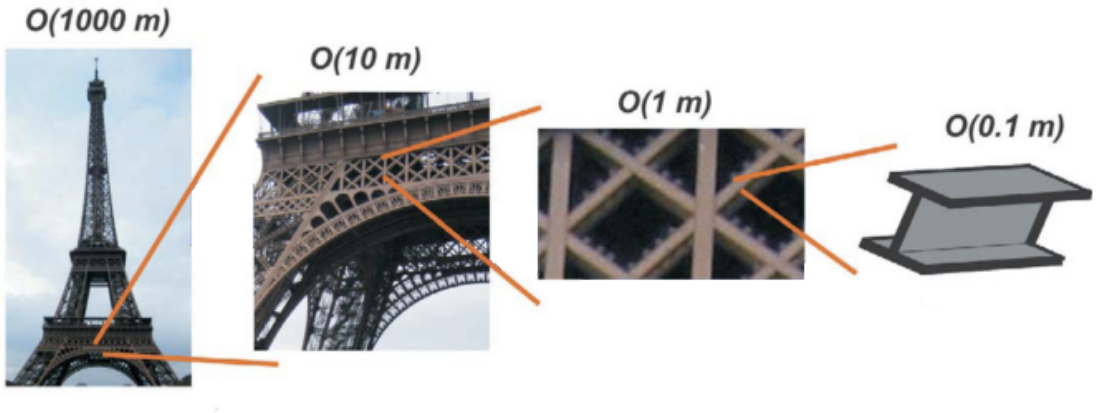

Figure 1.1 - Hierarchy in engineered structures. From [16].

2 Background and Literature Review

2.1 Water

2.1.1 The Water Molecule

The most fundamental entity of water is a combination of three atoms: two hydrogen atoms joined to an oxygen atom with strong covalent bonds. Figure 2.1 shows a water molecule, with the black lines representing covalent bonds. The length of the covalent bond is approximately one Angstrom.

Oxygen is more electronegative than hydrogen, so in a water molecule the oxygen atom will attract more electrons than the hydrogen atoms. Since electrons are negatively charged, the oxygen atom bears a partial negative charge (δ-) and the two hydrogen atoms bear a partial positive charge (δ+). This charge differential means the water molecule is *polar* [17-19].

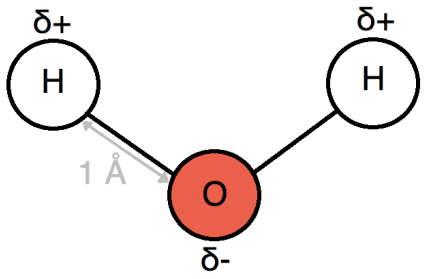

Figure 2.1 - A water molecule, showing one oxygen atom (red) and two hydrogen atoms (white), joined by covalent bonds of length approximately 1 Å, or 10^{-10}m.

2.1.2 Hydrogen Bonds

The hydrogen atom in one water molecule is attracted to the oxygen atom in another molecule, so neighboring molecules orient themselves following this attraction. This creates a hydrogen bond: an interaction of both the ionic and covalent kind, shown in Figure 2.2 [17, 18, 20].

Individually, hydrogen bonds are far weaker than fully covalent bonds, and are easily broken and reformed [18]. The strength of the hydrogen bond arises when large numbers of them can be formed, for instance, in bulk water or ice. The cumulative effect of the hydrogen bonds gives bulk water its unusual physical and chemical characteristics [17], and plays a tremendous role in the structure and function of proteins [20-25]. Hydrogen bonds do not exclusively bind water to water, they can connect water to other functional groups such as those containing oxygen,

nitrogen and sulfur, or connect these functional groups to each other [18]. Hydrogen bonds typically have length of 3 Å [26].

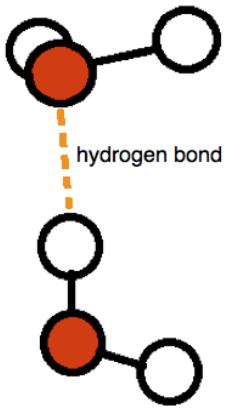

Figure 2.2 – The hydrogen atom in one water molecule is attracted to the oxygen atom in a neighboring water molecule.

2.2 Proteins

Proteins are biological molecules, made of long chains of amino acids, and are the principal structural element in a vast number of biological materials. In addition, they are key to providing biological function, and are responsible for a large number of tasks and processes necessary for the survival of living cells. Proteins assume a three dimensional structure that provides the ideal conditions for it to serve its function [2, 3, 7, 17, 18, 27].

2.2.1 Amino Acids

The basic constituents of proteins are amino acids. To form a protein, amino acids are arranged into long chains and joined by peptide bonds (polypeptide chains). There are 20 amino acids commonly found in proteins, listed in Figure 2.3. The properties of a protein depend highly on its constituent amino acids. The type, but also the sequence, the relative amounts of amino acids and the length of the polypeptide chain dictate the structure and function of the resulting protein. The 20 basic amino acids can create an extremely large variety of different proteins. The arrangement of amino acids along a protein chain is called the primary structure [9, 17, 18, 27, 28].

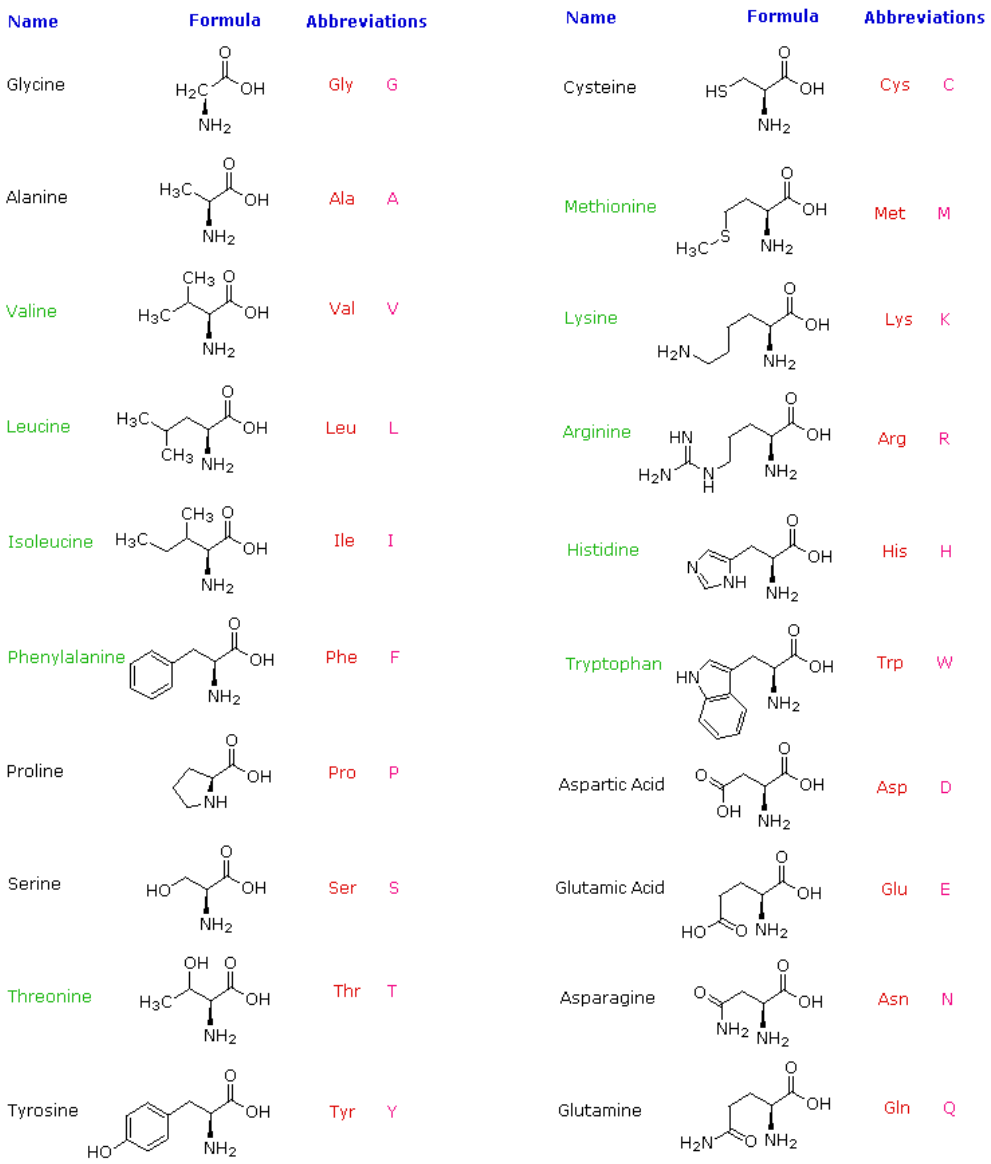

Figure 2.3 – List of 20 amino acids commonly found in proteins. From [29]

A polypeptide chain has a backbone and side groups, shown in Figure 2.4. The backbone is flexible, and allows the chain to assume an almost infinite variety of positions. The side groups are inherently hydrophilic or hydrophobic – usually both groups are found within one protein chain. The side groups determine the protein's three dimensional structure and its chemical characteristics [18].

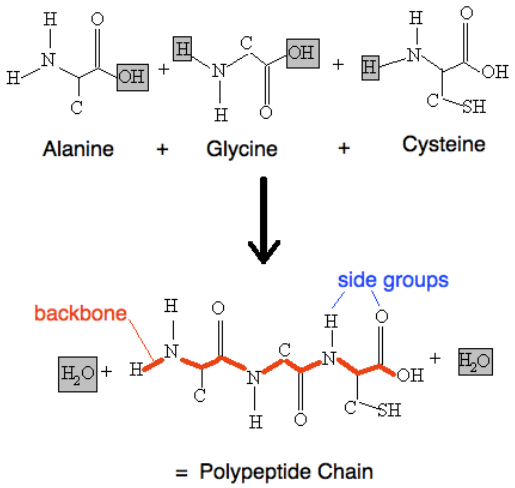

Figure 2.4 - An example of a polypeptide chain, showing the 3 constituent amino acids, the backbone and the side groups. Adapted from [30]

2.2.2 Hydrophobic and Hydrophilic Groups

The amino acids are inherently hydrophobic, hydrophilic or charged. Hydrophobic parts of the protein will reorganize themselves so they are not in contact with water. The hydrophilic residues assume positions where they can interact with the water, and in doing so effectively shield the hydrophobic residues from water, as shown in Figure 2.5. Thus, the hydrophobicity or hydrophilicity of a residue has a large influence on the ultimate conformation and structure assumed by the protein. The burial of the hydrophobic residues inside a protein is one of the main factors that affect protein folding [18, 20, 31].

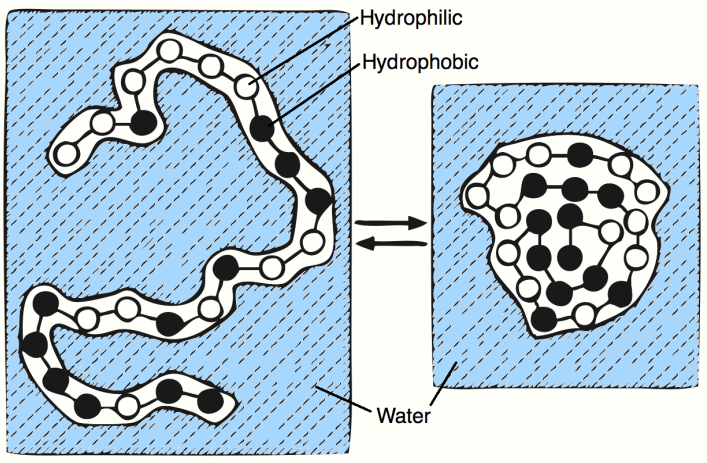

Figure 2.5 – The effect of hydrophilic and hydrophobic residues on the arrangement of a protein when exposed to water. Adapted from [32].

2.2.3 Secondary Structure

The local folding arrangement of chains of amino acids is called a protein's secondary structure [18]. These structures have a large role in the mechanical properties of the resulting bulk material. The three most common secondary structures are helices, beta-sheets and random coils. [2, 3, 7] Figure 2.6 shows commonly occurring secondary structures in protein materials.

The unique structure of a protein typically consists of many different types of secondary structure, the combination of which defines its mechanical properties on a larger scale. Secondary structures are created and stabilized by the formation of bonds between amino acids on the same or different chains. The bonds can be covalent bonds, van der Waals, or most commonly hydrogen bonds, and are influenced by hydrophilicity, hydrophobicity, and by the surrounding medium. They are key to maintaining the secondary structure and higher levels of hierarchy, holding together individual proteins and assemblies of proteins [6]. The arrangement that a protein assumes aims to maximize the amount of charged sites that are available for hydrogen bonding [6, 9, 16, 28]. The hydrogen bonds that are vital in maintaining the structure of proteins are weak enough for a single bond to be considered insignificant. However, when many of them act together they can provide stable secondary structure arrangements [6].

Helices are found in the areas of proteins where mechanical stability is required. Random coils are found in areas where extensibility is required. Beta sheets are common in materials that show remarkable resistance and strength against mechanical manipulation, such as spider dragline silk [6, 33]. Beta sheets are extensively hydrogen bonded, which is what gives them their strength in shear [6, 33].

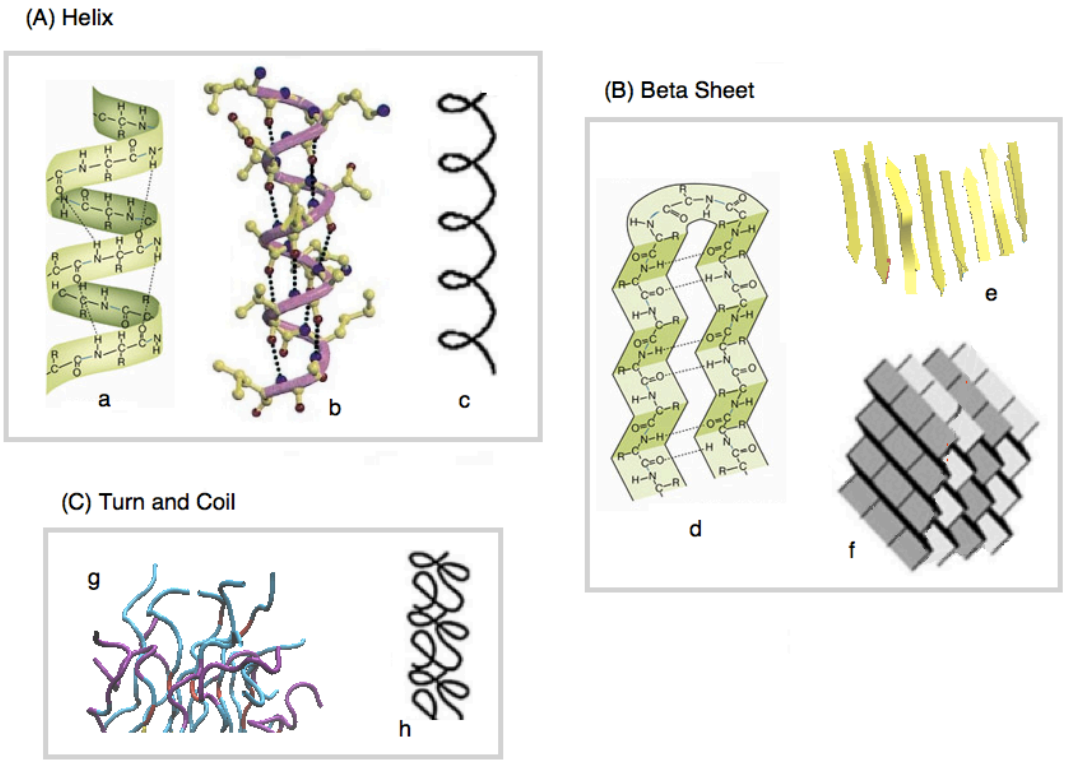

Figure 2.6 - Common secondary structures in protein materials: helices, beta sheets, turns and coils. a) alpha helix, b) 3_{10} helix, c) cartoon representation of a helix, d) beta pleated sheet, e) cartoon of antiparallel beta sheet, f) beta sheet crystal, g) turns (purple) and coils (blue) – turns have some ordered defined structure, and coils are random, h) cartoon of turns and coils. Adapted from [9, 34, 35]

2.3 Spider Dragline Silk

Spider dragline silk is a protein based fiber, used as the main structural component in spider webs and as a lifeline for the spider [36, 37]. Protein fibers are formed when protein strands self-assemble into intricate fibers structurally optimized to be strong along a particular direction. The polymer backbone is aligned with the direction where load capacity is needed so the covalent bonds in the backbone are fully used [9, 38].

One of the most studied silks is the dragline silk of the Nephila clavipes or golden orb web spider [39-42]. It is well understood because the dragline silk can be harvested – the spider is trapped and anesthetized, and forced spinning gives samples on which to perform experiments. In addition, its large size makes its glands easy to dissect [41, 43]. Figure 2.7 shows the Nephila clavipes in its web, and Figure 2.8 shows a scanning electron microscopy image of a dragline silk strand.

Figure 2.7 - Nephila clavipes spider in its web [44]

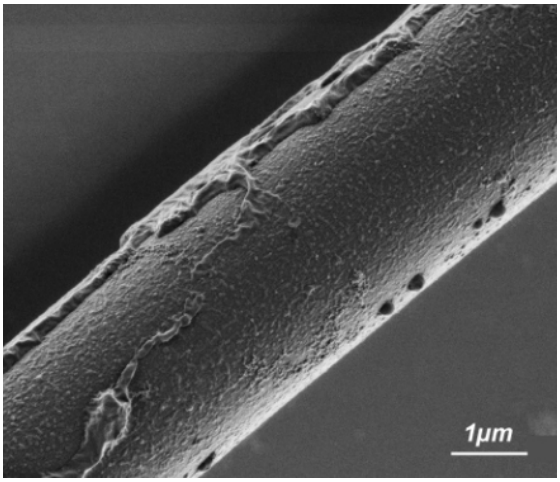

Figure 2.8 - Scanning electron microscopy image of Nephila clavipes dragline silk. Reprinted from [45] Copyright 2000, with permission from Wiley Periodicals Inc.

Apart from the dragline, Nephila clavipes can make up to six other types of silk, shown in Table 2.1 [3, 9, 39, 41, 46-48]. Each has a distinct role to play in the web architecture (see Figure 2.9) and hence the survival of the spider.

Table 2.1 – Different types of Nephila clavipes silk, their uses and properties.

Nephila clavipes silks and their uses		
Silk	**Use**	**Properties**
Major ampullate dragline	Web frame and radii	Stiff, strong, tough
Minor ampullate	Web reinforcement	Sticky, extensible
Flagelliform	Core fibers of adhesive spiral	Extensible, sticky, tough
Aggregate	Adhesive silk of spiral	Sticky, tough
Cylindrical	Cocoon, egg protection	Tough
Aciniform	Swathing and inner egg sack	Tough
Pyriform	Junction between fibers	Sticky, tough
[41, 46]	[41]	[47]

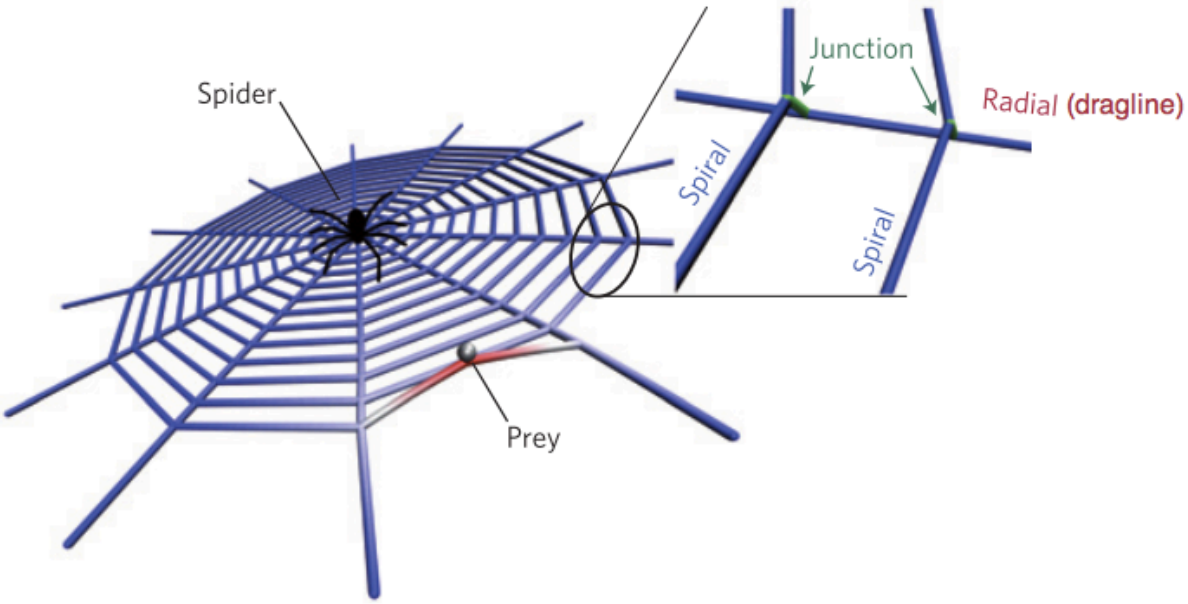

Figure 2.9 - Schematic of a spider web, indicating the radial, spiral and junction silk fibers. Reprinted from [49], copyright 2013, with permission from Nature Publishing Group.

2.3.1 Molecular Structure of Dragline Silk

A dragline silk fiber is made of protein, so its primary structure relies on the 20 amino acids as building blocks. Alanine, glycine, serine and proline feature heavily in dragline silk [50]. Silk is made up of repeating units, and has a distinct secondary, tertiary structures and higher level structures [39].

2.3.1.1 MaSp1 and MaSp2

Dragline silk is made of a combination of 2 different proteins, Major Ampullate Spindroin 1 (MaSp1) and Major Ampullate Spindroin 2 (MaSp2). Both MaSp1 and MaSp2 proteins have the same semi-crystalline repeat pattern (described below), but have different structure and thus distinct mechanical functions [39, 41, 51-56]. A Nephila clavipes dragline silk fiber has multiple layers or fibrils. The inner core is made of a mixture of MaSp1 and MaSp2, the outer layer is pure MaSp1 [57]. Approximately 60% to 80% of the total volume in a dragline silk fiber is MaSp1 [39, 56, 58, 59].

MaSp1 has a higher degree of crystallinity, and contains alanine and glycine repeats. The presence of alanine leads to the formation of regular repeated beta sheet crystals. MaSp1 contains barely

any proline [39, 46]. Conversely MaSp2 contains large amounts of proline, which decreases the degree of crystallinity and increases disorder. Strands that include proline tend to twist away from regular conformations and instead assume random coil shapes, so feature heavily in the amorphous sections of dragline silk. As a result MaSp2 has lower beta sheet content than MaSp1. Therefore MaSp2 is thought to contribute more to extensibility and elasticity, while MaSp1 helps in achieving strength [56].

2.3.1.2 A Semi-Crystalline Fiber

Dragline silk proteins contain two distinct repeating regions: stiff crystalline beta sheets, distributed among a softer amorphous network. [37, 39-41, 43, 55, 58, 60-67]. The beta sheets account for approximately 15 – 30% of the protein, and the amorphous accounts for the remaining 70 - 85% [40, 58, 62]. The structural hierarchy of spider dragline silk is shown in Figure 2.10.

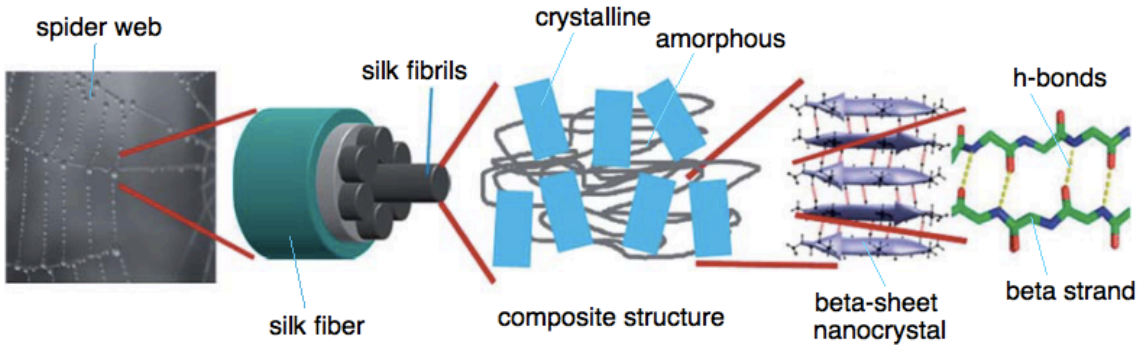

Figure 2.10 – Hierarchical structure of spider silk. Reprinted from [59] with permission from the Royal Society.

The beta sheets that occur in both MaSp1 and MaSp2 are mostly alanine [54, 63, 68], and are arranged in antiparallel layers and are highly cross-linked with hydrogen bonds. Beta sheets are hydrophobic, highly conserved and permanently ordered [39, 52, 53, 55, 56, 58, 69].

The amorphous region of dragline silk is mostly hydrophilic, and can be further separated into a permanently disordered section (20%) and a glycine-rich region that links the beta sheets and the random coils with relatively more order. The permanently disordered region forms random coils, and the glycine-rich region forms helices and beta turns [54]. The structure of the amorphous region is harder to characterize than the beta sheet, because of the lack of detail provided by experimental and imaging techniques. In general due to the lack of atomic resolution full-scale models, the link between the chemical structure, genetic makeup, atomic structure and

the properties at the macro scale remain unclear [52] and there is much to be learned about the properties of dragline silk [70].

2.3.1.3 Hydrogen Bonds and Silk Structure

The secondary structure of silk is held together with hydrogen bonds. It is the presence and action of hydrogen bonds that gives dragline silk its secondary structure and thus its physical behavior [71]. The beta sheets are strongly cross-linked with hydrogen bonds [39, 72, 73]. The regularity of their arrangement allows for multiple bonds which makes for a strong structure [55]. The alanine that makes up the beta sheets is hydrophobic and thus water is unable to penetrate and diffuse inside beta sheets, maintaining the stability of the crystals even in wet conditions.

In contrast to the beta sheets, the amorphous regions suffer from considerably weaker hydrogen bonds, because of the poor orientation of the strands [55]. However it is argued that even the strands in the amorphous section run parallel to the fiber and loading direction, implying some degree of order and function [53].

2.3.2 Mechanical Properties of Silk

Dragline silk is required to perform some specific functions: it must give structural integrity to the web under environmental loading and moving anchor points, it must allow for the absorption of kinetic energy when intercepting prey without an elastic recoil that would launch it away [74], it must support the spider's weight and restrict mid-air spinning [3, 50, 71]. The material that evolved to satisfy these requirements has attractive mechanical properties. [72]

Spider dragline silk has a strength to weight ratio several times greater than steel [36, 55, 75]. For a low-weight low-density fiber, its mechanical properties are truly impressive. It combines strength and elasticity currently unattainable by synthetic materials. In fact the toughness of dragline silk exceeds most natural and manmade fibers, it can absorb massive amounts of energy before rupture, and it becomes stiffer as it is stretched [3, 4, 8, 36, 37, 39, 40, 47-49, 53, 55, 61, 64, 65, 70, 76-78]. Spider dragline silk is therefore an ideal material to try and imitate [39, 49, 53, 79].

Dragline silk's toughness is due to its semi-crystalline molecular structure, in particular the repeating units, which arrange themselves into hierarchies [50, 59, 73]. The hierarchical nature of silk structure means that there are both crystalline and amorphous regions within a nanometer scale. This allows incoming load to be efficiently distributed between the two regions, and allow them to impart their respective mechanical properties to the fiber [71]. Beta sheets are strongest in shear, which justifies their orientation within the global silk fiber [6]. Extensibility in silk is due to the hidden length found in the amorphous regions, which have been shown to extend the time before total failure of a silk sample [39, 56].

2.3.3 Supercontraction

Water has the ability to fundamentally reorganize silk's molecular structure, causing dramatic changes in mechanical properties and physical characteristics. Dry silk is relatively stiff, and wet silk is much softer, more compliant and extensible. This occurs even under ambient conditions, where the humidity can range from 10% to 100% [39, 57, 58, 71, 80].

At high humidity, some spider dragline silks will shrink by up to 50% - this phenomenon is known as supercontraction, and is not fully understood. Upon wetting, an unrestrained silk fiber shows a significant contraction in length, and if constrained in a web it will generate significant stresses [55, 58, 59, 61, 75, 79]. Interesting is the threshold at which this occurs. At hydration levels below 70%, silk fibers show slight swelling upon exposure to water, but once the hydration exceeds 70% there is a fundamental reorganization of the amorphous structure [39, 55, 60, 75].

There is little agreement on whether supercontraction is an evolutionary feature [75] or a constraint [60]. It has been suggested that supercontraction is an evolutionary mechanism to keep the spider web taut under the additional loading of morning dew or precipitation [49].

Since the beta sheet crystals are hydrophobic, they do not undergo important structural changes when hydrated, so the supercontraction phenomenon occurs in the amorphous phase only [81]. Above a critical hydration level (70%), water molecules disturb the hydrogen bonds between strands in the amorphous structure and allow the to reorganize into a less ordered, more coiled, lower energy state [55, 60, 75]. Even at high levels of supercontraction and hydration, the beta sheet crystals remain intact and ordered because the water molecules can't penetrate them. However, their orientation relative to the bulk fiber does decrease [75]. Concurrently, the orientation of the disordered and glycine rich linker regions decreases [60]. The dramatic response of silk to water indicate that the dry fiber is frozen into a glassy state that is partially extended, and once exposed to water this glassy state relaxes and random coils form [64, 68]. The wetted silk turns into an elastomer [64, 81, 82]. Figure 2.11 shows the process of supercontraction from a molecular perspective. This is a hypothesized model of the interaction of water with dragline silk protein created by [55] based on experimental observations of mechanical performance of virgin and supercontracted silk, combined with the theoretical process of supercontraction obtained from literature.

Supercontraction arises due to the response of both the random coils and the relatively oriented glycine-rich linker regions that make up the amorphous region: it is argued that the phenomenon would not occur were one of these responses missing [55].

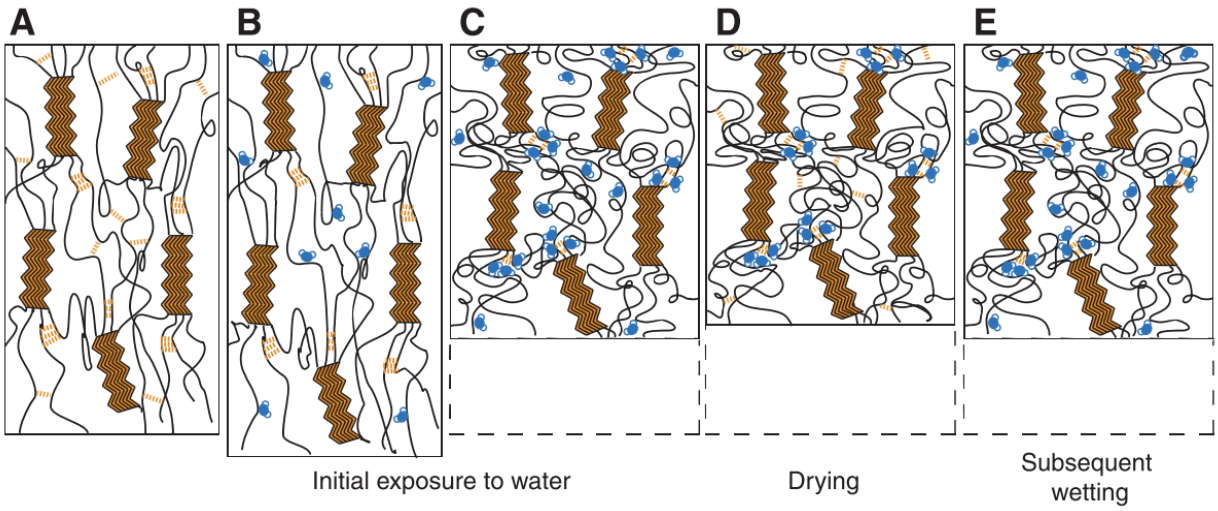

Figure 2.11 – The process of supercontraction, as understood from [55]. This is a hypothesized model of the interaction of water and spider dragline silk, based on experimental observation of mechanical performance of virgin and supercontracted silk, as well as theory from literature. (A) The beta sheets, represented by brown zigzags, distributed within the amorphous network. The triple gold lines represent the glycine-rich linker regions, and single gold lines represent weak individual hydrogen bonds in the amorphous region. (B) Water molecules, in blue, first enter silk where they interact with hydrophilic random coils in the amorphous region. Some hydrogen bonds are broken and the silk relaxes slightly. (C) Humidity exceeds 70% and water molecules penetrate the glycine-rich linker regions and break some of the hydrogen bonds. The strands reconfigure to lower energy state, which causes contraction in the fiber. (D) Upon drying, some water molecules remain permanently bound in the glycine-rich linker regions. Hydrogen bonds in amorphous structure reform, and the fiber shrinks and stiffens. (E) Subsequent wetting causes water to re-enter the amorphous structure and disrupt the hydrogen bonds again. Reprinted from [55].

3 Computational Methods and Tools

3.1 Why use Computational Methods

Computational approaches in materials science can complement experimental results, and if adequately parameterized, can be as valuable as a physical specimen (and play an important role in developing fundamental understanding of a material's behavior). Working from the fundamental structure, and using a computational approach to simulate the behavior of individual atoms and molecules, the material's behavior at the macro scale can be predicted from its chemical composition. The laws of physics are used to model the behavior of the fundamental structure under applied mechanical stimulation or external conditions [2]. Computational modeling of complex biological materials can help inform experiments, and if results are properly compared to experimental data, can be used as a predictive tool in the behavior of biological materials and many length and time scales. The use of bottom-up all-atomistic models of biological materials, built using basic chemical structure, has so far been a successful approach [5]. The combination of computation and experimental approaches may offer significant improvements in the design and understanding of both natural and synthetic materials [40].

3.2 Tools

3.2.1 Molecular Dynamics

Molecular dynamics (MD) is a method of computer simulation that applies the laws of mechanics to the study of molecules and their trajectories. Molecular dynamics uses Newtonian laws of motion ($F=ma$) to predict the movement of individual particles [83]. It considers atoms as point masses, and their interactions are described using force fields (such as CHARMM). This approach is used to study particles such as proteins and other biomolecules, as a large group of atoms that represent a material volume. Additionally, molecular dynamics is a tool that can relate deformations at the nano scale to macroscopic material properties, by informing coarser models and characterizing the various structural features of intermediate length scales [2]. The equations of motion are solved at every time step, to create a dynamic model of the material. Molecular dynamics is a very fine-scale all-atomistic modeling tool, and as such it is limited by processing power to small volumes and short time scales. Using molecular dynamics involves a trade-off between accuracy and computational efficiency [84], however, it can be used effectively to inform coarser scales and approaches [2, 7, 40].

Molecular dynamics is a useful tool to provide data on the behavior of materials at scales that cannot yet be reached by experiments [16, 85]. However, it is essential that molecular dynamics results be compared with experimental results, to either validate or invalidate the model that is used [85]. In this study, all molecular dynamics simulations were run using NAMD (a molecular dynamics program developed by [86]) with the CHARMM22 force field, and the results are visualized in VMD.

3.2.2 Visualization

Visual Molecular Dynamics, or VMD, is a tool that allows the visualization and analysis of complex biological molecules and their equilibration trajectories [87]. Among many other things, it can calculate the variation of the secondary structure of a molecule as it equilibrates. VMD uses the STRIDE algorithm to calculate secondary structure [73]. The STRIDE Algorithm [88, 89] uses hydrogen bond energy, protein backbone torsion angle and protein coordinates to predict secondary structure

3.2.3 CHARMM Force Field

The CHARMM force field is used in Molecular Dynamics to describe the forces between the atoms in a complex biological molecule. It provides a reasonable description of the behavior of proteins. The forces between atoms arise due to covalent interactions, and long-range electrostatic interaction such as van der Waals, ionic and hydrogen bonds. In CHARMM, bonds between atoms are modeled by springs, and are unable to break. As electrostatic interactions, hydrogen bonds are able to break and be reformed in the CHARMM force field [6].

3.2.4 Implicit and Explicit Water Models

Adding water to a molecular dynamics model requires a compromise, either the water molecules are modeled explicitly, or the effects of the presence of water are artificially added with an implicit water model. Explicit water is more accurate because the molecules are not restrained and can travel throughout the system in a realistic way. However, this carries with it extra computation expense and for large systems may make simulation times very long. Implicit water can reduce the computation time but at the expense of accuracy, since it represents the properties in an averaged manner [90]. However, it is impossible to view the manner in which water participates in protein folding using implicit water [91], and it is also impossible to trace the activity of single water molecules within the system. In the experimental section of this study, all water models used are in explicit form.

3.3 Silk Models

The experimental section describes simulations using a model of Nephila clavipes dragline silk developed by [52, 59, 73]. This section provides a brief summary of the steps involved in the model's development. The samples were built by [52] from the bottom-up, using the amino acid sequence of silk of the Nephila clavipes protein MaSp1 to create the strands, see Figure 3.1. The models were built to resolve a lack of atomic level descriptions of spider silk based on genetic makeup, chemical information and physical constraints [52].

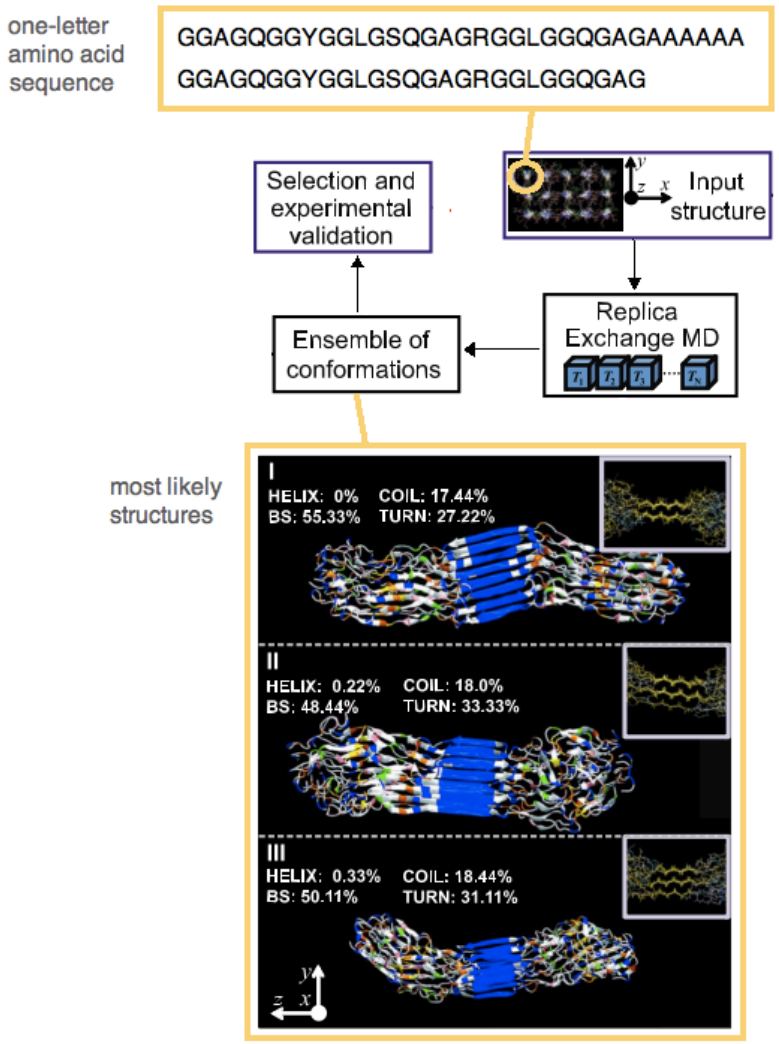

Figure 3.1 – Using REMD to find the likely conformations of a sample of MaSp1 built from strands of the amino acid sequence. The sample contains 15 strands. Reprinted from [52], copyright 2010, with permission from the American Institute of Physics.

The poly alanine repeat was shown to be optimized with 6 alanines [73]. These structures are shown in Figure 3.2. The model in the orange box was run through an explicit water simulation to obtain a more realistic protein tertiary structure and molecular conformation [73]. It is the result of this explicit water molecular dynamics simulation that serves as the starting point for the computational experiments detailed in the next chapter.

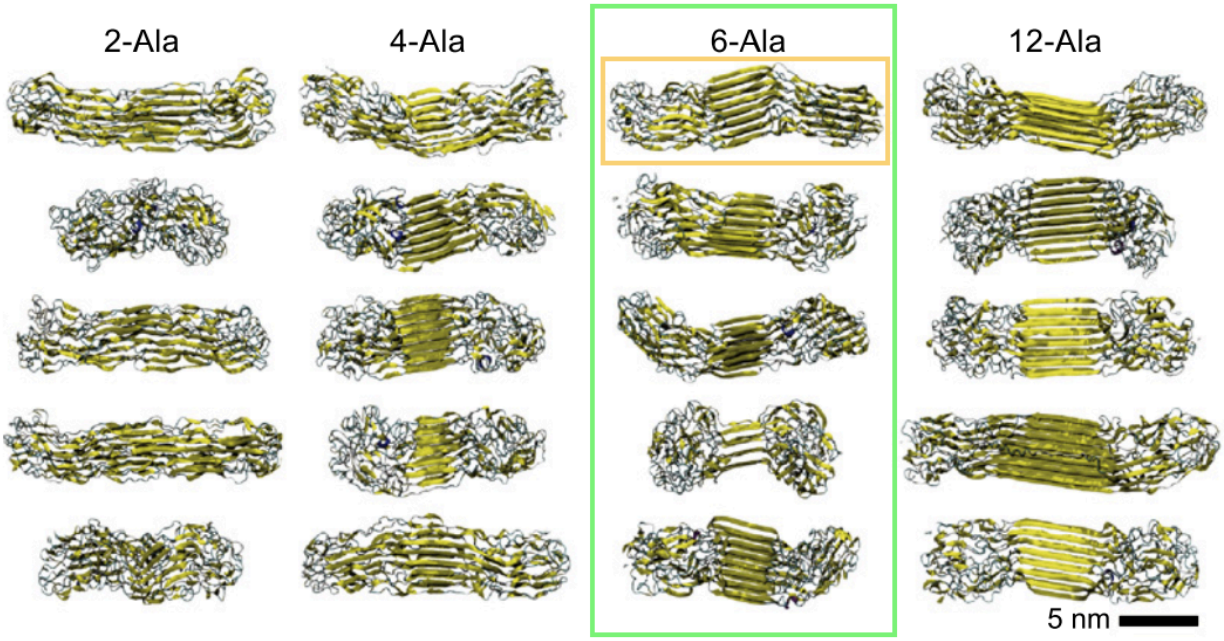

Figure 3.2 - Five most likely structures as computed by Replica Exchange Molecular Dynamics for 2, 4, 6 and 12 alanine samples. 6-ala is the optimum, marked in the green box. The orange box identifies the sample used in this study. Reprinted from [73], copyright 2012, with permission from Elsevier.

4 Experimental Section

This section describes two different experiments that aim to investigate supercontraction on a molecular dynamics model of spider dragline silk. The presence of water has been shown to dramatically affect the bulk properties of dragline silk, causing supercontraction in wet conditions. This response is due to water molecules altering the molecular structure of the silk, however the exact mechanisms of how this is translated to a response at the macro scale are still unclear.

With this in mind, two different experiments were performed. Both observe the interactions between a protein and explicit water molecules, modeled by TIP3[1]. The first experiment measures the effect of removing all water molecules from a large MaSp1 protein model containing over 9555 atoms. The second measures the effect of few water molecules on a small subsection of the silk protein sample, containing 227 atoms, in terms of its molecular structure and also how the water interacts with the protein on a smaller scale.

4.1 Experiment 1: MaSp1 in a Vacuum

MaSpI is a protein found in the dragline silk of the Nephila clavipes spider. Figures 4.1, 4.2 and 4.3 show the MaSp1 protein model used in this experiment, using colors to represent different aspects of the sample. Figure 4.1 shows the hydrophobic, hydrophilic and basic regions. Figure 4.2 shows the amino acids that make up the model, and Figure 4.3 shows the model's secondary structures.

[1] TIP3 is a type of explicit water molecule model, in which all the bond lengths and angles between the three atoms are constrained. This ensures the molecule remains defined as water, and does not break up into individual hydrogen and oxygen atoms.

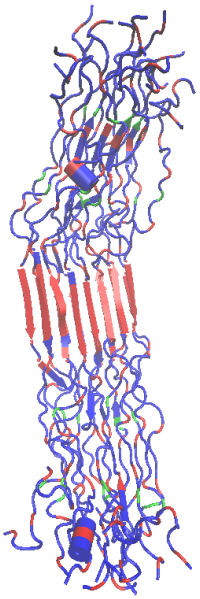

Figure 4.1 - MaSp1 colored by its affinity to water. Hydrophobic regions are in red, hydrophilic regions are blue, and basic regions are green. This image shows the beta sheet, center, is strongly hydrophobic and its structure is thus unlikely to be affected by the presence of water. The amorphous regions are mostly hydrophilic.

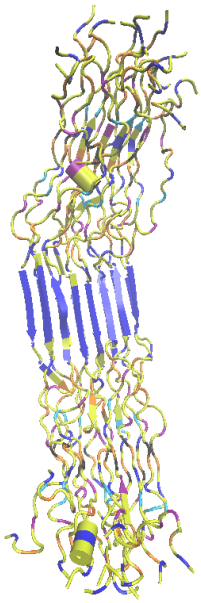

Figure 4.2 – MaSp1 colored by its constituent amino acids. Alanine is blue, arginine is light blue, glycine is yellow, glutamine is orange, leucine is purple, serine is black and tyrosine is dark orange. This image shows that the beta sheet is made of alanine.

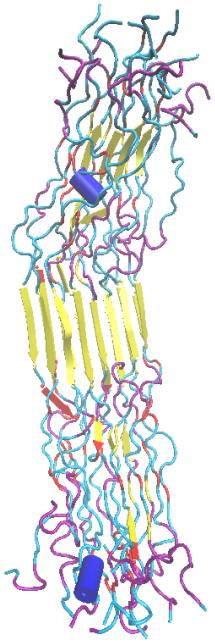

Figure 4.3 – MaSp1 colored by its secondary structure. Beta sheets are yellow, beta bridges are red, 3_{10} helices are blue, turns are purple, and coils are light blue. The model shows one large beta sheet in the center, and several other smaller ones scattered throughout. There are two 3_{10} helices in the amorphous region. The rest of the amorphous region is a combination of beta bridges, turns and coils.

4.1.1 Hypothesis

Water is unlikely to affect the beta-sheet crystals in MaSp1 due to the extensive hydrogen bonding between residues and the densely packed structure. Thus, water is expected to have an impact on the amorphous regions, whose disordered arrangement allows water molecules to enter and associate with the protein. It is anticipated that the removal of water causes an overall increase in the ordered secondary structure content. Water molecules break the weak hydrogen bonds in amorphous regions and cause the strands to coil. Removing the water should lower the content of coils, and increase the content of more ordered turns, beta sheets and helices (in particular the turns, because beta sheet and helices are often too well bonded for water to interfere with them in the first place).

This change in secondary structure is hypothesized to cause the bulk manifestation of supercontraction, namely a large change in length, in dragline silk. Since this experiment creates the conditions for the reversal of supercontraction, it is expected that the sample show an increase in volume, accompanied by the ordering of the secondary structure. The two amorphous regions and expected to increase in size, and the beta sheet is expected to stay constant. This is shown schematically in Figure 4.4.

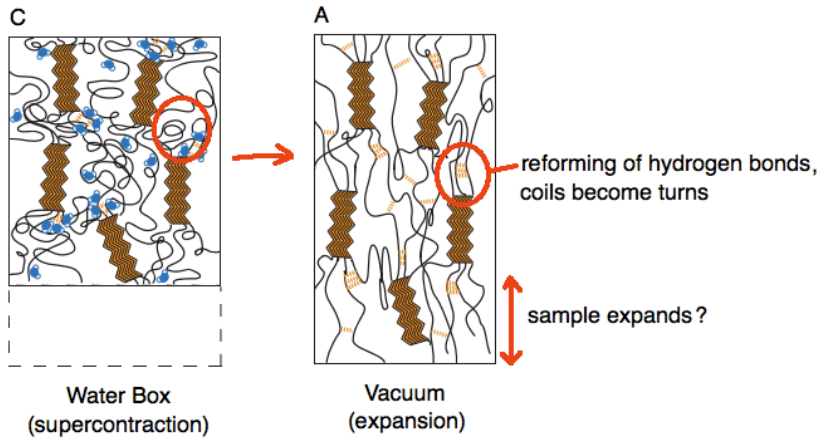

Figure 4.4 – Equilibration of MaSp1 in a vacuum is hypothesized to cause an expansion. The sample in the water box is assumed to have supercontracted. Removing all the water should show an expansion due to the ordering of the amorphous secondary structure and the reforming of hydrogen bonds, here shown by gold lines. Adapted from [55], full figure shown in Figure 2.11.

4.1.2 Method

The sample of MaSp1 previously equilibrated in an explicit water box [73] was stripped of all its water and allowed to equilibrate in a vacuum in a full atomistic simulation for 30 nanoseconds. The molecular dynamics program NAMD was used to run the full atomistic simulation, as per the method used by [73], in an NVT ensemble[2] with periodic boundary conditions, for 30 nanoseconds. To prevent image interactions, the periodic box wraps the protein by at least 10 Å. Equilibration is performed at 300 K and with Particle Mesh Ewald (PME) electrostatics[3]. The configuration file used is shown in Appendix 7.1.1. All results are compared with an equivalent experiment by [73] with the protein fully solvated in an explicit water box.

[2] In an *NVT* ensemble, the volume is kept constant and the pressure allowed to fluctuate
[3] PME allows the effective computing of long range electrostatic interactions, by calculating the exact Coulomb interactions (only possible with periodic boundary conditions).

4.1.3 Results

Root Mean Square Deviation

The 30 nanosecond NAMD simulation outputs a trajectory file of 600 frames of the protein's instantaneous position, so that the trajectory of the equilibration can be viewed. A flat Root Mean Square Deviation (RMSD) between atom positions suggests the structure has equilibrated. Figure 4.5 shows the RMSD over time for both the water box and the vacuum equilibration, and it was decided that for both cases the last 10 nanoseconds would be considered equilibrated. Note: the water box simulation was only run for 20 nanoseconds.

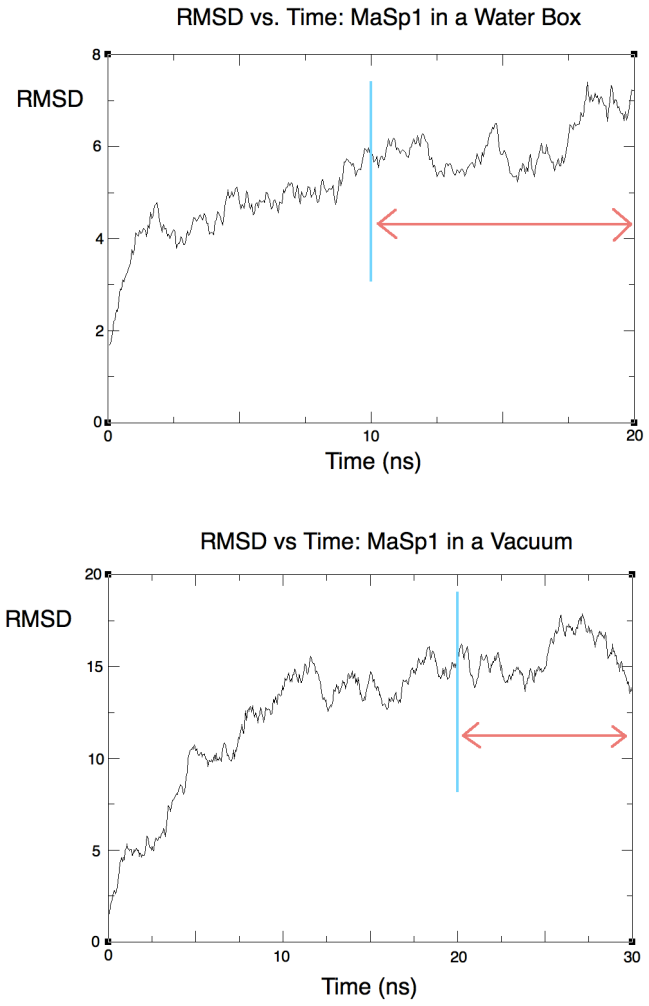

Figure 4.5 - RMSD of equilibration of MaSp1 in a water box (top) and vacuum (bottom), showing region over which RMSD is approximately flat and structure is approximately in equilibrium. For both simulations this corresponds to the last 10 nanoseconds.

Secondary Structure

The secondary structure of the last 10 nanoseconds was calculated using a script in VMD (shown in Appendix 7.1.2). The average and standard deviation for the percentages of helix, beta sheet, turn and coil in both the water box (blue) and vacuum (orange) simulations are shown in Figure 4.6

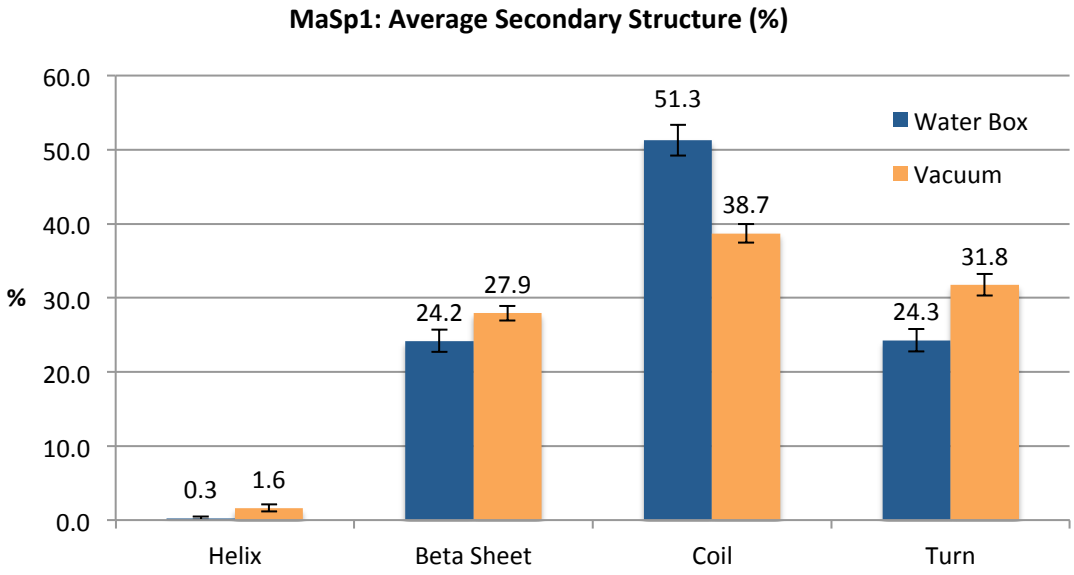

Figure 4.6 - Average secondary structure for last 10 nanoseconds of the equilibration trajectory for MaSp1 in a water box (blue) and a vacuum (orange) for the last 10 nanoseconds of equilibration. This image shows that the percentage of helices, beta sheets and turns all increase in a vacuum, and the percentage of coils decreases. This is consistent with the hypothesis: the percentage of ordered secondary structures has increased with the removal of water.

Figure 4.6 shows that the secondary structure content differs between MaSp1 in a water box and in a vacuum. The content of helices increases from 0.3% in the water box to 1.6% in the vacuum. The beta sheet content also increases from 24.2 in the water box to 27.9% in the vacuum. The coil content, which is the most disordered and random secondary structure, decreases from 51.3% in the water box to 38.7% in the vacuum. Finally, the turn content, which is a slightly more ordered amorphous secondary structure, increases from 24.3% in the water box to 31.8% in the vacuum.

These results overall support the hypothesis that the amount of ordered secondary structure (that is, beta sheets, helices and turns) should increase in the vacuum. Both beta sheet and helix content increasing suggests that the presence of explicit water does interfere with their bonds, albeit to a small extent. The removal of water allows these disrupted secondary structures to reform their hydrogen bonds, see Figure 4.7. The increase in turn content, by a larger extent, suggests the reformation of hydrogen bonds in coiled networks, causing an overall reorganization

of the structure. The loss in coil content also supports the fact that the presence of water destroys ordered secondary structures leading them to form random coils.

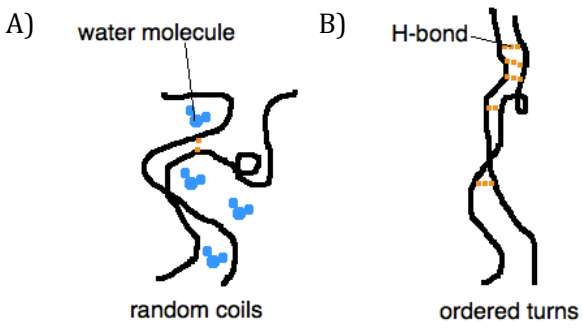

Figure 4.7 – A) The presence of water molecules causes the protein to form random coils, and in doing so break the hydrogen bonds between the two protein strands. B) The removal of water allows hydrogen bonds to reform between the strands, as they are no longer being interfered with by the water molecules. As a result the strands become more ordered and the coils become turn structures.

The population standard deviation of the secondary structure variation throughout the trajectory was measured using Excel, and are shown as error bars in Figure 4.6. The relatively small standard deviations suggest that the differences between the water box and the vacuum values are significant, since they appear outside the standard deviations of both sets of data.

Average Position

The coordinates of each atom in the protein in the last 10 nanoseconds were extracted and averaged, to find the average conformation of the protein equilibrated in both a water box and a vacuum. This was done because each frame represents an instantaneous conformation that, as it fluctuates about its equilibrium position, could turn out to be far from the average. Figure 4.8 shows the average position of MaSp1 in a water box (blue, left) and in a vacuum (orange, right). As shown, at a glance the conformations do not seem to differ widely. More detailed analyses were performed to measure the change in shape. The script used to find the average position using VMD is shown in Appendix 7.1.3.

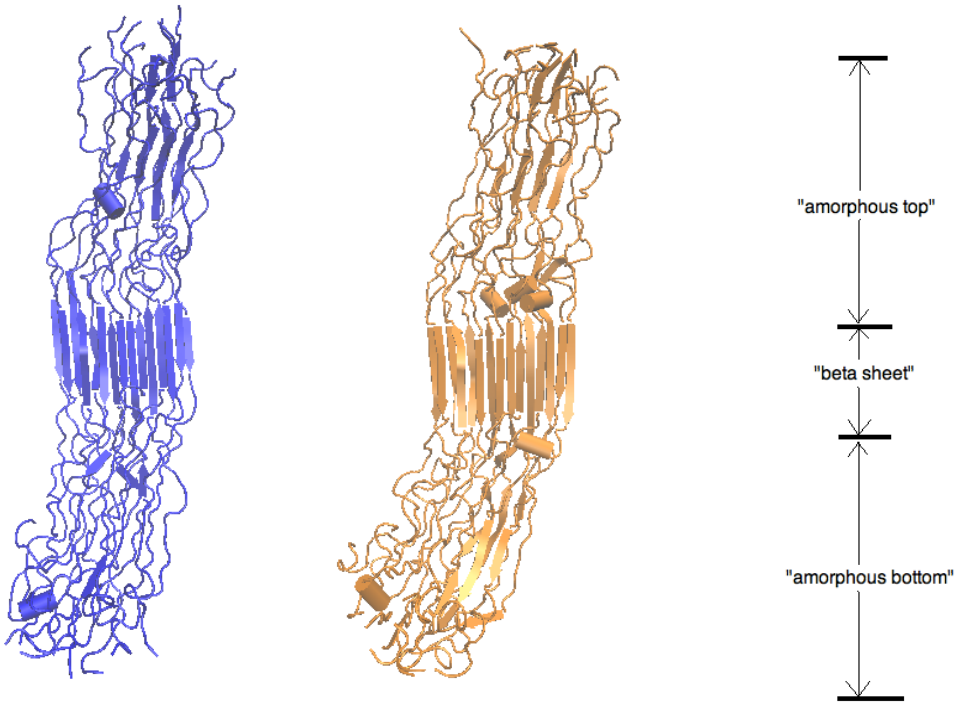

Figure 4.8 - MaSpI average equilibrated position in a water box (blue, left) and in a vacuum (orange, right). Labels show how the structures were split into three regions, amorphous top, amorphous bottom and beta sheet.

The structure was split into three sections: two amorphous (top and bottom) and one beta sheet, in order to clearly compare them in a water box and in a vacuum. The structures were separated based on their z-axis coordinates. The limit of the beta sheet was taken as the end of the mostly poly-alanine repeat, including both extended beta sheets and beta bridges so the "beta sheet" section contains AGAAAAAAGGA.

The average coordinates of the two amorphous regions and the beta sheet were plotted for both the water box (blue) and the vacuum (orange), and shown in Figures 4.9, 4.10 and 4.11.

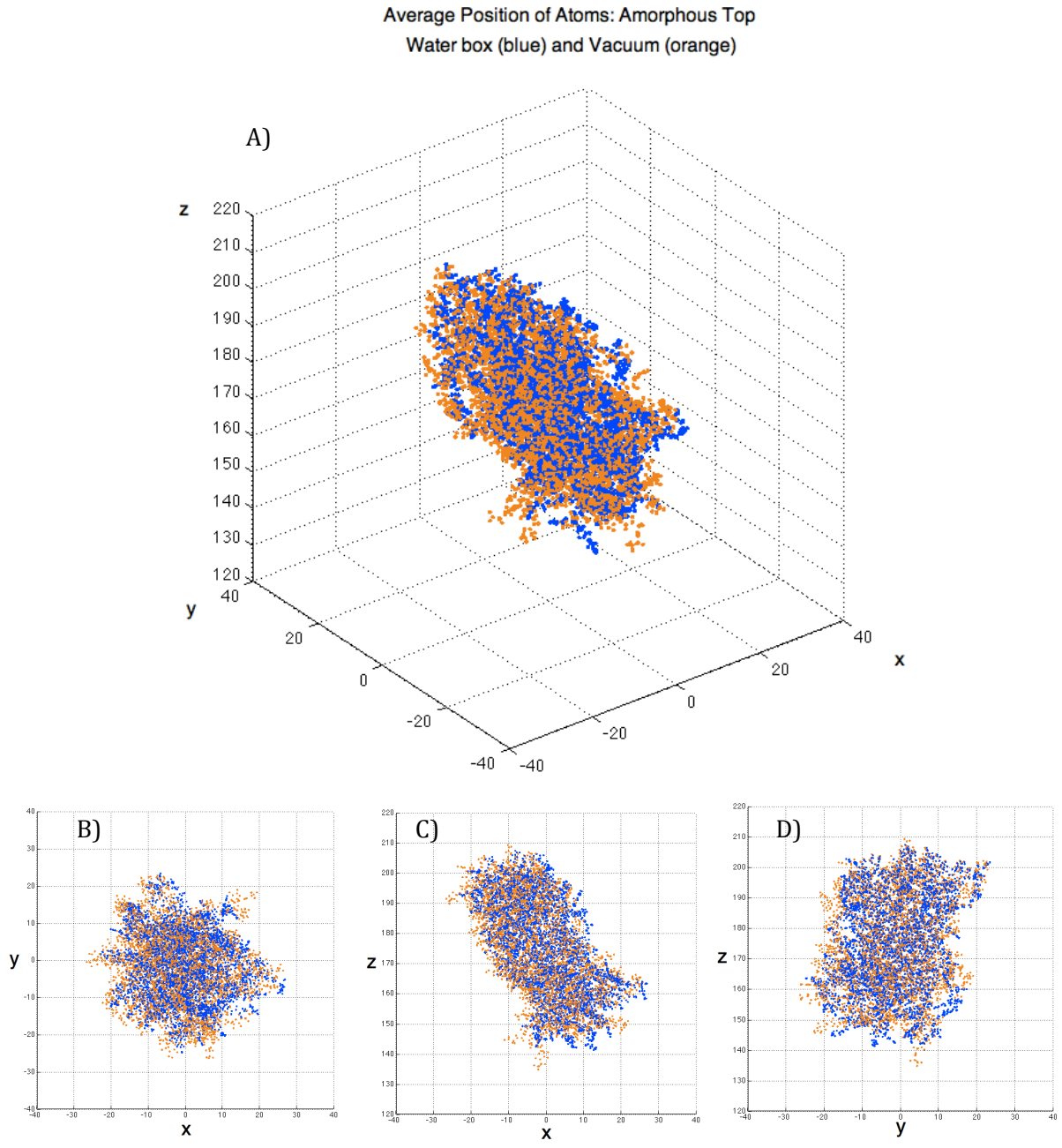

Figure 4.9 – Amorphous top region: average position of all atoms, in a water box (blue) and vacuum (orange). A) 3D view, B) top view *x-y* plane, C) side view *x-z* plane, D) side view *y-z* plane. All units in Å.

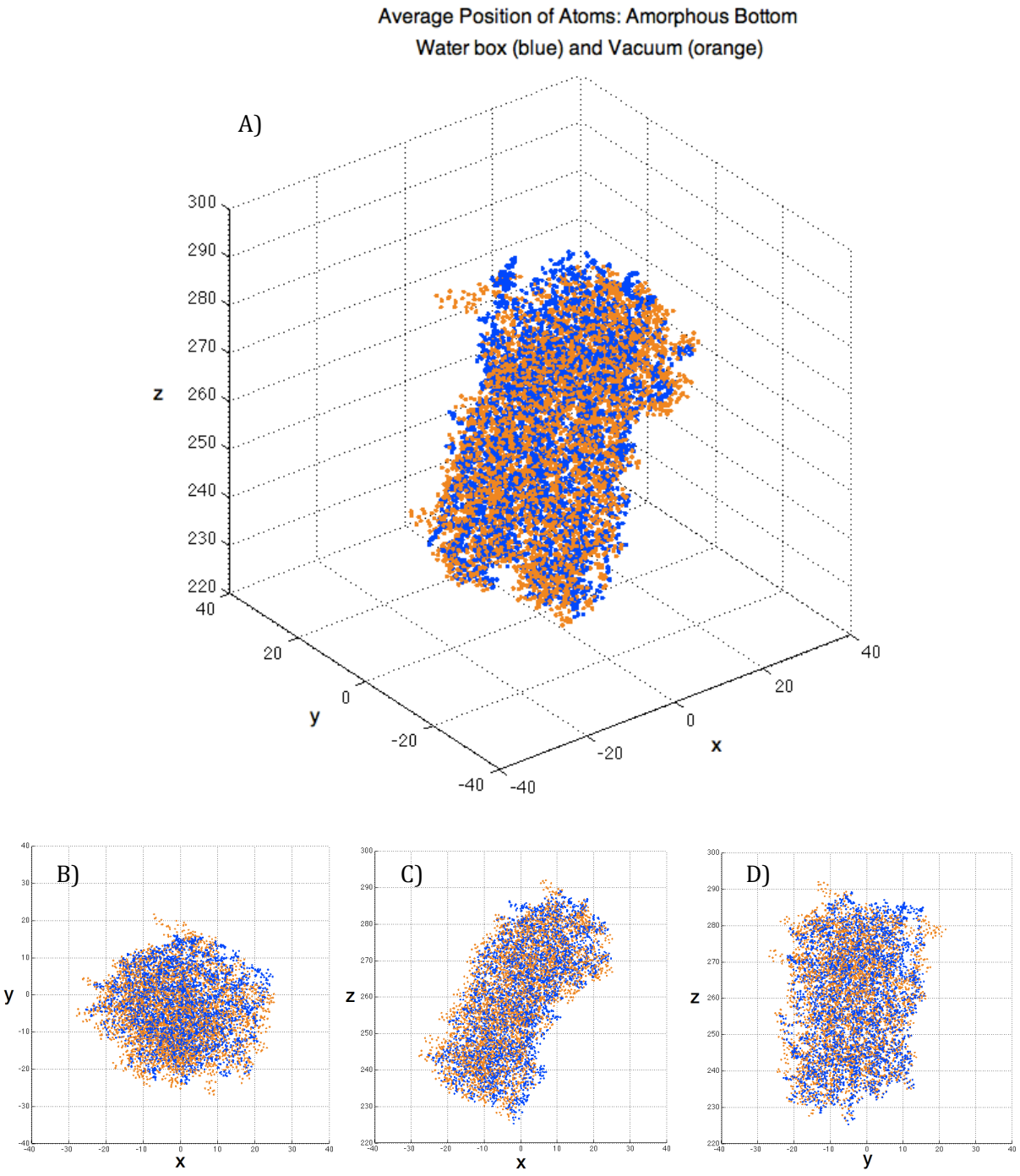

Figure 4.10 – Amorphous bottom region: average position of all atoms, in a water box (blue) and vacuum (orange). A) 3D view, B) top view *x-y* plane, C) side view *x-z* plane, D) side view *y-z* plane. All units in Å.

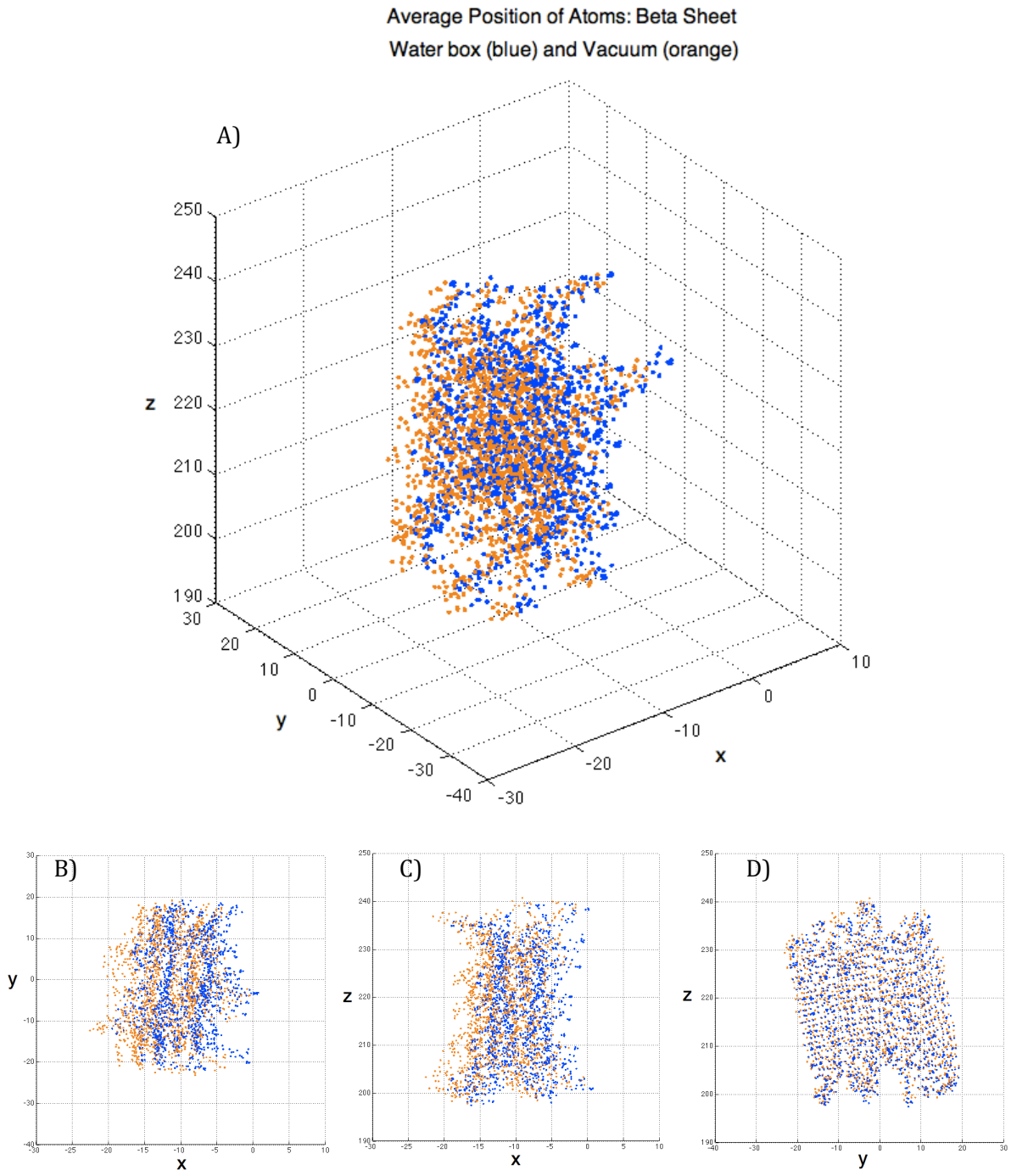

Figure 4.11 – Beta sheet region: average position of all atoms, in a water box (blue) and vacuum (orange). A) 3D view, B) top view *x-y* plane, C) side view *x-z* plane, D) side view *y-z* plane. All units in Å.

As shown in Figures 4.9, 4.10 and 4.11, the overall shapes described by the average positions of the atoms in the water box and the vacuum do not show the dramatic change expected in supercontraction. These results are not comparable by visual examination, so further analysis is necessary to describe any changes that occur.

Radii of Gyration and Equivalent Ellipsoids

The radius of gyration is the root mean square distance of all the atoms from the center of mass of the region. It is used here to quantify the size of the ensemble of atoms in the two amorphous regions and the beta sheet. The radius of gyration is calculated using a built-in function in VMD, details of which are in Appendix 7.1.4. The results are shown in Table 4.1.

The gyration tensor of each region was calculated using LAMMPS [92], script shown in Appendix 7.1.4, and plotted in MATLAB as an ellipsoid, in an attempt to give a representative volume of space taken up by the two amorphous regions and the beta sheet. The results are shown in Table 4.1. Figures 4.12, 4.13 and 4.14 show the equivalent ellipsoids for the water box equilibration (blue) and the vacuum equilibration (orange) of the three regions.

The gyration tensor S is

$$S = \begin{bmatrix} \boldsymbol{\lambda_x}^2 & 0 & 0 \\ 0 & \boldsymbol{\lambda_y}^2 & 0 \\ 0 & 0 & \boldsymbol{\lambda_z}^2 \end{bmatrix} \quad \textbf{(1)}$$

where λ_x, λ_y and λ_z are the principal moments of the gyration tensor.

The radius of gyration R is

$$R^2 = \boldsymbol{\lambda_x}^2 + \boldsymbol{\lambda_y}^2 + \boldsymbol{\lambda_z}^2 \quad \textbf{(2)}$$

The volume of the equivalent ellipsoid described by the gyration tensor is

$$V=(4/3)\pi\lambda_x\lambda_y\lambda_z \quad \textbf{(3)}$$

Table 4.1 –Gyration tensor and radius of gyration for amorphous bottom, amorphous top and beta sheet regions of MaSp1 in a water box and a vacuum

	Amorphous Top		Amorphous Bottom		Beta Sheet	
	Water Box	Vacuum	Water Box	Vacuum	Water Box	Vacuum
λ_x	9.20	9.51	9.74	10.30	3.82	3.86
λ_y	9.24	9.56	8.86	9.11	10.44	10.52
λ_z	16.15	15.90	14.96	14.46	10.58	10.58
Radius of gyration (Å)	19.94	19.97	20.76	20.85	15.32	15.39
Difference (%)	0.15		0.43		0.46	
Volume of ellipsoid	1831.66	1928.79	1720.72	1809.94	562.76	571.81
Difference (%)	5.30		5.19		1.61	

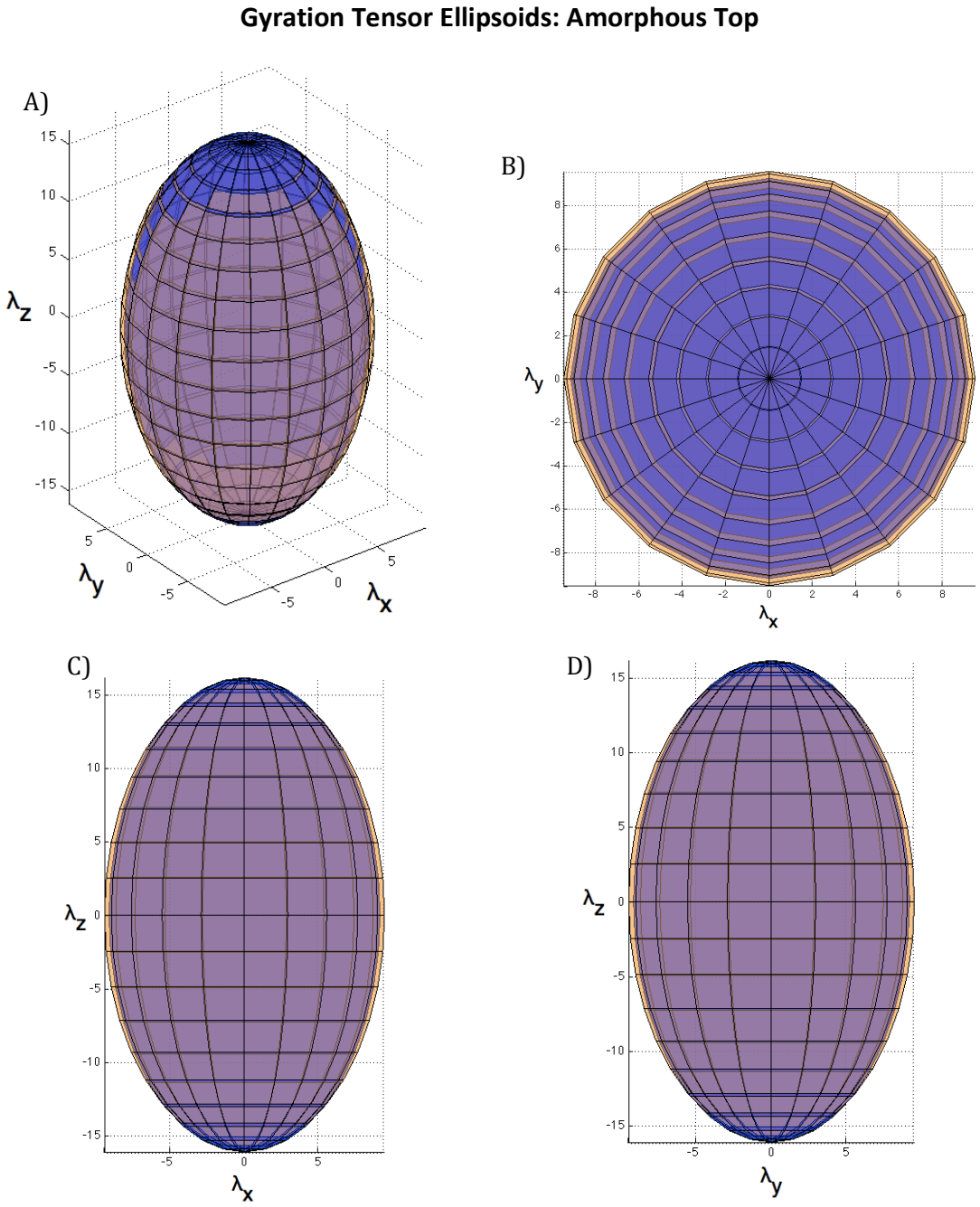

Figure 4.12 – Ellipsoids of the gyration tensors of the amorphous top region, for water box (blue) and vacuum (orange). A) Shows a 3D view, B) top view x-y plane, C) side view x-z plane and D) side view y-z plane. This image shows that the vacuum ellipsoid is slightly larger in the x and y directions, and the water box ellipsoid is larger in the z direction. All units in Å.

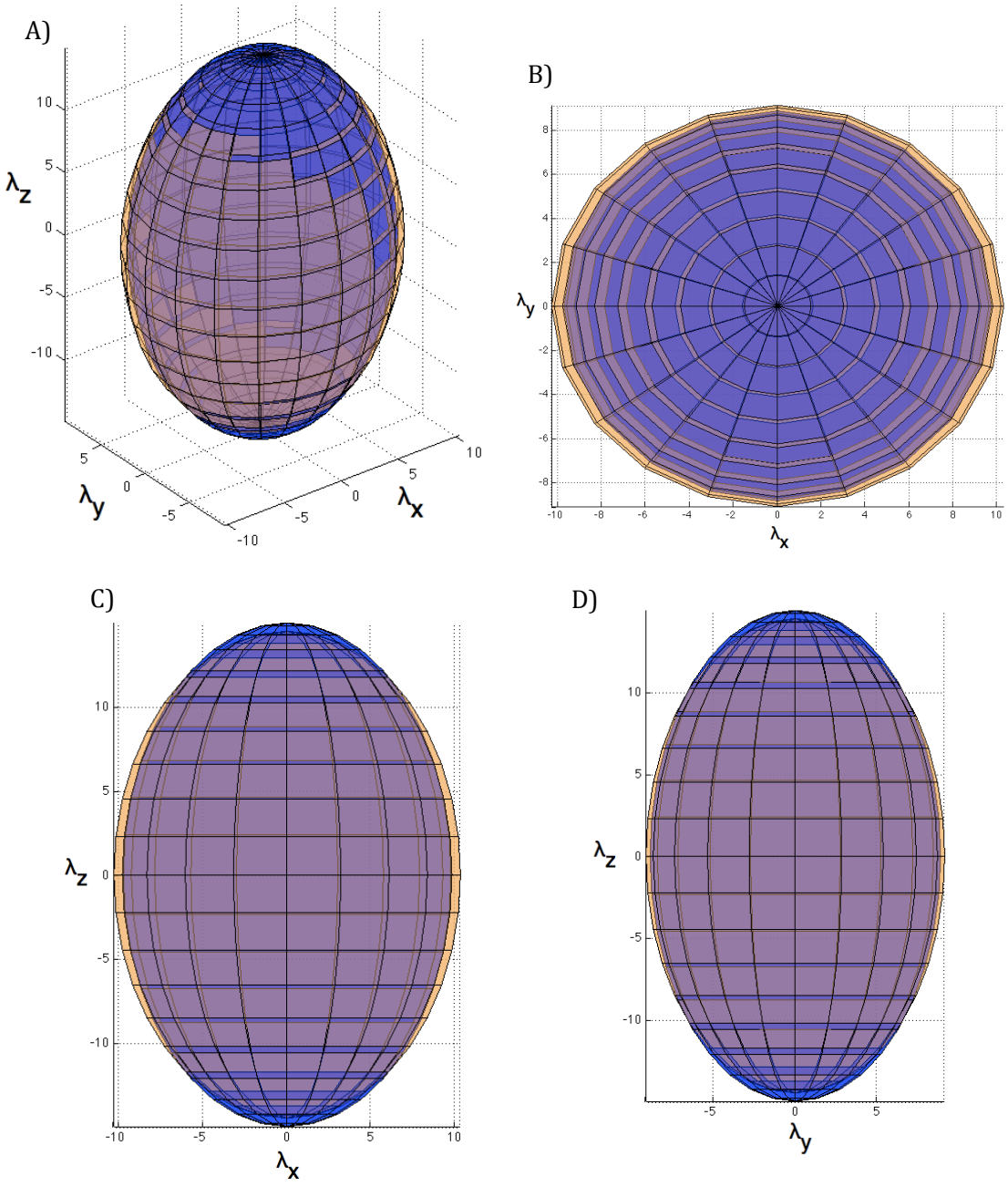

Figure 4.13 - Ellipsoids of the gyration tensors of the amorphous bottom region, for water box (blue) and vacuum (orange). A) Shows a 3D view, B) top view x-y plane, C) side view x-z plane and D) side view y-z plane. This image shows that the vacuum ellipsoid is slightly larger in the x and y directions, and the water box ellipsoid is larger in the z direction. All units in Å.

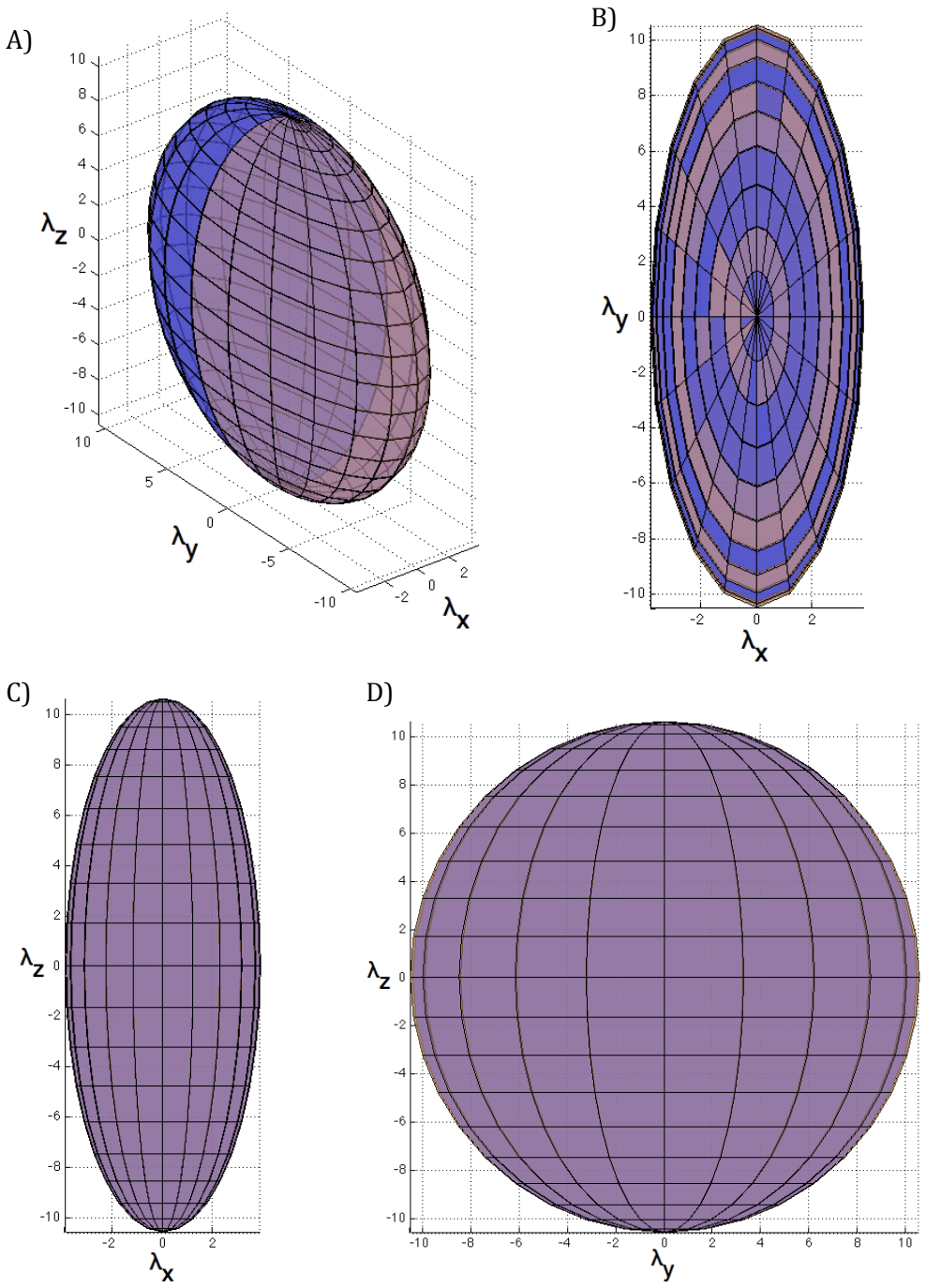

Figure 4.14 - Ellipsoids of the gyration tensors of the beta sheet, for water box (blue) and vacuum (orange). A) Shows a 3D view, B) top view x-y plane, C) side view x-z plane and D) side view y-z plane. This image shows there is no obvious difference between the size of the water box and vacuum ellipsoids. All units in Å.

Using the gyration tensor to draw an ellipsoid is an approximate way to measure the change in volume occupied by the two amorphous regions and the beta sheet. It is used here alongside the scalar radius of gyration measured by VMD to quantify the difference in size caused by the removal of water.

Table 4.1 shows that there is a small increase in volume of the ellipsoids of both amorphous regions. The top amorphous region increases in volume by 5.30 %, and the bottom amorphous region increases in volume by 5.19%, following the removal of water. This is shown also in the increase of the scalar radius of gyration, whose value for the top amorphous region increases by 0.15% and for the bottom amorphous region by 0.43%. It is likely that the calculation steps required to find the gyration tensor, combined with the multiplication required to calculate the volume of the ellipsoid has magnified the difference from below 0.5% to about 5%. In any case, both calculations show the amorphous regions increasing in volume in a vacuum, albeit by a small amount.

The volume of the ellipsoid for the beta sheet shows only a 1.6% increase in a vacuum, and the radius of gyration increases by 0.46%. The ellipsoid swells less than the two amorphous regions, supporting the hypothesis that water largely leaves the beta sheet region alone compared to the amorphous regions. The difference in radius of gyration is larger than the amorphous regions, however it is still below 0.5% and may not be relevant.

Comparing these values to those reported from experimental observations of supercontraction, it is clear that the molecular dynamics model does not show the dramatic change in size, of up to 50%, that is reported in the literature on supercontraction. There is a small increase in volume upon removal of water, however due to the fluctuations of the protein in equilibrium, and the manipulation of values required to calculate the ellipsoids, the swelling shown is not large enough to be conclusive.

4.1.4 Discussion

The conclusive result that can be drawn from this experiment is the change in average secondary structure between the water box and the vacuum. The percentages of all ordered structures increase and the percentage of disordered structures decreases when the water is removed. This supports the hypothesis that water molecules enter the amorphous regions of the dragline silk, and break the hydrogen bonds in this poorly bonded region, which causes a reorganization of the strands into random coiled structures. Their removal leads to the reformation of hydrogen bonds within the amorphous structure, turning random coils into more ordered turns.

The volumes occupied by the two amorphous regions and the beta sheet show a small swelling, between 0.5 and 5%, but this is not consistent with values reported in the literature on supercontraction of changes in length of up to 50%. In addition, the manipulation of data required to calculate the change in volume may have amplified changes, which make these results overall inconclusive.

In conclusion, the results of Experiment 1 show a change in molecular structure that is consistent with the theoretical model of supercontraction in the literature, namely an increase in ordered secondary structures and a corresponding decrease in disordered secondary structure. The results do not show the large change in volume that is expected with supercontraction: upon removal of water, the sample swells but only up to approximately 5%. The changes in molecular structure of the model do not translate to a bulk change in the size.

4.1.5 Outlook

The bulk response of a spider silk fiber to high humidity that is consistent with supercontraction remains to be shown in an all-atomistic model. The results shown above do not show the extreme contraction that was expected, however they do show a change in secondary structure consistent with literature and the hypothesis.

Further testing could be done on a more coarsely grained sample: it is possible that the actual size of the sample used above is so small that any change in volume is too small to be measured conclusively. Increasing the sample size while decreasing its resolution may allow larger size variations to be observed.

Supercontraction occurs in fibers of dragline silk where both MaSp1 and MaSp2 are present: thus testing the response of only MaSp1 to the removal of water may not show the bulk properties seen in experiments. This experiment could be repeated on MaSp2, since these proteins occur together in dragline silk, MaSp2 may have a different response due to the presence of different amino acids, namely proline. The mechanical properties (balance of strength and elasticity) of spider silk are due to the proper ratio of MaSp1 and MaSp2 [54]: thus is possible that the presence of MaSp2 has a large role to play in the macro scale manifestation of supercontraction.

4.2 Experiment 2: Glycine-Rich Linker Region

This second experiment is motivated by the suggestion that supercontraction occurs only above a threshold of 70% relative humidity [55] above which the water molecules are able to penetrate and disrupt the glycine-rich linker region. This experiment involves two short strands of spider silk, in the glycine-rich region that links the amorphous region and the beta sheet, shown in Figure 4.15 and referred to as the Snippet. This experiment was designed to simplify the system to fewer constituent parts, and at the same time investigating the effect of hydration at levels between 0 and 100%. The Snippet is two strands with sequence 1. GAGQGGYGGL and 2. QGGLGGRGAG, which represent the local glycine-rich linker region of a single longer, looped strand.

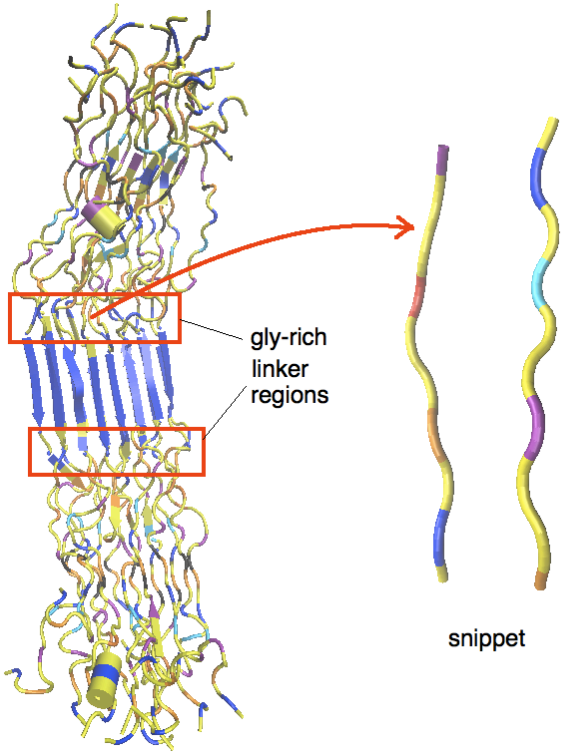

Figure 4.15 – MaSp1 colored by amino acid, showing the location of the glycine-rich linker regions and the origin of the snippet. The snippet consists of: glycine (yellow), alanine (blue), glutamine (orange), tyrosine (red), leucine (purple) and arginine (light blue).

The small glycine-rich linker regions, linking the beta sheets and the disordered amorphous regions have some order in their secondary structure, which can contain 3_{10} helices, beta bridges and turns. The secondary structures are not shown in Figure 4.15. Due to their relative order compared to the permanently disordered random coils, water molecules have a harder time breaking the secondary structure of the glycine-rich linker region. Above a certain relative

humidity (70%) it is hypothesized that the water can start entering and disordering the secondary structure by interfering with the hydrogen bonds and causing random coils to form.

4.2.1 Hypothesis

The experiment measures the effect of increasing water content on several properties of the protein sample: the secondary structure, the average distance of water molecules to the center of mass of the protein, the average size of the protein, and the solvent accessible surface area.

As discussed above for the larger MaSp1 sample, it is expected that as more water molecules are added, the secondary structure becomes more disordered: the percentage of helices, beta bridges and turns are reduced, the percentage of random coils is increased.

Since the glycine-rich region is mostly hydrophilic, water molecules should be attracted to and interact with the protein. The average distance between the water molecules and the protein is expected to decrease with increasing water content: since the presence of water destroys ordered secondary structures, a less-ordered structure should be able to contain more water molecules. However the presence of more water molecules may make the protein assume a more random coiled shape, which is likely to take up less volume and thus there may not be space for many water molecules near the center of the protein. Both the average distance of the water molecules to the center of the protein, and the size of the protein itself, are measured.

Finally, there are two possible scenarios for the effect of water on solvent accessible surface area. The first is that the water molecules create a very disordered random coil structure that collapses in on itself and decreases the available surface area accessible to the solvent. The second is that the water creates a secondary structure that is more random and more accessible to water molecules, maximizing the available surface area of the mostly hydrophilic protein to interaction with water.

4.2.2 Method

The snippet was created by manually altering the structure file of a larger silk strand (in PDB format). The resulting snippet is shown below, colored by residue name (Figure 4.16), and by its affinity to water (Figure 4.17).

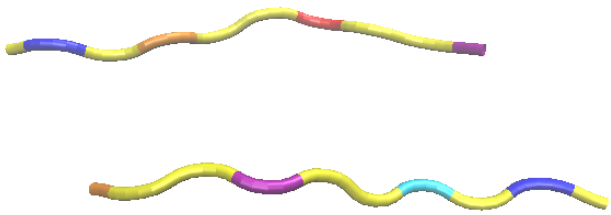

Figure 4.16 – Snippet colored according to residue name. Glycine is yellow, alanine is blue, glutamine is orange, tyrosine is red, leucine is purple, and arginine is light blue.

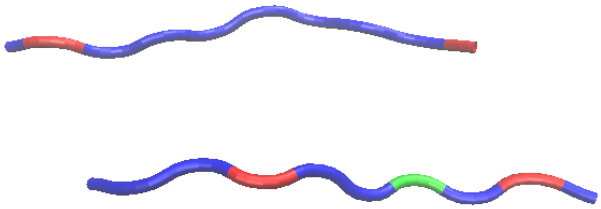

Figure 4.17 – Snippet colored according to its affinity with water. The hydrophilic regions are blue, hydrophobic is red and basic is green. The sample is mostly hydrophilic which suggests interaction with water.

Seven different samples were created from this snippet, each with ten more water molecules than the previous one, and labeled as follows: Snippet 0 (has no water), Snippet 10 (has 10 water molecules, and so on…), Snippet 20, Snippet 30, Snippet 40, Snippet 50 and Snippet 60. The samples were hydrated using a trial and error procedure in VMD, adjusting the box size and the padding around the protein until the desired number of water molecules appeared[4]. Only TIP3 explicit water was used.

The samples were equilibrated using NAMD with periodic boundary conditions for 30 nanoseconds. The aim is to keep *rigidbonds* off and with no bonded atoms excluded from non-bonded interactions. *Rigidbonds* in NAMD constrains the length of the bond between hydrogen atoms and the atom to which it is bonded. The small size of the sample should be able to support the additional computation required by a fully unconstrained system, and this should give a more accurate indication of the behavior of the water molecules. However, excluding no bonds caused the simulation to become unstable and to abort. In order for the simulation to run successfully, nearest neighbors directly connected via a linear bond were excluded from electrostatic

[4] This was done because VMD cannot create an automatic protein structure file for water molecules added manually to a PDB file.

interactions, which allowed *rigidbonds* to be off. This is specified using "exclude 1-2" in the Integrator Parameters of the NAMD configuration file. In addition, TIP3 water molecules include rigid bonds that cannot be altered. The configuration file used in all Snippet simulations is shown in Appendix 7.2.1.

4.2.3 Results

All Snippet simulations went from a parallel starting position, as shown in Figure 4.15, to a more coiled shape. Examples of the equilibrated structures are shown in Figure 4.18 for Snippet 0 and Snippet 60.

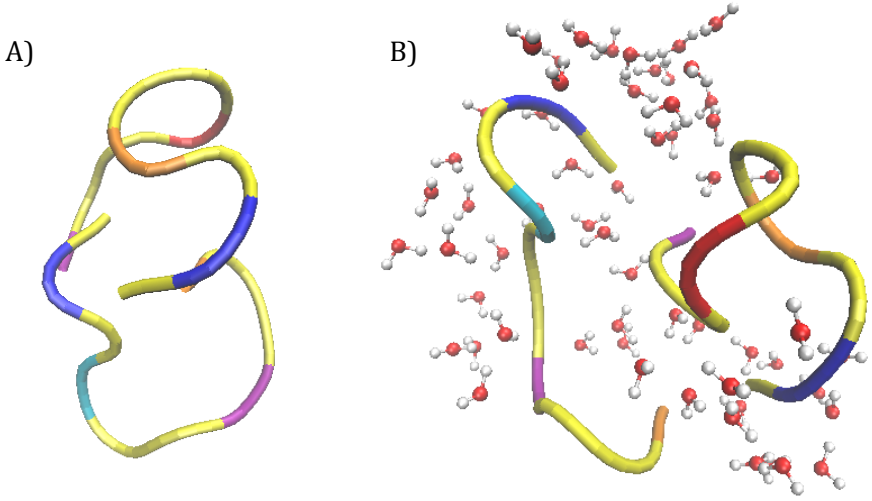

Figure 4.18 – A) Snippet 0 conformation after 30 nanoseconds, B) Snippet 60 conformation after 30 nanoseconds. All other Snippet samples (10 to 50) assume similar coiled shapes once allowed to relax from their straight starting conformations.

Root Mean Square Deviation

The root mean square deviation plots of all seven snippet simulations were created using VMD. These are shown in Figures 4.19 and 4.20. It was decided that the last 15 nanoseconds of all simulations would be considered when calculating the average secondary structure in the next section.

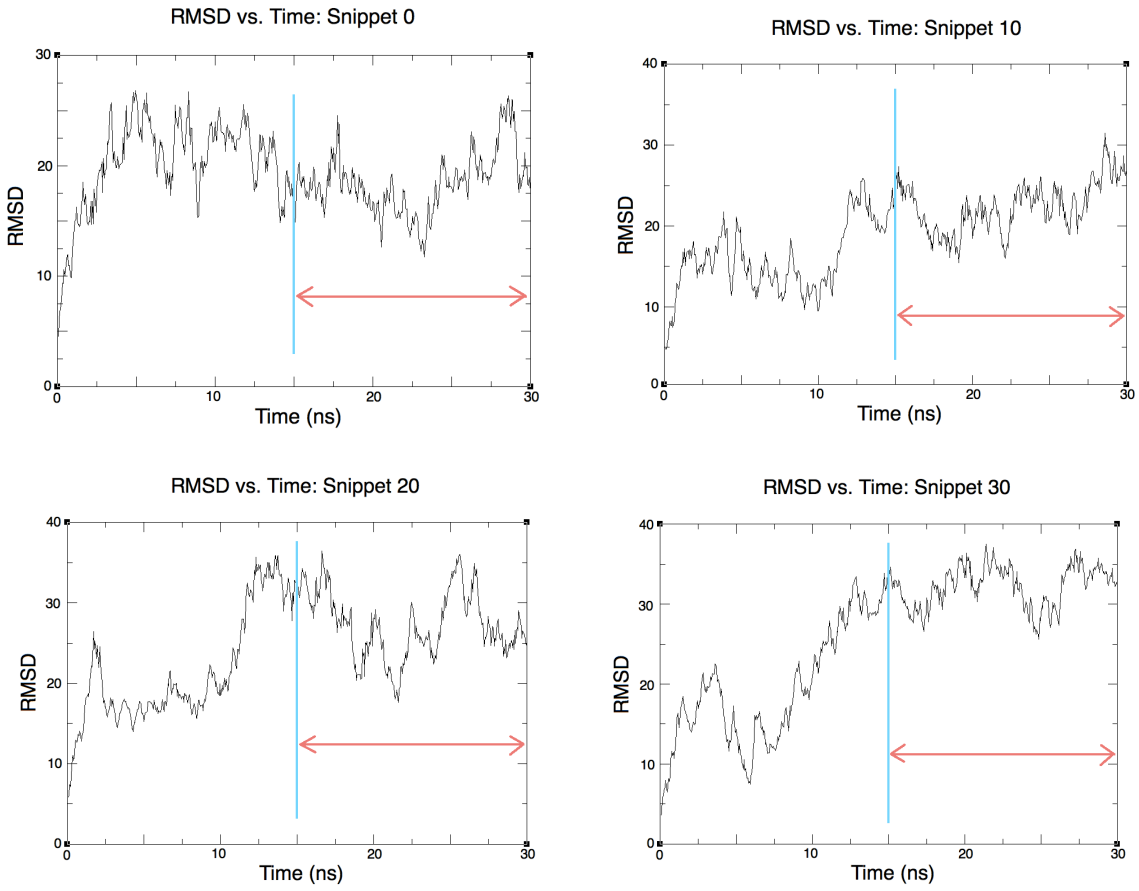

Figure 4.19 - Root mean square deviation versus time for Snippet 0, 10, 20 and 30, showing last 15 nanoseconds that are considered equilibrated.

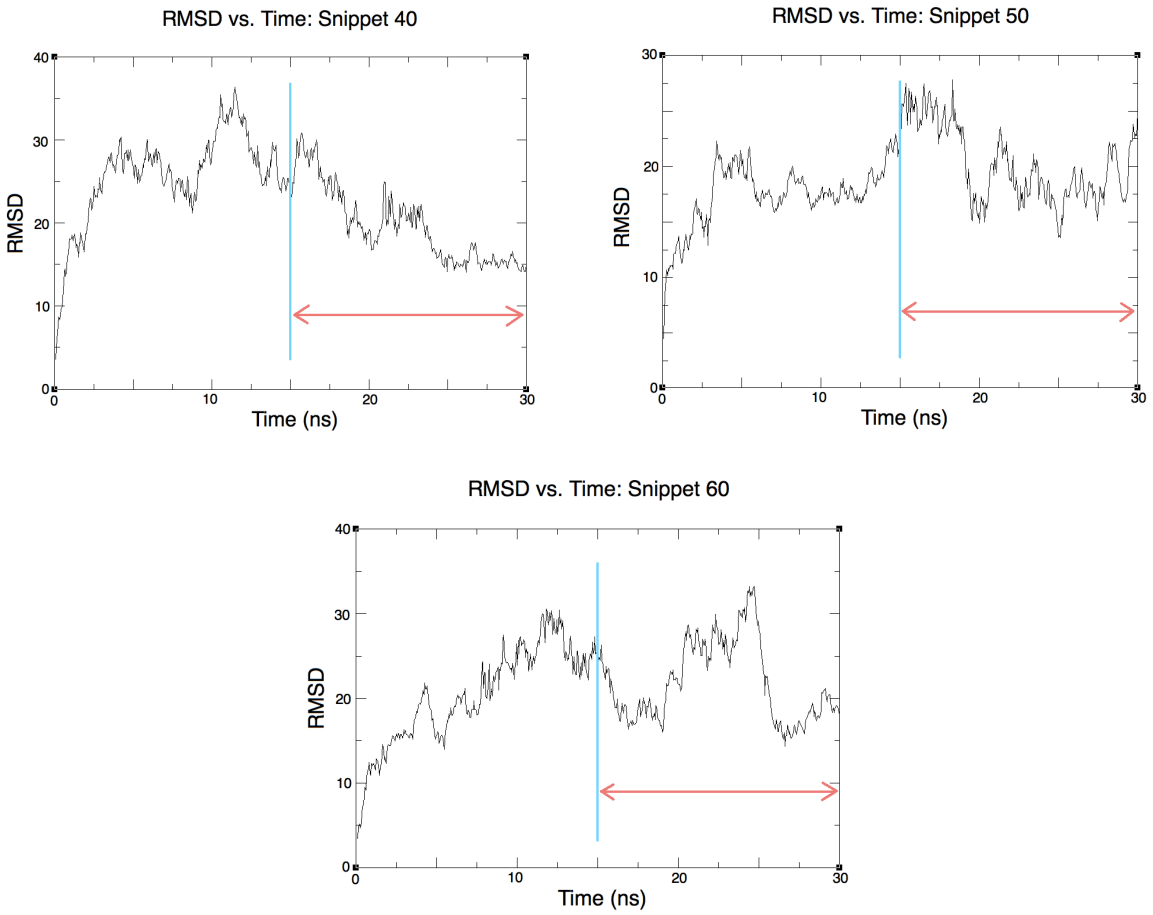

Figure 4.20 – Root mean square deviation for Snippet 40, 50 and 60, showing last 15 nanoseconds that are considered equilibrated.

Secondary Structure

The secondary structure over the whole trajectory of each snippet simulation was calculated in VMD, using the script shown in Appendix 7.1.2. Values for the last 15 nanoseconds were averaged, to find the average secondary structure for each level of hydration. These results are shown in Figure 4.21.

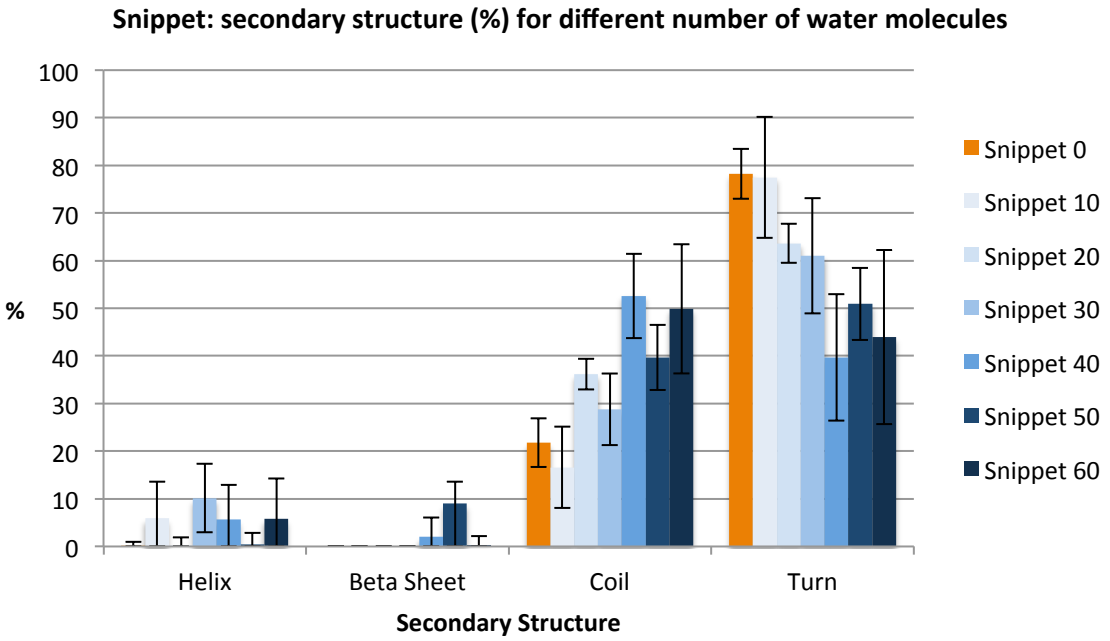

Figure 4.21 – Snippet: average secondary structure (%) for different number of water molecules, for the last 15 nanoseconds of the equilibration.

The helix content remains relatively constant between approximately 0 and 15% including error. This is constant with the hypothesis because the solvation level is probably not high enough to cause interference with the helices, since they are an ordered secondary structure, and thus more highly bonded and less easily disrupted than the turns.

The beta sheet content seems to increase with increasing water content. However, the script used to calculate secondary structure includes beta bridges in the script's definition of beta sheet. Beta bridges occur when a single hydrogen bond links two atoms in a beta-sheet shape. Since the snippet sample is not large enough to form beta sheets, the formation of beta bridges could be simply an accident where the strands have coiled and made a hydrogen bond that happens to be a beta bridge.

The percentage of coils increases, from approximately 20% to 50%, which is consistent with the hypothesis. Consequently the percentage of turns decreases, from 78% to 42%. This can be

explained by the water molecules associating with the relatively ordered yet weakly bonded turn structures and breaking their hydrogen bonds, allowing them to reconfigure into coils: a lower order, lower energy state. These results are consistent with the results observed in Experiment 1, which showed an increase in ordered secondary structure with the removal of water.

It is noted here that the standard deviations are significant. Due to the small size of the sample, the simulation trajectory shows large fluctuation, which seems to be echoed in the variation of secondary structure.

Distance of Water Molecules to Center of Protein

To get an estimate of both the motion of the water molecules through the equilibration, and the location in which they end up on average, the distance of each water molecule to the geometrical center of the protein was measured using the script shown in Appendix 7.2.2.

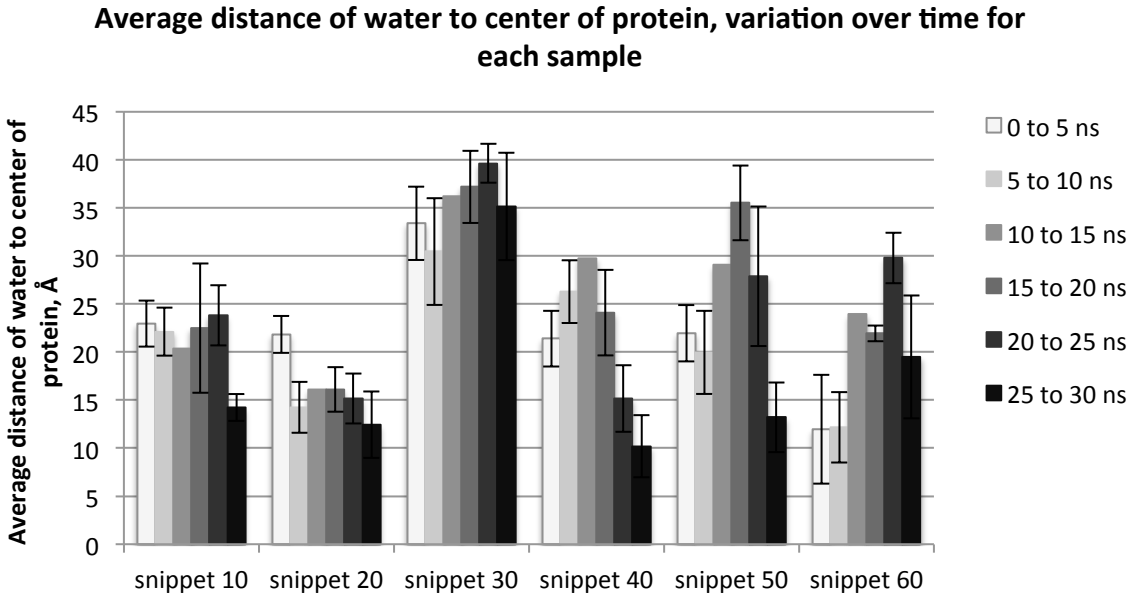

Figure 4.22 – Average distance of water molecules to center of protein, averaged in 5 nanosecond increments, for all snippet samples that contain water

The script calculates the distance between each water molecule and the geometrical center of the protein. The average value per time step was found to get an instantaneous average of the proximity of water to protein. These values were then grouped together in 5 nanosecond increments, to show how the average distance between water molecules and the center of the protein changes for each sample over the trajectory, and the data is shown in Figure 4.22.

Snippet 10 and 20 both show a decrease in average water distance to center of protein, both from approximately 23 Å to 13 Å. This can be interpreted as water molecules getting closer to the protein as the equilibration progresses, and is to be expected as the protein is mostly hydrophilic and there is plenty of space for the water molecules not to interfere with each other as they travel closer to the protein.

Snippet 30 has water molecules that are consistently furthest away from the protein. The average distance separating them increases by approximately 3 Å over the course of the equilibration. This sample seems to be an outlier since it follows neither trends nor average values of the remaining samples.

Snippet 40 and 50 show wider fluctuation in their average water distances. Snippet 40 shows an increase from 22 Å to a maximum of 30 Å, followed by a decrease to 10 Å. Snippet 50 shows a similar pattern, an initial increase from 22 to 35 Å, followed by a decrease to 13 Å. The larger fluctuations of these more hydrated samples could indicate a longer and more complicated equilibration, due to the larger number of molecules present. They do show an eventual increase in proximity between the protein and the water molecules, suggesting that after an initial reorganization where the water molecules may impede each other's progress, they are overall attracted to the protein.

Snippet 60 initially show an increase from 12 to 30 Å followed by a decrease to 20 Å. This is overall an increase in average distance between water and protein throughout the trajectory. This may signify two things: first that the large numbers of water molecules simply do not fit near the protein, which increases the average distance of all the water molecules. Second, the highly solvated model may have caused a collapse in secondary structure of the protein which has pushed out the water molecules forcing them to arrive in locations altogether further from the center of the protein.

Comparing the size of the standard deviation between the less hydrated and the more hydrated samples does not show much difference. A large standard deviation indicates both a large variation in the position of each water molecule in one timeframe, and a large fluctuation of the average position over time. Taking an average of an average is not ideal, however this data shows so much variation in any way it is presented that it is difficult to draw any meaningful conclusion about the bulk behavior of water molecules around the protein.

The average of this data was found for each hydration level, for the last 15 nanoseconds. The results are shown in Figure 4.23.

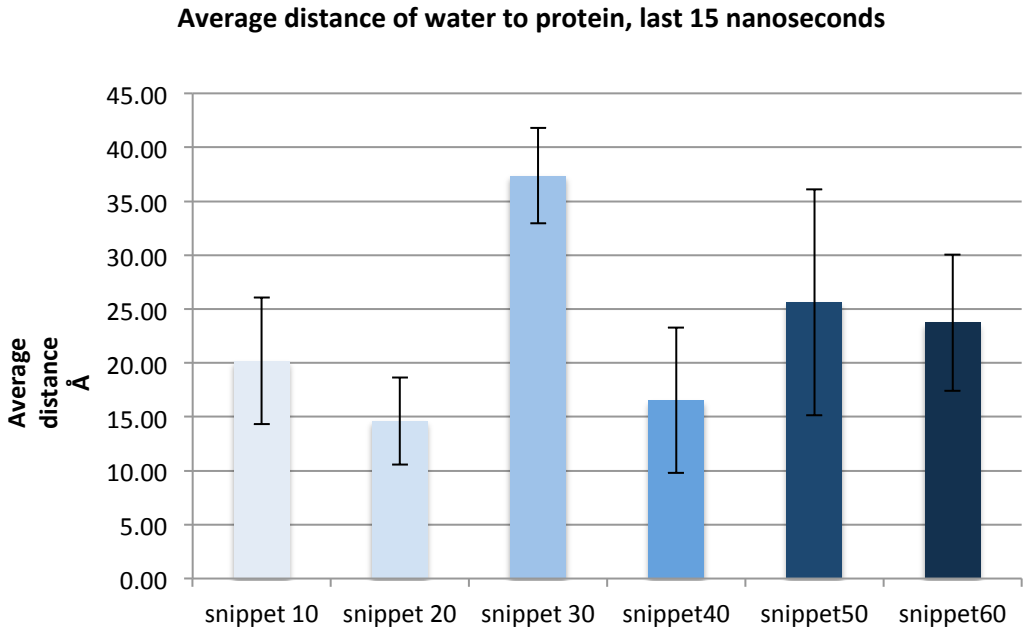

Figure 4.23 – Average distance of water to protein over last 15 nanoseconds

The averages for the last 10 nanoseconds do not show a definite trend, and have high standard deviations, shown here by the error bars. Thus, this data is not conclusive.

Size of Protein

To measure of the volume occupied by the protein, the distance of each atom to the instantaneous geometric center of the protein was measured for each frame, for each snippet sample (using the script shown in Appendix 7.2.3). An average per frame was found, and these results are shown in the plots below.

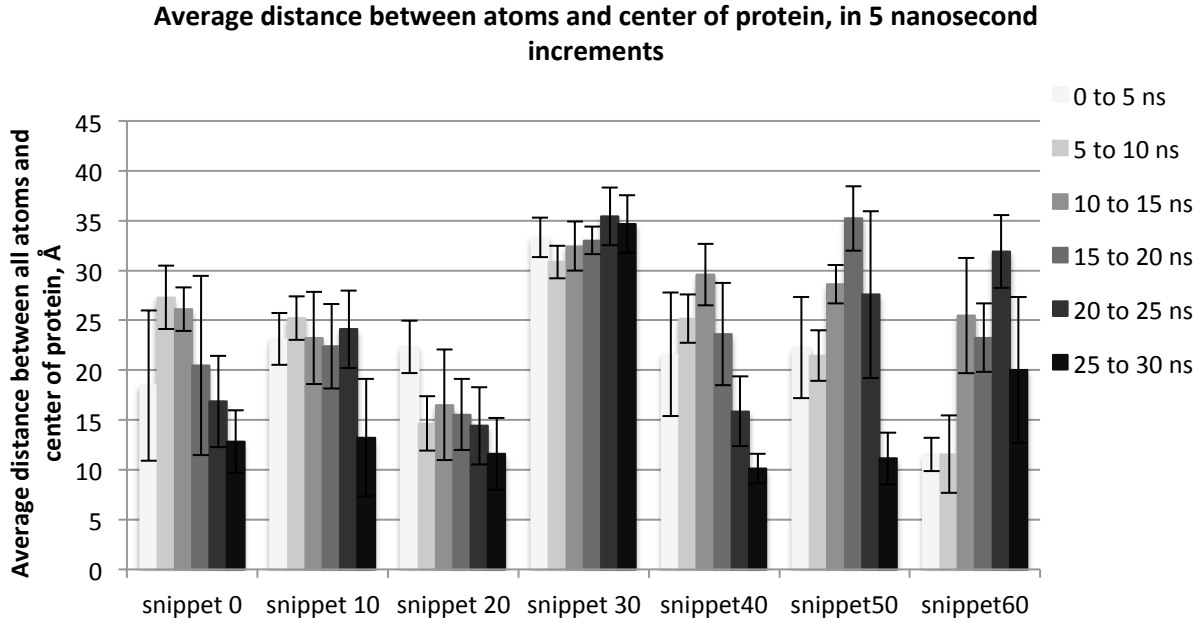

Figure 4.24 – Average distance of atoms to center of protein, in 5 nanosecond increments

Figure 4.24 shows a representation of the size of the protein only, in 5 nanosecond increments for each hydration level. When all the snippets are compared to each other, there is no clear conclusion and the presence of increasing amounts of water does not seem to correlate well with the size of the protein. However upon further inspection there are some trends that can be identified from these results. First of all, Snippet 30 is almost constantly the highest in the plot, meaning it is in general the largest or most spaced out conformation. The smallest one overall is Snippet 20. Unfortunately the distances have a large fluctuation, which is shown in the error bars, so in an effort to compare the structures to each other, the average over the last 15 nanosecond (and corresponding standard deviation) were found and are plotted in Figure 4.25.

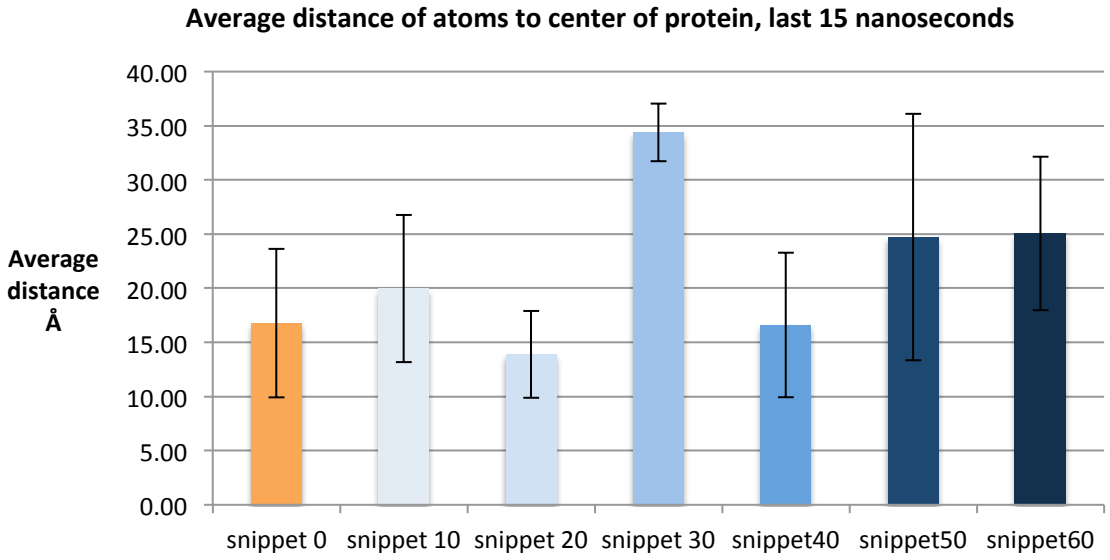

Figure 4.25 – Average distance of atoms to center of protein, for the last 15 nanoseconds

Figure 4.25 confirms the suggestion that Snippet 30 is consistently on average the largest. It also shows the smallest scatter, so fluctuates the least. If Snippet 30 is omitted, the remaining samples show some sort of upward trend, but since the error bars are so large, it is difficult to draw a conclusion from this data.

The average sizes of the proteins can be compared to the secondary structure data. Higher levels of hydration correspond to less ordered structures, which in theory take up a smaller volume. This data suggests an upward trend in protein size, which does not corroborate the hypothesis.

Further experimentation could run these simulations for a much longer time, or perhaps under constrained boundary conditions to represent the behavior of the glycine rich linker region in a more realistic setting.

Solvent Accessible Surface Area

Solvent accessible surface area (SASA) measures the amount of the protein that is accessible to interact with the surrounding solvent. SASA is measured using a built-in function in VMD. The radius of each atom can be extended by a certain length to define a sphere within which to measure exposure to solvent. This radius can be estimated when looking at the approximate radius of a solvent molecule: a value of 1.4 A has been reported in the literature [31, 93, 94] so this is the value chosen for the SASA calculated here.

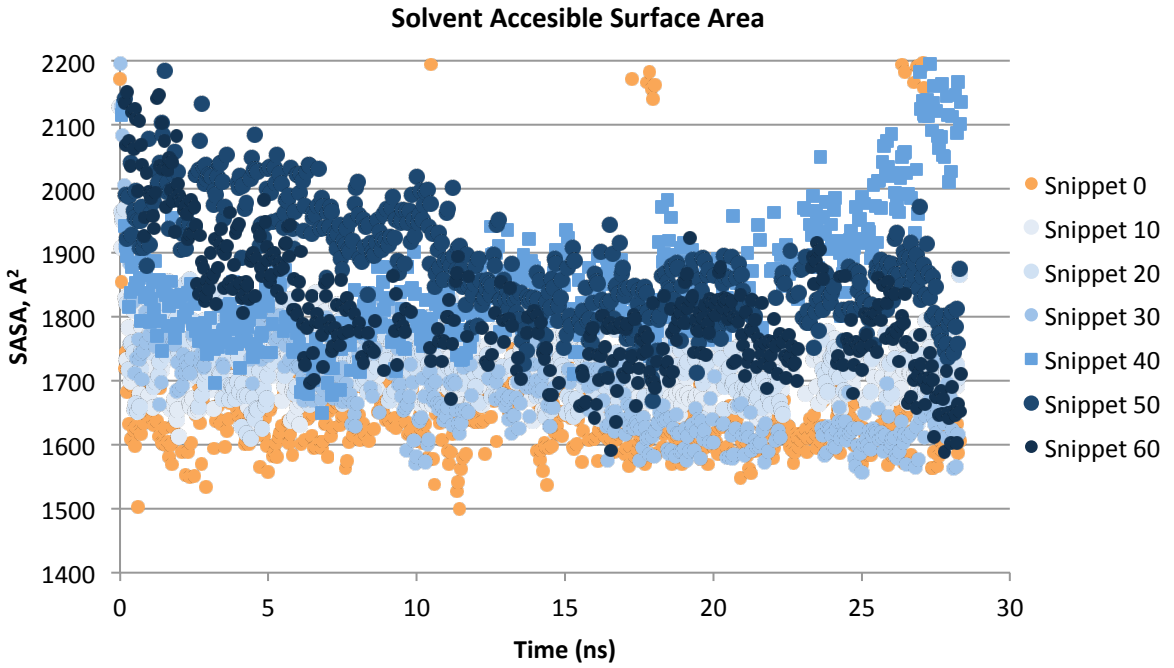

Figure 4.26 – Solvent accessible surface area of each Snippet over time

Figure 4.26 shows the solvent accessible surface area as it varies throughout the equilibration for all seven Snippet samples. The plot shows some noise, but for the purpose of discussion the few outlying values at the top of the plot are not considered and are not shown in Figure 4.26. For all cases, the SASA decreases in the early stages of the equilibration. This signifies the initial reorganization of the protein structure into a more collapsed folded state from its straighter starting point. It should be noted that the starting point corresponds to the conformation of the glycine-rich linker region within a larger silk model. Once it is unrestrained as in this experiment, it is expected to crumple substantially (which is shown to occur in all the samples, even in the absence of water, as shown in Figure 4.18). After about 10 nanoseconds the SASA stays relatively constant on average for the seven samples, except for an increase near the end for Snippet 40.

The results suggest that Snippet 0 has the least solvent-accessible surface area, and the most is seen in Snippet 40, 50 and 60. This could suggest that the presence of water entices the protein to assume a shape that is compatible with the presence of water – in other words, water makes a space for itself within and around the protein. This is compared with the results of the secondary structure: with increasing water content, the average secondary structure leans more towards random coils than ordered structures. Random coils are known to be less well bonded and thus more available for water to associate with. The SASA findings corroborate this argument. More water present near a protein could encourage it to assume a shape that includes and encourages the presence of water in the protein structure.

4.2.4 Discussion

Experiment 2 measures four different aspects of a small sample of spider dragline silk, under seven different hydration conditions. The results are not all conclusive, however some insights can be drawn. In general, a large fluctuation of the systems during equilibration lead to results being plagued with large standard deviations and sudden spikes in values. The configuration files of the simulations write .dcd files (trajectory files) only every 25000 time steps, which is equivalent to every 0.05 nanoseconds. Further simulation could be done with a reduced .dcd frequency, in order to create smoother visualizations of the protein activity. However this would not affect the values obtained, it could only serve to smooth out the plots.

Overall the most conclusive data can be drawn from the secondary structure content, as in Experiment 1. The simulations show an overall decrease in ordered secondary structure, which includes helices, beta sheets and turns, and a corresponding increase in the number of random coils, with increasing water content. This is consistent with both the literature and the results obtained in MaSp1. Supercontraction is proposed to involve the breaking of secondary structures in the glycine-rich linker region modeled by the snippet: this is supported by results here. However, literature suggests that the dominant secondary structure found in these linker regions are 3_{10} helices, and it is their deconstruction in high humidity that contributes to the macroscopic effect of supercontraction. There is very little content of 3_{10} helices in the snippet samples, as they are too short to create a real helix. Further study could be done using longer samples of the linker region that have the potential to form large helix structures.

The solvent accessible surface area (SASA) measurements show that SASA increases with water content. This suggests that water creates a secondary structure and conformation that allows for its presence. This may be due to the hydrophilic residues assuming shapes that try to maximize contact with water – if water is not present, the residues will instead attract themselves which could lead to a reduction in the solvent accessible surface area. In any case, this data suggests that water influences the structure and shape taken by the protein as it folds.

Finally the data on average distance of water molecules to the center of the protein, and average size of the protein only, both do not yield clear conclusions. This is most likely due to the large fluctuations shown by the trajectories, and could also be due to fact that the equilibrating protein crosses the periodic boundaries in some cases. It possible that crossing the periodic boundary causes errors in the vector distance between two atoms as calculated by VMD.

4.2.5 Outlook

The first step in further experimentation would be setting up a dragline silk sample of an intermediate size, preferably with the potential to form a clear secondary structure, and test it with different hydration levels as done for the snippet. This could show whether the breaking of ordered secondary structure leads to volume changes in hydration levels between 0 and 100%.

It could be possible that not enough water molecules were used in the snippet samples to reach the threshold of supercontraction – further experimentation could run the same snippet samples for a longer time and with more water molecules.

5 Conclusion

Spider dragline silk is a protein-based fiber that supercontracts by up to 50% upon exposure to high humidity. Water is known to affect the crystalline and amorphous regions of dragline silk differently, due to the extensive hydrogen bonding of the crystals they are not able to be penetrated by water and thus the bulk of the response is due to changes in the amorphous structure. Water is thought to penetrate into the amorphous structure, breaking hydrogen bonds and causing an entropic reorganization of the strands into a less ordered, random coil state. This molecular change is translated to the macro scale as a dramatic retraction in length [55, 58, 59, 61, 75, 79].

While supercontraction has been shown experimentally, the process by which the molecular structure changes and how this echoes up length scales to create the dramatic contraction are not well understood. The use of full-atomistic molecular dynamics models can allow investigation of a length scale that is not yet accessible by physical methods. By altering the water content, it is expected that the molecular dynamics model would exhibit supercontraction. Two experiments are designed to test this theory. Experiment 1 removes all the water from a sample of dragline protein MaSp1, to induce a reversal of supercontraction. Experiment 2 investigates the effect of varying water content on a small subsection of the amorphous region of MaSp1, on secondary structure, size and solvent accessible surface area.

The literature on supercontraction suggests that it is manifested at the nano scale by a change in secondary structure. More specifically, this change involves the loss of ordered structures including turns, helices, and to a lesser extent, beta sheets, and the gain of disordered random coils. This change has been shown both in Experiment 1, on a large sample of the protein MaSp1, and in Experiment 2, on a much smaller sub-sample of MaSp1. In this respect, both experiments were consistent with the literature.

However, neither experiment managed to show the bulk effects of supercontraction in a conclusive manner, and could be due to many reasons. The multitude of assumptions and simplifications made in molecular dynamics simulations could have led to errors, complications and the mistranslation of molecular structure to physical shape. The samples may be just too small for any volumetric change to be measurable. Both experiments are inherently unrealistic because they both involve a full or partial vacuum – thus making it difficult to compare the results qualitatively with experimental findings in literature.

In this study, molecular dynamics demonstrates supercontraction to a certain extent. While the change in molecular structure and order is clear, the study fails to show the bulk change in volume on the nano scale consistent with the physical expression of supercontraction on a macroscopic length scale.

Possible further experimentation and approaches to demonstrating supercontraction on a molecular scale include:

- Creating a sample with less resolution: it is possible that the actual size of the MaSp1 sample used in Experiment 1 is so small that any change in volume is too small to be measured conclusively. Increasing the sample size while decreasing its resolution may allow larger size variations to be observed.
- Experiment 1 could be repeated on a similar full resolution sample of MaSp2. Since these proteins occur together in dragline silk, MaSp2 may have a different response due to the presence of different amino acids, namely proline. MaSp2 could have a large role to play in supercontraction.
- A larger Snippet sample could be created, preferably with the potential to form a clear secondary structure, and test it with different hydration levels as done for the Snippet in Experiment 2. This could show whether the breaking of ordered secondary structure such as helices, beta sheets and turns leads to volume changes in hydration levels between 0 and 100%.
- Finally, it is possible that not enough water molecules were used in the snippet samples to reach the threshold of supercontraction – further experimentation could run the same snippet samples for a longer time and with more water molecules.

6 References

1. Buehler, M.J., *Tu(r)ning weakness to strength.* Nano Today, 2010. **5**(5): p. 379-383.

2. Cranford, S.W. and M.J. Buehler, *Biomateriomics*2012, Dordrecht: Springer Netherlands.

3. Meyers, M.A., et al., *Biological materials: Structure and mechanical properties.* Progress in Materials Science, 2008. **53**(1): p. 1-206.

4. Currie, H.A., et al., *Genetically Engineered Chimeric Silk-Silver Binding Proteins.* Advanced Functional Materials, 2011. **21**(15): p. 2889-2895.

5. Espinosa, H.D., et al., *Merger of structure and material in nacre and bone – Perspectives on de novo biomimetic materials.* Progress in Materials Science, 2009. **54**(8): p. 1059-1100.

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

7. Buehler, M.J. and Y.C. Yung, *Deformation and failure of protein materials in physiologically extreme conditions and disease.* Nature Materials, 2009. **8**(3): p. 175-88.

8. Fratzl, P. and R. Weinkamer, *Nature's hierarchical materials.* Progress in Materials Science, 2007. **52**(8): p. 1263-1334.

9. Calvert, P., *Encyclopedia of Materials: Science and Technology*. 2nd ed2001.

10. Ji, B. and H. Gao, *Mechanical properties of nanostructure of biological materials.* Journal of the Mechanics and Physics of Solids, 2004. **52**(9): p. 1963-1990.

11. Chen, P.-Y., et al., *Structure and mechanical properties of selected biological materials.* Journal of the Mechanical Behavior of Biomedical Materials, 2008. **1**(3): p. 208-226.

12. Wegst, U.G.K. and M.F. Ashby, *The mechanical efficiency of natural materials.* Philosophical Magazine, 2004. **84**(21): p. 2167-2186.

13. Cranford, S.W., et al., *Materiomics: An -omics Approach to Biomaterials Research.* Advanced Materials, 2013: p. 1-23.

14. Vincent, J.F.V. and D.L. Mann, *Systematic technology transfer from biology to engineering.* Philosophical Transactions of the Royal Society A, 2002. **360**(1791): p. 159-173.

15. Vincent, J.F.V., *Structural Biomaterials*1990, Princeton, NJ: Princeton University Press.

16. Buehler, M.J. and T. Ackbarow, *Fracture mechanics of protein materials.* Materials Today, 2007. **10**(9): p. 46-58.

17. McKee, T. and J.R. McKee, *Biochemistry: The Molecular Basis of Life* 3rd ed2003, New York City: McGraw-Hill.

18. Pratt, C.W. and K. Cornely, *Essential Biochemistry*2004, Hoboken, NJ: John Wiley & Sons.

19. Bagchi, B., *From anomalies in neat liquid to structure, dynamics and function in the biological world.* Chemical Physics Letters, 2012. **529**: p. 1-9.

20. Degrève, L., et al., *On the Role of Water in the Protein Activity.* Brazilian Journal of Physics, 2004. **34**(1): p. 102-115.

21. Denisov, V.P. and B. Halle, *Protein hydration dynamics in aqueous solution.* Faraday Discussions, 1996. **103**: p. 227-244.

22. Levy, Y. and J.N. Onuchic, *Water and proteins: A love-hate relationship.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(10): p. 3325-3326.

23. Kim, B. and F. Hirata, *Structural fluctuation of protein in water around its native state: A new statistical mechanics formulation.* Journal of Chemical Physics, 2013. **138**(5): p. 054108-11.

24. Swenson, J., et al., *Properties of hydration water and its role in protein dynamics.* Journal of Physics: Condensed Matter, 2007. **19**(20): p. 205109-18.

25. Prasad, B.V.L.S. and K. Suguna, *Role of water molecules in the structure and function of aspartic proteinases.* Acta Crystallographica Section D Biological Crystallography, 2002. **58**(2): p. 250-259.

26. Harris, T.K. and A.S. Mildvan, *High-precision measurement of hydrogen bond lengths in proteins by nuclear magnetic resonance methods.* Proteins, 1999. **35**(3): p. 275-282.

27. Bao, G. and S. Suresh, *Cell and molecular mechanics of biological materials.* Nature Materials, 2003. **2**(11): p. 715-725.

28. Aftabuddin, M.D. and S. Kundu, *Hydrophobic, Hydrophilic, and Charged Amino Acid Networks within Protein.* Biophysical Journal, 2007. **93**(1): p. 225-231.

29. Michigan State University. *Proteins, Peptides & Amino Acids*. n.d. [cited 2013 18 April]; Available from: http://www2.chemistry.msu.edu/faculty/reusch/virttxtjml/proteins.htm.

30. Karchin, R. *Hidden Markov Models and Protein Sequence Analysis*. n.d. [cited 2013 29 April]; Available from: http://compbio.soe.ucsc.edu/ismb99.handouts/KK185FP.html.

31. Durham, E., et al., *Solvent accessible surface area approximations for rapid and accurate protein structure prediction.* Journal of Molecular Modeling, 2009. **15**(9): p. 1093-1108.

32. Huang, D.M. and D. Chandler, *Temperature and length scale dependence of hydrophobic effects and their possible implications for protein folding.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(15).

33. Colgin, M.A. and R.V. Lewis, *Spider minor ampullate silk proteins contain new repetitive sequences and highly conserved non-silk-like "spacer regions".* Protein Science, 1998. **7**: p. 667-672.

34. ABCTE. *Proteins*. Module 3: Basic Science, Biology Core 2007 [cited 2013 May 8]; Available from: http://www.abcte.org/files/previews/biology/s3_p2.html.

35. *Alpha-Helix, 310 Helix, Pi Helix (Molecular Biology)*. what-when-how [cited 2013 May 8]; Available from: http://what-when-how.com/molecular-biology/alpha-helix-310-helix-and-pi-helix-molecular-biology/.

36. Vollrath, F., B. Madsen, and Z. Shao, *The effect of spinning conditions on the mechanics of a spider's dragline silk.* Proceedings of the Royal Society B, 2001. **268**: p. 2339-2346.

37. Grubb, D.T. and G. Ji, *Molecular chain orientation in supercontracted and re-extended spider silk.* International Journal of Biological Macromolecules, 1999. **24**(2-3): p. 203-210.

38. Wong Po Foo, C., et al., *Novel nanocomposites from spider silk-silica fusion (chimeric) proteins.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(25): p. 9428-9433.

39. Bratzel, G. and M.J. Buehler, *Molecular mechanics of silk nanostructures under varied mechanical loading.* Biopolymers, 2012a. **97**(6): p. 408-417.

40. Gronau, G., et al., *A review of combined experimental and computational procedures for assessing biopolymer structure-process-property relationships.* Biomaterials, 2012. **33**(33): p. 8240-55.

41. Hinman, M.B., J.A. Jones, and R.V. Lewis, *Synthetic spider silk: a modular fiber.* Trends in Biotechnology, 2000. **18**: p. 374-379.

42. Lewis, R.V., *Spider silk: the unraveling of a mystery.* Accounts of Chemical Research, 1992. **25**(392-398).

43. Xu, M. and R.V. Lewis, *Structure of a protein superfiber: spider dragline silk.* Proceedings of the National Academy of Sciences of the United States of America, 1990. **87**(18): p. 7120-7124.

44. Patel, V., *Nephila clavipes male and female*, 2007: Wikimedia Commons.

45. Augsten, K., P. Mühlig, and C. Herrmann, *Glycoproteins and skin-core structure in Nephila clavipes spider silk observed by light and electron microscopy.* Scanning, 2000. **22**(1): p. 12-15.

46. Brooks, A.E., et al., *An Investigation of the Divergence of Major Ampullate Silk Fibers from Nephila clavipes and Argiope aurantia.* Biomacromolecules, 2005. **6**: p. 3095-3099.

47. Belton, D.J., et al., *Silk-silica composites from genetically engineered chimeric proteins: materials properties correlate with silica condensation rate and colloidal stability of the proteins in aqueous solution.* Langmuir, 2012b. **28**(9): p. 4373-4381.

48. Vollrath, F. and D.P. Knight, *Liquid crystalline spinning of spider silk.* Nature, 2001. **410**(6828): p. 541-548.

49. Qin, Z. and M.J. Buehler, *Spider silk: webs measure up.* Nature Materials, 2013. **12**(3): p. 185-187.

50. Ko, F.K., *Engineering Properties of Spider Silk Fibers*, in *Natural Fibers, Plastics and Composites*, F.T. Wallenberger and N.E. Weston, Editors. 2004, Kluwer Academic Publishers.

51. Hinman, M.B. and R.V. Lewis, *Isolation of a Clone Encoding a Second Dragline Silk Fibroin - Nephila clavipes dragline silk is a two-protein fiber.* Journal of Biological Chemistry, 1992. **267**(21): p. 19320-19324.

52. Keten, S. and M.J. Buehler, *Atomistic model of the spider silk nanostructure.* Applied Physics Letters, 2010a. **96**(15).

53. van Beek, J.D., et al., *The molecular structure of spider dragline silk: Folding and orientation of the protein backbone.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(16): p. 10266-10271.

54. Brooks, A.E., et al., *Distinct contributions of model MaSp1 and MaSp2 like peptides to the mechanical properties of synthetic major ampullate silk fibers as revealed in silico.* Journal of Nanotechnology, Science and Applications, 2008. **1**: p. 9-16.

55. Blackledge, T.A., et al., *How super is supercontraction? Persistent versus cyclic responses to humidity in spider dragline silk.* Journal of Experimental Biology, 2009. **212**: p. 1981-1989.

56. Renault, A., et al., *Surface properties and conformation of Nephila clavipes spider recombinant silk proteins at the air-water interface.* Langmuir, 2009. **25**(14): p. 8170-8180.

57. Brown, C.P., et al., *The critical role of water in spider silk and its consequence for protein mechanics.* Nanoscale, 2011. **3**(9): p. 3805-3811.

58. Guan, J., F. Vollrath, and D. Porter, *Two mechanisms for supercontraction in Nephila spider dragline silk.* Biomacromolecules, 2011. **12**(11): p. 4030-4035.

59. Keten, S. and M.J. Buehler, *Nanostructure and molecular mechanics of spider dragline silk protein assemblies.* Journal of the Royal Society Interface, 2010c. **7**(53): p. 1709-1721.

60. Liu, Y., Z. Shao, and F. Vollrath, *Relationships between supercontraction and mechanical properties of spider silk.* Nature Materials, 2005a. **4**(12): p. 901-905.

61. Agnarsson, I., et al., *Spider silk as a novel high performance biomimetic muscle driven by humidity.* Journal of Experimental Biology, 2009b. **212**(13): p. 1990-1994.

62. Gillespie, D.B., C. Viney, and P. Yager, *Raman Spectroscopic Analysis of the Secondary Structure of Spider Silk Fiber*, in *Silk Polymers*1993, American Chemical Society.

63. Thiel, B.L., K.B. Guess, and C. Viney, *Non-Periodic Lattice Crystals in the Hierarchical Microstructure of Spider (Major Ampullate) Silk.* Biopolymers, 1997. **41**(7703-719).

64. Elices, M., et al., *The hidden link between supercontraction and mechanical behavior of spider silks.* Journal of the Mechanical Behavior of Biomedical Materials, 2011. **4**(5): p. 658-669.

65. Gosline, J.M., M.E. DeMont, and M.W. Denny, *The structure and properties of spider silk.* Endeavour, 1986. **10**(1): p. 37-43.

66. Hayashi, C.Y., N.H. Shipley, and R.V. Lewis, *Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins.* International Journal of Biological Macromolecules, 1999. **24**: p. 271-275.

67. Termonia, Y., *Molecular Modeling of Spider Silk Elasticity.* Macromolecules, 1994. **24**: p. 7378-7381.

68. Grubb, D.T. and L.W. Jelinski, *Fiber Morphology of Spider Silk: The Effects of Tensile Deformation.* Macromolecules, 1997. **30**(96): p. 2860-2867.

69. Canabady-Rochelle, L.L.S., et al., *Bioinspired silicification of silica-binding peptide-silk protein chimeras: comparison of chemically and genetically produced proteins.* Biomacromolecules, 2012. **13**(3): p. 683-690.

70. Koski, K.J., et al., *Non-invasive determination of the complete elastic moduli of spider silks.* Nature Materials, 2013. **12**(3): p. 1-6.

71. Vollrath, F. and D. Porter, *Spider silk as a model biomaterial.* Applied Physics A, 2006. **82**(2): p. 205-212.

72. An, B., et al., *Inducing β-Sheets Formation in Synthetic Spider Silk Fibers by Aqueous Post-Spin Stretching.* Biomacromolecules, 2011. **12**: p. 2375-2381.

73. Bratzel, G. and M.J. Buehler, *Sequence-structure correlations in silk: Poly-Ala repeat of N. clavipes MaSp1 is naturally optimized at a critical length scale.* Journal of the Mechanical Behavior of Biomedical Materials, 2012b. **7**: p. 30-40.

74. Tatham, A.S. and P.R. Shewry, *Elastomeric proteins: biological roles, structures and mechanisms.* Trends in Biochemical Sciences, 2000. **25**(11): p. 567-571.

75. Agnarsson, I., et al., *Supercontraction forces in spider dragline silk depend on hydration rate.* Zoology, 2009a. **112**(5): p. 325-331.

76. Turner, J. and C. Karatzas, *Advanced Spider Silk Fibers By Biomimicry*, in *Natural Fibers, Plastics and Composites*2004, Kluwer Academic Publishers.

77. Kümmerlen, J., et al., *Local Structure in Spider Dragline Silk Investigated by Two-Dimensional Spin-Diffusion Nuclear Magnetic Resonance.* Macromolecules, 1996. **29**(8): p. 2920-2928.

78. Sponner, A., et al., *Characterization of the Protein Components of Nephila Clavipes Dragline Silk.* Biochemistry, 2005. **44**(16): p. 4727-4736.

79. Vollrath, F., D. Porter, and C. Holland, *The Science of Silks.* MRS Bulletin, 2013. **38**(1): p. 73-80.

80. Liu, Y., Z. Shao, and F. Vollrath, *Extended wet-spinning can modify spider silk properties.* Chemical Communications, 2005b(19): p. 2489-2491.

81. Ene, R., P. Papadopoulos, and F. Kremer, *Supercontraction in Nephila spider dragline silk – Relaxation into equilibrium state.* Polymer, 2011. **52**(26): p. 6056-6060.

82. Gosline, J.M., M.W. Denny, and M.E. DeMont, *Spider silk as rubber.* Nature, 1984. **309**: p. 551-552.

83. Sagui, C. and T.A. Darden, *Molecular dynamics simulations of biomolecules: long-range electrostatic effects.* Annual Review of Biophysics and Biomolecular Structure, 1999. **28**: p. 155-179.

84. Piana, S., et al., *Evaluating the effects of cutoffs and treatment of long-range electrostatics in protein folding simulations.* PloS ONE, 2012. **7**(6): p. e33918.

85. van Gunsteren, W.F. and H.J.C. Berendsen, *Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry.* Angewandte Chemie International Edition, 1990. **29**: p. 992-1023.

86. Phillips, J.C., et al., *Scalable Molecular Dynamics with NAMD.* Journal of Computational Chemistry, 2005. **26**: p. 1781-1802.

87. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual Molecular Dynamics.* Journal of Molecular Graphics, 1996. **14**: p. 33-38.

88. Frishman, D. and P. Argos, *Knowledge-Based Protein Secondary Structure Assignment.* Proteins, 1995. **23**(4): p. 566-579.

89. Heinig, M. and D. Frishman, *STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins.* Nucleic Acids Research, 2004. **32**: p. W500-2.

90. Isgro, T., et al., *NAMD Tutorial*, 2012, University of Illinois at Urbana-Champaign.

91. Matysiak, S., P.G. Debenedetti, and P.J. Rossky, *Role of Hydrophobic Hydration in Protein Stability: A 3D Water-Explicit Protein Model Exhibiting Cold and Heat Denaturation.* Journal of Physical Chemistry B, 2012. **116**: p. 8095-8104.

92. Plimpton, S., *Fast Parallel Algorithms for Short-Range Molecular Dynamics.* Journal of Computational Physics, 1995. **117**: p. 1-42.

93. Lee, B. and F.M. Richards, *The Interpretation of Protein Structures: Estimation of Static Accessibility.* Journal of Molecular Biology, 1971. **55**(3): p. 379-400.

94. Hasel, W., T.F. Hendrickson, and W.C. Still, *A Rapid Approximation to the Solvent Accessible Surface Areas of Atoms.* Tetrahedron Computer Methodology, 1988. **1**(2): p. 103-116.

95. Falsafi, S. *Average position of atoms in a trajectory*. 2011 October 27 2011 [cited 2013 March 19]; Available from: http://www.ks.uiuc.edu/Research/vmd/mailing_list/vmd-l/18952.html.

7 Appendix

7.1 Experiment 1: MaSp1 in a Vacuum

7.1.1 NAMD Configuration File

```
#############################################################
## JOB DESCRIPTION                                         ##
#############################################################

# Minimization and Equilibration of
# MaSp1 with no water - variable pressure constant volume


#############################################################
## ADJUSTABLE PARAMETERS                                   ##
#############################################################

structure          masp1.psf
coordinates        masp1.pdb
set temperature    300
set outputname     masp1_eq2
firsttimestep      0

#############################################################
## SIMULATION PARAMETERS                                   ##
#############################################################

# Input
paraTypeCharmm        on
parameters          par_all27_prot_lipid.inp
temperature         $temperature

# Force-Field Parameters
exclude             scaled1-4
1-4scaling          1.0
cutoff              12.
switching           on
switchdist          10.
pairlistdist        13.5

# Integrator Parameters
timestep            2.0  ;# 2fs/step
rigidBonds          all  ;# needed for 2fs steps
nonbondedFreq       1
```

```
fullElectFrequency  2
stepspercycle       10

# Constant Temperature Control
langevin            on    ;# do langevin dynamics
langevinDamping     5     ;# damping coefficient (gamma) of 5/ps
langevinTemp        $temperature
langevinHydrogen    off    ;# don't couple langevin bath to hydrogens

# Periodic Boundary Conditions
cellBasisVector1    69.2690    0.   0.
cellBasisVector2     0.   64.6500   0.
cellBasisVector3     0.    0   172.2230
cellOrigin          -3.07813   -2.95906  217.39705
wrapAll             on
margin            3.0

# PME (for full-system periodic electrostatics)
PME                 yes
PMEGridSizeX        72
PMEGridSizeY        72
PMEGridSizeZ        180

# Constant Pressure Control (variable volume)
# useGroupPressure      yes ;# needed for rigidBonds
# useFlexibleCell       no
# useConstantArea       no
#langevinPiston        on
#langevinPistonTarget  1.01325 ;#  in bar -> 1 atm
#langevinPistonPeriod  100.
#langevinPistonDecay   50.
#langevinPistonTemp    $temperature

# Output
outputName          $outputname
restartfreq         25000     ;# 25000steps = every 0.05 ns
dcdfreq             25000
xstFreq             25000
outputEnergies      25000
outputPressure      25000


#############################################################
## EXTRA PARAMETERS                                        ##
#############################################################

#############################################################
## EXECUTION SCRIPT                                        ##
#############################################################
```

```
# Minimization
# minimize            25000
reinitvels          $temperature
run 15000000 ;# 30ns
```

Reference:
[73].

7.1.2 Secondary Structure of a Trajectory

Calculates the secondary structure of a .dcd file

```
set all [atomselect top all]
set mol [molinfo top]
set nf [molinfo $mol get numframes]
set log [open allratio.txt w]
puts $log "Frame \t HELIX \t BS \t COIL \t TURN"

for {set i 0} {$i < $nf} {incr i 1} {
animate goto $i
display update ui
$all frame $i
$all update
mol ssrecalc top
#set nhb [llength [lindex [measure hbonds 4.0 40 $prot] 0]]
set TT [format "%5.2f" [llength [[atomselect top "name CA"] get
resname]]]
set HL [format "%5.2f" [llength [[atomselect top "helix and name CA"]
get resname]]]
set BS [format "%5.2f" [llength [[atomselect top "betasheet and name
CA"] get resname]]]
set RC [format "%5.2f" [llength [[atomselect top "coil and name CA" ]
get resname]]]
set TR [format "%5.2f" [llength [[atomselect top "turn and name CA" ]
get resname]]]
set HLP [format "%5.2f" [expr 100*$HL/$TT]]
set BSP [format "%5.2f" [expr 100*$BS/$TT]]
set RCP [format "%5.2f" [expr 100*$RC/$TT]]
set TRP [format "%5.2f" [expr 100*$TR/$TT]]

puts $log "$i \t $HLP \t $BSP \t $RCP \t $TRP"
#puts "$i \t $nhb"
flush $log
}

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

Reference: Chia-Ching Chou, 2013

7.1.3 Average Position of Equilibrated Structure

```
# VMD .tcl script used to
# calculate average positions of all atoms

set sel [atomselect top protein]
set avg [measure avpos $sel]
set coords [$sel get {x y z}]
$sel set {x y z} $avg
$sel writepdb avgpdb.pdb
$sel set {x y z} $coords
$sel delete
```

Reference: [95]

7.1.4 Radius of Gyration and Gyration Tensor

Radius of Gyration in VMD

```
# Calculating the radius of gyration in VMD

set a [atomselect top protein]
measure rgyr $a
```

Gyration Tensor in Lammps

```
# Created by charmm2lammps v1.8.1 on Tue Apr 16 15:22:47 EDT 2013

units           real
neigh_modify    delay 2 every 1

atom_style      full
bond_style      harmonic
angle_style     charmm
dihedral_style  charmm
improper_style  harmonic

pair_style      lj/charmm/coul/long 8 10
pair_modify     mix arithmetic
kspace_style    pppm 1e-4

read_data       abot_masp1_230600.data

special_bonds   charmm
fix             1 all nve
fix             2 all shake 1e-6 500 0 m 1.0
velocity        all create 0.0 12345678 dist uniform

thermo          1
thermo_style    multi
timestep        0.5

compute rgyr all gyration
fix 111 all ave/time 1 1 1 c_rgyr[1] c_rgyr[2] c_rgyr[3] file rgyr.out

run             1
```

7.2 Experiment 2: Glycine-Rich Linker Region (Snippet)

7.2.1 Configuration File for all Snippet Simulations

```
#############################################################
## JOB DESCRIPTION                                         ##
#############################################################

# Minimization and Equilibration of
# Snippet0 (and Snippet10 – 60, replace file names as necessary)

#############################################################
## ADJUSTABLE PARAMETERS                                   ##
#############################################################

structure          snippet0.psf
coordinates        snippet0.pdb
set temperature    300
set outputname     snippet0_eq7
firsttimestep      0

#############################################################
## SIMULATION PARAMETERS                                   ##
#############################################################

# Input
paraTypeCharmm      on
parameters          par_all27_prot_lipid.inp
temperature         $temperature

# Force-Field Parameters
exclude             1-2
# 1-4scaling          1.0
cutoff              12.
switching           on
switchdist          10.
pairlistdist        13.5
splitpatch     hydrogen

# Integrator Parameters
timestep            1.0  ;# 1fs/step
rigidBonds          none  ;# needed for 2fs steps
nonbondedFreq       2
fullElectFrequency  2
stepspercycle       10

# Constant Temperature Control
langevin            on    ;# do langevin dynamics
langevinDamping     5     ;# damping coefficient (gamma) of 5/ps
```

```
langevinTemp        $temperature
langevinHydrogen    off    ;# don't couple langevin bath to hydrogens

# Periodic Boundary Conditions
cellBasisVector1    61.958    0.   0.
cellBasisVector2     0.   35.258   0.
cellBasisVector3     0.    0   28.923
cellOrigin          -127.375   -0.854  0.940
wrapAll             on
margin          3.0

# PME (for full-system periodic electrostatics)
PME                 yes
PMEGridSizeX        64
PMEGridSizeY        36
PMEGridSizeZ        30

# Constant Pressure Control (variable volume)
useGroupPressure      no ;# needed for rigidBonds
useFlexibleCell       no
useConstantArea       no

langevinPiston        off
# langevinPistonTarget  1.01325 ;#  in bar -> 1 atm
# langevinPistonPeriod  100.
# langevinPistonDecay   50.
# langevinPistonTemp    $temperature

# Output
outputName          $outputname
restartfreq         25000     ;# 25000steps = every 0.05 ns
dcdfreq             25000
xstFreq             25000
outputEnergies      25000
outputPressure      25000


#############################################################
## EXTRA PARAMETERS                                        ##
#############################################################


#############################################################
## EXECUTION SCRIPT                                        ##
#############################################################

# Minimization
minimize            25000
reinitvels          $temperature
run 15000000 ;# 30ns
```

7.2.2 Distance of Water Molecules to Center of Protein

```
# To measure the distance of all water molecules to center of
protein,
# using indices defined by pdb file

set all [atomselect top all]
set prot [atomselect top protein]
set w [atomselect top "index 200 to 286"]
set mol [molinfo top]
set nf [molinfo $mol get numframes]
set nw [molinfo $mol get numatoms]
set log [open waterdistance.txt w]

for {set i 0} {$i <= $nf} {incr i 1} {
      set c [measure center $prot]
      #set coordc [lindex [$c get {x y z}] 0]
      puts $log "frame $i"
      for {set k 227} {$k < $nw} {incr k 1} {
            set w [atomselect top "index $k"]
            set d [measure center $w]
            set dist [vecdist $c $d]
            puts $log "$k $dist"
      }
}
```

7.2.3 Distance of all Protein Atoms to Center of Protein

```
# To measure the distance of all atoms to center of protein,
# using indices defined by pdb file

set all [atomselect top all]
set prot [atomselect top protein]
#set w [atomselect top "index 200 to 286"]
set mol [molinfo top]
set nf [molinfo $mol get numframes]
set nw [molinfo $mol get numatoms]
set log [open distance_all.txt w]

for {set i 0} {$i <= $nf} {incr i 1} {
      set prot [atomselect top protein frame $i]
      set c [measure center $prot]
      #set coordc [lindex [$c get {x y z}] 0]
      puts $log "frame $i"
      for {set k 0} {$k < 227} {incr k 1} {
            set w [atomselect top "index $k"]
            set d [measure center $w]
            set dist [vecdist $c $d]
            puts $log "$k $dist"
      }
}
```